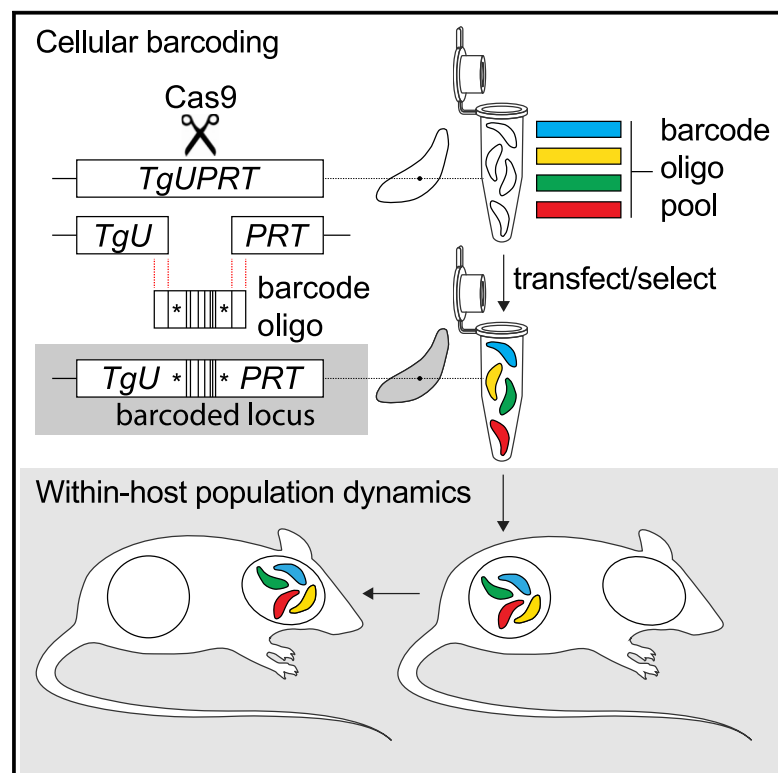# Cellular barcoding of protozoan pathogens reveals the within-host population dynamics of *Toxoplasma gondii* host colonization

Graphical abstract

Authors

Ceire J. Wincott, Gayathri Sritharan, Henry J. Benns, ..., Calvin Tiengwe, Sarah E. Ewald, Matthew A. Child

Correspondence

se2s@virginia.edu (S.E.E.),
m.child@imperial.ac.uk (M.A.C.)

In brief

Cellular barcoding of pathogens enable colonization of the host organism to be quantitatively investigated. Wincott et al. establish a versatile CRISPR-based method to barcode *Toxoplasma gondii* and *Trypanosoma brucei* and present their discovery that colonization of the murine host brain by *T. gondii* is a surprisingly permissive process.

Highlights

- *Toxoplasma gondii* and *Trypanosoma brucei* can be cellularly barcoded

- A simple method generates complex barcoded populations from single transfections

- Complex barcoded populations of *T. gondii* are genetically stable *in vitro*

- Cellular barcoding reveals promiscuous host-brain colonization by *T. gondii*

CellPress

Cell Reports Methods

Report

Cellular barcoding of protozoan pathogens reveals the within-host population dynamics of *Toxoplasma gondii* host colonization

Ceire J. Wincott,[1] Gayathri Sritharan,[1,2] Henry J. Benns,[1,4] Dana May,[7] Carla Gilabert-Carbajo,[1] Monique Bunyan,[3] Aisling R. Fairweather,[1] Eduardo Alves,[1] Ivan Andrew,[5] Laurence Game,[5] Eva-Maria Frickel,[3,6] Calvin Tiengwe,[1] Sarah E. Ewald,[7,*] and Matthew A. Child[1,8,*]

[1]Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK
[2]Department of Biological Sciences, Birkbeck, University of London, Mallet Street, Bloomsbury, London WC1E 7HX, UK
[3]Host-Toxoplasma Interaction Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1 1BF, UK
[4]Department of Chemistry, Imperial College London, White City Campus, London W12 0BZ, UK
[5]UKRI London Institute of Medical Sciences Genomics Laboratory, Shepherd's Bush, London W12 0NN, UK
[6]Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston B15 2TT, UK
[7]Department of Microbiology, Immunology and Cancer Biology at the Carter Immunology Center, University of Virginia School of Medicine, Charlottesville, VA 22908, USA
[8]Lead contact
*Correspondence: se2s@virginia.edu (S.E.E.), m.child@imperial.ac.uk (M.A.C.)
https://doi.org/10.1016/j.crmeth.2022.100274

MOTIVATION *Toxoplasma gondii*'s chronic infection of the host is associated with brain colonization and parasite life cycle stage differentiation, but how this affects *T. gondii*'s within-host population dynamics is unclear. Cellular barcoding has been used to provide these insights for viruses and bacteria, but these methods have not been widely adapted to eukaryotic parasites such as *T. gondii*. Such knowledge can reveal selection bottlenecks, advancing our understanding of how this aspect of the within-host-pathogen interaction influences pathogenesis.

SUMMARY

Cellular barcoding techniques are powerful tools to understand microbial pathogenesis. However, barcoding strategies have not been broadly applied to protozoan parasites, which have unique genomic structures and virulence strategies compared with viral and bacterial pathogens. Here, we present a CRISPR-based method to barcode protozoa, which we successfully apply to *Toxoplasma gondii* and *Trypanosoma brucei*. Using libraries of barcoded *T. gondii*, we evaluate shifts in the population structure from acute to chronic infection of mice. Contrary to expectation, most barcodes were present in the brain one month post-intraperitoneal infection in both inbred CBA/J and outbred Swiss mice. Although parasite cyst number and barcode diversity declined over time, barcodes representing a minor fraction of the inoculum could become a dominant population in the brain by three months post-infection. These data establish a cellular barcoding approach for protozoa and evidence that the blood-brain barrier is not a major bottleneck to colonization by *T. gondii*.

INTRODUCTION

Microbial colonization of tissues within the host organism is a key feature of host-pathogen interactions and is often responsible for the pathology of the associated disease (Ribet and Cossart, 2015). For example, colonization of the brain by the eukaryotic protozoan parasites such as *Toxoplasma gondii*, *Trypanosoma brucei*, and *Plasmodium* spp. can lead to encephalitis in immune compromised or severely infected hosts (Luft et al., 1993). In the case of *T. gondii* infection, the acute phase of infection, whether initiated via the natural, per oral route or intraperitoneal (i.p.) inoc-

ulation, is characterized by rapid tachyzoite replication and dissemination to nearly every tissue in the host. The chronic phase of infection is accompanied by systemic immune clearance and parasite conversion to slower growing, encysted bradyzoites that persist in skeletal muscle and the brain (Wohlfert et al., 2017). The "immune privileged" quality of the chronic infection tissue niches such as the brain coupled with the protective properties of the *T. gondii* cyst wall are essential for transmission to the next host via predation. This is a unique aspect of *T. gondii*'s life cycle compared with related protozoa (Dubey, 1997; Wohlfert et al., 2017). For example, colonization of the

brain niche by the distantly related protozoan *T. brucei brucei* contributes to trypanosomiasis pathology (Rodgers, 2010). However, for *T. brucei*, this is likely a dead-end niche, as evidenced by the severe, rapidly lethal pathology and no means of transmission to the tsetse fly or mosquito vector once in the CNS.

For *T. gondii* infection, it follows that intraspecific competition prior to and during brain colonization is expected to have an impact on long-term success of a clone. Despite advances in our molecular understanding of tachyzoite-to-bradyzoite differentiation (Barrett et al., 2019; Waldman et al., 2020), our comprehension of *T. gondii*'s within-host population dynamics remains limited. One critical function of tissue barriers, including the blood-brain barrier (BBB) and the intestinal epithelial barrier, is to physically restrict pathogen access (Elsheikha and Khan, 2010; Kim, 2008). Notably, stringent bottlenecks are experienced by the related parasite *Plasmodium* spp. during transmission to and from the mosquito host, with these points of population restriction key targets for transmission-blocking therapeutic intervention strategies (Graumans et al., 2020; Sinden, 2017). It is anticipated that the BBB imposes a selection bottleneck upon the *T. gondii* population as the chronic infection is established. However, we lack tools to determine if the brain is colonized infrequently by a small number of parasites that subsequently expand within the tissue niche or if the host brain broadly permissive to colonization by *T. gondii*.

DNA-based cellular barcoding has been instrumental for understanding how the host and the infection influences genetic complexity of pathogen populations, and colonization dynamics (Blundell and Levy, 2014; Kebschull and Zador, 2018). Early studies using restriction site-tagged poliovirus identified a bottleneck limiting the genetic diversity of viral quasi-species transmitted to the murine brain (Pfeiffer and Kirkegaard, 2006). Furthermore, studies using wild-type isogenic tagged strains (WITS) of *Salmonella* have provided insights into selection bottlenecks experienced during gastro-intestinal and invasive bacterial colonization (Grant et al., 2008; Lim et al., 2014). The population structure of *Salmonella* infection was found to experience dramatic selective bottlenecks during the colonization of the gut niche, with the predominant colonizing strain also the dominant strain transmitted by super shedders (Lam and Monack, 2014). The generation of WITS requires the insertion of DNA barcodes into the genome of the infectious agent (Grant et al., 2008). Distinct from applications in which barcodes are used to track mutant populations, these cellular barcodes allow the complexity of the population to be closely monitored and mapped over the course of an infection using quantitative next-generation sequencing (NGS) approaches. Unexpectedly, cellular barcoding has not been widely used to study the within-host population structure of wild-type protozoan pathogens.

RESULTS

Protozoan pathogens can be cellularly barcoded with a simple CRISPR-based strategy

Inspired by WITS (Grant et al., 2008), we sought to establish a versatile system to cellularly barcode eukaryotic pathogens. We considered that our approach should (1) modify cells to harbor single barcodes at an identical position within a non-essential genomic locus and (2) include a selection strategy to ensure that a successfully barcoded cell could be enriched to represent the entire population. This would ensure that the barcoded population was isogenic apart from the unique barcode identifier. Established CRISPR-Cas9 tools for *T. gondii* and *T. brucei* (Rico et al., 2018; Shen et al., 2014; Sidik et al., 2014) were adapted to mediate site-specific recombination of DNA barcodes into non-essential genes that encode negative selectable markers in *T. gondii* tachyzoites or *T. brucei* trypomastigotes (Figure 1A). Notably, *T. brucei* naturally lacks non-homologous end-joining (NHEJ) machinery, and *T. gondii* parasites deficient for NHEJ (RHΔ*ku80*) were used to increase recombination efficiency (Huynh and Carruthers, 2009). Parasites were co-transfected with a plasmid encoding both the Cas9 nuclease and guide RNA (gRNA) scaffold (Shen et al., 2014), and a unique 60 nt single-stranded donor template encoding the barcode. Successful targeting of Cas9 and genomic integration of the barcode disrupted the *UPRT* coding sequence, conferring resistance to the prodrug 5-fluorodeoxyuridine (FUDR) (Donald and Roos, 1995). To prevent further modification of the *UPRT* locus following a single barcode integration event, our strategy also deleted both the protospacer DNA sequence recognized by the CRISPR gRNA and the protospacer adjacent motif (PAM). We confirmed barcode integration and the expected genomic modification by Sanger sequencing of the *UPRT* locus in the drug-resistant parasite population (Figure 1B). A similar cellular barcoding strategy was applied to the bloodstream form *T. brucei* whereby Cas9 was targeted to the haploid *AAT6* locus (Rico et al., 2018), a single-copy non-essential gene that confers sensitivity to eflornithine (Vincent et al., 2010) (Figure 1A). Parasites stably expressing Cas9 were co-transfected with *AAT6*-targeting sgDNAs and a double-stranded barcoding oligo. Successful barcode integration into the *AAT6* locus conferred resistance to eflornithine. Sanger sequencing confirmed the expected genomic rearrangement and barcoding of the *T. brucei* genome (Figure 1B). These data indicate our barcoding strategy is an efficient tool to insert unique a barcode sequence at specific genomic positions in two divergent eukaryotic pathogens.

Barcode alleles can be identified and quantified in complex populations

To examine the stability of individual barcodes across a population of uniquely labeled parasites and to evaluate the sensitivity of barcode detection, we focused on *T. gondii*. First, we generated a plate-mapped library of 96 uniquely barcoded parasites lines using multiplexed transfection (Figure 1C). We reasoned that this multiplexed strategy would be useful for future chemical or genetic screening applications applied to each of the 96 uniquely barcoded strains prior to pooling. To quantify the relative representation of individual barcodes after pooling, an NGS pipeline was developed. DNA was extracted from the pooled parasite library, the barcoded region of the UPRT locus was amplified for Illumina sequencing, and barcodes representation was quantified using Galaxy (Figures 1C and S1A) (Afgan et al., 2018; Smith et al., 2009). Our pipeline successfully identified all 96 uniquely barcoded strains from the pooled population (Figure 1D). Notably, wells from one row of the 96-well plate were
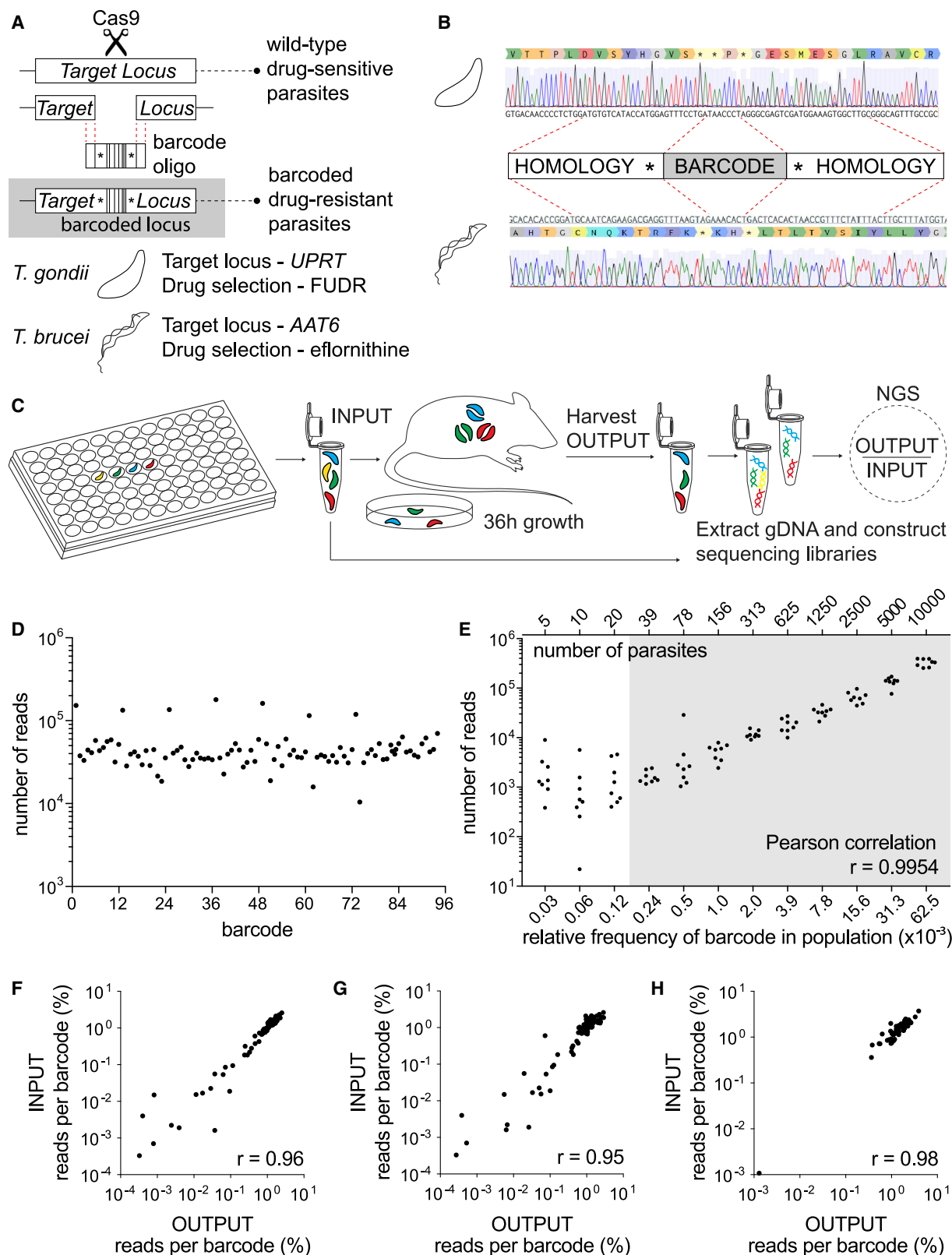
(legend on next page)

consistently more highly represented in the pooled population, reflecting a technical challenge of the 96-well transfection method (Figure 1D). Although unplanned, this observation suggested that the NGS readout was highly sensitive to differences in barcode frequency in the pooled population. This was confirmed experimentally, as greater variation in barcode representation was observed between biological replicate transfections (Figures S1B and S1C) than between technical replicates of DNA indexed for NGS sequencing (Figures S1B, S1D, and S1E), underscoring the reproducibility of the NGS pipeline. We anticipate that the sources of variability could include parasite number and stochastic DNA extraction variability.

We next tested if we could identify groups of barcodes present at different frequencies within a single complex population. We generated a 2-fold dilution series of the pooled parasite population prior to isolating DNA for NGS sequencing. A positive correlation was observed between the number of reads (read output) and the number of parasites in the input (Figure 1E; r = 0.9954). This confirmed that barcodes could be successfully identified and reliably quantified within libraries of at least 96-member complexity.

Multiplex barcoded parasite libraries are stably maintained *in vitro* and *in vivo*

To test whether the complexity of the barcode libraries was stably maintained *in vitro* the pooled library of barcoded parasites was serially passed through human foreskin fibroblasts (HFFs). Lysed-out parasite cultures were sampled every passage (~36 h) for a period of six passages, equivalent to six lytic growth cycles (invasion, replication, egress). Relative to the input, the genetic complexity of the barcode population *in vitro* was remarkably stable after one lytic cycle or six passages (Figures 1F and 1G). We next tested our ability to propagate and recover the barcode library *in vivo* within a murine host using a pooled library of 63 barcodes (in this pilot experiment, some wells were not efficiently transfected). An inoculum of 2 × 10⁶ parasites was injected intraperitoneally into C57BL/6 mice. After 36 h, parasites were isolated from the peritoneal cavity by lavage and compared with the input population. A strong positive correlation was

observed between the input and peritoneal exudate populations (Figure 1H; Pearson correlation coefficient [PCC] r = 0.98), demonstrating that the genetic complexity of the multiplex barcode library was stable over the first 36 h of *in vivo* infection.

Libraries of barcode alleles can be generated by a "one-pot" transfection method

One challenge of the multiplex platform arose from experiment-to-experiment variability in transfection efficiency across individual wells (Figures 1D and 1H). Our CRISPR-Cas9 strategy was designed to ensure the integration of a single barcode into the *UPRT* locus by simultaneously deleting the protospacer and PAM motifs recognized by the CRISPR gRNA. We hypothesized that a barcoded library of parasites could be generated by a single transfection with the Cas9-gRNA plasmid and a pooled library of oligonucleotide repair templates using a widely available cuvette-based transfection apparatus (Figure 2A). This "one-pot" method was tested on type I RHΔ*ku80* and type II Pru-Δ*ku80* parasite strains in parallel, using the same mixed pool of 96 barcode oligo repair templates. FUDR-resistant parasite populations were enriched, genomic DNA isolated, UPRT locus amplicons prepared, and NGS libraries sequenced. All 96 barcodes were identified in the sequence reads generated from the UPRT locus amplicon (Figure 2B). When the distribution of barcodes in RHΔ*ku80* and PruΔ*ku80* were compared directly, the two independently transfected parasite strains exhibited correlated frequency distributions for all but the least abundant barcodes, with a Pearson correlation coefficient of 0.7 (Figure 2B). We confirmed the technical reproducibility of NGS runs by comparing technical replicates of indexed DNA amplicons (Figure S2A), consistent with previous reports (Robinson et al., 2014). We concluded that differences in barcode frequency most likely reflect minor variations in the relative abundance of each barcode oligonucleotide template within the pool used for transfection, rather than a negative impact of some barcodes on parasite fitness. Supporting this notion, the population was remarkably stable after 28 days of growth *in vitro* (Figure S2B), indicating that there were no subtle fitness defects associated with the integration of specific barcodes, with low-frequency

---

Figure 1. Protozoan pathogens can be cellularly barcoded using a simple CRISPR-based strategy

(A) Schematic of cellular barcoding strategy. A CRISPR guide RNA targets Cas9 to the *UPRT* locus in *T. gondii* or the *AAT6* locus in *T. b. brucei*, where Cas9 endonuclease activity introduces a double-strand break. This break is repaired by homologous DNA repair using the co-transfected 60-mer barcode oligo donor template.

(B) Sanger sequencing confirmation that barcode integration results in disruption of the *UPRT* in *T. gondii* or *AAT6* in *T. b. brucei* coding sequence and destruction of the protospacer DNA sequence and PAM.

(C) Multiplexed transfection strategy and NGS pipeline to generate barcoded libraries of *T. gondii* tachyzoites for *in vitro* and *in vivo* studies.

(D) Complex libraries containing 96 uniquely barcoded strains can be produced and the relative frequency of each barcode quantified by NGS. Scatterplot presents read counts for individual barcodes.

(E) Individual barcodes can be detected at low frequencies, and read depth is a sensitive proxy for parasite number. Scatterplot presents the number of reads for a serial 2-fold dilution series of known numbers of parasites and the relative frequency of individual barcodes in the population. The shaded area indicates the data used to calculate the PCC provided: r = 0.9954, n = 9, p < 0.0001 (two-tailed).

(F and G) The population structure of multiplexed barcode libraries is stable *in vitro*. Scatter plot comparing 96 individual barcode frequencies within the INPUT pooled library population and the OUTPUT sample harvested after one ~36 h lytic growth cycle (F; PCC r = 0.96, n = 96, p < 0.0001 [two-tailed]) or six lytic growth cycles (G; PCC r = 0.95, n = 96, p < 0.0001 [two-tailed]).

(H) The population structure of multiplexed barcode libraries is stable during the early stage of *in vivo* infection. Scatter plot comparing 63 individual barcode frequencies within a pooled library population for an intraperitoneal inoculum of 2 × 10⁶ (INPUT) and the OUTPUT sample harvested from the peritoneal cavity of C57BL6 mice 36 h post-infection. PCC r = 0.98, n = 63, p < 0.0001 (two-tailed).

See also Figure S1.
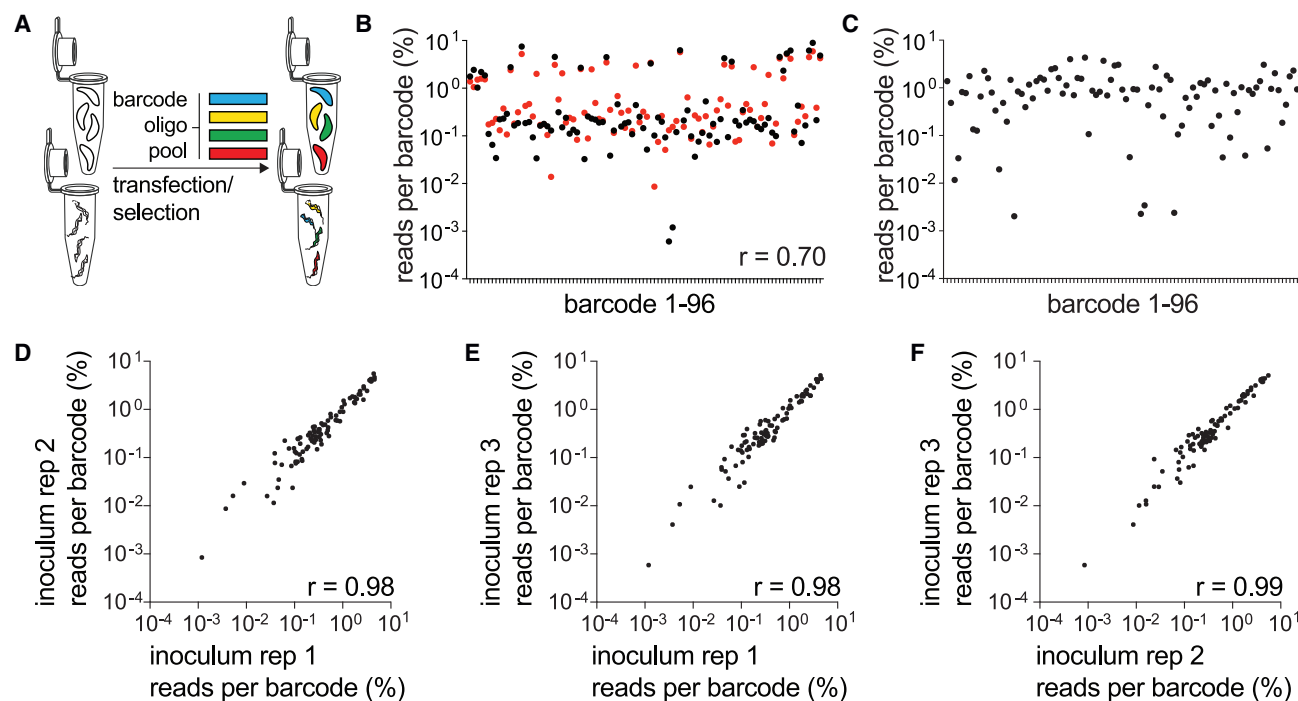
Figure 2. Barcoded libraries can be generated via one-pot transfections

(A) Schematic of the one-pot transfection method to generate barcoded parasite libraries.

(B) Overlay of the relative percent frequency of each barcode in an RHΔ*ku80* population (black dots) or a PruΔ*ku80* population (red dots) generated from a common pool of barcode oligos by one-pot transfection. The scatter plot is distributed according to barcode identifier (1–96), and PCC compares the two populations, r = 0.70, n = 96, p < 0.0001 (two-tailed).

(C) Data from one-pot transfections with *T. brucei*. Scatter plot presents relative percentage frequency of barcodes within transfected parasites, distributed according to barcode identifier (1–96).

(D–F) One-pot transfections using PruΔ*ku80* were diluted to a founder population of 37,000 parasites and re-expanded in HFFs prior to NGS. All 96 barcodes were identified in each sample, and the relative percentage frequency of barcodes was highly correlated in pairwise comparison for each inoculum sample: (D) inoculum 1 versus 2, PCC r = 0.98, n = 96, p < 0.0001 (two-tailed); (E) inoculum 1 versus 3, PCC r = 0.98, n = 96, p < 0.0001 (two-tailed); (F) inoculum 2 versus 3, PCC r = 0.99, n = 96, p < 0.0001 (two-tailed). PCC values provided on scatter plots indicate degree of correlation between populations being compared. See also Figures S1 and S2.

barcodes stably maintained. Stochastic variation was therefore the likely driver of the limited variation observed. It is notable that a higher level of biological reproducibility was observed for one-pot library transfections than the original multiplexed transfection approach (Figure S1C). We next tested the one-pot strategy in *T. brucei*. We were similarly able to produce complex 96-member libraries of uniquely barcoded strains from a single transfection (Figure 2C). The one-pot transfection protocol was used for all subsequent experiments.

Cellular barcodes reveal the population structure of a *T. gondii* infection *in vivo*

The transition from acute to chronic infection corresponds with the spatial redistribution of *T. gondii* to skeletal muscle and the CNS (Wohlfert et al., 2017). This is accompanied by tachyzoite differentiation into bradyzoites cysts, which are necessary for parasite transmission (Barrett et al., 2019). The restrictive nature of the BBB is well documented in other infection models (Profaci et al., 2020). We hypothesized that these spatial, temporal, and developmental transitions would impose bottlenecks upon the parasite population represented in the brain at chronic infection.

We sought to test this by infecting murine hosts intraperitoneally with the pooled library of PruΔ*ku80* parasites stably expressing 96 neutral barcode alleles.

First, we confirmed that diluting the PruΔ*ku80* library to a dose tolerated by mice (and therefore appropriate for achieving a chronic infection) did not disrupt the population structure in a way that could confound the interpretation of *in vivo* experiments. Three inoculum samples of 37,000 viable tachyzoites (determined by plaque assay) were plated on HFFs and re-expanded in tissue culture followed by DNA isolation for NGS sequencing. In each inoculum sample, all 96 barcodes were detected by NGS sequencing, and pairwise comparisons of each sample were strongly correlated (Figures 2D–2F), confirming that *in vitro* expansion of *in vivo* samples would minimally affect library composition.

An *in vivo* experiment was conceived to study the effect of murine host colonization. This would probe the within-host population genetics of the *T. gondii* infection as it progressed from the initial acute infection in the peritoneum to the chronic infection in the brain (Figure 3A). To determine if the parasite population was stable early in acute infection, parasites were isolated from the
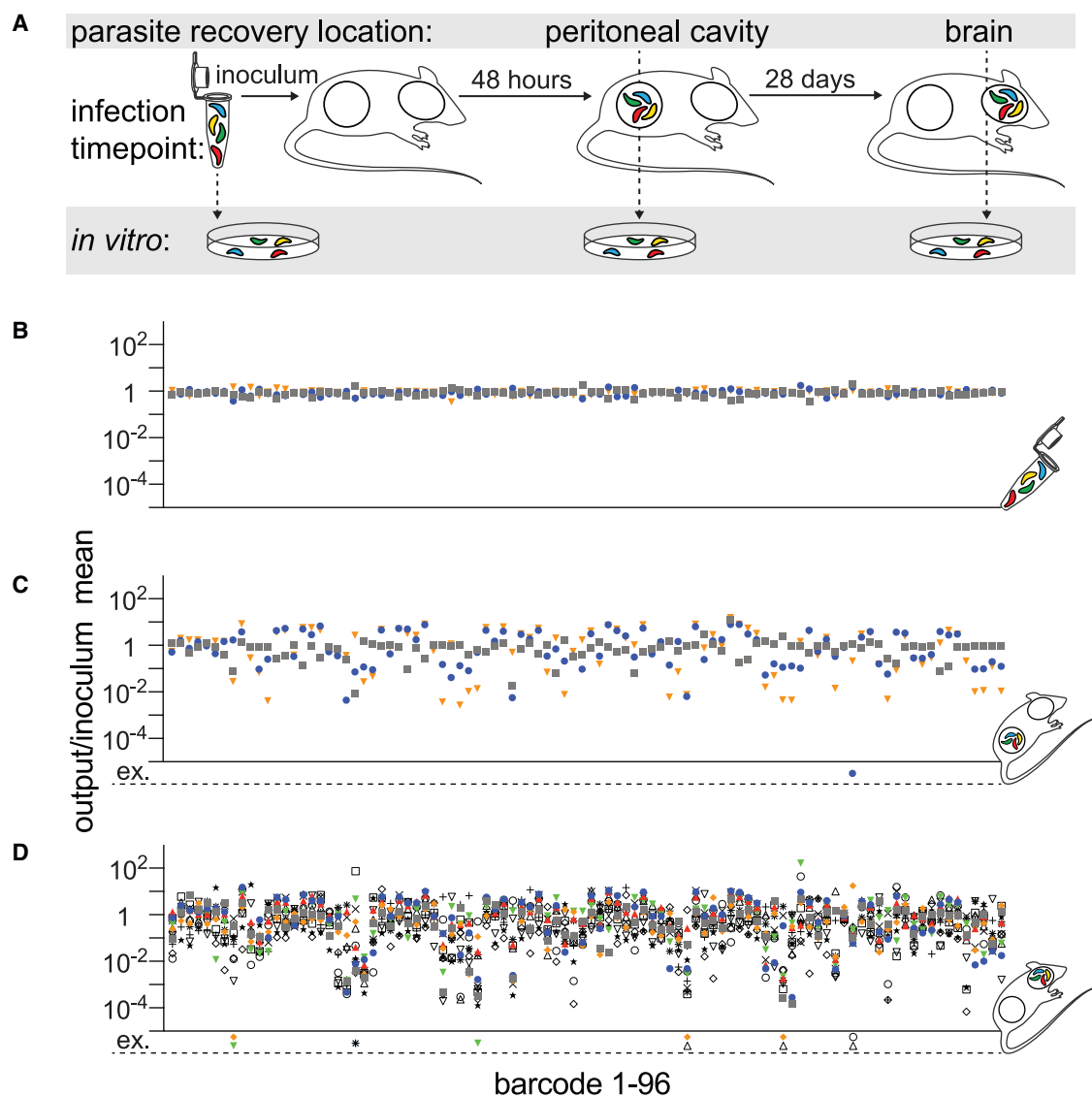
Figure 3. The murine host brain is permissively colonized by *T. gondii*

(A) Schematic of an experiment to identify changes in the *T. gondii* population structure over the course of a 28-day *in vivo* infection. CBA/J mice were infected with an inoculum of 37,000 tachyzoites injected intraperitoneally. The infectious population structure was monitored in peritoneal exudates at 48 h post-infection (early acute) or in the brain at the onset of the chronic phase (28 days post-infection). Parasites were re-expanded in tissue culture prior to NGS sequencing.

(B–D) Scatterplot represent individual barcode frequencies in each sample relative to the mean of the inoculum in the inoculum replicate (B) n = 3 replicates, the peritoneal cavity at 48 h (C) n = 3 mice, and the brain at 28 days (D); n = 14 mice. Barcode extinctions (ex.) are indicated below the x axis in the corresponding position to the absent barcoded strain (see also Table S1). Each symbol represents an individual inoculum or mouse.

See also Table S1.

peritoneal cavity of some mice at 48 h post-infection and expanded *in vitro*. Although barcode extinction was rare (one barcode in a single mouse), the relative frequency of barcodes in each peritoneal isolate was unique to each animal (Figures 3B and 3C). This observation supports a model in which the initial selective sweep at the onset of acute infection is stochastic and emphasized the need to consider each host organism as a unique environment. After 28 days, the brains of chronically infected mice were isolated, and parasites expanded *in vitro*. Unexpectedly, the majority of barcodes were detected in

each host brain (Figure 3D; Table S1), consistent with a minimal founder effect on the genetic diversity of the parasite population colonizing the CNS. The cumulative extinction frequency across all 14 mice was 0.007 (10 of 1,344 total barcodes), and lost barcodes were most often represented at low abundance within the inoculum (Table S3). Of note, these extinction events were observed in the NGS runs with the lowest total read counts, suggesting that extinctions could be due to reduced sequence sampling depth rather than a true absence of the specific barcoded parasite(s) in the brain (Table S3). Together, these data indicate

that any selection bottleneck experienced by parasites during the colonization of the brain niche must be broad.

Stochastic selection and host genotype shape the *T. gondii* population structure colonizing the brain

We next sought to determine how host genetic background influences the dynamics of parasite brain colonization over time. To do this we infected Swiss Webster mice, which are outbred to maximize genetic diversity and heterozygosity, and inbred CBA/J mice (Figure 4A) (Watson and Davis, 2019). Mean parasite cyst burden was not significantly different between CBA/J mice and Swiss Webster mice at one month post-infection, indicating that a similar number of parasites were accessing the brain niche (Figure 4B). As expected, cyst burden declined over time, but the reduction in mean cyst number was similar between mouse genotypes (McGovern et al., 2020; Watson and Davis, 2019).

To evaluate the parasite population structure within these hosts, we used Cavalli Sforza's chord distance calculation (Cavalli-Sforza and Edwards, 1967). Chord distance is a geometric measure of population distance. It provides a measurement of genetic divergence between two populations expressed between zero and one, with "zero" indicating that the genetic structures of the two populations are identical and "one" indicating maximum genetic divergence. Although chord distance does not provide absolute quantification of the width of any bottleneck, it can be used to infer changes in the genetic structure of a population between points A and B. In CBA/J mice (Figure 4C) there was minimal distance separating the inoculum from the peritoneal exudate population 72 h post-infection. A significant increase in distance was observed between the inoculum and the one-month brain samples. Consistent with the decline in cyst number over time, chord distance was greater at three months post-infection in CBA/J brain. Outbred Swiss Webster mice also exhibited little shift in chord distance from the inoculum to the peritoneal exudate and a significant shift in distance from the inoculum to the one-month post-infection brain samples (Figure 4D). However, the individual Swiss mice exhibited greater variation in parasite population structure at each chronic time point than CBA/J mice. Unlike CBA/J mice, Swiss mice supported parasite populations with a similar median chord distance between inoculum and one- or three-month post-infection brains.

Cumulatively, our data indicated that following an i.p. infection the murine brain was unexpectedly permissive to colonization (Figure 3D), and parasite population structure was dynamic within individual hosts (Figures 4C and 4D). In CBA/J mice, the genetic distance separating the brain niche parasite population from the inoculum was greater at three months versus one month post-infection (Figure 4C). With this in mind, we analyzed our data to understand if the barcoded lineages that were the most abundant in the inoculum had a competitive advantage for long-term persistence. In CBA/J mice, dominant barcode alleles at three months post-infection (Figure 4E, top row) corresponded to parasite populations that were among the top 35% of reads (top six most abundant barcodes) in the inoculum. Although this trend was similar at one month post-infection, several CBA/J mice harbored a dominant barcode lineage represented in the bottom 25% of reads in the inoculum (Figure 4E outset in the inoculum pie chart). In agreement with the chord distance analysis, individual Swiss mice had diverse dominant population structures at three months post-infection (Figure 4E, bottom row). Although dominant barcodes in the Swiss brains were typically derived from lineages represented within the top 75% of reads in the inoculum (Figure 4E, most abundant 15 barcodes), this was not always the case. For example, one mouse was dominated by a barcode that was present at low frequency in the inoculum, while another mouse exhibited a population structure in which barcode frequency was more evenly distributed than in the inoculum or peritoneal parasite isolates. This heterogeneity was also reflected in the barcodes that dominated Swiss brains at one month post-infection, suggesting that even relatively low-frequency strains can establish and maintain persistent infection in a manner that likely depends on stochastic variables and host genotype.

DISCUSSION

Brain residency has long been appreciated as a strategy for pathogenesis exploited by *T. gondii* to evade sterilizing immunity, as well as an evolutionary strategy to increases the likelihood of transmission via predation. Feline consumption of high-fat, energy-rich infected neuronal tissue of prey, including mice, provides *T. gondii* with a route back into the definitive feline host (Dubey, 1997; Plantinga et al., 2011). A means to evade any restrictive bottleneck when colonizing the brain niche would be consistent with this evolutionary strategy, supporting maximal transmission of genetic diversity into the feline host to contribute to recombination in the subsequent sexual cycle. Additionally, *T. gondii* can infect most warm-blooded animals, indicating that intermediate host-specific selection environments will frequently and unpredictably change, favoring different phenotypes at different times. It therefore likely benefits the parasite to maintain genetic diversity and phenotypic plasticity to ensure survival in varied future host species.

A key role of tissue barriers is to restrict pathogen access to the underlying tissue niche (Kim, 2008). Unexpectedly our findings imply multiple unique *T. gondii* colonization events from circulation into the brain parenchyma, rather than rare brain invasions coupled with expansion within that niche. During intraperitoneal infection, the loss of barcodes is stochastic, and most barcodes are identified in the brain parenchyma at one month post-infection (Figure 3). Chord distance indicated a change in the genetic structure of the parasite population in the CNS at one month post-infection (Figures 4C and 4D). In CBA/J mice and Swiss mice, the dominant barcodes in the brain following intraperitoneal infection tended to be those most frequently observed in the inoculum, but most barcodes were still detected, indicating limited selective pressure imparted by the BBB. There were exceptions to this, particularly in outbred Swiss mice, in which minor constituents of the inoculum were able to dominate the brain population in some individuals. Although this finding is intriguing, a greater number of infections (requiring more complex statistical methods) will need to be done before definitive conclusions can be made. It is also important to acknowledge that our data represent the sum of two differentiations: tachyzoite-to-bradyzoite *in vivo*, followed by bradyzoite-to-tachyzoite *in vitro*.
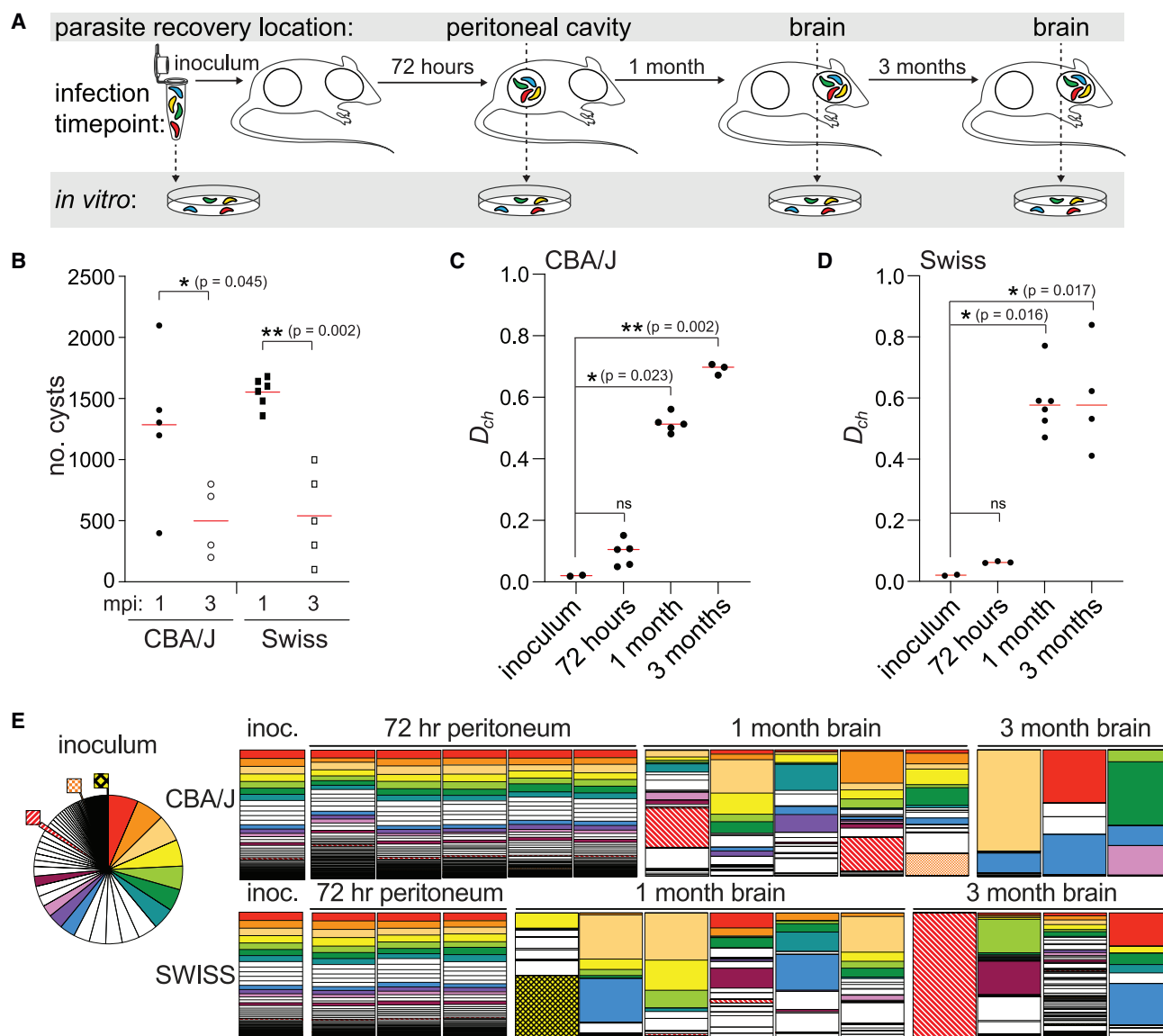
Figure 4. The population structure of the *T. gondii* chronic infection is dynamic within the murine host brain niche

(A) Schematic of an experiment investigating changes in the *T. gondii* population structure over the course of a three-month *in vivo* infection in inbred CBA/J and outbred Swiss Webster mice. Mice were intraperitoneally injected with 12,000 tachyzoites and the *T. gondii* population structure was evaluated at an early acute (72 h in peritoneal exudate), early chronic (one month in the brain), and late chronic (three months in the brain) infection time points.

(B) Bradyzoite cyst counts from *T. gondii* infected brains isolated from CBA/J and Swiss Webster mice at one and three months post-infection (mpi). Significance was tested using one-way ANOVA.

(C and D) Scatterplot of chord distances (D_{ch}) calculated from barcode representation in the inoculum relative to 48-h peritoneal samples, 1-month post-infection brains, or 3-month post-infection brains in CBA/J (C) or Swiss Webster mice (D). Significance relative to inoculum was tested using Kruskal-Wallis one-way ANOVA, with the Mann-Whitney test used for pairwise comparisons.

(E) Parts-of-whole charts representing the relative frequency of each barcode within the mouse samples. Barcodes are ranked in descending order of abundance in the inoculum mean (inoculum mean n = 3, represented as a pie chart and bar charts labeled "inoc."). A color was assigned to a barcode if it was the dominant barcode in any brain sample or if it represented greater than 20% of all reads in any brain sample. Each color is unique to one barcode. Any barcode that dominated a mouse brain and was represented in the bottom 25% of inoculum reads is outset in the inoculum pie chart for clarity and shaded brown, gray, or black in parts-of-whole charts.

We used intraperitoneal infection because it facilitates a precise quantification of barcode frequency and coverage across the inoculum dose. The inoculum dose was chosen to achieve at least 100× coverage of all barcodes, thereby accounting for possible differences in viability of tachyzoites *in vitro* versus *in vivo*. This is not possible when using bradyzoite cysts, as each cyst, which must be harvested from infected mice, has a variable parasite number. In the future, it will be interesting to

investigate how host colonization following oral infection shapes the population structure of the parasite, and how it differs from the intraperitoneal route. It should also be noted that the results of our study do not mean that there is no bottleneck for infection of the brain niche, simply that with the number of barcode markers used in this study, we were not able to determine the bottleneck width. On the basis that all 96 barcodes were identified within the brains of the infected hosts, we estimate the minimum size of the bottleneck to be at least 96 parasites.

The observations made in this study would have not been possible without a means to barcode this eukaryotic pathogen. Cellular barcoding has provided critical insights into the infection biology of viruses such as poliovirus (Kuss et al., 2008; Pfeiffer and Kirkegaard, 2006) and bacteria such as *Salmonella* (Grant et al., 2008; Kaiser et al., 2013; Lam and Monack, 2014; Lim et al., 2014) and *Escherichia coli* (Hullahalli and Waldor, 2021). We anticipate that barcoded *T. gondii* strains will have similar far-reaching applications. Although barcode sequencing (barseq) strategies have been widely used for eukaryotic pathogen phenotypic screens (Alsford et al., 2011; Baker et al., 2021; Beneke et al., 2019; Beneke and Gluenz, 2020; Bushell et al., 2017; Sidik et al., 2016), to date the use of cellular barcodes to study the within-host infectious population structure of wild-type eukaryotic pathogens has been limited. A notable example being the use of eight uniquely barcoded *T. brucei* strains to study colonization of the tsetse fly and subsequent survival of these strains within the bloodstream of the murine host (Oberle et al., 2010). The increased number of cellular barcodes afforded by our simple approach will provide an even greater opportunity for *T. brucei* researchers to interrogate colonization the host organism, such as the colonization of subcutaneous fat and bone marrow (Mugnier et al., 2015).

The versatility of the strategies presented will allow researchers to work with individual barcoded strains in isolation prior to pooling (via plate-based library generation) or with complex libraries of barcoded strains generated through our one-pot approach. We believe that this will expand the application of our approach beyond the population genetic studies described herein. For example, barcoded cancer cell lines have been successfully used to multiplex an *in vivo* drug screen (Gruner et al., 2016). Importantly, this allowed the authors to chemically interrogate metastatic seeding, a unique feature of the *in vivo* disease that cannot be accurately represented *in vitro*. It should be noted that disruption of the *UPRT* locus has been documented to negatively affect cyst burden, which would suggest that our data might underrepresent barcode diversity in the CNS. However, all barcodes are inserted at the same position, so the cross-comparison of populations is internally controlled. Our simple oligo barcoding strategy can be applied to systems accessible to CRISPR that contains endogenous or engineered negative selection markers, which we successfully demonstrated with *T. brucei*. In principle, our approach allows parallel integration of even greater numbers of unique barcodes, which will increase precision in future host-pathogen population genetic studies. An increased number of cellular barcodes combined with methods such as sequence tag analysis of microbial populations (STAMP) (Abel et al., 2015) or the recently published update STAMPR

(Hullahalli et al., 2021) will make it possible to quantify the absolute founder population number present within the host brain. Combined with these methods, our cellular barcoding approach will allow researchers to probe the within-host population genetics of tissue colonization during a *T. gondii* infection with an unprecedented degree of resolution.

Limitations of the study

Our methodology currently requires that bradyzoite cysts spontaneously re-differentiate *in vitro* following isolation from the host. This second stage conversion is required to expand parasite numbers to sufficient quantities for our NGS workflow, but it should be noted that it might influence the frequency distribution of barcodes in the final population.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - *Toxoplasma gondii* cell culture
  - Animal work
  - *Trypanosome brucei brucei* cell culture
- METHOD DETAILS
  - Generation of barcoded *Toxoplasma gondii*
  - Generation of barcoded *Trypanosome brucei brucei*
  - NGS library preparation
  - *In vivo* experiments
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bottleneck analysis and chord distance calculations
  - Pearson correlation coefficients, significance and error

REFERENCES

Abel, S., Abel zur Wiesch, P., Chang, H.H., Davis, B.M., Lipsitch, M., and Waldor, M.K. (2015). Sequence tag-based analysis of microbial population dynamics. Nat. Methods 12, 223–226. https://doi.org/10.1038/nmeth.3253.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46, W537–W544. https://doi.org/10.1093/nar/gky379.

Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berriman, M., Hertz-Fowler, C., and Horn, D. (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. Genome Res. 21, 915–924. https://doi.org/10.1101/gr.115089.110.

Baker, N., Catta-Preta, C.M.C., Neish, R., Sadlova, J., Powell, B., Alves-Ferreira, E.V.C., Geoghegan, V., Carnielli, J.B.T., Newling, K., Hughes, C., et al. (2021). Systematic functional analysis of Leishmania protein kinases identifies regulators of differentiation or survival. Nat. Commun. 12, 1244. https://doi.org/10.1038/s41467-021-21360-8.

Barrett, M.P., Kyle, D.E., Sibley, L.D., Radke, J.B., and Tarleton, R.L. (2019). Protozoan persister-like cells and drug treatment failure. Nat. Rev. Microbiol. 17, 607–620. https://doi.org/10.1038/s41579-019-0238-x.

Beneke, T., Demay, F., Hookway, E., Ashman, N., Jeffery, H., Smith, J., Valli, J., Becvar, T., Myskova, J., Lestinova, T., et al. (2019). Genetic dissection of a Leishmania flagellar proteome demonstrates requirement for directional motility in sand fly infections. PLoS Pathog. 15, e1007828. https://doi.org/10.1371/journal.ppat.1007828.

Beneke, T., and Gluenz, E. (2020). Bar-seq strategies for the LeishGEdit toolbox. Mol. Biochem. Parasitol. 239, 111295. https://doi.org/10.1016/j.molbiopara.2020.111295.

Blundell, J.R., and Levy, S.F. (2014). Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. Genomics 104, 417–430. https://doi.org/10.1016/j.ygeno.2014.09.005.

Bushell, E., Gomes, A.R., Sanderson, T., Anar, B., Girling, G., Herd, C., Metcalf, T., Modrzynska, K., Schwach, F., Martin, R.E., et al. (2017). Functional profiling of a plasmodium genome reveals an abundance of essential genes. Cell 170, 260–272.e8. https://doi.org/10.1016/j.cell.2017.06.030.

Cavalli-Sforza, L.L., and Edwards, A.W. (1967). Phylogenetic analysis. Models and estimation procedures. Am. J. Hum. Genet. 19, 233–257.

Donald, R.G., and Roos, D.S. (1995). Insertional mutagenesis and marker rescue in a protozoan parasite: cloning of the uracil phosphoribosyltransferase locus from Toxoplasma gondii. Proc. Natl. Acad. Sci. USA 92, 5749–5753.

Dubey, J.P. (1997). Tissue cyst tropism in Toxoplasma gondii: a comparison of tissue cyst formation in organs of cats, and rodents fed oocysts. Parasitology 115, 15–20. https://doi.org/10.1017/s0031182097008949.

Elsheikha, H.M., and Khan, N.A. (2010). Protozoa traversal of the blood-brain barrier to invade the central nervous system. FEMS Microbiol. Rev. 34, 532–553. https://doi.org/10.1111/j.1574-6976.2010.00215.x.

Grant, A.J., Restif, O., McKinley, T.J., Sheppard, M., Maskell, D.J., and Mastroeni, P. (2008). Modelling within-host spatiotemporal dynamics of invasive bacterial disease. PLoS Biol. 6, e74. https://doi.org/10.1371/journal.pbio.0060074.

Graumans, W., Jacobs, E., Bousema, T., and Sinnis, P. (2020). When is a plasmodium-infected mosquito an infectious mosquito? Trends Parasitol. 36, 705–716. https://doi.org/10.1016/j.pt.2020.05.011.

Grüner, B.M., Schulze, C.J., Yang, D., Ogasawara, D., Dix, M.M., Rogers, Z.N., Chuang, C.H., McFarland, C.D., Chiou, S.H., Brown, J.M., et al. (2016). An in vivo multiplexed small-molecule screening platform. Nat. Methods 13, 883–889. https://doi.org/10.1038/nmeth.3992.

Hullahalli, K., Pritchard, J.R., and Waldor, M.K. (2021). Refined quantification of infection bottlenecks and pathogen dissemination with STAMPR. mSystems 6, e0088721. https://doi.org/10.1128/mSystems.00887-21.

Hullahalli, K., and Waldor, M.K. (2021). Pathogen clonal expansion underlies multiorgan dissemination and organ-specific outcomes during murine systemic infection. Elife 10, e70910. https://doi.org/10.7554/eLife.70910.

Huynh, M.H., and Carruthers, V.B. (2009). Tagging of endogenous genes in a Toxoplasma gondii strain lacking Ku80. Eukaryot. Cell 8, 530–539. https://doi.org/10.1128/EC.00358-08.

Kaiser, P., Slack, E., Grant, A.J., Hardt, W.D., and Regoes, R.R. (2013). Lymph node colonization dynamics after oral Salmonella Typhimurium infection in mice. PLoS Pathog. 9, e1003532. https://doi.org/10.1371/journal.ppat.1003532.

Kebschull, J.M., and Zador, A.M. (2018). Cellular barcoding: lineage tracing, screening and beyond. Nat. Methods 15, 871–879. https://doi.org/10.1038/s41592-018-0185-x.

Kim, K.S. (2008). Mechanisms of microbial traversal of the blood-brain barrier. Nat. Rev. Microbiol. 6, 625–634. https://doi.org/10.1038/nrmicro1952.

Kuss, S.K., Etheredge, C.A., and Pfeiffer, J.K. (2008). Multiple host barriers restrict poliovirus trafficking in mice. PLoS Pathog. 4, e1000082. https://doi.org/10.1371/journal.ppat.1000082.

Lam, L.H., and Monack, D.M. (2014). Intraspecies competition for niches in the distal gut dictate transmission during persistent Salmonella infection. PLoS Pathog. 10, e1004527. https://doi.org/10.1371/journal.ppat.1004527.

Lim, C.H., Voedisch, S., Wahl, B., Rouf, S.F., Geffers, R., Rhen, M., and Pabst, O. (2014). Independent bottlenecks characterize colonization of systemic compartments and gut lymphoid tissue by salmonella. PLoS Pathog. 10, e1004270. https://doi.org/10.1371/journal.ppat.1004270.

Luft, B.J., Hafner, R., Korzun, A.H., Leport, C., Antoniskis, D., Bosler, E.M., Bourland, D.D., 3rd, Uttamchandani, R., Fuhrer, J., Jacobson, J., et al. (1993). Toxoplasmic encephalitis in patients with the acquired immunodeficiency syndrome. Members of the ACTG 077p/ANRS 009 Study Team. N. Engl. J. Med. 329, 995–1000. https://doi.org/10.1056/NEJM199309303291403.

McGovern, K.E., Cabral, C.M., Morrison, H.W., and Koshy, A.A. (2020). Aging with Toxoplasma gondii results in pathogen clearance, resolution of inflammation, and minimal consequences to learning and memory. Sci. Rep. 10, 7979. https://doi.org/10.1038/s41598-020-64823-6.

Mugnier, M.R., Cross, G.A.M., and Papavasiliou, F.N. (2015). The in vivo dynamics of antigenic variation in Trypanosoma brucei. Science 347, 1470–1473. https://doi.org/10.1126/science.aaa4502.

Oberle, M., Balmer, O., Brun, R., and Roditi, I. (2010). Bottlenecks and the maintenance of minor genotypes during the life cycle of Trypanosoma brucei. PLoS Pathog. 6, e1001023. https://doi.org/10.1371/journal.ppat.1001023.

Pfeiffer, J.K., and Kirkegaard, K. (2006). Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. Proc.

Natl. Acad. Sci. USA *103*, 5520–5525. https://doi.org/10.1073/pnas.0600834103.

Plantinga, E.A., Bosch, G., and Hendriks, W.H. (2011). Estimation of the dietary nutrient profile of free-roaming feral cats: possible implications for nutrition of domestic cats. Br. J. Nutr. *106*, S35–S48. https://doi.org/10.1017/S0007114511002285.

Profaci, C.P., Munji, R.N., Pulido, R.S., and Daneman, R. (2020). The blood-brain barrier in health and disease: important unanswered questions. J. Exp. Med. *217*, e20190062. https://doi.org/10.1084/jem.20190062.

Ribet, D., and Cossart, P. (2015). How bacterial pathogens colonize their hosts and invade deeper tissues. Microbes Infect. *17*, 173–183. https://doi.org/10.1016/j.micinf.2015.01.004.

Rico, E., Jeacock, L., Kovářová, J., and Horn, D. (2018). Inducible high-efficiency CRISPR-Cas9-targeted gene editing and precision base editing in African trypanosomes. Sci. Rep. *8*, 7960. https://doi.org/10.1038/s41598-018-26303-w.

Robinson, D.G., Chen, W., Storey, J.D., and Gresham, D. (2014). Design and analysis of Bar-seq experiments. G3 (Bethesda) *4*, 11–18. https://doi.org/10.1534/g3.113.008565.

Rodgers, J. (2010). Trypanosomiasis and the brain. Parasitology *137*, 1995–2006. https://doi.org/10.1017/S0031182009991806.

Shen, B., Brown, K.M., Lee, T.D., and Sibley, L.D. (2014). Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9. mBio *5*, e01114–01114. https://doi.org/10.1128/mBio.01114-14.

Sidik, S.M., Hackett, C.G., Tran, F., Westwood, N.J., and Lourido, S. (2014). Efficient genome engineering of Toxoplasma gondii using CRISPR/Cas9. PLoS One *9*, e100450. https://doi.org/10.1371/journal.pone.0100450.

Sidik, S.M., Huet, D., Ganesan, S.M., Huynh, M.H., Wang, T., Nasamu, A.S., Thiru, P., Saeij, J.P.J., Carruthers, V.B., Niles, J.C., and Lourido, S. (2016). A genome-wide CRISPR screen in Toxoplasma identifies essential apicomplexan genes. Cell *166*, 1423–1435.e12. https://doi.org/10.1016/j.cell.2016.08.019.

Sinden, R.E. (2017). Developing transmission-blocking strategies for malaria control. PLoS Pathog. *13*, e1006336. https://doi.org/10.1371/journal.ppat.1006336.

Smith, A.M., Heisler, L.E., Mellor, J., Kaper, F., Thompson, M.J., Chee, M., Roth, F.P., Giaever, G., and Nislow, C. (2009). Quantitative phenotyping via deep barcode sequencing. Genome Res. *19*, 1836–1842. https://doi.org/10.1101/gr.093955.109.

Vincent, I.M., Creek, D., Watson, D.G., Kamleh, M.A., Woods, D.J., Wong, P.E., Burchmore, R.J.S., and Barrett, M.P. (2010). A molecular mechanism for eflornithine resistance in African trypanosomes. PLoS Pathog. *6*, e1001204. https://doi.org/10.1371/journal.ppat.1001204.

Waldman, B.S., Schwarz, D., Wadsworth, M.H., 2nd, Saeij, J.P., Shalek, A.K., and Lourido, S. (2020). Identification of a master regulator of differentiation in Toxoplasma. Cell *180*, 359–372.e16. https://doi.org/10.1016/j.cell.2019.12.013.

Watson, G.F., and Davis, P.H. (2019). Systematic review and meta-analysis of variation in Toxoplasma gondii cyst burden in the murine model. Exp. Parasitol. *196*, 55–62. https://doi.org/10.1016/j.exppara.2018.12.003.

Wohlfert, E.A., Blader, I.J., and Wilson, E.H. (2017). Brains and brawn: toxoplasma infections of the central nervous system and skeletal muscle. Trends Parasitol. *33*, 519–531. https://doi.org/10.1016/j.pt.2017.04.001.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
DH5α strain *E. coli*	NEB	Cat#C2987I
XL1 Blue	Agilent	Cat#200249
Critical commercial assays		
DNeasy Blood & Tissue Kit	Qiagen	Cat#69504
Kapa HiFi HotStart Readymix	Roche Diagnostics	Cat#07958927001
Nextera XT Index Kit	Illumina	Cat# TG-131-2001
Qubit™ 1X dsDNA HS Assay Kits	Invitrogen	Cat# Q33230
D1000 Tapestation Screen Tape & Reagents	Agilent	Cat# 5067-5582 Cat# 5067-5583
Deposited data		
All NGS Raw Reads	This paper	NCBI SRA BioProject ID: PRJNA839765
Deposited NGS raw reads	This paper	https://doi.org/10.6084/m9.figshare.20357913
Experimental models: Cell lines		
human foreskin fibroblasts	ATCC (American Type Culture Collection)	HFF-1 ATCC® SCRC-1041™
Experimental models: Organisms/strains		
T. gondii RHΔ*ku80*	Moritz Treeck	N/A
T. gondii PruΔ*ku80*	Moritz Treeck	N/A
T. brucei brucei Lister 427	Gloria Rudenko	N/A
Mouse C57BL/6	Jackson Laboratories	RRID:IMSR_JAX:000664Info
Mouse CBA/J	Jackson Laboratories	RRID:IMSR_JAX:000656
Mouse Swiss Webster	Jackson Laboratories	RRID:IMSR_JAX:034608Info
Oligonucleotides		
Toxoplasma gondii Barcoding Oligos	Life Technologies Ltd	Table S2
Trypanosome brucei brucei Barcoding Oligos	Life Technologies Ltd	Table S2
Recombinant DNA		
pSAG1::Cas9-U6::sgUPRT vector	Moritz Treeck	Addgene #54467
pPOTv4	Gloria Rudenko	N/A
Software and algorithms		
Prism	Graphpad	https://www.graphpad.com/scientific-software/prism/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Matthew Child (m.child@imperial.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All next-generation sequencing data have been deposited at NCBI Sequence Read Archive and are publicly available as of the date of publication. Fastq files for the bar-seq analysis and raw counts in this study have been deposited in the Short Read

Archive (SRA) at https://www.ncbi.nlm.nih.gov/sra/ under BioProject ID SRA: PRJNA839765 and Figshare (https://doi.org/10.6084/m9.figshare.20357913).

- This paper does not report original code.
- Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Toxoplasma gondii cell culture

T. gondii parasite strains were maintained by serial passage in confluent human foreskin fibroblasts (HFF-1 ATCC® SCRC-1041™). HFFs were cultured at 37°C with 5% CO2 in Dulbecco's Modified Eagle's medium supplemented with 10% foetal bovine serum and 2 mM L-glutamine. Tachyzoites were harvested via mechanical syringe lysis of heavily infected HFFs through a 25-gauge needle. *RH*Δku80 parasites were used for *in vivo* and *in vitro* studies. *Pru*Δku80 parasites were used in *in vivo* experiments where chronic infections were established.

Animal work

Six-week old female C57BL/6 or CBA/J mice were purchased from Jackson Laboratories. Mice were acclimated for seven days prior to infection. Six- to eight-week old female Swiss Webster mice (originally form Jackson Laboratories) were obtained from the University of Virginia Centre for Comparative Medicine foster and sentinel colony. For studies using CBA/J and Swiss Webster mice, the animal protocols were approved by the University of Virginia Institutional Animal Care and Use Committee (protocol # 4107-12-18). All animals were housed and treated in accordance with AAALAC and IACUC guidelines at the University of Virginia Veterinary Centre for Comparative Medicine. The procedures involving C57BL/6 (multiplex experiment) mice were approved by the local ethical committee of the Francis Crick Institute Ltd, Mill Hill Laboratory and are part of a project license approved by the Home Office, UK, under the Animals (Scientific Procedures) Act 4001986.

Trypanosome brucei brucei cell culture

We used a bloodstream form Lister 427 strain *T. brucei brucei* for all experiments. The cells were cultured at 37°C, 5% CO2 in HMI-9 medium supplemented with 20% heat inactivated Foetal Bovine Serum, 100 U/ml penicillin and 100 μg/ml streptomycin (Gibco).

METHOD DETAILS

Generation of barcoded Toxoplasma gondii

Generation of barcoded *T. gondii* strains and libraries: 60-nucleotide single-stranded oligos were designed to include a unique six nucleotide barcode sequence flanked by a stop codon and homology regions on either side. Barcodes were designed using the DNA barcode designer and decoder, nxcode (http://hannonlab.cshl.edu/nxCode/nxCode/main.html). The sequences of all oligos within the 96-member library can be found in key resources table. Barcoded libraries of tachyzoites were generated using two alternative strategies: For strategy A, 96 independent transfections were carried out in 16 well Nucleocuvette strips. 10 μg of the pSAG1::Cas9-U6::sgUPRT vector and 10 μg of the barcode oligo (equivalent to an~1:160molar ratio of plasmid to oligo) were co-transfected into approximately 1×10^6 extracellular tachyzoites using the 4D-Nucleofector X Unit programme F1-115 (Lonza). 24 hours post-transfection, transgenic barcoded parasites were selected for using 5 μM 5'-fluro-2'-deoxyuridine (FUDR). Barcoded strains were independently maintained, and only pooled just prior to use. For strategy B, a single "one-pot" transfection was carried out. An oligo library pool containing roughly equal amounts of all barcode oligos was prepared. The ratio of the pSAG1::Cas9-U6::sgUPRT vector to the total oligo pool was the same as in strategy A, though here the final concentration of any single oligo within the pool was~100-fold less. Transfection and selection were performed as for A, with the complex barcoded strain library generated and maintained as a single population.

Generation of barcoded Trypanosome brucei brucei

60-nucleotide double-stranded oligos were designed to include a unique six nucleotide barcode sequence (bold, upper case) flanked by a stop codon (lower case) and 24 bp homology regions on either side:

TGCAATCAGAAGACGAGGTTTAAGtag**AAACAC**tgaCTCACACTAACCGTTTCGATTTAC.

Amino acid transporter *AAT6* (Tb927.8.5450) was selected as a suitable locus for barcode integration. DNA encoding the sgRNA sequence targeting the *AAT6* locus (GTTTAAGTTCACATTGTCGC) was generated by PCR as described in (Rico et al., 2018), ethanol precipitated, and 10 μg was mixed with 10 ng pre-annealed oligonucleotides. The mixture (20 μl total volume) was added to ~10^7 Cas9 expressing bloodstream form Lister 427 *T. b. brucei* cells in 100 μl Amaxa buffer and electroporated using Amaxa Nucleofector IIb (Lonza) program X-001. Transfected cells were immediately added to pre-warmed HMI-9 medium containing 270 μM eflornithine. Transfected cell cultures were passaged under selection every two days. Drug resistant parasites were harvested seven days after transfection for genomic DNA isolation. PCR amplicons encompassing the barcoded region of the *AAT6* locus (from four independent transfections) were generated using ORF-specific PCR primer sequences (5′ to 3′) ATGAGAGAGCCGATACAAACTTCAAC and

TCAGAGTTCAGCAATGACGCTG. Barcode integration was confirmed by Sanger sequencing of these amplicons. For the one-pot transfection strategy, complementary single-stranded barcoding oligos were annealed to produce 96 unique double stranded barcoding repair templates. Annealed barcoding oligos were then pooled, and used in a single transfection as described above

NGS library preparation

Frozen cell pellets of parasites were thawed to room temperature and genomic DNA extracted using the DNeasy Blood & Tissue Kit (Qiagen). Genomic DNA libraries were prepared following the 16S Metagenomic Sequencing Library Preparation guide (Illumina). In brief, an~300 base pair amplicon region encompassing the 6 nt barcode sequence was amplified (30 cycles) from the barcoded UPRT locus in *Toxoplasma gondii* using primer sequences (5′ to 3′)

    TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtggatgtgtcataccatggagtttcctg   and
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGtgttttagtgtaacaaagtggacagcagc.

    The barcoded AAT6 locus in *Trypanosome brucei brucei* was amplified using primer sequences (5′ to 3′)

    TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGatgcaatcagaagacgaggtttaagtag
    GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGagcgagttcaatactaacaataaccagg.

    These primer sequences include the specified Illumina adapter overhang sequences (bold, uppercase). AMPure XP beads were used to purify the resulting PCR product. An indexing PCR (10 cycles) was carried using the purified product as the template to add dual indices and sequencing adapters to the amplicon using the Nextera XT Index Kit (Illumina). Indexed libraries were then cleaned using AMPure XP beads and quantified on the Quantus Fluorometer using the QuantiFluor ONE dsDNA System (Promega). Amplicons were purity-checked and sized on a TapeStation using D1000 ScreenTape System (Agilent). For each NGS run, typically 8 to 25 uniquely indexed libraries were pooled at equimolar concentrations for multiplexed outputs on either an Illumina MiSeq or NextSeq sequencer using the MiSeqV3 PE 75 bp kit or NextSeq 500/550 Mid Output v2.5 PE 75 bp kit respectively. PhiX DNA spike-in of 20% was used in all NGS runs. Following acquisition, sequencing data was demultiplexed and total sample reads extracted from fastq files using the Galaxy web platform (www.usegalaxy.org). Within Galaxy, sequencing reads were concatenated, trimmed, and split into the respective barcodes. Phred QC scores for all NGS runs were >30 with the exception of a single run used for analysis of technical and biological replicates, which still gave an acceptable score of 28. Following trimming to the appropriate 6 nt region a stringent barcode mismatch tolerance of 0% was applied, typically resulting in 10-15% of total reads being discarded. Barcode read data was analysed using Prism 8 and correlations coefficients calculated within the software using Pearson analysis. Testing the sensitivity of the NGS pipeline (Figure 1E), a 96 well plate of 96 uniquely barcoded strains was set up (using transfected drug resistant parasites), with 10,000 parasites/well. For the plate (12 columns x 8 rows), a serial two-fold dilution was performed across the 12 columns, and all rows in the final column pooled after the final dilution. Genomic DNA was then prepared from the final pooled sample and processed for NGS as described.

In vivo experiments

For intraperitoneal infection the pooled barcode parasite library was expanded on HFFs in a T175 flask. Once full parasite vacuoles were observed, parasites were scraped and syringe lysed, counted on a haemocytometer and diluted to an inoculum of 37,000 viable parasites (data represented in Figure 3) or 12,000 viable parasites (data represented in Figure 4) in 200 μL of PBS per mouse. The numbers of viable parasites in the IP infection inoculums were determined by plaque assay. At the time of inoculation 2×10^6 parasites were frozen as an initial population control. In addition, three inoculum control samples were expanded immediately on HFF T25 flasks. After 48 or 72 hours three to five mice were euthanized to isolate parasites in the peritoneal exudate. Specifically, 10 mL of PBS was injected by 25G needle into the peritoneal cavity, mice were rocked vigorously, and peritoneal fluid removed by syringe. Parasites and exudate cells were washed twice in 10 mL of media containing penicillin/streptomycin, pelleted at 1,500 rpm and plated on HFFs T25 flasks. Parasites were harvested when they approached full lysis of the monolayer pelleted and frozen for genomic DNA isolation. After 28 days or three months, the remaining mice were euthanized. Carcasses were incubated in 20% bleach for 10 minutes and the brain was excised in the biosafety cabinet under sterile conditions. To isolate parasites the brains were mashed though a 70 μm filter using 25mL PBS with 5% FBS and penicillin/streptomycin. Brain mash was pelleted for 10 minutes at 1,500 rpm, washed twice with PBS and penicillin/streptomycin then plated on HFF monolayers in T75 flasks. After 36 hours, media was changed to remove debris. Parasites were harvested by syringe lysis when the HFF monolayer was nearly lysed out (approximately two weeks), pelleted and frozen for genomic DNA isolation. To confirm cyst formation in the brain at one month (28 days) or three months post infection,1/50th of the mash was reserved, fixed in 4% paraformaldehyde for 15 minutes then stained with a 1:500 dilution of dolichos bifluorus agglutinin conjugated to FITC in PBS (Vector Labs). FITC-positive cysts were confirmed by fluorescence and morphology under 20x magnification and the total cyst burden per brain was back-calculated.

QUANTIFICATION AND STATISTICAL ANALYSIS

Bottleneck analysis and chord distance calculations

Genetic selection bottlenecks experienced within the murine host were estimated by calculating changes in the relative frequencies of barcodes within dynamic T. gondii populations in relation to the starting population in the inoculum. The following equations 26 were used to calculate chord distance:

$$D_{ch} = \frac{2\sqrt{2}}{\pi}\sqrt{1 - \cos\theta}$$

$$\cos\theta = \sum_{i=1}^{k}\sqrt{f_{p1,i}f_{p2,i}}$$

D_{ch} = chord distance
k = number of distance alleles
$f_{p1,i}$ = frequency of allele i in population 1
$f_{p2,i}$ = frequency of allele i in population 2

Pearson correlation coefficients, significance and error
All other statistical analysis was conducted on GraphPad, Prism®.