Title: The host brain is permissive to colonization by *Toxoplasma gondii*

Authors: Wincott, CJ.[1], Sritharan, G.[1,2], Bunyan, M.[3], Alves, E.[1], Benns, HJ.[1,4], Frickel, EM.[3,5], Ewald, SE.[6], Child, MA.[1*]

* indicates corresponding author

Author affiliations:

[1] Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

[2] Department of Biological Sciences, Birkbeck, University of London, Malet 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] Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston B15 2TT, UK

[6] Department of Microbiology, Immunology and Cancer Biology at the Carter Immunology Center, University of Virginia School of Medicine, Charlottesville VA 228908 USA

Summary

Pathogenic infections and the diseases they cause are defined by invasion and colonization of distinct host cell types and tissue niches. In the case of viruses and bacteria, molecular and cellular barcoding has shaped our understanding of within-host pathogen population dynamics, and informed therapeutic intervention strategies. Host brain colonization is a clinically untreatable feature of persistent infection by the eukaryotic pathogen *Toxoplasma gondii,* and the process remains poorly understood. The host blood-brain barrier is expected to physically restrict parasite colonization of this tissue niche and force the infection through a selection bottleneck, however tools and technologies to test this hypothesis have not been available. Here, we have developed a simple CRISPR-based method to barcode *Toxoplasma* parasites, and then used complex libraries of barcoded parasites to define how the different phases of an infection shape the pathogen population structure. Unexpectedly, we have discovered that the murine host brain does not restrict parasite colonization, with the population structure predominantly shaped by a bottleneck experienced during the acute phase of infection. These data support an evolutionary strategy to maximize genetic diversity of parasite persister cells within the intermediate host brain for subsequent transmission into the definitive feline host.

Main text

Microbial colonization of tissues and organs within the host organism is a key feature of host-pathogen interactions, and often responsible for the pathology of the associated disease (Ribet and Cossart, 2015). For example, colonization of the brain by the eukaryotic protozoan parasite *Toxoplasma gondii* during the chronic phase of an infection can lead to toxoplasmic encephalitis in immunocompromised hosts (Luft et al., 1993). The spatial redistribution of a *T. gondii* infection as it transitions from the broadly distributed acute to the skeletal muscle and brain-associated chronic phase is accompanied by cellular differentiation; the parasite converts from the tachyzoite to the slower growing encysted bradyzoite (Wohlfert et al., 2017). Bradyzoite differentiation in niches like the brain is essential for transmission, and therefore competition prior to and during brain colonization is expected to have an impact on long-term success of a clone. Significant advances have been made in understanding how tachyzoites differentiate into persister-like bradyzoites (Barrett et al., 2019), with the transcription factor BFD1 recently shown to be a master developmental regulator (Waldman et al., 2020). In contrast, the effect of this developmental transition upon

the parasite's population structure during colonization of the immune-privileged brain niche remains an intangible mystery. One critical function of the blood-brain-barrier (BBB) is to physically restrict the access of pathogens such as *T. gondii* to the brain (Kim, 2008, Elsheikha and Khan, 2010). Therefore, it is anticipated that the BBB imposes a selection bottleneck upon the *T. gondii* population as the chronic infection is established, but we lack the tools to determine if this is true.

DNA-based molecular barcoding has been instrumental for understanding how the host and the infection process 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 quasispecies transmitted to the murine brain (Pfeiffer and Kirkegaard, 2006). Furthermore, studies using Wild-type Isogenic Tagged Strains (WITS) of *Salmonella* have provided key insights into selection bottlenecks experienced during colonization of distinct tissues and organs across the different phases of the infection (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 predominant colonizing strain also the dominant strain transmitted by super shedders (Lam and Monack, 2014). The generation of WITS requires the insertion of molecular barcodes into the genome of the infectious agent (Grant et al., 2008). These cellular barcodes then function as neutral alleles, allowing the complexity of the population to be closely monitored and mapped over the course of an infection using quantitative next-generation sequencing (NGS) approaches. Molecular barcoding has provided unique insights into the population structures of both viral and prokaryotic infectious disease, but due to technical limitations has not been used to study the within-host population structure of eukaryotic pathogens.

Here, we define how the acute-to-chronic transition and brain colonization by *T. gondii* shapes the active infection population structure of this eukaryotic pathogen. We establish a simple, scalable CRISPR-based single-strand oligo recombineering strategy to efficiently generate libraries of molecularly barcoded *T. gondii* strains. In the absence of environmental pressures these barcode libraries are stable, enabling quantitative analysis of the effect of the intact host organism environment upon parasite population dynamics. We then use these libraries to investigate the influence of the host organism upon the infection population structure of *T. gondii,* revealing that the brain niche is unexpectedly permissive to colonization by this ubiquitous pathogen.

Results

Toxoplasma tachyzoites can be molecularly barcoded with a simple CRISPR-based strategy

Inspired by WITS (Grant et al., 2008), we first sought to establish a system to molecularly barcode the eukaryotic pathogen, *T. gondii*. A CRISPR-Cas9 single-strand oligonucleotide recombineering strategy was designed for site-specific integration of molecular barcodes into the genome of *Toxoplasma* tachyzoites (**Fig. 1a**). Parasites lacking the non-homologous end joining pathway (Huynh and Carruthers, 2009) (RHΔ*ku80*) were co-transfected with a single plasmid encoding both the Cas9 nuclease and guide RNA (gRNA) scaffold (Shen et al., 2014), and a unique single-stranded oligo donor template encoding the molecular barcode. Targeting of the Cas9 nuclease to the *UPRT* locus promoted efficient integration of a 60-nucleotide single-stranded donor template. Barcode integration disrupted the *UPRT* coding sequence, conferring resistance to the prodrug 5-fluorodeoxyuridine (FUDR) and enabling positive selection of successfully barcoded parasite strains (Donald and Roos, 1995). Our strategy also deleted both the protospacer DNA sequence recognized by the CRISPR gRNA and the protospacer adjacent motif (PAM), preventing further modification of the site following a single barcode integration event. Sanger sequencing of an amplicon derived from the *UPRT* locus in the drug-resistant parasite population confirmed the expected genomic rearrangement following barcode oligo integration (**Fig. 1b**). This demonstrated that individual parasites encode single barcodes within amplicons derived from the *UPRT* locus, and validated the use of short single-stranded donor templates for HDR in *Toxoplasma*. This strategy allows for unique single barcodes to be inserted at the same genomic position, enabling population genetic studies.

Barcode alleles can be identified and quantified in complex populations

Following successful design and implementation of our molecular barcoding strategy, a multiplexed platform was established for 96-well plate-based transfections. This enabled parallel production of plate-mapped 96-member libraries of molecularly barcoded parasites. To quantify the relative representation of individual barcodes within a complex 96-member library pool we employed an NGS pipeline based on amplicon deep sequencing of the barcoded *UPRT* locus, combined with barcode counting (Smith et al., 2009) (**Fig. 1c**). A specific amplicon encompassing the molecular barcode was amplified from parasite genomic DNA (**Fig. S1**). The purified amplicon was indexed and sequenced on an Illumina platform (MiSeq

or NextSeq), and the data processed using Galaxy (Afgan et al., 2018). We first tested our ability to identify all 96 variants within a single population. Following multiplexed transfections, all 96 uniquely barcoded strains were combined to create a mixed library pool and then processed within our pipeline. We successfully identified all 96 uniquely barcoded strains (**Fig. 1d**). Biological replicates of our multiplexed transfections, and technical replicates with unique NGS libraries prepared from the same genomic material were compared (**Fig. S2a**). We observed the highest variation between independent biological experiments (**Fig. S2b**). Independently indexed libraries generated from single genomic samples were well correlated with the relative representation of individual barcodes within complex populations, indicating excellent technical reproducibility (**Fig. S2c and d**).

To determine the sensitivity of the NGS readout we defined the relationship between parasite number, barcode frequency within the complex population, and NGS reads. Using a plate-based two-fold dilution series of the barcoded strain library followed by NGS and barcode counting, we observed a positive correlation between parasite input and read output from 10,000 parasites/barcode down to a lower limit of ~39 parasites/barcode, with a Pearson correlation coefficient (PCC) r = 0.9954 within this range (**Fig. 1e**). This confirmed that barcodes could be successfully identified and reliably quantified at relative frequencies as low as 0.0002 (0.02%). Combined these data demonstrated that our pipeline was able to identify individual barcodes within libraries of at least 96-member complexity with high sensitivity.

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* in the absence of host-organism selective pressure, the pooled library of barcoded parasites was serially passaged through human foreskin fibroblasts (HFFs). Lysed-out parasite cultures were sampled every passage (~36 hours) for a period of six passages, equal to a minimum of six complete lytic growth cycles (invasion, replication, egress). The genetic complexity of the barcode population *in vitro* was remarkably stable, with the PCC testing the relationship between barcode abundance at the beginning and end of each 36-hour passage consistently >0.96 (**Fig. 2a, S3a-e)**. This was emphasized by the comparison of the first and last passages (P1 *vs.* P6), which had a PCC of 0.95, demonstrating that the genetic complexity of the 96-member barcode library was stably maintained over multiple rounds of lytic growth (**Fig. S3f**).

Following confirmation of *in vitro* stability, we tested our ability to propagate and recover the barcode library *in vivo* within a murine host. For the purposes of this pilot experiment we focused on a 36-hour intra peritoneal infection, reasoning that host immune pressure would be low at this early infection timepoint. Thus the relative stability of the population *in vivo* would be minimally influenced by the host immune system that might complicate interpretation. Four separate inoculums of the complex barcode library pool (0.25×10^6, 0.5×10^6, 1×10^6 or 2×10^6 parasites) were injected intraperitoneally into C57BL/6 mice, and the infection allowed to proceed for 36 hours before retrieval of parasites from the peritoneal cavity. For this pilot experiment, the poor growth of a subset of transfected parasites limited the barcode library complexity within the inoculum to 63 strains. Comparison of the infection input with the retrieved output samples showed positive correlation for the highest inoculum, PCC = 0.98 (**Fig. 2b**). This demonstrated that for the highest inoculum and within a 36-hour *in vivo* infection window, the genetic complexity of the barcode library was stable and could be successfully retrieved from the murine host. Combined these data confirmed that, in the absence of significant selective pressures, barcode library complexity was stably maintained *in vitro* and *in vivo*. Reduced correlation observed for the lower dose infections may be attributed to an artificial bottleneck created by inoculum dilution and not to any host-derived process (**Fig. S4a-c)**.

Following demonstration that barcoded libraries of parasites could be successfully generated by multiplexed transfection, we tested the limits of our method. We hypothesized that the CRISPR-Cas9 strategy used to promote integration of barcode oligos by HDR could support simultaneous integration of multiple individual barcodes across an entire population, within a single transfection. This would simplify complex barcode library production into a single "one-pot" transfection using widely available cuvette-based transfection apparatus. To test this, a single transfection was performed on a single isolated population of parasites using the CRISPR-Cas9 approach described, but with all 96 HDR barcoding oligos combined into a single pool. This pooled transfection was performed in parallel in representative type I and type II parasite strains (RHΔ*ku80* and PruΔ*ku80* respectively), using the same mixed pool of 96 barcode oligo repair templates. FUDR-resistant parasite populations were established, genomic DNA isolated, amplicons prepared, and NGS libraries sequenced. We identified all 96 barcodes within the sequence reads generated from the integration-specific amplicon (**Fig. S5a, b**). Variation between the relative frequencies of individual barcode alleles within the population was observed, indicating the possibility of an unanticipated barcode

effect upon parasite fitness. However, comparison of the two independently transfected parasite strains exhibited correlated frequency distributions for the different barcodes (**Fig. 2c, S5c**). We therefore concluded that differences were likely due to the high sensitivity of the readout identifying minor differences in the relative abundance of each barcode HDR template within the pool used for transfection. These data also indicated a higher level of biological reproducibility for one-pot pool transfected libraries than those generated using the original multiplexed transfection approach (**Fig. S2b**). One-pot pool-transfected barcoded populations were used for all subsequent experiments.

Cellular barcodes reveal the population structure of a eukaryotic pathogen infection

The transition of the *T. gondii* infection *in vivo* from the acute to the chronic stage is typified by the spatial redistribution of the parasite into skeletal muscle and the central nervous system (Wohlfert et al., 2017). Colonization of these environments is accompanied by the differentiation of parasites from the tachyzoite lifecycle form into slower growing encysted bradyzoites, which is important for parasite transmission (Barrett et al., 2019). The restrictive nature of the BBB is well documented (Profaci et al., 2020) and we hypothesized that the spatial, temporal, and developmental transitions occurring during the acute-to-chronic infection *in vivo* would impose bottlenecks upon the parasite population represented in the brain. We sought to test this by quantifying changes in the relative frequency distribution of our 96 neutral barcode alleles. An infection study was conceived to assess population structure changes over the course of a full infection, focused on determining the width of the bottleneck experienced by the parasite population as it colonizes the brain of the CBA/J murine host (**Fig. 2d**). We compared changes in the frequency distribution of barcodes in the inoculum, at the acute phase of infection in the peritoneal cavity (at 48 hours) and once chronic infection was established in the brain at day 28. Comparison of independent replicate NGS runs for individual libraries indicated a very high degree of reproducibility (**Fig. S5d**), allowing us to confidently compare samples multiplexed across different NGS runs. Considering the 36-hour pilot infection data using lower inoculums (**Fig. S4**), we anticipated that the dilution of parasite cultures for the inoculum could affect relative frequencies of barcodes within the population and influence our interpretation of data outputs. To control for this, the initial transfected pool was compared to inoculum samples (37,000 tachyzoites) re-expanded *in vitro* (**Fig. 2d**). The artificial bottleneck imposed by dilution led to changes in the relative frequencies of individual barcode, observed through a loss of correlation between the samples (**Fig. S5e**). Although relative

frequencies changed, overall library coverage was retained with all 96 barcodes still represented within the population. Independently expanded populations of the same 37,000 parasite inoculum were highly correlated with one another (**Fig. S5f-h**), confirming that *in vitro* expansion of *in vivo* samples would minimally affect library composition. At 48 hours, shifts in the relative frequencies of barcodes within the peritoneal exudate populations were compared to the averaged barcode representation in the inoculum samples (**Fig. 2e, f**), and these changes were unique to each animal. This indicated that stochastic selective pressures likely drive initial selective sweeps during the early stages of the acute infection, and emphasized the need to consider each host organism as a unique environment. After 28 days, brains of chronically infected mice were isolated, and parasites expanded *in vitro*. Unexpectedly, most barcodes were identified in each individual host brain, with extinction events infrequently observed (**Fig. 2g, Table S1**). At the 28-day timepoint the cumulative extinction frequency across all 14 mice was 0.007 (10/1344), with extinction events associated with lower frequency barcodes within the starting infection population (**Table S3)**. We also noted that extinction events were observed in NGS runs with lower 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**). This indicated that any selection bottleneck experienced by parasites during the colonization of the brain niche must be broad. Taken together these data suggest that with an inoculum of 37,000 parasites (roughly 370 parasites representing each barcode) limited founder effects act on the genetic diversity of the parasite population colonizing the brain.

The brain niche is permissive to colonization by T. gondii

To better understand the process of colonization, we sought to quantify the effective population size in the brain and estimate the width of the acute-to-chronic infection bottleneck. Sequence tag-based analysis of microbial populations (STAMP) (Abel et al., 2015) frames molecular barcodes as neutral alleles, enabling classical population genetic theory and equations for effective population size estimates (N_e) (Charlesworth, 2009) to be applied to NGS datasets. The relative frequencies of all barcodes in a complex population are converted into a single metric of population diversity, which provides a direct estimation of founder population size, and approximates bottleneck width ($N_e \approx N_b$). Due to fundamental differences in virulence of *Toxoplasma* strains (Behnke et al., 2011, Reikvam and Lorentzen-Styr, 1976), we anticipated that the absolute founder number would be strongly influenced by the genetic background of both the parasite and

the host (Watson and Davis, 2019). In light of this we sought to quantify relative changes in N_b as an indication of populations having been through a genetic selection bottleneck, analysing our dataset with the equations defined in the STAMP methodology (Abel et al., 2015). Our initial analyses indicated that compared to the infection inoculum, the population within the brain had experienced a restrictive bottleneck resulting in an ~100-fold reduction in N_b (**Fig. 2h**). Future experiments including further calibration of N_b are necessary before any conclusive statement regarding the absolute founder population number can be accurately made (Abel et al., 2015). Despite this, the consistency in the calculated N_b between the different individual host animals was remarkable. To further assess how N_b changed over the course of the infection we performed the same analysis of read data obtained from parasites populations within the peritoneal cavity after 48 hours, retrieved as the acute infection is being established. Surprisingly, the change in N_b at this early acute phase of the infection relative to the inoculum was similar that seen at 28 days (**Fig. 2h**). For one infection, the change in N_b was intermediate to that observed for the other mice in the group. Due to the terminal nature of parasite retrieval from the host, it was not possible to directly compare the relative frequency distribution of barcodes for a single inoculum replicate at the two points of infection in a single animal. However, on the basis of these data we hypothesize that the most restrictive bottleneck experienced by *T. gondii* during the course of an infection is within the early acute phase, with the BBB imposing negligible further restriction upon the parasite population diversity during colonization of the brain niche. This interpretation of the relative N_b values for the different stages of the infection is consistent with the detection of most barcoded parasite strains within the brains of chronically infected mice (**Fig. 2g**). Calculation of genetic distance (chord distance) is an additional way to quantify the genetic similarity of complex populations derived from one another, such as those obtained from different tissues and phases of infection (Abel et al., 2015, Cavalli-Sforza and Edwards, 1967). Bottlenecks lead to substantial genetic drift between populations, with greater chord distances indicative of a loss of population "relatedness". Applying these equations, we found that there was no significant difference in chord distances between the populations at the different stages of infection relative to the starting population (**Fig. 2i**). These data indicate that there is no further genetic drift experienced by the parasite population following the initial acute phase as the brain is colonized. This further supports our data suggesting an absence of restrictive genetic selection bottlenecks during the establishment of the chronic phase of the *T. gondii* infection.

Discussion

A key role of the BBB is to restrict pathogen access to the immune-privileged brain niche (Kim, 2008). Unexpectedly our findings imply multiple, unique colonization events occur in the brain during chronic infection. Neuronal tissue is rich in fats, and in the case of prey species such as mice, represents a high-energy nutritional food component in the diets of predator species such as felids (Plantinga et al., 2011). Feline consumption of high-energy infected neuronal tissue of prey provides *T. gondii* with a route back into the definitive feline host (Dubey, 1997), and is an evolutionary strategy which increases the likelihood of feline transmission. 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. From the evolutionary perspective of the parasite this would be highly advantageous. The synergistic alignment of this macro scale ancient predator-prey relationship with the molecular features of the *T. gondii* host-pathogen interaction is truly remarkable.

The discoveries made in this study would have not been possible without a means to barcode this eukaryotic pathogen. Molecular 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, Lam and Monack, 2014, Lim et al., 2014, Kaiser et al., 2013), and we anticipate that barcoded *Toxoplasma* strains will have similar far reaching application. While barcode sequencing strategies have been leveraged for eukaryotic pathogen phenotypic screens (Alsford et al., 2011, Bushell et al., 2017, Sidik et al., 2016), to our knowledge this is the first use of cellular barcodes to study the within-host infectious population structure of a eukaryotic pathogen. Our simple oligo barcoding strategy can be applied to any system accessible to CRISPR that contains endogenous or engineered negative selection markers. Our approach allows for parallel integration of even greater numbers of unique barcodes which, combined with STAMP (Abel et al., 2015), will provide increased precision in future host-pathogen population genetic studies and quantification of the absolute founder population number. With particular application to this complex host-pathogen interaction, researchers can now probe the within-host population genetics of the infection with an unprecedented degree of molecular resolution.

Methods

Parasite 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. RHΔ*ku80* parasites were used for *in vivo* and *in vitro* studies. PruΔ*ku80* parasites were used in in vivo experiments where chronic infections were established. Parasite strains were received as a kind gift from Dr Moritz Treeck.

Generation of barcoded *T. gondii* strains and libraries: 60-nucleotide single strand oligos were designed to include a unique 6 nucleotide barcode sequence flanked by a stop codon and homology regions on either side. The sequences of all oligos within the 96-member library can be found in **Table S2.** 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 (Shen et al., 2014) and 10 µg of the barcode oligo (equivalent to an ~1:160 molar 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 oligo pool was the same as in strategy A, though here the final concentration of any single oligo within the pool will be ~100-fold less. Transfection and selection was performed as for A, with the complex barcoded strain library generated and maintained as a single population.

NGS library preparation: Frozen cell pellets of extracellular tachyzoites 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 containing the 6 nt barcode sequence was amplified from the barcoded UPRT locus using primer sequences (5′ to 3′)

TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGtggatgtgtcataccatggagtttcctg and **GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**tgttttagtgtaacaaagtggacagcagc. 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 was carried using the purified product as the template to ligate 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. Following acquisition, sequencing data was demultiplexed and total sample reads extracted from fastq files using the Galaxy web platform. Sequencing reads were then 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.

Mice: 6-week old female C57BL/6 or CBA/J mice were purchased from Jackson Laboratories. Mice were acclimated for 7 days prior to infection. For studies using CBA/J 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 Service Center. The procedures involving C57BL/6 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 1986.

Mouse infections: 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 in 200 µL of PBC per mouse. The viability of parasites in the inoculum was confirmed 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 hours three 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 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 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 2 weeks), pelleted and frozen for genomic DNA isolation. To validate cyst formation in the brain at 28 days 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 under 20x magnification.

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. Calculations of bottleneck width (N_b) were performed according to method outlined by Abel *et al* (Abel et al., 2015) using the following equations:

$$\hat{F} = \frac{1}{k}\sum_{i=1}^{k}\frac{(f_{i,s} - f_{i,0})^2}{f_{i,0}(100 - f_{i,0})}$$

$$N_b \approx N_e = \frac{g}{\hat{F} - \frac{1}{S_0} - \frac{1}{S_s}}$$

k = number of distinct alleles
$f_{i,0}$ = frequency of allele i at time 0
$f_{i,s}$ = frequency of allele i at sampling
g = number of generations during competitive growth
S_0 = sample size at time 0
S_s = sample size at sampling

The following equations (Cavalli-Sforza and Edwards, 1967) 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 distinct alleles
$f_{P1,i}$ = frequency of allele i in population 1
$f_{P2,i}$ = frequency of allele i in population 2

References

ABEL, S., ABEL ZUR WIESCH, P., CHANG, H. H., DAVIS, B. M., LIPSITCH, M. & WALDOR, M. K. 2015. Sequence tag-based analysis of microbial population dynamics. *Nat Methods,* 12**,** 223-6, 3 p following 226.

AFGAN, E., BAKER, D., BATUT, B., VAN DEN BEEK, M., BOUVIER, D., CECH, M., CHILTON, J., CLEMENTS, D., CORAOR, N., GRUNING, B. A., GUERLER, A., HILLMAN-JACKSON, J., HILTEMANN, S., JALILI, V., RASCHE, H., SORANZO, N., GOECKS, J., TAYLOR, J., NEKRUTENKO, A. & BLANKENBERG, D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res,* 46**,** W537-W544.

ALSFORD, S., TURNER, D. J., OBADO, S. O., SANCHEZ-FLORES, A., GLOVER, L., BERRIMAN, M., HERTZ-FOWLER, C. & HORN, D. 2011. High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res,* 21**,** 915-24.

BARRETT, M. P., KYLE, D. E., SIBLEY, L. D., RADKE, J. B. & TARLETON, R. L. 2019. Protozoan persister-like cells and drug treatment failure. *Nat Rev Microbiol,* 17**,** 607-620.

BEHNKE, M. S., KHAN, A., WOOTTON, J. C., DUBEY, J. P., TANG, K. & SIBLEY, L. D. 2011. Virulence differences in Toxoplasma mediated by amplification of a family of polymorphic pseudokinases. *Proc Natl Acad Sci U S A,* 108**,** 9631-6.

BLUNDELL, J. R. & LEVY, S. F. 2014. Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. *Genomics,* 104**,** 417-30.

BUSHELL, E., GOMES, A. R., SANDERSON, T., ANAR, B., GIRLING, G., HERD, C., METCALF, T., MODRZYNSKA, K., SCHWACH, F., MARTIN, R. E., MATHER, M. W., MCFADDEN, G. I.,

PARTS, L., RUTLEDGE, G. G., VAIDYA, A. B., WENGELNIK, K., RAYNER, J. C. & BILLKER, O. 2017. Functional Profiling of a Plasmodium Genome Reveals an Abundance of Essential Genes. *Cell,* 170**,** 260-272 e8.

CAVALLI-SFORZA, L. L. & EDWARDS, A. W. 1967. Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet,* 19**,** 233-57.

CHARLESWORTH, B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet,* 10**,** 195-205.

DONALD, R. G. & 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 U S A,* 92**,** 5749-53.

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 (Pt 1)**,** 15-20.

ELSHEIKHA, H. M. & KHAN, N. A. 2010. Protozoa traversal of the blood-brain barrier to invade the central nervous system. *FEMS Microbiol Rev,* 34**,** 532-53.

GRANT, A. J., RESTIF, O., MCKINLEY, T. J., SHEPPARD, M., MASKELL, D. J. & MASTROENI, P. 2008. Modelling within-host spatiotemporal dynamics of invasive bacterial disease. *PLoS Biol,* 6**,** e74.

HUYNH, M. H. & CARRUTHERS, V. B. 2009. Tagging of endogenous genes in a Toxoplasma gondii strain lacking Ku80. *Eukaryot Cell,* 8**,** 530-9.

KAISER, P., SLACK, E., GRANT, A. J., HARDT, W. D. & REGOES, R. R. 2013. Lymph node colonization dynamics after oral Salmonella Typhimurium infection in mice. *PLoS Pathog,* 9**,** e1003532.

KEBSCHULL, J. M. & ZADOR, A. M. 2018. Cellular barcoding: lineage tracing, screening and beyond. *Nat Methods,* 15**,** 871-879.

KIM, K. S. 2008. Mechanisms of microbial traversal of the blood-brain barrier. *Nat Rev Microbiol,* 6**,** 625-34.

KUSS, S. K., ETHEREDGE, C. A. & PFEIFFER, J. K. 2008. Multiple host barriers restrict poliovirus trafficking in mice. *PLoS Pathog,* 4**,** e1000082.

LAM, L. H. & MONACK, D. M. 2014. Intraspecies competition for niches in the distal gut dictate transmission during persistent Salmonella infection. *PLoS Pathog,* 10**,** e1004527.

LIM, C. H., VOEDISCH, S., WAHL, B., ROUF, S. F., GEFFERS, R., RHEN, M. & PABST, O. 2014. Independent bottlenecks characterize colonization of systemic compartments and gut lymphoid tissue by salmonella. *PLoS Pathog,* 10**,** e1004270.

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.

PFEIFFER, J. K. & KIRKEGAARD, K. 2006. Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. *Proc Natl Acad Sci U S A,* 103**,** 5520-5.

PLANTINGA, E. A., BOSCH, G. & 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 Suppl 1**,** S35-48.

PROFACI, C. P., MUNJI, R. N., PULIDO, R. S. & DANEMAN, R. 2020. The blood-brain barrier in health and disease: Important unanswered questions. *J Exp Med,* 217.

REIKVAM, A. & LORENTZEN-STYR, A. M. 1976. Virulence of different strains of Toxoplasma gondii and host response in mice. *Nature,* 261**,** 508-9.

RIBET, D. & COSSART, P. 2015. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect,* 17**,** 173-83.

SHEN, B., BROWN, K. M., LEE, T. D. & SIBLEY, L. D. 2014. Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9. *mBio,* 5**,** e01114-14.

SIDIK, S. M., HUET, D., GANESAN, S. M., HUYNH, M. H., WANG, T., NASAMU, A. S., THIRU, P., SAEIJ, J. P., CARRUTHERS, V. B., NILES, J. C. & LOURIDO, S. 2016. A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential Apicomplexan Genes. *Cell,* 166**,** 1423-1435 e12.

SMITH, A. M., HEISLER, L. E., MELLOR, J., KAPER, F., THOMPSON, M. J., CHEE, M., ROTH, F. P., GIAEVER, G. & NISLOW, C. 2009. Quantitative phenotyping via deep barcode sequencing. *Genome Res,* 19**,** 1836-42.

WALDMAN, B. S., SCHWARZ, D., WADSWORTH, M. H., 2ND, SAEIJ, J. P., SHALEK, A. K. & LOURIDO, S. 2020. Identification of a Master Regulator of Differentiation in Toxoplasma. *Cell,* 180**,** 359-372 e16.

WATSON, G. F. & 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.

WOHLFERT, E. A., BLADER, I. J. & WILSON, E. H. 2017. Brains and Brawn: Toxoplasma Infections of the Central Nervous System and Skeletal Muscle. *Trends Parasitol,* 33**,** 519-531.

Acknowledgements:

This work was supported by grants NC/S001239/1 from the NC3Rs (to C.J.W and M.A.C), and 202553/Z/16/Z from the Wellcome Trust & Royal Society (to M.A.C.). This work was also supported by the Francis Crick Institute (FC001076 to EMF), which receives its core funding from Cancer Research UK, the UK Medical Research Council, and the Wellcome Trust. EMF was supported by a Senior Wellcome Trust Fellowship (217202/Z/19/Z). The Imperial BRC Genomics Facility provided resources and support that contributed to the research results reported within this paper. The Imperial BRC Genomics Facility is supported by NIHR funding to the Imperial Biomedical Research Centre. Ivan Andrew, JP Haywood and Laurence Game at the UKRI London Institute of Medical Sciences Genomics Laboratory. We would also like to thank Dr Ellen M. McDonagh for critical reading of the manuscript.

Author Contributions: Conception/design of the work: EF, SEE, MAC. Acquisition/analysis/interpretation of data: CJW, GS, MB, EDSA, HJB, SEE, MAC. Manuscript drafting and revision: CJW, GS, EDSA, HJB, EF, SEE, MAC

Competing interests: The authors declare no competing interests.

Materials & Correspondence: All correspondence and material requests should be addressed to MAC, m.child@imperial.ac.uk

Figure 1

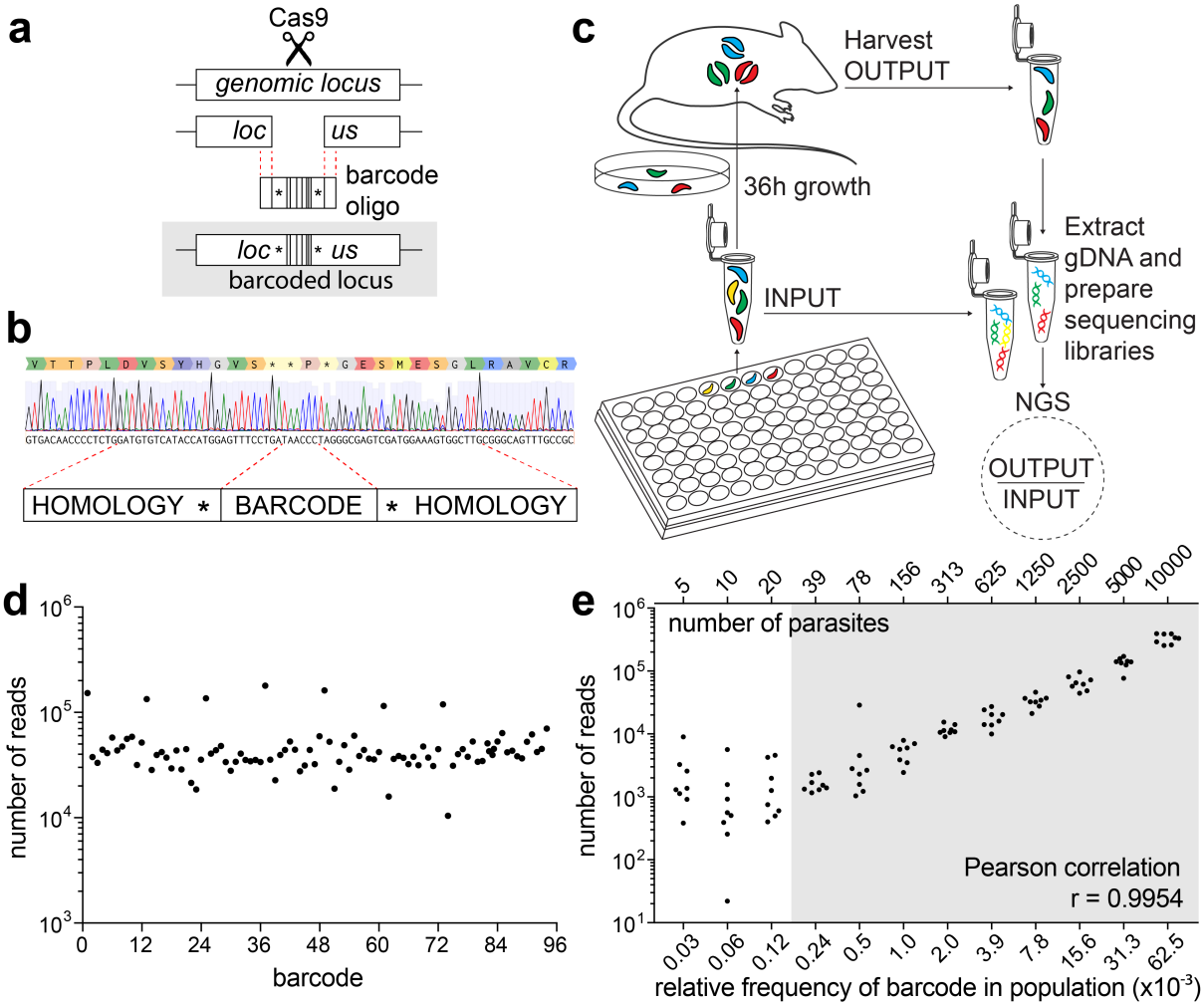


Figure 1: *T. gondii* tachyzoites can be molecularly barcoded using a simple CRISPR-based strategy. a) Schematic of molecular barcoding strategy. A CRISPR guide RNA targets Cas9 to the *UPRT* locus, where Cas9 endonuclease activity introduces a double strand break. This break is repaired by HDR, with the co-transfected 60mer barcode oligo used as the donor template. **b**) Sanger sequencing confirmation that barcode integration results in the simultaneous disruption of the *UPRT* coding sequence, 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, with individual barcodes identified and quantified within NGS datasets. Scatter plot 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. Scatter plot presents number of reads for a serial two-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 (two-tailed) = <0.0001.

Figure 2

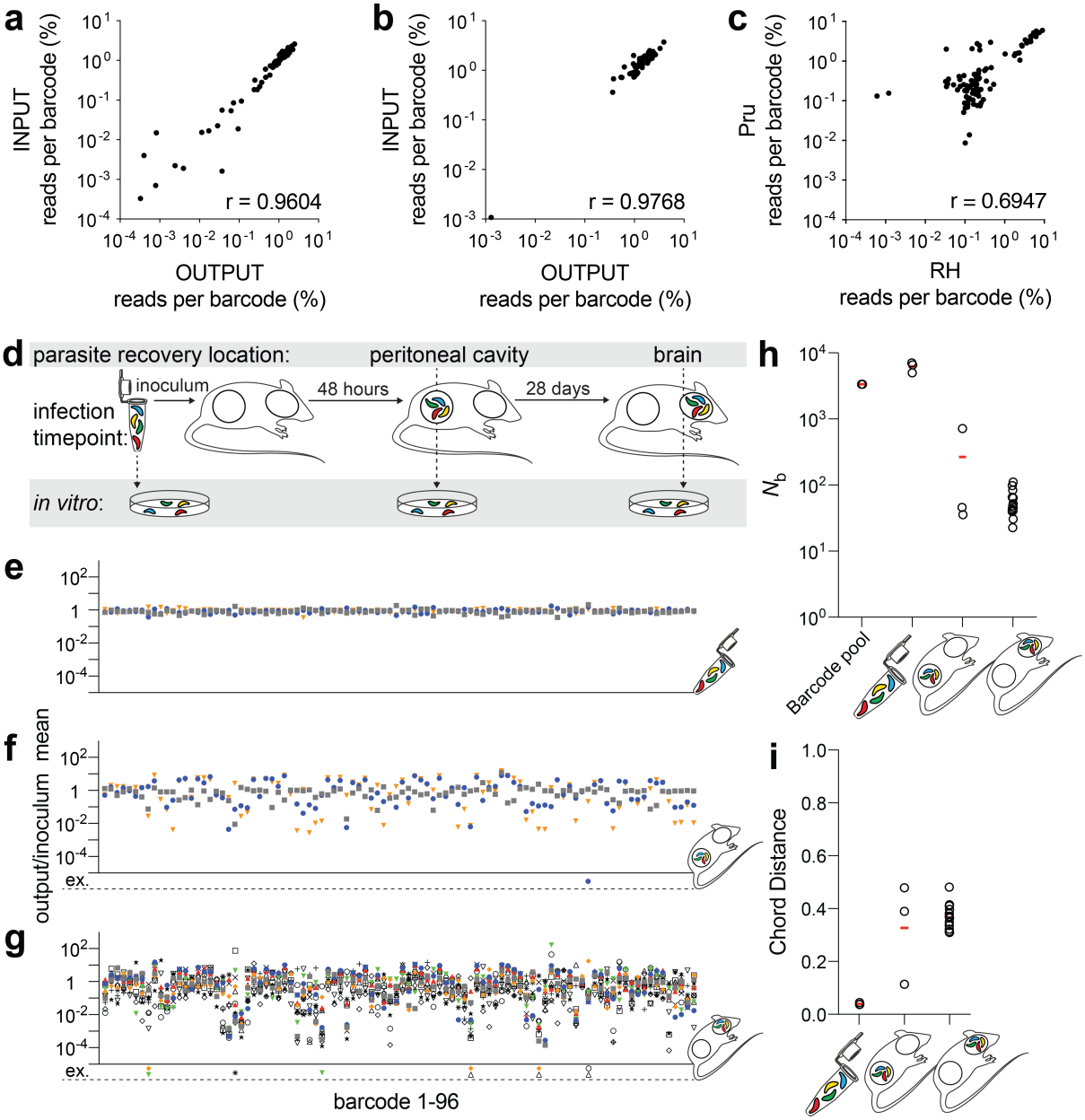


Figure 2: Barcoded libraries of *T. gondii* parasites are stably maintained *in vitro* and *in vivo*, and can be generated via one-pot transfections. a) The population structure of complex barcode libraries is stable *in vitro*. Scatter plot comparing individual barcode frequencies within the complex pooled library population for an INPUT sample (used for the *in vitro* infection) and the OUTPUT sample harvested following a single lytic growth cycle (~36 hours). PCC r = 0.96, n = 96, P (two-tailed) = <0.0001. **b)** The population structure of complex barcode libraries is stable during early stages of an *in vivo* infection. Scatter plot comparing individual barcode frequencies within a pooled library population for an INPUT sample (the peritoneal infection inoculum) and the OUTPUT sample harvested from the peritoneal cavity following 36 hours of

546 growth. PCC r = 0.98, n = 63, P (two-tailed) = <0.0001. **c**) Scatter plot of relative barcode frequencies
547 comparing barcoded libraries of parasites generated for two different *T. gondii* strains (RHΔ*ku80* and
548 PruΔ*ku80*) where the one-pot transfection was performed using the same pool of barcoding oligos. PCC r =
549 0.69, n = 96, P (two-tailed) = <0.0001. PCC values provided on scatter plots indicate degree of correlation
550 between different populations being compared. **d)** Schematic of an experiment investigating changes in the
551 population structure of *T. gondii* over the course of a 28-day *in vivo* infection. The infection is artificially
552 initiated within the peritoneal cavity (, and the infectious population structure monitored at an early acute (48
553 hours) and chronic phase (28 days). The timepoints are compared to the inoculum mean to identify host
554 bottlenecks restricting diversity. **e-g**) Scatter plot of changes in individual barcode frequencies relative to the
555 inoculum mean in the inoculum (**e**), the peritoneal cavity at 48 hours, n = 3 (**f**), and the brain at 28 days, n =
556 14 (**g**). Barcode extinctions (ex.) are indicated below the x-axis in the corresponding position to the absent
557 barcoded strain (see also table S1). **h**) Scatter plot of calculated N_b values for parasite populations presents
558 relative changes in total population barcode frequencies during the establishment of the acute and chronic
559 infection in the peritoneal cavity and brain respectively. **i**) Scatter plot of calculated chord distances presents
560 the relatedness and genetic distance between two distinct population samples (acute and chronic) relative to
561 the starting population in the inoculum.
562
563
564
565
566
567
568
569
570
571
572
573
574

Supplemental Material

Table S1

Infection timepoint:	48 hours			28 days													
Mouse #	1	2	3	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Barcode #	-	79	-	-	-	8, 60, 71	-	8, 36	-	-	79	-	60, 71, 79	-	-	-	22
Total read #	$14x10^6$	$0.59x10^6$	$2.3x10^6$	$8.3x10^6$	$10x10^6$	$0.62x10^6$	$6.3x10^6$	$0.46x10^6$	$8.0x10^6$	$3.3x10^6$	$1.8x10^6$	$4.1x10^6$	$0.73x10^6$	$5.0x10^6$	$1.7x10^6$	$4.2x10^6$	$0.92x10^6$

Table S1: Details of barcode extinctions occurring during the different phases of infection. Barcode numbers and the mouse in which the extinction was observed are noted. Extinctions are defined by an absence of the barcode sequence within the processed NGS read data.

Figure S1

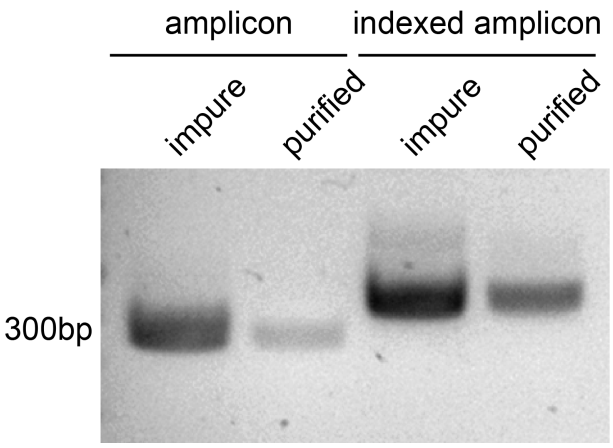
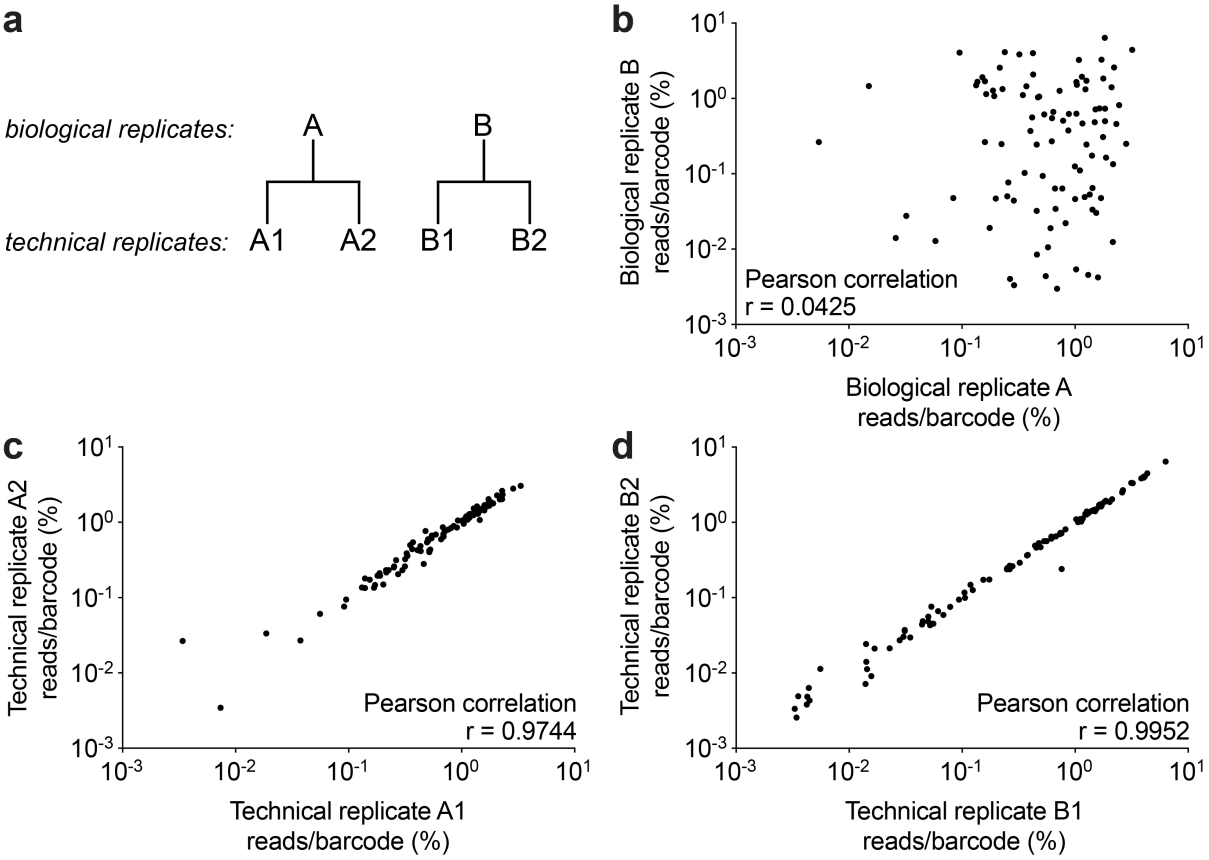


Figure S1: Construction of NGS libraries for amplicon deep sequencing. A single ~300 bp amplicon was amplified from genomic DNA, and purified (amplicon, lanes 1 and 2). The purified amplicon was then indexed and re-purified prior to quantification and sizing (indexed amplicon, lanes 3 and 4).

595 **Figure S2**



596

597

598 **Figure S2: Analysis of experimental variation for barcoded parasite libraries and NGS pipeline. a)**
599 Strategy to identify the primary major source of variation within the multiplexed transfection barcoding
600 strategy and NGS pipeline. Scatter plots present data comparing the percentage representation of individual
601 barcodes within library pools for (**b**) biological replicates, PCC r = 0.0425, n = 96, P (two-tailed) = 0.6807
602 (n.s.), and technical replicates (**c**) PCC r = 0.9744, n = 96, P (two-tailed) = <0.0001 , and (**d**) PCC r =
603 0.9952, n = 96, P (two-tailed) = <0.0001. PCC values provided on scatter plots indicate degree of correlation
604 between different populations being compared.

605

606

607

608

609

610

611

Figure S3

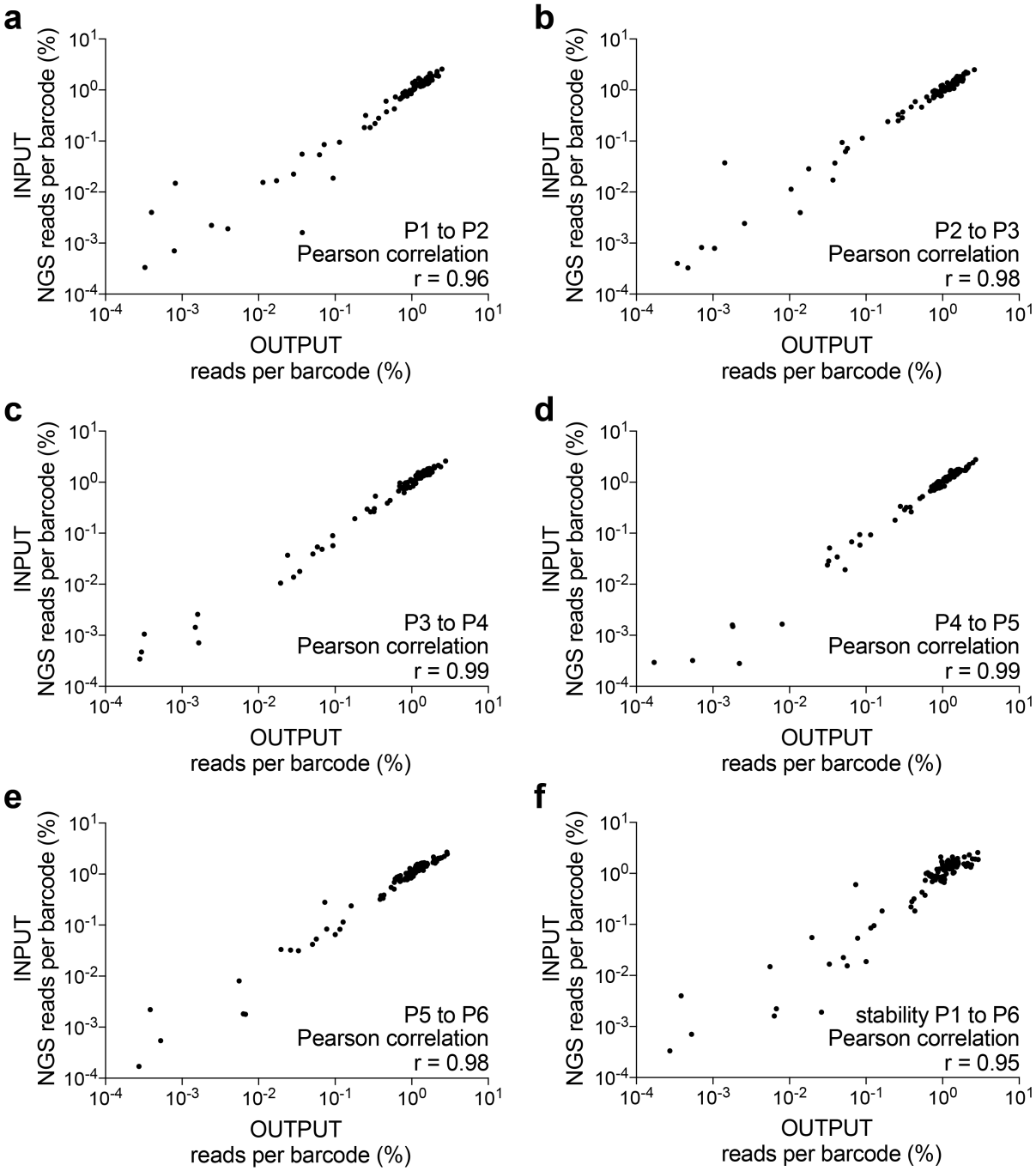


Figure S3: Assessment of barcoded strain library *in vitro* stability. A barcoded library of parasites was serially passaged through HFF host cells, and sampled following the completion of each lytic cycle (~36 hours/passage). Scatter plots present data comparing the percentage representation of individual barcodes within library pools for the passages indicated on the plots. (**a)** is a repeat of data presented in figure 2a, and shown here for completeness and comparison alongside the entire serial passage experiment dataset. PCC values provided on scatter plots indicate degree of correlation between different populations being

compared: PCC **a**) r = 0.96, n = 96, P (two-tailed) = <0.0001. **b**) r = 0.98, n = 96, P (two-tailed) = <0.0001. **c**) r = 0.99, n = 96, P (two-tailed) = <0.0001. **d**) r = 0.99, n = 96, P (two-tailed) = <0.0001. **e**) r = 0.98, n = 96, P (two-tailed) = <0.0001. **f**) r = 0.95, n = 96, P (two-tailed) = <0.0001.

Figure S4

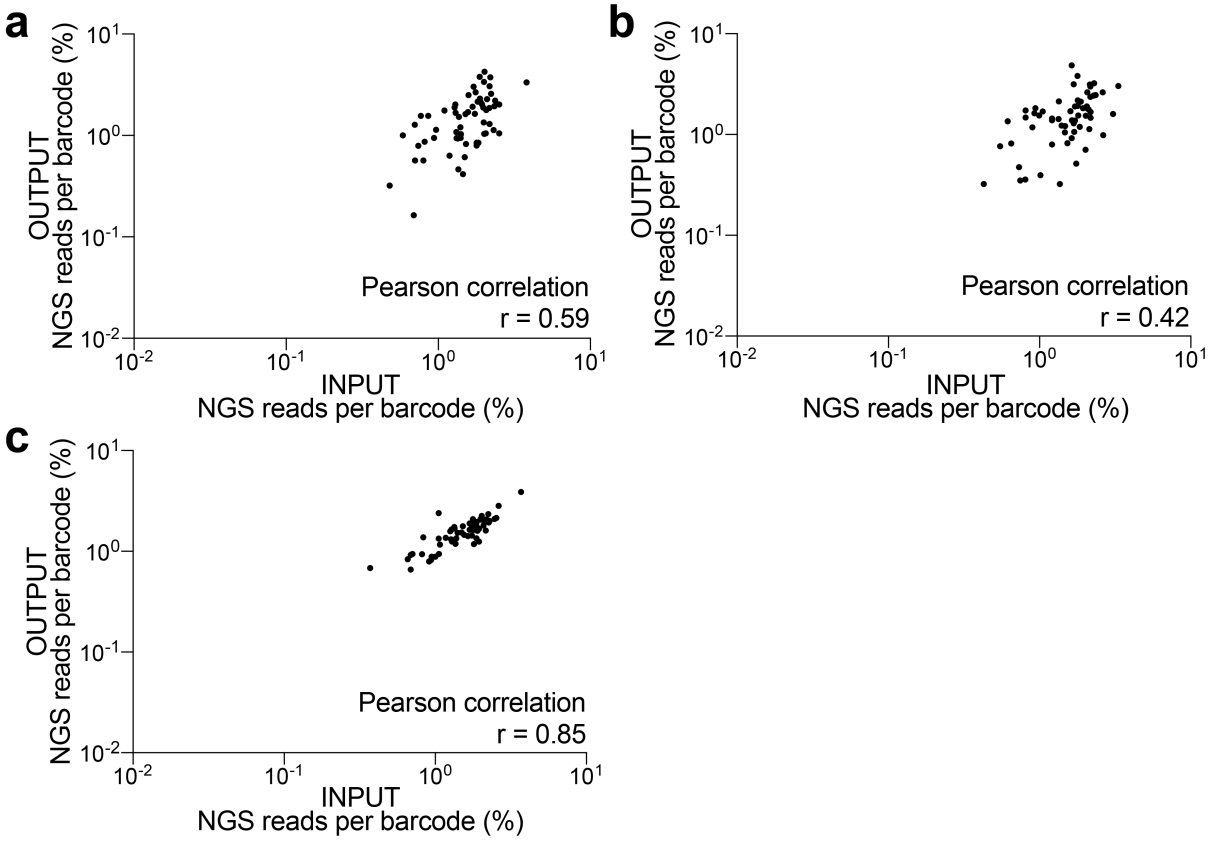


Figure S4: Assessment of *in vivo* stability of barcoded libraries of parasites. Three separate inoculums of the same complex barcode library pool were injected intraperitoneally into mice, allowed to grow for 36 hours, and then retrieved. Scatter plots present data for the percentage representation of individual barcodes within the library pools for parasite infection inoculums (INPUT) compared with the corresponding population retrieved from the peritoneal cavity of the infected host (OUTPUT) for (**a**) 0.25×10^{6} parasites, PCC r = 0.59, n = 62, P (two-tailed) = <0.0001. **b**) 0.5×10^{6} parasites, PCC r = 0.62, n = 62, P (two-tailed) = 0.0006. **c**) 1×10^{6} parasites, PCC r = 0.85, n = 62, P (two-tailed) = <0.0001. PCC values provided on scatter plots indicate degree of correlation between populations being compared.

Figure S5

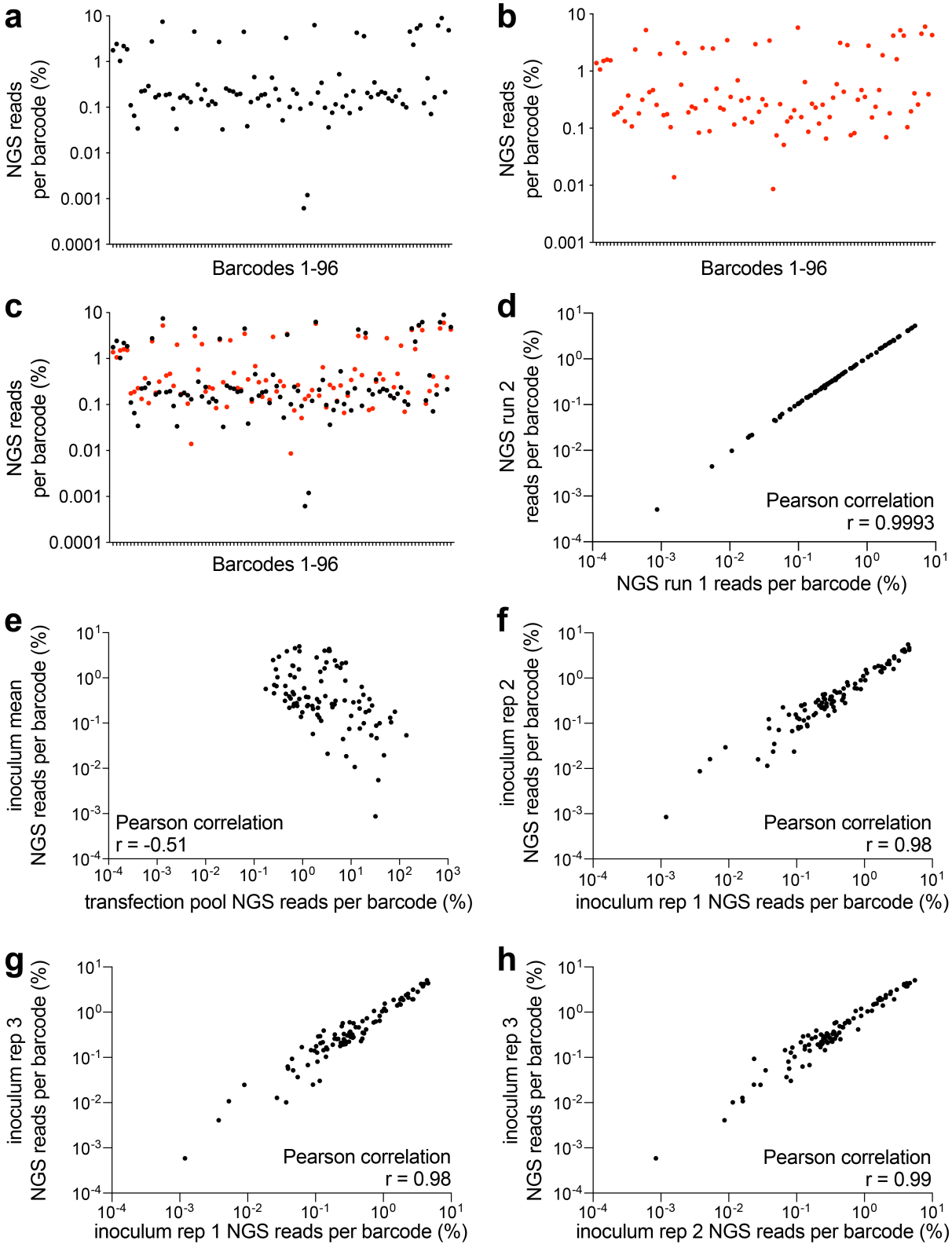


Figure S5: Analysis of molecularly barcoded *T. gondii* libraries generated using a one-pot transfection approach. One-pot transfections using (**a**) RHΔ*ku80* or (**b**) PruΔ*ku80* yield a complex library of barcoded strains with complete representation of all 96 oligos included in the transfection pool. **c**) Overlay of data

673 from the different strain transfections (**b** and **c**). Scatter plots in **a-c** present relative percentage frequency of
674 barcodes within the population, distributed according to barcode identifier (1-96). **d**) Scatter plot of relative
675 percentage frequency of barcodes within a single genomic sample, processed on independent NGS runs.
676 PCC r = 0.99, n = 96, P (two-tailed) = <0.0001. **e-h**) Scatter plots comparing relative percentage frequency
677 of barcode for the one-pot transfection pool versus the expanded inoculum mean (**e**) PCC r = -0.50, n = 96, P
678 (two-tailed) = <0.0001, and comparison of each of the three independently *in vitro*-expanded biological
679 replicates of the inoculum used for the mouse infection (**f**) PCC r = 0.98, n = 96, P (two-tailed) = <0.0001;
680 (**g**) PCC r = 0.98, n = 96, P (two-tailed) = <0.0001; (**h**) PCC r = 0.99, n = 96, P (two-tailed) = <0.0001. PCC
681 values provided on scatter plots indicate degree of correlation between populations being compared.
682