Video Article

QTL Mapping and CRISPR/Cas9 Editing to Identify a Drug Resistance Gene in *Toxoplasma gondii*

Bang Shen[1], Robin H. Powell[2], Michael S. Behnke[2]

[1] State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University

[2] Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University

Correspondence to: Bang Shen at shenbang@mail.hzau.edu.cn, Michael S. Behnke at mbehnke@lsu.edu

Abstract

Scientific knowledge is intrinsically linked to available technologies and methods. This article will present two methods that allowed for the identification and verification of a drug resistance gene in the Apicomplexan parasite *Toxoplasma gondii*, the method of Quantitative Trait Locus (QTL) mapping using a Whole Genome Sequence (WGS) -based genetic map and the method of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 -based gene editing. The approach of QTL mapping allows one to test if there is a correlation between a genomic region(s) and a phenotype. Two datasets are required to run a QTL scan, a genetic map based on the progeny of a recombinant cross and a quantifiable phenotype assessed in each of the progeny of that cross. These datasets are then formatted to be compatible with R/qtl software that generates a QTL scan to identify significant loci correlated with the phenotype. Although this can greatly narrow the search window of possible candidates, QTLs span regions containing a number of genes from which the causal gene needs to be identified. Having WGS of the progeny was critical to identify the causal drug resistance mutation at the gene level. Once identified, the candidate mutation can be verified by genetic manipulation of drug sensitive parasites. The most facile and efficient method to genetically modify *T. gondii* is the CRISPR/Cas9 system. This system comprised of just 2 components both encoded on a single plasmid, a single guide RNA (gRNA) containing a 20 bp sequence complementary to the genomic target and the Cas9 endonuclease that generates a double-strand DNA break (DSB) at the target, repair of which allows for insertion or deletion of sequences around the break site. This article provides detailed protocols to use CRISPR/Cas9 based genome editing tools to verify the gene responsible for sinefungin resistance and to construct transgenic parasites.

Video Link

The video component of this article can be found at https://www.jove.com/video/55185/

Introduction

Host range determines the extent of a parasites' prevalence. Some parasites have very specific host requirements that limit the area from which they are found, others are generalists. One such generalist is *Toxoplasma gondii (T. gondii)*. This parasite is found worldwide as it can infect all mammals and many birds. Humans are also susceptible and it is estimated that approximately 1/3 of the global population has been infected. Fortunately, a robust immune response normally controls the growth of the parasite, but in situations where the immune system is compromised the parasite can grow unchecked and cause diseases, often encephalic. Also, this parasite can cause congenital diseases if previously uninfected women are infected during pregnancy as they lack immune memory to quickly limit the spread of the parasite. Additionally, there is a burden of ocular toxoplasmosis that can result in vision loss [1]. For these reasons *T. gondii* has become a focus of study, and due to the many molecular methods developed for its study, a model for Apicomplexan parasites. Two methods that will be discussed here are Quantitative Trait Locus (QTL) mapping and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 gene editing. QTL mapping and CRISPR/Cas9 editing are, respectively, forward and reverse genetic approaches that have been used in previous studies to identify and/or characterize *T. gondii* virulence genes. Here, these methods are combined to identify and confirm the function of a sinefungin resistance (SNF[r]) gene, TgME49_290860, and its orthologs (annotated as *SNR1*) [2].

Although *T. gondii* can infect a wide variety of intermediate hosts, it must pass through and infect the intestinal epithelial cells of a felid to complete the life cycle. Cats are the definitive hosts of the parasite where the sexual stages are produced and genetic recombination via meiosis occurs. To conduct a QTL study one must create a genetic cross, and in the case of *Toxoplasma* this means passing two different parasite strains that differ in a phenotypic trait, the parental stains, through a cat to produce recombinant progeny [3]. Before being fed to cats, the parental strains are made resistant to separate drugs to allow for more efficient identification of recombinants by double drug selection of the progeny [4]. Three drugs have been used in *T. gondii* for this purpose; fluorodeoxyribose (FUDR) for which uracil phosphoribosyl transferase (*UPRT*) is the resistance gene [5], adenosine arabinoside (ARA) for which Adenosine Kinase (*AK*) is the resistance gene [6], and sinefungin (SNF) for which the resistance gene was unknown [7]. Several genetic crosses have been created for *T. gondii*, but only the 24 progeny of the ME49-FUDR[r] X VAND-SNF[r] cross were genotyped using whole genome sequencing (WGS) [8]. This opened the possibility of mapping and identifying the SNF resistance gene using this cross as the VAND parent was made sinefungin resistant by chemical mutagenesis of the drug sensitive VAND

(VAND-SNF[s]) strain, and the VAND reference genome was sequenced using the VAND-SNF[s] strain thus allowing for the identification of all polymorphisms between the progeny WGS and the VAND-SNF[s] reference genome, including the inherited parental VAND-SNF[r] mutation that rendered some of the progeny sinefungin resistant.

In order to identify the causal single nucleotide polymorphism (SNP) in the SNF[r] progeny, several computational based open source resources can be used to analyze the data. To create the genetic map for the ME49-FUDR[r] X VAND-SNF[r] cross the REDHORSE software suite [9] was developed that uses WGS alignments of the parents and progeny to accurately detect the genomic positions of genetic crossovers. This mapping information can then be combined with the phenotypic data (SNF[r] in the progeny) to format a dataset compatible with the 'qtl' package [10] in the R statistical programming software where a QTL scan can be run to reveal significant loci correlated with the phenotype. To identify the causal SNP located within the QTL locus, the WGS reads of the progeny can be individually aligned to the sinefungin sensitive VAND genome using the Bowtie 2 alignment program [11] from which SNPs can be called using the VarScan mpileup2snp variant caller program [12]. Using these SNPs, the QTL locus can then be scanned for polymorphisms that are present in the SNF[r] but not in the SNF[s] progeny. With the causal SNP identified in the coding region of a gene, genetic modification of the candidate SNF[r] gene can be performed in a SNF[s] strain to verify the drug resistance function.

The CRISPR/Cas9 genome editing system was recently established in *Toxoplasma*[13], which added important tools for the exploration of the complex biology of this parasite, particularly for genetic studies in non-laboratory adapted strains. Because of the highly active Non-Homologous End Joining (NHEJ) activity in WT *Toxoplasma* cells, targeted genome modification is difficult to achieve since exogenously introduced DNA is randomly integrated to the genome at extremely high frequency [14]. To increase the success rate of locus specific modification, different approaches have been employed to increase the efficiency of homologous recombination and/or decrease the NHEJ activity [15,16]. One of these approaches is the CRISPR/Cas9 system. Compared to other methods, the CRISPR/Cas9 system is efficient at introducing site-specific modifications and is easy to design [13,17,18]. In addition, it can be used in any *Toxoplasma* strain without extra modification to the parasite [13,19].

The CRISPR/Cas9 system originated from the adaptive immune system of *Streptococcus pyogenes*, which uses it to defend the invasion of mobile genetic elements such as phages [20,21,22]. This system utilizes the RNA-guided DNA endonuclease enzyme Cas9 to introduce double-strand DNA break (DSB) into the target, which is subsequently repaired either by the error-prone NHEJ to inactivate target genes through short indel mutations, or by homology directed recombination to alter the target locus exactly as designed [23,24]. The target specificity is determined by a small RNA molecule named single guide RNA (gRNA), which contains an individually designed 20 nt sequence that has 100% homology to the target DNA [22]. The gRNA molecule also contains signatures recognized by Cas9 that guide the nuclease to the target site, which includes a special Protospacer Adjacent Motif (PAM, sequence is 'NGG') [25,26]. Therefore, the gRNA molecule and the PAM sequence work together to determine the Cas9 cleavage site in the genome. One can easily change the gRNA sequence to target different sites for cleavage.

When the CRISPR/Cas9 system was first developed in *Toxoplasma*, a single plasmid expressing the Cas9 nuclease and the gRNA molecule was used to introduce DSB at the targeting site [13,17]. It has been shown that the CRISPR/Cas9 system drastically increases the efficiency of site-specific genome modification, not only by homologous recombination, but also non-homologous integration of exogenous DNA [13]. It does this in wildtype strains that contain NHEJ activity. Therefore, this system can be used in essentially any *Toxoplasma* strain for efficient genome editing. In a typical experiment, the target specific CRISPR plasmid and the DNA fragment used to modify the target are co-transfected into parasites. If the DNA fragment used to modify the target contains homologous sequences to the target locus, homologous recombination can be used to repair the DSB introduced by CRISPR/Cas9 to allow precise modification of the target. On the other hand, if the introduced DNA fragment does not contain homologous sequence, it can still be integrated into the CRISPR/Cas9 targeting site. The latter is often used to disrupt genes by insertion of selectable markers or to complement mutants at loci that allow negative selection [13]. Here the *SNR1* locus serves as an example to show how CRISPR/Cas9 can be used for gene disruption and the generation of transgenic parasites.

Protocol

1. Assess SNF[r] in the Progeny of the ME49-FUDR[r] x VAND-SNF[r] Cross

NOTE: *T. gondii* is an obligate intracellular parasite and the tachyzoite stage readily grows in tissue culture.

1. To grow the *T. gondii* parasite, maintain them in confluent Human Foreskin Fibroblast (HFF) monolayer culture seeded to T25 flasks with Dulbecco's Modified Eagle's Medium (DMEM) media supplemented with 10% Fetal Bovine Serum (FBS) (D10 media) at 37 °C and 5% CO_2. NOTE: See Protocol 1.1 Note in **Supplementary File 1**.
2. To test individual progeny clones for sinefungin resistance, grow each clone to high parasite density in a T25 HFF culture flask for 2-3 d, until majority of the parasites begin to lyse the host cells.
3. Scrape the monolayer with a cell scraper and pass the parasite solution through a 10 mL syringe/22 G blunt needle 2-3x to lyse the host cells releasing parasites. The solution can be pulled back up into the syringe for multiple needle passages.
4. Filter the parasite solution into a new conical tube through a membrane with the pore size of 3 µm to remove HFF cells and cellular debris.
5. Count the parasites on a hemocytometer and determine the number of parasites per mL.
6. Using a pipettor, pass 2.5×10^5 parasites to a new T25 HFF culture flask containing D10 media with 0.3 µM working concentration of the sinefungin drug.
7. Grow the parasites at 37 °C and 5% CO_2 for 7-10 d.
8. Score the progeny for the growth phenotype in sinefungin drug. Observe the monolayer under an inverted phase-contrast microscope and score 0 for no growth (sinefungin sensitive), score 1 for growth and lysis of monolayer (sinefungin resistant). NOTE: See Protocol 1.8 Note in **Supplementary File 1**.

2. Run a QTL Scan of the SNF[r] Phenotype in R/qtl

NOTE: See Protocol 2 Note in **Supplementary File 1**.

1. Install the R programming language software on a local computer. See [27].
2. Run R and install the package 'qtl'. Either do this from the R GUI interface option 'Packages->Install package(s)', or run the following command from the R command line (the first ">" symbol in each example denotes the beginning of a command, not to be copied):
    >install.packages("qtl")
    NOTE: Once R and the 'qtl' package have been installed, there are two ways to run R/qtl, from the R command line or using a graphical user interface (GUI) program called J/qtl [28] : see [29] to download J/qtl. This protocol will provide the R/qtl command line syntax with occasional reference to the equivalent function in J/qtl.
3. Load the 'qtl' package:
    > library(qtl)
    NOTE: See Protocol 2.3 Note in **Supplementary File 1** for dataset format and downloads.
4. **Load the dataset file into R/qtl.**
    1. For "csv" format (See: Data File 1):
        >SNFR <- read.cross(format="csv", file="$PATH/Rqtl-SNFR.csv", genotypes=c("0", "1"), na.strings=c("-"), convertXdata=FALSE)
    2. Or for "gary" format (See: Data File 2):
        >SNFR <- read.cross(format="gary", dir="$PATH", chridfile="chrid.txt", mnamesfile="mname.txt", mapfile="markerpos.txt", genfile="genotype.txt", phefile="phenos.txt", pnamesfile="phenonames.txt", convertXdata=FALSE)
        Where $PATH is the directory path to the file(s) containing the qtl dataset. For example: /Users/HotDiggityDog/QTLfiles
        NOTE: The files can also be loaded into J/qtl with the previous commands using the 'Insert Comment or Command' option, or load the data with the GUI 'File->Load Cross Data' option.

5. Calculate probabilities for the map:
    >SNFR <- calc.genoprob(SNFR, step=2.0, off.end=0.0, error.prob=1.0E-4, map.function="haldane", stepwidth="fixed")
6. Run a single scan using a binary distribution of the sinefungin resistance phenotype across all chromosomes:
    > SNFR.scan <- scanone(cross=SNFR, chr=c("Ia", "Ib", "II", "III", "IV", "V", "VI", "VIIa", "VIIb", "VIII", "IX", "X", "XI", "XII"), pheno.col=c(1), model="binary", method="em")
    NOTE: If running the VIR phenotype use the 'model="normal"' distribution option.
7. Run 1000 permutations to calculate significance thresholds and then attribute the permutations to the SNFR.scan variable:
    > SNFR.scan.permutations <- scanone(cross=SNFR, chr=c("Ia", "Ib", "II", "III", "IV", "V", "VI", "VIIa", "VIIb", "VIII", "IX", "X", "XI", "XII"), pheno.col=c(1), model="binary", method="em", n.perm=1000, perm.Xsp=FALSE, verbose=FALSE)
    >attr(SNFR.scan, "pheno.col") <- c(1)
    NOTE: Steps 2.5 through 2.7 can be run in J/qtl with the 'Analysis->Main Scan-> Run One QTL Genome Scan' option.
8. **Plot the scan results:**
    >plot(SNFR.scan, gap=0, bandcol="grey")
    NOTE: The plot produced in J/qtl is interactive.
    1. Plot the scan results for just chromosome IX, the chromosome with the highest peak:
        >plot(SNFR.scan, chr=c("IX"), gap=0, bandcol="grey", show.marker.names=TRUE)
9. Show all the marker positions and LOD scores from the single scan:
    >SNFR.scan
10. Show the threshold results of the permutation test:
    >summary(SNFR.scan.permutations)
11. Show only those markers greater than the alpha=.05 threshold value (taken from the previous output):
    >SNFR.scan[SNFR.scan$lod >2.68,]
12. Show the markers with the highest LOD score on each chromosome:
    >summary(SNFR.scan)
13. Show those markers that have the max LOD score value (taken from the previous output):
    >SNFR.scan[SNFR.scan$lod >=6.57,]
    NOTE: See Protocol 2.13 Note in **Supplementary File 1**.

3. Identify the Causal SNF[r] Mutation using WGS Reads from the Progeny

NOTE: See Protocol 3 Note in **Supplementary File 1**.

1. **Align WGS reads of individual progeny to the sinefungin sensitive VAND (VAND-SNF[s]) parental genome.**
    NOTE: Download the VAND reference genome (SNF[s]) from NCBI Assembly AEYJ00000000.2 and the WGS reads (.sra files) for the 24 progeny from NCBI SRA PRJNA258152. Also, download the following programs: Bowtie2 [11], NCBI SRA Toolkit, SAMtools [30], VarScan [12], and MUMmer [31].
    1. Create a Bowtie 2 compatible index from the VAND reference genome FASTA file using the bowtie2-build program, here naming the index "VAND":
        >bowtie2-build GCA_000224845.2_TGVAND_v2_genomic.fna VAND
    2. Convert the SRA formatted WGS read files to FASTQ files using fastq-dump with the --split-files option for paired reads (the SRR1555372.sra file for progeny P1_14VB will be used as an example):
        >fastq-dump --split-files SRR1555372.sra
        NOTE: Run this and all the following commands in Protocol 3.1 for all 24 progeny.
    3. Align the WGS reads to the reference genome using bowtie2 with output to a SAM (.sam) file and the --end-to-end option:

```
>bowtie2 -x VAND -1 SRR155372_1.fastq -2 SRR1555372_2.fastq -S SRR155372.sam --end-to-end
```

4. Convert the .sam file to a BAM (.bam) file:
```
>samtools view -bS SRR155372.sam > SRR155372.bam
```
5. Sort the .bam file:
```
>samtools sort SRR155372.bam SRR155372.sort
```
6. Index the sorted .bam file:
```
>samtools index SRR155372.sort.bam
```

2. **Call SNPs for the progeny using VarScan**
   NOTE: See Protocol 3.2 Note in **Supplementary File 1**.
   1. Convert all indexed BAM (.sort.bam) files into one pileup formatted file using samtools mpileup with the VAND genome reference file, all progeny .sort.bam files, and output to a file named AllProgeny-mpileup.txt:
   ```
   >samtools mpileup -f GCA_000224845.2_TGVAND_v2_genomic.fna SRR1555599.sort.bam SRR1555372.sort.bam
   SRR1555660.sort.bam SRR1555672.sort.bam SRR1556052.sort.bam SRR1556122.sort.bam SRR1556192.sort.bam
   SRR1556194.sort.bam SRR1556195.sort.bam SRR1556199.sort.bam SRR1556200.sort.bam SRR1556202.sort.bam
   SRR1556203.sort.bam SRR1556274.sort.bam SRR1556276.sort.bam SRR1556278.sort.bam SRR1556395.sort.bam
   SRR1556396.sort.bam SRR1556397.sort.bam SRR1556398.sort.bam SRR1556399.sort.bam SRR1556400.sort.bam
   SRR1556401.sort.bam SRR1556402.sort.bam > AllProgeny-mpileup.txt
   ```
   2. Call all SNPs using the VarScan mpileup2snp program using the AllProgeny-mpileup.txt file, minimum read coverage of 5, minimum variant frequency across reads of .8, a p-value of .01, and output to a file names AllProgeny-SNPs.txt:
   ```
   >java -jar VarScan.jar mpileup2snp AllProgeny-mpileup.txt --min-coverage 5 --min-var-freq .8 --p-value .01 > AllProgeny-SNPs.txt
   ```
   NOTE: The AllProgeny-SNPs.txt file is a tab delimited text file that will be used in Protocol 3.4 to locate the causal SNP.

3. **Find VAND genome coordinates that correspond to coordinates of the ME49 based QTL locus**
   NOTE: See Protocol 3.3 Note in **Supplementary File 1**.
   1. Use the MUMmer nucmer program to align the ME49 genome (available for download from ToxoDB.org [32]) and the VAND reference genome file (output goes to the out.delta file):
   ```
   >nucmer ToxoDB-28_TgondiiME49_Genome.fasta GCA_000224845.2_TGVAND_v2_genomic.fna
   ```
   2. Use the MUMmer show-coords program to obtain the coordinates of the two genome alignments, output to a file named ME49vsVAND-coords.txt:
   ```
   >show-coords out.delta > ME49vsVAND-coords.txt
   ```
   NOTE: The ME49vsVAND-coords.txt file can be used to find the VAND contigs and positions that correspond to the QTL locus; markers MV359 to MV366 located on ME49 chromosome IX between 3,187,537 to 4,202,258 bp. There are two VAND contigs that span the corresponding QTL locus; KN044604.1:657,441-1 bp and KN042501.1:430,910-41,870 bp (both contigs align reverse to the ME49 chromosome).

4. **Assess progeny SNPs located within the QTL locus for SNPs inherited only in the SNF^r progeny**
   NOTE: See Protocol 3.4 Note in **Supplementary File 1**.
   1. Review Data File 3 to identify the causal mutation located within the QTL locus. Scan the data for a pattern where the SNF[r] progeny have a SNP and the SNF[s] do not (feasible because progeny SNPs were obtained by comparison to the VAND-SNF[s] reference genome). This should only result in one position with this pattern, position 348130 on VAND contig KN042501.1.
   2. See row 6604 in Data File 3 (**Figure 3**).
   NOTE: See Protocol 3.4.2 Note in **Supplementary File 1**.

4. Verification of Identified Hits by CRISPR/Cas9-mediated Gene Inactivation.

NOTE: To confirm the causal SNP identified by QTL mapping and WGS SNP analysis, the corresponding genetic changes need to be made in a WT SNF[s] background and the resulting phenotype examined. In the case of sinefungin resistance, the responsible mutation results in inactivation of the *SNR1* gene by early termination [2]. Therefore, disruption of *SNR1* can be used for confirmation. Here a detailed protocol is provided for using CRISPR/Cas9 induced indel mutations to disrupt *SNR1*, to demonstrate its involvement in sinefungin resistance (**Figure 4**).

1. *SNR1* **specific CRISPR plasmid construction**
   NOTE: See Protocol 4.1 Note in **Supplementary File 1** to obtain plasmids and maps.
   1. Obtain the genomic sequence for the target gene (*SNR1*, TGME49_290860) from ToxoDB [32].
   2. Use online tools such as E-CRISP [33] to design target specific gRNAs. Do not include the PAM sequence (NGG) in the gRNA.
   NOTE: See Protocol 4.1.2 Note in **Supplementary File 1**.
   3. Synthesize primers to construct the target specific CRISPR plasmid.
      1. According to the design from step 4.1.2, synthesize two primers to change the *UPRT* targeting gRNA in the original CRISPR plasmid to the selected gRNA sequence by site-directed mutagenesis.
      NOTE: The sequences of these two primers are as follow: gRNA-Fw: NNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAGC (N20 is the target specific gRNA sequence) and gRNA-Rv: AACTTGACATCCCCATTTAC [2].
   4. Perform the site-directed mutagenesis reactions.
      1. Using the *UPRT* targeting CRISPR plasmid (pSAG1::CAS9-U6::sgUPRT) as template and the two primers listed above, perform site-directed mutagenesis reactions according to the manufacturer's instructions.
      NOTE: Use the following cycling conditions for PCR: 98 °C for 1 min, followed by 25 cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 5 min, followed by final extension for 2 min at 72 °C.

5. Transform the mutagenesis products containing the GOI-specific CRISPR plasmids to *E. coli* competent cells by chemical transformation according to the supplier's protocol (see **Tables of Materials**). Grow the transformants on lysogeny broth (LB) plates containing ampicillin (100 µg/mL) at 37 °C. Subsequently, pick 2-4 clones and individually grow them in 5 mL LB medium supplemented with 100 µg/mL ampicillin for 12-16 h.
    1. Extract plasmids from the cultures using a DNA isolation kit and analyze them by DNA gel electrophoresis to check the size of the plasmids (expected plasmid size is 9674 bp, use the original CRISPR plasmid as control). Use the M13-Rev primer (5'-CAGGAAACAGCTATGACC) to sequence the plasmids to confirm the target specific gRNA sequence.
        1. Once positive clones are identified, store both plasmid DNA at -20 °C and corresponding bacterial culture at -80 °C for future use.
        NOTE: The CRISPR plasmid to be used for transfection should be dissolved in deionized water and the concentration determined by spectrophotometry or an alternative method.

2. **Parasite transfection.**
   NOTE: General methods for culturing *T. gondii* in HFF cells in D10 medium have been described previously [34].
    1. Two to three days before transfection, add enough parasites to a T25 flask containing a confluent HFF cell monolayer (0.5-1 x 10^6 for type 1 strains, 1-2 x 10^6 for other strains) to achieve 70 - 80% HFF cell lysis within 2-3 d.
    2. Examine the culture under an inverted phase contrast microscope, when the HFF monolayer is 70-80% lysed by the parasites. Gently remove the medium with a pipette and wash the cells off the flask surface with 5 mL cytomix buffer (120 mM KCl, 0.15 mM $CaCl_2$, 10 mM K_2HPO_4/KH_2PO_4 pH = 7.6, 25 mM HEPES, 2 mM EDTA, 5 mM $MgCl_2$, pH = 7.6). Subsequently, transfer the parasite solution to a 15-mL conical tube.
       NOTE: See Protocol 4.2.2 Note in **Supplementary File 1**.
    3. Pass the parasite solution through a 10 mL syringe/22 G blunt needle 2-3x.
    4. Filter the parasite solution into a new conical tube through a membrane with the pore size of 3 µm to remove HFF cells and cellular debris.
    5. Pellet the filtered parasites by centrifugation at 400 x g for 10 min.
    6. Pour off the supernatant and resuspend the pelleted parasites with 10 mL cytomix buffer, take a 10 µL aliquot to determine the parasite concentration with a hemocytometer and pellet the rest by centrifugation at 400 × g for 10 min.
    7. Remove supernatant and resuspend the pellet in cytomix buffer to obtain a density of 4 x 10^7 parasites/mL.
    8. In a 4 mm gap cuvette, mix 250-300 µL parasite solution (1-1.3 x 10^7 parasites) with 7.5 µg CRISPR plasmid, 6 µL ATP (100mM), and 6 µL glutathione (GSH) (250 mM).
    9. Electroporate the parasites [35]. Include a separate electroporation as a negative control in which a CRISPR plasmid targets elsewhere, such as the *UPRT* targeting CRIPSR plasmid.
       NOTE: The electroporation settings depend on the device. For the electroporator listed in the Materials use the 4 mm gap cuvettes. The following protocol is recommended: 1,700 V, 176 µs of pulse length, 2 pulses with 100 ms interval.

3. **Determine the frequency of sinefungin resistance after CRISPR targeting.**
    1. Seed HFF cells to coverslips placed in 24-well plates 3-4 d before electroporation so that they are confluent by the time to perform transfection in 4.2.
        1. Trypsinize the HFF cells from one confluent T25 flask and subsequently add 12 mL D10 medium and mix well.
        2. Add a coverslip to each well of a 24-well plate and aliquot 500 µL HFF cell solution to each well. Incubate the plate at 37 °C for 3 4 d to let the cells grow until confluent.
           NOTE: Cells from one T25 flask are enough to seed one 24-well plate.
        3. Add 50 µL of electroporated parasite solution from step 4.2.9 into one well of 24-well plate containing a coverslip seeded with confluent HFF cells. Transfer the rest of electroporated parasite solution into a T25 flask seeded with confluent HFF cells.

    2. Assess the efficiency of transfection.
        1. Grow the cells in the 24-well plate for 24 h and subsequently fix the cells (host cells and parasites) with 500 µl 4% formaldehyde to check the expression of Cas9-GFP by immunofluorescent assay (IFA).
        2. Probe the cells with an anti-GFP antibody and a *Toxoplasma* specific antibody (such as anti-TgALD) to label total parasites. See the antibody product sheet for the recommended dilution (usually, 1:1,000 is sufficient for IFAs). If the antibodies are unconjugated, use two different secondary antibodies conjugated with fluorescent dyes to label the primary antibodies.
        3. Observe the labeled cells on a fluorescent microscope with filters appropriate for the secondary fluorescent dyes. Obtain transfection efficiency by dividing the number of GFP positive parasites by the number of total parasites.
           NOTE: The two antibodies used for IFA should come from two different host species, for example using the combination of mouse anti-GFP and rabbit anti-TgALD.

    3. Determine the frequency of sinefungin resistance in transfected parasites.
        1. For the transfected parasites cultured in T25 flasks, grow them for 2-3 d until natural egress. Then collect the parasites, blunt needle lyse host cells to release intracellular parasites, and purify them by filtration through membranes with pore size of 3 µm. Count the parasites with a hemocytometer to estimate the density.
        2. Add purified parasites into 6-well plates seeded with confluent HFF cells: for the first row (3 wells), add 200 parasites/well and grow them in regular D10 medium (2 mL/well); for the second row, add 5000 parasites/well and grow them in D10 medium containing 0.3 µM sinefungin.
        3. Put the plates in a 5% CO_2 incubator and grow the parasites for at 37 °C 8-10 d without disturbance to allow plaques to form. Fix the samples with 70% ethanol (2 mL/well) and stain the monolayer with 0.1% crystal violet (2 mL/well). Wash the wells with water to visualize the plaques (clear zones) formed by parasite growth.
           NOTE: The negative control should be processed the same way side by side.

    4. Calculate the rate of CRISPR induced sinefungin resistance.

June 2017 | 124 | e55185 | Page 5 of 12

1. Obtain the average number (X) of plaques from wells containing no drug and calculate the parasite viability as X/200. Obtain the average number (Y) of plaques from wells containing sinefungin to calculate the rate of CRISPR induced sinefungin resistance as follow:

   Rate of CRISPR induced resistance = 200×Y/ (5000 × X × transfection efficiency)

   NOTE: Transfection efficiency is obtained from 4.3.2.

4. **Examine the indel mutations induced by CRISPR/Cas9**
   1. Grow the electroporated parasites from section 4.2.9 in a T25 flask seeded with HFF cells for 2 d at 37 °C. Then, replace the medium with 5 mL D10 medium containing 0.3 µM sinefungin (selection medium).
      1. Keep the parasites in selection medium for at least 3 passages until the resistant pool becomes stable (indicated by the absence of parasite growth in the negative control group but robust parasite growth in the experimental group).

   2. Subclone the sinefungin resistant pool to obtain clonal strains.
      1. Collect freshly egressed parasites, purify by 3 µm membrane filtration and subclone into 96-well plates seeded with confluent HFF cells in 150 µL D10 medium. Grow the subcloning cultures in a CO_2 incubator at 37 °C for 7 - 10 d without disturbing the plates.

         NOTE: See Protocol 4.4.2.1 Note in S**upplementary File 1**.

   3. Check the 96-well plates under an inverted phase-contrast microscope to look for wells that contain only one plaque. Mark such wells and subsequently transfer the cells from each well to 24-well plates seeded with HFF cells using a pipette.
   4. When 80-90% of the HFF cells in a well are lysed, harvest the parasites (about 500 µL) and pass 50 µL parasite solution to a new well to maintain the strain. Use the rest (≈450 µL) to extract genomic DNA for PCR amplification.
      1. For genomic DNA isolation from the SNF[r] clone, pellet the parasites (small amounts of HFF cell contamination is tolerable) at 1,000 x g for 10 min. Wash the pelleted parasites with phosphate-buffered saline (PBS) once and pellet them again. Isolate genomic DNA from pelleted cells using a commercial kit or by boiling. Resuspend parasites in 50-100 µL PBS.

         NOTE: Perform PCR to obtain a fragment of the *SNR1* gene for sequencing. Use the following primers to amplify the SNR1 locus[2]: SNR1-Amp-Fw: 5' CCGACCACAA CAATTTTC and SNR1-Amp-Rv: 5'GACGTGATTCACTTTTTTACAGACAGAC. Use high fidelity DNA polymerases. Amplify the WT locus for control purposes.

   5. Perform the PCR with 20 µL reaction volume and run it with the following program: 98 °C for 1 min, followed by 30 cycles of 98 °C for 10 s, 55 °C for 30 s and 72 °C for 2.5 min, followed by final extension for 2 min at 72 °C.

      NOTE: Sequence the PCR products using the following primers: SNR1-Seq1: 5' GCC ACA TGC TTT AGC GTG, SNR1-Seq2: 5' TCC TCT CCA TCA CGG GTT GG, SNR1-Seq3: 5' GCA AGA GCC GCG TGA CG, SNR1-Seq4: 5' CTC TCC CGC GGT CGA G, SNR1-Seq5: 5' GCA CCG TCC GCA AGC, and SNR1-Seq6: 5' CCG GAA GGT GAA TCG TTC TTC.
   6. Compare the sequences of *SNR1* gene from sinefungin resistant mutants to that of the WT strain using a sequence alignment tool.

      NOTE: See **Supplementary File 2** Protocol 5 for details on the utilization of negative selection at the *SNR1* locus for genetic complementation or transgenic strain construction.
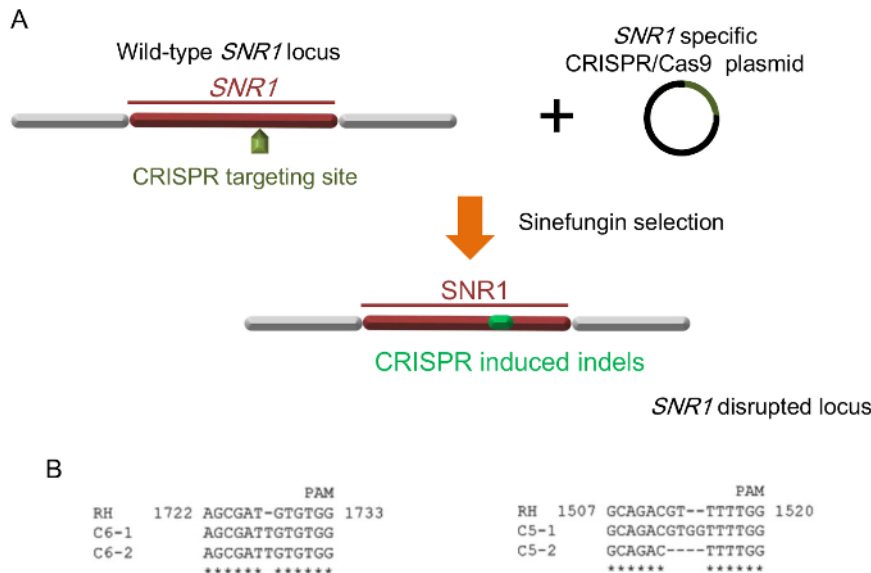
Representative Results

This article outlines in detail several methods that can be used in succession to identify a gene responsible for drug resistance (**Figure 1**). The 24 progeny of the ME49-FUDR[r] X VAND-SNF[r] cross were assessed for resistance to the drug sinefungin as described in Protocol 1. Using the genetic map and the SNF[r] phenotype of the progeny, a QTL scan was run in R/qtl - Protocol 2 (**Figure 2**). This resulted in one significant peak on chromosome IX spanning approximately 1 Mbp. It is in this region that the causal mutation is located.

To identify the causal mutation, WGS reads from the progeny were aligned to the VAND-SNF[s] reference genome using Bowtie2 - Protocol 3.1, SNPs were called using VarScan mpileup2snp - Protocol 3.2, and the QTL locus in the VAND genome was identified using MUMmer - Protocol 3.3. Progeny SNPs within the QTL locus were extracted and scanned for a pattern where the SNF[r] progeny have a SNP and the SNF[s] do not, which is feasible because progeny SNPs were obtained by comparison to the VAND-SNF[s] reference genome (**Figure 3**). Only one SNP matched this pattern that results in an early stop codon in a putative amino acid transporter gene named *SNR1*.

Confirmation that *SNR1* is the SNF[r] resistance gene was performed using the CRISPR/Cas9 system. A new CRISPR/Cas9 plasmid engineered for *T. gondii* gene editing was made containing a gRNA targeting the *SNR1* gene near the SNF[r] mutation identified in the progeny - Protocol 4 (**Figure 4A**). The *SNR1* targeting CRISPR/Cas9 plasmid was electroporated into a SNF[s] WT parasite strain and resistant mutants were obtained when cultured in 0.3 µM sinefungin. No SNF[r] parasites were obtained when electroporated with the UPRT targeting CRISPR/Cas9 plasmid. Several SNF[r] CRISPR mutants were cloned and the region around the *SNR1* gRNA target was sequenced. Each mutant had an indel that disrupted the coding sequence of the *SNR1* gene (**Figure 4B**). This method can also be used to insert a targeting construct into the *SNR1* locus via NHEJ (**Figure 5),** or by HR (**Figure 6**) - **Supplemental File 2**.

Figure 1: Schematic Workflow for Experiments Outlined in the Protocols. The main steps in identifying and confirming the SNF[r] gene using the ME49-FUDR[r] X VAND-SNF[r] cross are shown with reference to the appropriate Protocols outlined in this article. Please click here to view a larger version of this figure.

Figure 2: R/qtl Commands to run QTL Scan on the SNF[r] Phenotype. The commands as outlined in Protocol 2 were run in R using the qtl package. Representative commands and plot are shown. Please click here to view a larger version of this figure.

Figure 3: Location of the Causal SNP. Progeny WGS reads were aligned to the VAND-SNF[s] reference genome with Bowtie 2, SNPs were called with VarScan, and the corresponding VAND QTL locus identified with MUMmer. SNPs located within the QTL locus were imported into a spreadsheet and the causal SNP was identified. SNF[r] progeny (yellow), SNF[s] progeny (green), SNF[r] SNP (red) (see **Data File 3**). Please click here to view a larger version of this figure.

Figure 4: Confirming *SNR1* as SNF[r] Gene using CRISPR/Cas9. (A) CRISPR/Cas9 induced indel mutations (green) in the *SNR1* locus. **(B)** Representative indels in *SNR1* caused by two different *SNR1* specific CRISPR plasmids (C5 and C6 respectively). RH is the WT strain and C5 and C6 are SNF[r] mutants. (B is taken from reference [2]). Please click here to view a larger version of this figure.
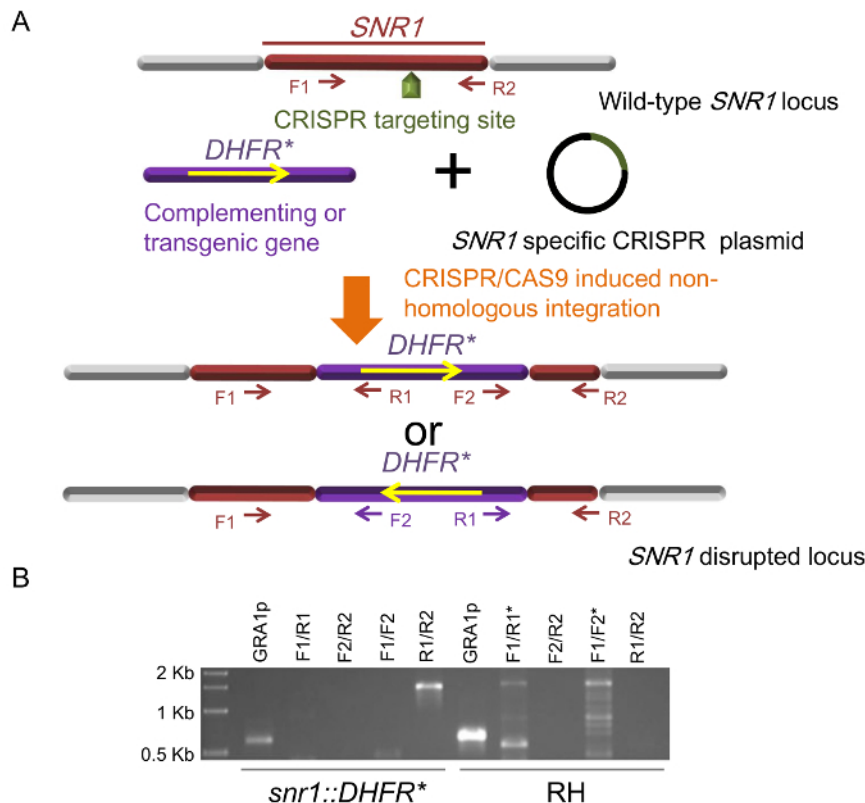
Figure 5. Insertional Mutagenesis using CRISPR/Cas9. (A) Insertion of complementing or transgenic genes into the *SNR1* locus by CRISPR/Cas9 mediated site-specific integration. Two possible orientations of the inserted *DHFR** mini gene and the primers used for identification, F1/2 and R1/2 represent the priming sites of oligos used in diagnostic PCRs. **(B)** Diagnostic PCRs of one *snr1::DHFR** clone, RH is used as a WT control. PCR of the GRA1p is included as a control checking the quality of genomic DNA as templates. Asterisks indicate lanes with unspecific bands (Figure 5B is taken from reference [2]). Please click here to view a larger version of this figure.
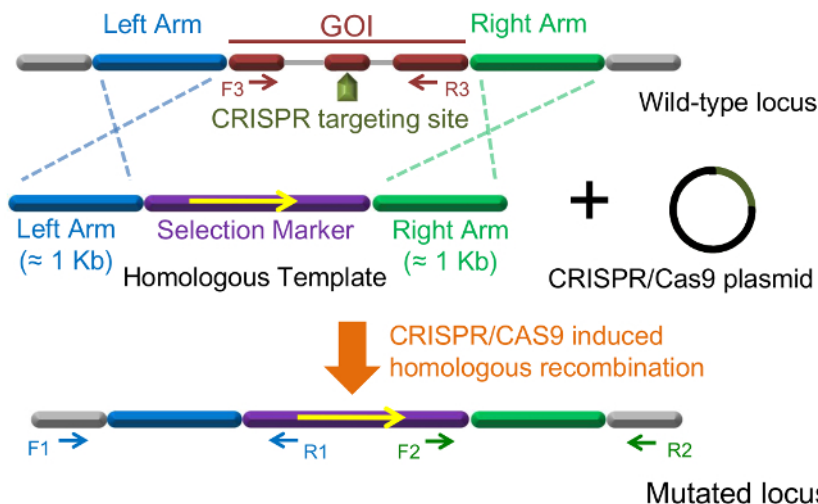
Figure 6: Gene Knockout using CRISPR/Cas9. A classic design for CRISPR/Cas9 mediated homologous gene replacement in *T. gondii*. The orientation of inserted transgene is fixed in this case. F1/R1,F2/R2 and F3/R3 represent the priming sites of oligos used in diagnostic PCRs. Please click here to view a larger version of this figure.

Data File 1: ME49-FUDR^r X VAND-SNF^r Cross File for use in R/qtl, "csv" Format. One .csv file that can be loaded into R/qtl with the format="csv" option. Please click here to download this file.

Data File 2: (chrid txt, genotype txt, markerpos.txt, mname.txt, phenonames.txt, and phenos.txt). ME49-FUDR^r X VAND-SNF^r Cross Files for use in R/qtl, "gary" Format. Six .txt files that can be loaded in R/qtl with the format="gary" option. Please click here to download this file.

Data File 3. Spreadsheet with Progeny SNPs across the SNF^r QTL Locus. Contains one worksheet with the progeny SNPs, and a second worksheet with the phenotypic data of the parents and progeny of the cross. Please click here to download this file.

Supplemental File 1: Additional Protocol Details. Please click here to download this file.

Supplemental File 2: Protocol 5 - Utilization of Negative Selection at the *SNR1* Locus for Genetic Complementation or Transgenic Strain Construction. Please click here to download this file.

Discussion

These protocols present several methods that when combined allow for the identification of a drug resistance gene in *T. gondii*. Two in particular were integral to the project, the relatively seasoned method of QTL mapping and the recently developed method of CRISPR/Cas9 gene editing. Lander and Botstein published their influential paper in 1989 demonstrating QTL mapping which correlates genetic loci with phenotypes [36]. More recently in 2012, Dounda and Charpentier described the CRISPR/Cas9 editing system in *Streptococcus pyogenes* [22] that was quickly adapted as a genetic tool in many different models, including *T. gondii*[13,17]. Both methods were useful here, where QTL mapping defined the locus containing the drug resistance mutation that was ultimately identified using WGS based SNP detection, and CRISPR/Cas9 editing provided the means to confirm *SNR1* is the sinefungin drug resistance gene.

The ME49-FUDR^r X VAND-SNF^r cross [8] was originally developed to interrogate a virulence phenotype [19], but the parental strains also happened to have an additional phenotypic difference for which the causal gene was not known, sinefungin resistance in the VAND parent. This highlights one benefit of crosses in that they can be repurposed when additional phenotypic differences in the parents are observed. This was the case for another *Toxoplasma* cross, type 2 x type 3, where several genes involved in virulence were found by mapping multiple phenotypes [37,38,39,40]. To date, four different *T. gondii* crosses have been described and used to map genes responsible for phenotypes [8,37,41,42], all of which have the potential to be reused to map new phenotypes for which the genetic basis is unknown. Along these lines, the genomes for 62 *T. gondii* strains representing the known global genetic diversity have been sequenced [43]. New crosses could be made from this pool for strains that differ in interesting phenotypes. Having extolled the advantages of QTL mapping, it needs to be said that generating a cross isn't a minor undertaking. There are other methods that can be used to identify causal genes. One powerful technique uses chemical mutagenesis to create mutants that can be screened for phenotypes. To find the causal gene, mutants can either be complemented with cosmid libraries [44] or genome resequencing methods can be used to find the causal mutation [45]. For more on this, see two JoVE articles by Coleman *et al.* and Walwyn *et al.* that outline these approaches [46,47].

Many of the steps leading to the identification of the causal SNF^r mutation (Protocols 2 and 3) rely on computational methods conducted with software that is freely available for academic use. Detailed commands for each step are provided and when run with the proper files will allow the user to recreate the datasets necessary to find the causal SNF^r SNP. Keep in mind that some of the syntax in the commands refer to filenames or directory structure ($PATH) that can be modified to the user preference. Although the commands given here certainly do not exhaust the ways one can analyze a cross with QTL analysis and WGS based SNP identification, they are comprehensive enough to repeat the experiment described in this article and should allow the user to become more familiar with how these approaches are utilized in a step by step fashion.

Although QTL mapping and WGS sequence based SNP detection were sufficient to identify a candidate SNF[r] gene, additional experiments are needed to confirm its role in drug resistance. This can be convincingly shown through gene disruption or knockout techniques, both of which can be achieved using CRISPR/Cas9 gene editing. Detailed methods for using CRISPR/Cas9 to either generate indel mutations or insert transgenic constructs into a target gene in *T. gondii* are given. The targeting specificity which CRISPR/Cas9 provides increases the efficiency of gene editing over traditional methods. Also, this increased efficiently has made it possible to use non-laboratory adapted strains for genetic studies which were previously difficult to modify [13]. Even though in its infancy, CRISPR/Cas9 has already been used to make gene disruptions [2,13,17,18,48,49], tag genes [17], and make gene knockouts [16,19,50,51,52,53,54] in *T. gondii*, promising to be a useful tool for many other studies in the future.

The discovery that *SNR1* inactivation leads to sinefungin resistance makes the *SNR1* locus a promising site for transgene insertion or genetic complementation. To maximize the success of using CRISPR/Cas9 mediated gene targeting to direct the integration of transgene into the *SNR1* locus, the following aspects should be considered during the experimental design. First, a selection marker is recommended to be included in the transgenic construct to increase the efficiency of strain construction. If a selection marker is included, both positive and negative selection can be used, resulting in almost 100% of the doubly selected parasites being transgenic with the GOI integrated at the *SNR1* locus. In contrast, if the transgenic construct does not contain additional selectable traits and relies on negative selection at the *SNR1* locus, the efficiency of successful transgenesis largely depends on the efficiency of cotransfection of the CRISPR plasmid and the transgenic construct.

Second, although CRISPR/Cas9 mediated site-specific integration of a non-homologous DNA fragment is frequently used for complementation and transgenesis, it should be noted that the orientation of insertion cannot be guaranteed in such cases as either direction is possible. This may pose problems to some applications. For example, to complement a mutant with different gene alleles, it is difficult to ensure that all alleles are inserted in the same orientation, and discrepancies in orientation may cause expression differences. For this type of application, DNA constructs with sequences homologous to the *SNR1* locus are recommended, which will drive the proper integration orientation (**Figure 6**).

Third, during transfection, the ratio between the CRISPR plasmid and the transgenic DNA molecule is critical for successful transgenic strain construction. This ratio needs to be adjusted according to the selection strategies. The following guidelines are recommended: 1) If the transgenic construct contains a drug resistant marker and the corresponding drug is the only selection used to generate transgenic parasites, *i.e.* sinefungin is not used, the suggested molar ratio between the transgenic construct and the CRISPR plasmid is 1:5. Using more CRISPR plasmid in this case increases the likelihood that parasites receiving the transgenic construct will also receive the CRISPR plasmid, therefore the drug resistant parasites are more likely to have the marker inserted at the CRISPR targeting site. If the ratio is reversed, most parasites that receive the transgenic construct won't get the CRISPR plasmid. As a consequence, the vast majority of drug resistant parasites obtain the transgenic construct through random integration not associated with CRISPR/CAS9 mediated site-specific insertion. 2) If the transgenic construct contains a drug resistant marker and the corresponding drug is used along with sinefungin to select transgenic parasites, the suggested molar ratio between the transgenic construct and the CRISPR plasmid is 1:1. This strategy provides the highest efficiency of transgenic strain construction. 3) If the transgenic construct does not contain a selectable maker, relying on negative selection by sinefungin alone to obtain transgenic parasites, the suggested molar ratio between the transgenic construct and the CRISPR plasmid is 5:1. The rationale for this design is the same as in the first guideline above. Since both pyrimethamine and sinefungin were used for selection in Protocol 5, the ratio between *DHFR** mini gene and the *SNR1* targeting CRISPR plasmid was set as 1:1.

Taken together, the methods outlined here have a level of detail that was not possible to convey in the original publication that identified *SNR1* [2]. These protocols; specifically, the command line syntax, sequential layout of the programs utilized, and the use of CRISPR/Cas9 should aid future endeavors to identify new genes responsible for phenotypes.

Disclosures

The authors declare that they have no competing financial interests.

Acknowledgements

References

1. Dubey, J. P. The history of Toxoplasma gondii--the first 100 years. *J Eukaryot Microbiol.* **55** (6), 467-475 (2008).
2. Behnke, M. S., Khan, A., & Sibley, L. D. Genetic Mapping Reveals that Sinefungin Resistance in Toxoplasma gondii Is Controlled by a Putative Amino Acid Transporter Locus That Can Be Used as a Negative Selectable Marker. *Eukaryot Cell.* **14** (2), 140-148 (2015).
3. Dubey, J. P. History of the discovery of the life cycle of Toxoplasma gondii. *Int J Parasitol.* **39** (8), 877-882 (2009).
4. Pfefferkorn, L. C., & Pfefferkorn, E. R. Toxoplasma gondii: genetic recombination between drug resistant mutants. *Exp Parasitol.* **50** (3), 305-316 (1980).
5. Donald, R. G., & Roos, D. S. Insertional mutagenesis and marker rescue in a protozoan parasite: cloning of the uracil phosphoribosyltransferase locus from Toxoplasma gondii. *Proc Natl Acad Sci U S A.* **92** (12), 5749-5753 (1995).
6. Sullivan, W. J., Jr. *et al.* Insertional tagging of at least two loci associated with resistance to adenine arabinoside in Toxoplasma gondii, and cloning of the adenosine kinase locus. *Mol Biochem Parasitol.* **103** (1), 1-14 (1999).
7. Khan, A. *et al.* Composite genome map and recombination parameters derived from three archetypal lineages of Toxoplasma gondii. *Nucleic Acids Res.* **33** (9), 2980-2992 (2005).
8. Khan, A. *et al.* NextGen sequencing reveals short double crossovers contribute disproportionately to genetic diversity in Toxoplasma gondii. *BMC Genomics.* **15** (1), 1168 (2014).

9.   Shaik, J. S., Khan, A., Beverley, S. M., & Sibley, L. REDHORSE-REcombination and Double crossover detection in Haploid Organisms using next-geneRation SEquencing data. *BMC Genomics.* **16** (1), 133 (2015).

10. Arends, D., Prins, P., Jansen, R. C., & Broman, K. W. R/qtl: high-throughput multiple QTL mapping. *Bioinformatics.* **26** (23), 2990-2992 (2010).

11. Langmead, B., & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* **9** (4), 357-359 (2012).

12. Koboldt, D. C. *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22** (3), 568-576 (2012).

13. Shen, B., Brown, K. M., Lee, T. D., & Sibley, L. D. Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9. *MBio.* **5** (3), e01114-01114 (2014).

14. Fox, B. A., Ristuccia, J. G., Gigley, J. P., & Bzik, D. J. Efficient gene replacements in Toxoplasma gondii strains deficient for nonhomologous end joining. *Eukaryot Cell.* **8** (4), 520-529 (2009).

15. Huynh, M. H., & Carruthers, V. B. Tagging of endogenous genes in a Toxoplasma gondii strain lacking Ku80. *Eukaryot Cell.* **8** (4), 530-539 (2009).

16. Wang, J. L. *et al.* The Past, Present, and Future of Genetic Manipulation in Toxoplasma gondii. *Trends Parasitol.* (2016).

17. Sidik, S. M., Hackett, C. G., Tran, F., Westwood, N. J., & Lourido, S. Efficient genome engineering of Toxoplasma gondii using CRISPR/Cas9. *PLoS One.* **9** (6), e100450 (2014).

18. Sugi, T., Kato, K., & Weiss, L. M. An improved method for introducing site-directed point mutation into the Toxoplasma gondii genome using CRISPR/Cas9. *Parasitol Int.* (2016).

19. Behnke, M. S. *et al.* Rhoptry Proteins ROP5 and ROP18 Are Major Murine Virulence Factors in Genetically Divergent South American Strains of Toxoplasma gondii. *PLoS Genet.* **11** (8), e1005434 (2015).

20. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* **315** (5819), 1709-1712 (2007).

21. Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature.* **468** (7320), 67-71 (2010).

22. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337** (6096), 816-821 (2012).

23. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science.* **339** (6121), 819-823 (2013).

24. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science.* **339** (6121), 823-826 (2013).

25. Jiang, F., Zhou, K., Ma, L., Gressel, S., & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science.* **348** (6242), 1477-1481 (2015).

26. Nishimasu, H. *et al.* Crystal Structure of Staphylococcus aureus Cas9. *Cell.* **162** (5), 1113-1126 (2015).

27. The R Project for Statistical Computing. *https://www.r-project.org*. (2016).

28. Smith, R., Sheppard, K., DiPetrillo, K., & Churchill, G. Quantitative trait locus analysis using J/qtl. *Methods Mol Biol.* **573** 175-188 (2009).

29. The Churchill Group. *J/qtl. http://churchill.jax.org/software/jqtl.shtml.* (2013).

30. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* **25** (16), 2078-2079 (2009).

31. Kurtz, S. *et al.* Versatile and open software for comparing large genomes. *Genome Biol.* **5** (2), R12 (2004).

32. Gajria, B. *et al.* ToxoDB: an integrated Toxoplasma gondii database resource. *Nucleic Acids Res.* **36** (Database issue), D553-556 (2008).

33. Heigwer, F., Kerr, G., & Boutros, M. E-CRISP: fast CRISPR target site identification. *Nat Methods.* **11** (2), 122-123 (2014).

34. Roos, D. S. Molecular genetic tools for the identification and analysis of drug targets in Toxoplasma gondii. *Curr Top Microbiol Immunol.* **219** 247-259 (1996).

35. Soldati, D., & Boothroyd, J. C. Transient transfection and expression in the obligate intracellular parasite Toxoplasma gondii. *Science.* **260** (5106), 349-352 (1993).

36. Lander, E. S., & Botstein, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics.* **121** (1), 185-199 (1989).

37. Saeij, J. P. *et al.* Polymorphic secreted kinases are key virulence factors in toxoplasmosis. *Science.* **314** (5806), 1780-1783 (2006).

38. Saeij, J. P. *et al.* Toxoplasma co-opts host gene expression by injection of a polymorphic kinase homologue. *Nature.* **445** (7125), 324-327 (2007).

39. Rosowski, E. E. *et al.* Strain-specific activation of the NF-kappaB pathway by GRA15, a novel Toxoplasma gondii dense granule protein. *J Exp Med.* **208** (1), 195-212 (2011).

40. Reese, M. L., Zeiner, G. M., Saeij, J. P., Boothroyd, J. C., & Boyle, J. P. Polymorphic family of injected pseudokinases is paramount in Toxoplasma virulence. *Proc Natl Acad Sci U S A.* **108** (23), 9625-9630 (2011).

41. Taylor, S. *et al.* A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen Toxoplasma gondii. *Science.* **314** (5806), 1776-1780 (2006).

42. Behnke, M. S. *et al.* Virulence differences in Toxoplasma mediated by amplification of a family of polymorphic pseudokinases. *Proc Natl Acad Sci U S A.* **108** (23), 9631-9636 (2011).

43. Lorenzi, H. *et al.* Local admixture of amplified and diversified secreted pathogenesis determinants shapes mosaic Toxoplasma gondii genomes. *Nat Commun.* **7** 10147 (2016).

44. Gubbels, M. J. *et al.* Forward genetic analysis of the apicomplexan cell division cycle in Toxoplasma gondii. *PLoS Pathog.* **4** (2), e36 (2008).

45. Farrell, A. *et al.* A DOC2 protein identified by mutational profiling is essential for apicomplexan parasite exocytosis. *Science.* **335** (6065), 218-221 (2012).

46. Coleman, B. I., & Gubbels, M. J. A genetic screen to isolate Toxoplasma gondii host-cell egress mutants. *J Vis Exp.* (60) (2012).

47. Walwyn, O. *et al.* Forward genetics screens using macrophages to identify Toxoplasma gondii genes important for resistance to IFN-gamma-dependent cell autonomous immunity. *J Vis Exp.* (97) (2015).

48. Rugarabamu, G., Marq, J. B., Guerin, A., Lebrun, M., & Soldati-Favre, D. Distinct contribution of Toxoplasma gondii rhomboid proteases 4 and 5 to micronemal protein protease 1 activity during invasion. *Mol Microbiol.* **97** (2), 244-262 (2015).

49. Varberg, J. M., Padgett, L. R., Arrizabalaga, G., & Sullivan, W. J., Jr. TgATAT-Mediated alpha-Tubulin Acetylation Is Required for Division of the Protozoan Parasite Toxoplasma gondii. *mSphere.* **1** (1) (2016).

50. Zheng, J., Jia, H., & Zheng, Y. Knockout of leucine aminopeptidase in Toxoplasma gondii using CRISPR/Cas9. *Int J Parasitol.* **45** (2-3), 141-148 (2015).

51. Long, S., Wang, Q., & Sibley, L. D. Analysis of Noncanonical Calcium-Dependent Protein Kinases in Toxoplasma gondii by Targeted Gene Deletion Using CRISPR/Cas9. *Infect Immun.* **84** (5), 1262-1273 (2016).

52. Wang, K. *et al.* Identification of Novel O-Linked Glycosylated Toxoplasma Proteins by Vicia villosa Lectin Chromatography. *PLoS One.* **11** (3), e0150561 (2016).

53. Yang, M., Zheng, J., Jia, H., & Song, M. Functional characterization of X-prolyl aminopeptidase from Toxoplasma gondii. *Parasitology.* 1-7 (2016).
54. Zhang, W. *et al.* Analysis of the virulence determination mechanisms in a local Toxoplasma strain (T.gHB1) isolated from central China. *Parasitol Res.* (2016).