1   **Long title:**

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

3   *gondii* host colonization

4

5   **Short title:**

6   *The murine host brain is promiscuously colonized by Toxoplasma gondii*

7

8   **Authors:** Ceire J. Wincott[1], Gayathri Sritharan[1,2], Henry J. Benns[1,4], Farzana B. Liakath[7], Carla

9   Gilabert-Carbajo[1], Monique Bunyan[3], Aisling R. Fairweather[1], Eduardo Alves[1], Ivan Andrew[5],

10 Laurence Game[5], Eva M. Frickel[3,6], Calvin Tiengwe[1], Sarah E. Ewald[7*], Matthew A. Child[1*].

11

12 * indicates co-corresponding authorship

13

14 **Author affiliations:**

15 [1] Department of Life Sciences, Imperial College London, South Kensington Campus, London SW7
16 2AZ, UK

17 [2] Department of Biological Sciences, Birkbeck, University of London, Malet Street, Bloomsbury
18 London WC1E 7HX, UK

19 [3] Host-Toxoplasma Interaction Laboratory, The Francis Crick Institute, 1 Midland Road, London NW1
20 1BF, UK

21 [4] Department of Chemistry, Imperial College London, White City Campus, London W12 0BZ, UK

22 [5] UKRI London Institute of Medical Sciences Genomics Laboratory, Shepherd's Bush, London W12
23 0NN, UK

24 [6] Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Edgbaston
25 B15 2TT, UK

26 [7] Department of Microbiology, Immunology and Cancer Biology at the Carter Immunology Center,
27 University of Virginia School of Medicine, Charlottesville VA 228908 USA

28

29

30

31

Abstract

Molecular barcoding techniques have emerged as powerful tools to understand microbial pathogenesis. However, barcoding strategies have not been extended to protozoan parasites, which have unique genomic structures and virulence strategies compared to viral and bacterial pathogens. Here, we present a versatile CRISPR-based method to barcode protozoa, which we successfully apply to *Toxoplasma gondii* and *Trypanosoma brucei*. The murine brain is an important transmission niche for *T. gondii,* and brain persistence is a clinically untreatable feature of infection. The blood-brain barrier is expected to physically restrict parasite colonization of this niche, resulting in a selection bottleneck. 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 that represented a minor fraction of the inoculum could become a dominant population in the brain by three months post-infection. Together, these data establish the first, robust molecular barcoding approach for protozoa and evidence that the blood-brain barrier does not represent 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 often responsible for the pathology of the associated disease[1]. 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[2]. Similarly, colonization of the brain niche by the distantly related protozoan *Trypanosoma brucei brucei* is a major contributor to trypanosomiasis pathology[3]. In the case of *T. gondii* infection, the spatial redistribution is linked to parasite differentiation. The acute phase of infection, whether initiated via the natural, per oral route or intra peritoneal inoculation, 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[4]. The 'immune privileged' quality of the chronic infection tissue niches like the brain coupled

2

60 with the protective properties of the *T. gondii* cyst wall are essential for transmission. It follows that

61 intraspecific competition prior to and during brain colonization is expected to have an impact on long-

62 term success of a clone. Despite recent advances in our molecular understanding of tachyzoite-to-

63 bradyzoite differentiation[5,6], our comprehension of *T. gondii*'s within-host population dynamics

64 remains limited. One critical function of tissue barriers, including the blood-brain-barrier (BBB) and

65 the intestinal epithelial barrier, is to physically restrict pathogen access[7,8]. Therefore, it is anticipated

66 that the BBB imposes a selection bottleneck upon the *T. gondii* population as the chronic infection is

67 established. However, we lack tools to determine if the brain is colonized infrequently by a small

68 number of parasites that subsequently expand within the tissue niche or if the host brain broadly

69 permissive to colonization by *T. gondii*.

70   DNA-based molecular barcoding has been instrumental for understanding how the host and the

71 infection process influences genetic complexity of pathogen populations, and colonization dynamics[9,10].

72 Early studies using restriction site-tagged poliovirus identified a bottleneck limiting the genetic

73 diversity of viral quasi-species transmitted to the murine brain[11]. Furthermore, studies using Wild-type

74 Isogenic Tagged Strains (WITS) of *Salmonella* have provided key insights into selection bottlenecks

75 experienced during gastro-intestinal and invasive bacterial colonization[12,13]. The population structure

76 of *Salmonella* infection was found to experience dramatic selective bottlenecks during the colonization

77 of the gut niche, with the predominant colonizing strain also the dominant strain transmitted by super

78 shedders[14]. The generation of WITS requires the insertion of molecular barcodes into the genome of the

79 infectious agent[12]. These cellular barcodes then function as neutral alleles, allowing the complexity of

80 the population to be closely monitored and mapped over the course of an infection using quantitative

81 next-generation sequencing (NGS) approaches. Unfortunately, due to technical limitations molecular

82 barcoding has not been amenable to study the within-host population structure of protozoan pathogens.

83   Here, we establish two simple, scalable CRISPR-based oligo recombineering strategies to

84 efficiently generate libraries of molecularly barcoded *T. gondii* strains. We demonstrate the versatility

85 of our method by successfully barcoding the bloodstream form of *T. brucei.* Taking advantage of tools

86 to infect, monitor and re-isolate *T. gondii* from the natural murine host, we define how the acute-to-

87 chronic transition and brain colonization by *T. gondii* shapes the population structure of this eukaryotic

88   pathogen. In tissue culture infection and in the first days of *in vivo* infection, we confirm that barcoded

89   parasite libraries are stable, enabling quantitative analysis of the effect of the intact host organism

90   environment upon parasite population dynamics. We then use these libraries to investigate the influence

91   of the host organism upon the infection population structure of *T. gondii* revealing that colonization of

92   the murine brain niche is not restricted by a stringent bottleneck at the blood-brain barrier.

93

94   **Results**

95   ***Protozoan pathogens can be molecularly barcoded with a simple CRISPR-based strategy***

96   Inspired by WITS[12], we sought to establish a versatile system to molecularly barcode eukaryotic

97   pathogens. Previously established CRISPR-Cas9 tools for *T. gondii*[15,16] were adapted to mediate site-

98   specific recombination of molecular barcodes into non-essential genes that encode positive selectable

99   markers in *T. gondii* tachyzoites or *T. brucei* trypomastigotes (**Fig. 1A**). To increase recombination

100   efficiency, *T. gondii* lacking the non-homologous end-joining pathway (RHΔ*ku80*) were used[17].

101   Parasites were co-transfected with a plasmid encoding both the Cas9 nuclease and guide RNA (gRNA)

102   scaffold[15], and a unique 60-nucleotide single-stranded donor template encoding the molecular barcode.

103   Successful targeting of Cas9 and genomic integration of the barcode disrupted the *UPRT* coding

104   sequence, conferring resistance to the prodrug 5-fluorodeoxyuridine (FUDR)[18]. To prevent further

105   modification of the *UPRT* locus following a single barcode integration event our strategy also deleted

106   both the protospacer DNA sequence recognized by the CRISPR gRNA and the protospacer adjacent

107   motif (PAM). We confirmed barcode integration and the expected genomic by Sanger sequencing of

108   the *UPRT* locus in the drug-resistant parasite population (**Fig. 1B**). A similar molecular barcoding

109   strategy was applied to the bloodstream form *T. brucei* where Cas9 was targeted to the *AAT6* locus[19], a

110   single-copy non-essential gene that confers sensitivity to eflornithine[20] (**Fig. 1A**). Parasites stably

111   expressing Cas9 were co-transfected with *AAT6*-targeting sgDNAs and a double-stranded barcoding

112   oligo. Successful barcode integration into the *AAT6* locus conferred resistance to eflornithine. Sanger

113   sequencing confirmed the expected genomic rearrangement and barcoding of the *T. brucei* genome

114   (**Fig. 1B**). These data indicate our barcoding strategy is an efficient tool to insert unique a barcode

115   sequence at specific genomic positions in two divergent eukaryotic pathogens.

4

116

Barcode alleles can be identified and quantified in complex populations

117

118 To examine the stability of individual barcodes across a population of uniquely labeled parasites and to

119 evaluate the sensitivity of barcode detection, we focused on *T. gondii*. First, we generated a plate-

120 mapped library of 96 uniquely barcoded parasites lines using multiplexed transfection (**Fig. 1C**). We

121 reasoned that this multiplexed strategy would be useful for future chemical or genetic screening

122 applications applied to each of the 96 uniquely barcoded strains prior to pooling. To quantify the relative

123 representation of individual barcodes after pooling an NGS pipeline was developed. DNA was extracted

124 from the pooled parasite library, the barcoded region of the UPRT locus was amplified for Illumina

125 sequencing, and barcodes representation quantified using Galaxy (**Fig. 1C** and **Fig. S1A**)[21,22]. Our

126 pipeline successfully identified all 96 uniquely barcoded strains from the pooled population (**Fig. 1D**).

127 Notably, wells from one row of the 96 well plate were consistently more highly represented in the

128 pooled population, reflecting a technical challenge of the 96-well transfection method (**Fig. 1D**). While

129 unplanned, this observation suggested that the NGS readout was highly sensitive to differences in

130 barcode frequency in the pooled population. This was confirmed experimentally, as greater variation in

131 barcode representation was observed between biological replicate transfections (**Fig. S1B-C**) than

132 between technical replicates of DNA indexed for NGS sequencing (**Fig. S1B and D-E**), underscoring

133 the reproducibility of the NGS pipeline.

134 To determine the sensitivity of the NGS readout we generated a two-fold dilution series of the

135 pooled parasite population prior to isolating DNA for NGS sequencing. A positive correlation was

136 observed between the number of reads (read output) and the number of parasites in the input from

137 10,000 parasites/barcode down to 39 parasites/barcode (**Fig. 1E,** r = 0.9954). This confirmed that

138 barcodes could be successfully identified and reliably quantified within libraries of at least 96-member

139 complexity at relative frequencies as low as 0.0002 (0.02%).

140

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

141

142 To test whether the complexity of the barcode libraries was stably maintained *in vitro* the pooled library

143 of barcoded parasites was serially passed through human foreskin fibroblasts (HFFs). Lysed-out

144 parasite cultures were sampled every passage (~36 hours) for a period of six passages, equivalent to six

145 lytic growth cycles (invasion, replication, egress). Relative to the input, the genetic complexity of the

146 barcode population *in vitro* was remarkably stable after one lytic cycle or six passages (**Fig. 1F-G**). We

147 next tested our ability to propagate and recover the barcode library *in vivo* within a murine host using a

148 pooled library of 63 barcodes (in this pilot experiment some wells were not efficiently transfected). An

149 inoculum of 2×10^6 parasites was injected intraperitoneally into C57BL/6 mice. After 36 hours parasites

150 were isolated from the peritoneal cavity by lavage and compared to the input population. A strong

151 positive correlation was observed between the input and peritoneal exudate populations (**Fig. 1H**, PCC

152 r = 0.98), demonstrating that the genetic complexity of the multiplex barcode library was stable over

153 the first 36 hours of *in vivo* infection.

154

155 ***Libraries of barcode alleles can be generated by a 'one-pot' transfection method***

156 One challenge of the multiplex platform arose from experiment-to-experiment variability in transfection

157 efficiency across individual wells (**Fig. 1D and 1H**). Our CRISPR-Cas9 strategy was designed to ensure

158 the integration of a single barcode into the UPRT locus by simultaneously deleting the protospacer and

159 PAM motifs recognized by the CRISPR gRNA. We hypothesized that a barcoded library of parasites

160 could be generated by a single transfection with the Cas9-gRNA plasmid and a pooled library of

161 oligonucleotide repair templates using a widely available cuvette-based transfection apparatus (**Fig.**

162 **2A**). This 'one-pot' method was tested on type I RHΔ*ku80* and type II PruΔ*ku80* parasite strains in

163 parallel, using the same mixed pool of 96 barcode oligo repair templates. FUDR-resistant parasite

164 populations were enriched, genomic DNA isolated, UPRT locus amplicons prepared, and NGS libraries

165 sequenced. All 96 barcodes were identified in the sequence reads generated from the UPRT locus

166 amplicon (**Fig. 2B**). When the distribution of barcodes in RHΔ*ku80* and PruΔ*ku80* were compared

167 directly, the two independently transfected parasite strains exhibited correlated frequency distributions

168 for all but the least abundant barcodes, with a Pearson correlation coefficient of 0.7 (**Fig. 2B**). Similar

169 to the multiplex method we confirmed the technical reproducibility of NGS runs by comparing technical

170 replicates of indexed DNA amplicons (**Fig. S2**). We concluded that differences in barcode frequency

171 most likely reflect minor variations in the relative abundance of each barcode oligonucleotide template

172   within the pool used for transfection, rather than a negative impact of some barcodes on parasite fitness.

173   It is notable that a higher level of biological reproducibility was observed for one-pot library

174   transfections than the original multiplexed transfection approach (**Fig. S1C**). We next tested the one-

175   pot strategy in *T. brucei*. We were similarly able to produce complex 96-member libraries of uniquely

176   barcoded strains from a single transfection (**Fig. 2C**). The one-pot transfection protocol was used for

177   all subsequent experiments.

178

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

180   The transition from acute to chronic infection corresponds with the spatial redistribution of *T. gondii* to

181   skeletal muscle and the central nervous system[4]. This is accompanied by tachyzoite differentiation into

182   bradyzoites cysts, which are necessary for parasite transmission[5]. The restrictive nature of the BBB is

183   well documented in other infection models[23]. We hypothesized that these spatial, temporal, and

184   developmental transitions would impose bottlenecks upon the parasite population represented in the

185   brain at chronic infection. We sought to test this by infecting murine hosts intraperitoneally with the

186   pooled library of PruΔ*ku80* parasites stably expressing 96 neutral barcode alleles.

187   First, we confirmed that diluting the PruΔ*ku80* library to a dose tolerated by mice did not disrupt

188   the population structure in a way that could confound the interpretation of *in vivo* experiments. Three

189   inoculum samples of 37,000 viable tachyzoites (determined by plaque assay) were plated on HFFs and

190   re-expanded in tissue culture followed by DNA isolation for NGS sequencing. In each inoculum

191   sample, all 96 barcodes were detected by NGS sequencing and pairwise comparisons of each sample

192   were strongly correlated (**Fig. 2D-F**), confirming that *in vitro* expansion of *in vivo* samples would

193   minimally affect library composition.

194   An *in vivo* experiment was conceived to study the effect of murine host colonization. This

195   would probe the within-host population genetics of the *T. gondii* infection as it progressed from the

196   initial acute infection in the peritoneum to the chronic infection in the brain (**Fig. 3A**). To determine if

197   the parasite population was stable early in acute infection, parasites were isolated from the peritoneal

198   cavity of some mice at 48 hours post-infection and expanded in vitro. Although barcode extinction was

199   rare (one barcode in a single mouse), the relative frequency of barcodes in each peritoneal isolate was

7

200   unique to each animal (**Fig. 3B-C**). This observation supports a model where the initial selective sweep

201   at the onset of acute infection is stochastic and emphasized the need to consider each host organism as

202   a unique environment. After 28 days, the brains of chronically infected mice were isolated, and parasites

203   expanded *in vitro*. Unexpectedly, the majority of barcodes were detected in each host brain (**Fig. 3D,**

204   **Table S1**), consistent with a minimal founder effect on the genetic diversity of the parasite population

205   colonizing the central nervous system. The cumulative extinction frequency across all 14 mice was

206   0.007 (10 of 1,344 total barcodes) and lost barcodes were most often represented at low abundance

207   within the inoculum (**Table S3**). Of note, these extinction events were observed in the NGS runs with

208   the lowest total read counts, suggesting that extinctions could be due to reduced sequence sampling

209   depth rather than a true absence of the specific barcoded parasite(s) in the brain (**Table S3**). Together,

210   these data indicate that any selection bottleneck experienced by parasites during the colonization of the

211   brain niche must be broad.

212

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

214   We next sought to determine how host genetic background influences the dynamics of parasite brain

215   colonization over time. To do this we infected Swiss Webster mice, which are outbred to maximize

216   genetic diversity and heterozygosity, and inbred CBA/J mice (**Fig. 4A**)[24]. Mean parasite cyst burden

217   was not significantly different between CBA/J mice and Swiss Webster mice at one-month post-

218   infection, indicating that a similar number of parasites were accessing the brain niche (**Fig. 4B**). As

219   expected, cyst burden declined over time, but the reduction in mean cyst number was similar between

220   mouse genotypes[24,25].

221   To evaluate the parasite population structure within these hosts we used Cavalli Sforza's chord

222   distance calculation[26]. Chord distance is a geometric measure of population distance. While chord

223   distance does not provide absolute quantification of the width of any bottleneck, it can be used to infer

224   changes in the genetic structure of a population between points A and B. In CBA/J mice (**Fig. 4C**) there

225   was minimal distance separating the inoculum from the peritoneal exudate population 72 hours post-

226   infection. A significant increase in distance was observed between the inoculum and the one-month

227   brain samples. Consistent with the decline in cyst number over time, chord distance was greater at three

228   months post-infection in CBA/J brain. Outbred Swiss Webster mice also exhibited little shift in chord

229   distance from the inoculum to the peritoneal exudate and a significant shift in distance from the

230   inoculum to the one-month post-infection brain samples (**Fig. 4D**). However, the individual Swiss mice

231   exhibited greater variation in parasite population structure at each chronic time point than CBA/J mice.

232   Unlike CBA/J mice, Swiss mice supported parasite populations with a similar median chord distance

233   between inoculum and one or three months post-infection brains.

234   Cumulatively, our data indicated that following an IP infection the murine brain was

235   unexpectedly permissive to colonization (**Fig. 3D**) and parasite population structure was dynamic

236   within individual hosts (**Fig. 4C-D**). In CBA/J mice, the chord distance was greater at three months

237   post-infection than at one-month post-infection (**Fig. 4C**), so we became interested in understanding if

238   the barcoded lineages that were the most abundant in the inoculum had a competitive advantage for

239   long-term persistence. In CBA/J mice, dominant barcode alleles at three months post-infection (**Fig.**

240   **4E**, top row) corresponded to parasite populations that were among the top 35% of reads (top six most

241   abundant barcodes) in the inoculum. While this trend was similar at the one-month post-infection time

242   point, several CBA/J mice harboured a dominant barcode lineage that corresponded to one represented

243   in the bottom 25% of reads in the inoculum (**Fig. 4E** outset in the inoculum pie chart). In agreement

244   with the chord distance analysis results, individual Swiss mice had diverse dominant population

245   structures at three months post-infection (**Fig. 4E**, bottom row). Dominant barcodes in the Swiss brains

246   were often but not always derived from lineages represented within the top 75% of reads in the inoculum

247   (**Fig. 4E** most abundant 15 barcodes), however, this was not always the case. For example, one mouse

248   was dominated by a barcode that was present at low frequency in the inoculum, while another mouse

249   exhibited a population structure where barcode frequency was more evenly distributed than in the

250   inoculum or peritoneal parasite isolates. This heterogeneity was also reflected in the barcodes that

251   dominated Swiss brains at one-month post-infection, suggesting that even relatively low-frequency

252   members of a founder population can establish and maintain persistent infection in a manner that likely

253   depends on stochastic variables and host genotype.

254

255   **Discussion**

256   Brain residency has long been appreciated as a strategy for pathogenesis exploited by *T. gondii*

257   to evade sterilizing immunity, as well as an evolutionary strategy to increases the likelihood of

258   transmission via predation. Feline consumption of high-fat, energy rich infected neuronal tissue of prey,

259   including mice, provides *T. gondii* with a route back into the definitive feline host[27,28]. A means to evade

260   any restrictive bottleneck when colonizing the brain niche would be consistent with this evolutionary

261   strategy, supporting maximal transmission of genetic diversity into the feline host to contribute to

262   recombination in the subsequent sexual cycle. Additionally, *T. gondii* can infect most warm-blooded

263   animals, indicating that intermediate host-specific selection environments will frequently and

264   unpredictably change, favouring different phenotypes at different times. It therefore likely benefits the

265   parasite to maintain genetic diversity and phenotypic plasticity to ensure survival in varied future host

266   species.

267   A key role of tissue barriers is to restrict pathogen access to the underlying tissue niche[7].

268   Unexpectedly our findings imply multiple, unique *T. gondii* colonization events from circulation into

269   the brain parenchyma, rather than rare brain invasions coupled with expansion within that niche. During

270   intraperitoneal infection, the loss of barcodes is stochastic, and most barcodes are identified in the brain

271   parenchyma at one month post-infection (**Fig. 3**). Chord distance indicated a change in the genetic

272   structure of the parasite population in the CNS at one month post-infection (**Fig. 4C-D**). In CBA/J mice

273   and Swiss mice, the dominant barcodes in the brain following intraperitoneal infection tended to be

274   those most frequently observed in the inoculum, however, most barcodes were still detected indicating

275   limited selective pressure imparted by the BBB. However, there were exceptions to this rule,

276   particularly in outbred Swiss mice, where minor constituents of the inoculum were able to dominate the

277   brain population in some individuals.

278   We employed intraperitoneal infection because it facilitates a precise quantification of barcode

279   frequency and coverage across the inoculum dose. This is not possible using bradyzoite cysts as each

280   cyst, which must be harvested from infected mice, has a variable parasite number. In the future, it will

281   be interesting to investigate how host colonization following oral infection shapes the population

282   structure of the parasite, and how it differs from the intraperitoneal route. It should also be noted that

283   the results from our study do not mean that there is *no* bottleneck for infection of the brain niche, simply

284   that with the number of barcode markers used in this study we were not able to determine the bottleneck

285   width.

286   The observations made in this study would have not been possible without a means to barcode

287   this eukaryotic pathogen. Molecular barcoding has provided critical insights into the infection biology

288   of viruses such as poliovirus[11,29], bacteria such as *Salmonella*[12-14,30], and *Escherichia coli*[31]. We

289   anticipate that barcoded *T. gondii* strains will have similar far-reaching applications. While barcode

290   sequencing strategies have been leveraged for eukaryotic pathogen phenotypic screens[32-34], to our

291   knowledge this is the first use of cellular barcodes to study the within-host infectious population

292   structure of a eukaryotic pathogen. The versatility of the strategies presented will allow researchers to

293   work with individual barcoded strains in isolation prior to pooling (*via* plate-based library generation),

294   or with complex libraries of barcoded strains generated through our one-pot approach. We believe this

295   will expand the application of our approach beyond the population genetic studies described herein. It

296   should be noted that disruption of the *UPRT* locus has been documented to negatively impact cyst

297   burden, which would suggest that our data may *underrepresent* barcode diversity in the CNS. However,

298   all barcodes are inserted at the same position, so the cross-comparison of populations is internally

299   controlled in this study. Our simple oligo barcoding strategy can be applied to systems accessible to

300   CRISPR that contains endogenous or engineered negative selection markers, which we successfully

301   demonstrated with the first cellular barcoding of *T. brucei*. In principle, our approach allows for parallel

302   integration of even greater numbers of unique barcodes, which will provide increased precision in future

303   host-pathogen population genetic studies. An increased number of molecular barcodes combined with

304   methods such as Sequence Tag Analysis of Microbial Populations (STAMP)[35], or the recently published

305   update STAMPR[36], will make it possible to quantify the absolute founder population number present

306   within the host brain. Combined with these methods, our molecular barcoding approach will allow

307   researchers to probe the within-host population genetics of tissue colonization during a *T. gondii*

308   infection with an unprecedented degree of molecular resolution.

309

310   **Methods**

311   **Parasite cell culture:** *T. gondii* parasite strains were maintained by serial passage in confluent human

312   foreskin fibroblasts (HFF-1 ATCC® SCRC-1041™). HFFs were cultured at 37°C with 5% CO2 in

313   Dulbecco's Modified Eagle's medium supplemented with 10% foetal bovine serum and 2 mM L-

314   glutamine. Tachyzoites were harvested via mechanical syringe lysis of heavily infected HFFs through

315   a 25-gauge needle. RHΔ*ku80* parasites were used for *in vivo* and *in vitro* studies. PruΔ*ku80* parasites

316   were used in *in vivo* experiments where chronic infections were established. Parasite strains were

317   received as a kind gift from Dr. Moritz Treeck.

318

319   **Generation of barcoded *T. gondii* strains and libraries:** 60-nucleotide single-stranded oligos were

320   designed to include a unique six nucleotide barcode sequence flanked by a stop codon and homology

321   regions on either side. Barcodes were designed using the DNA barcode designer and decoder, nxcode

322   (http://hannonlab.cshl.edu/nxCode/nxCode/main.html). The sequences of all oligos within the 96-

323   member library can be found in **Table S2.** Barcoded libraries of tachyzoites were generated using two

324   alternative strategies: For strategy A, 96 independent transfections were carried out in 16 well

325   Nucleocuvette strips. 10 µg of the pSAG1::Cas9-U6::sgUPRT vector[15] and 10 µg of the barcode oligo

326   (equivalent to an ~1:160 molar ratio of plasmid to oligo) were co-transfected into approximately 1×10^6

327   extracellular tachyzoites using the 4D-Nucleofector X Unit programme F1-115 (Lonza). 24 hours post-

328   transfection, transgenic barcoded parasites were selected for using 5 µM 5'-fluro-2'-deoxyuridine

329   (FUDR). Barcoded strains were independently maintained, and only pooled just prior to use. For

330   strategy B, a single "one-pot" transfection was carried out. An oligo library pool containing roughly

331   equal amounts of all barcode oligos was prepared. The ratio of the pSAG1::Cas9-U6::sgUPRT vector

332   to the total oligo pool was the same as in strategy A, though here the final concentration of any single

333   oligo within the pool was ~100-fold less. Transfection and selection was performed as for A, with the

334   complex barcoded strain library generated and maintained as a single population.

335

336   **Generation of barcoded *T. b. brucei* strains:** 60-nucleotide double-stranded oligos were designed to

337   include a unique six nucleotide barcode sequence (bold, upper case) flanked by a stop codon (lower

338   case) and 24 bp homology regions on either side:

339   TGCAATCAGAAGACGAGGTTTAAGtag**AAACAC**tgaCTCACACTAACCGTTTCGATTTAC.

340   Amino acid transporter *AAT6* (Tb927.8.5450) was selected as a suitable locus for barcode integration.

341   DNA encoding the sgRNA sequence targeting the *AAT6* locus (GTTTAAGTTCACATTGTCGC) was

342   generated by PCR as described in[19,37], ethanol precipitated, and 10 µg was mixed with 10 ng pre-

343   annealed oligonucleotides. The mixture (20 µl total volume) was added to ~10^7 Cas9 expressing

344   bloodstream form Lister 427 *T. b. brucei* cells in 100 µl Amaxa buffer and electroporated using Amaxa

345   Nucleofector IIb (Lonza) program X-001. Transfected cells were immediately added to pre-warmed

346   HMI-9 medium containing 270 µM eflornithine. Transfected cell cultures were passaged under

347   selection every two days. Drug resistant parasites were harvested seven days after transfection for

348   genomic DNA isolation. PCR amplicons encompassing the barcoded region of the *AAT6* locus (from

349   four independent transfections) were generated using ORF-specific PCR primer sequences (5´ to 3´)

350   ATGAGAGAGCCGATACAAACTTCAAC and TCAGAGTTCAGCAATGACGCTG. Barcode

351   integration was confirmed by Sanger sequencing of these amplicons. For the one-pot transfection

352   strategy, complementary single-stranded barcoding oligos were annealed to produce 96 unique double

353   stranded barcoding repair templates. Annealed barcoding oligos were then pooled, and used in a single

354   transfection as described above.

355

356   **NGS library preparation:** Frozen cell pellets of extracellular tachyzoites were thawed to room

357   temperature and genomic DNA extracted using the DNeasy Blood & Tissue Kit (Qiagen). Genomic

358   DNA libraries were prepared following the 16S Metagenomic Sequencing Library Preparation guide

359   (Illumina). In brief, an ~300 base pair amplicon region encompassing the 6 nt barcode sequence was

360   amplified (30 cycles) from the barcoded *UPRT* locus using primer sequences (5´ to 3´)

361   **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**tggatgtgtcataccatggagtttcctg   and

362   **GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**tgttttagtgtaacaaagtggacagcagc.   These

363   primer sequences include the specified Illumina adapter overhang sequences (bold, uppercase).

364   AMPure XP beads were used to purify the resulting PCR product. An indexing PCR (10 cycles) was

365   carried using the purified product as the template to add dual indices and sequencing adapters to the

366   amplicon using the Nextera XT Index Kit (Illumina). Indexed libraries were then cleaned using AMPure

13

367   XP beads and quantified on the Quantus Fluorometer using the QuantiFluor ONE dsDNA System

368   (Promega). Amplicons were purity-checked and sized on a TapeStation using D1000 ScreenTape

369   System (Agilent). For each NGS run, typically 8 to 25 uniquely indexed libraries were pooled at

370   equimolar concentrations for multiplexed outputs on either an Illumina MiSeq or NextSeq sequencer

371   using the MiSeqV3 PE 75 bp kit or NextSeq 500/550 Mid Output v2.5 PE 75 bp kit respectively. PhiX

372   DNA spike-in of 20% was used in all NGS runs. Following acquisition, sequencing data was

373   demultiplexed and total sample reads extracted from fastq files using the Galaxy web platform

374   (www.usegalaxy.org). Within Galaxy, sequencing reads were concatenated, trimmed, and split into the

375   respective barcodes. Phred QC scores for all NGS runs were >30 with the exception of a single run used

376   for analysis of technical and biological replicates, which still gave an acceptable score of 28. Following

377   trimming to the appropriate 6 nt region a stringent barcode mismatch tolerance of 0% was applied,

378   typically resulting in 10-15% of total reads being discarded. Barcode read data was analysed using

379   Prism 8 and correlations coefficients calculated within the software using Pearson analysis. Testing the

380   sensitivity of the NGS pipeline (Fig. 1e), a 96 well plate of 96 uniquely barcoded strains was set up

381   (using transfected drug resistant parasites), with 10,000 parasites/well. For the plate (12 columns x 8

382   rows), a serial two-fold dilution was performed across the 12 columns, and all rows in the final column

383   pooled after the final dilution. Genomic DNA was then prepared from the final pooled sample, and

384   processed for NGS as described.

385

386   **Mice:** Six-week old female C57BL/6 or CBA/J mice were purchased from Jackson Laboratories. Mice

387   were acclimated for seven days prior to infection. Six to eight-week old female Swiss Webster mice

388   (originally form Jackson Laboratories) were obtained from the University of Virginia Center for

389   Comparative Medicine foster and sentinel colony. For studies using CBA/J and Swiss Webster mice,

390   the animal protocols were approved by the University of Virginia Institutional Animal Care and Use

391   Committee (protocol # 4107-12-18). All animals were housed and treated in accordance with AAALAC

392   and IACUC guidelines at the University of Virginia Veterinary Centre for Comparative Medicine. The

393   procedures involving C57BL/6 (multiplex experiment) mice were approved by the local ethical

394   committee of the Francis Crick Institute Ltd, Mill Hill Laboratory and are part of a project license

395   approved by the Home Office, UK, under the Animals (Scientific Procedures) Act 1986.

396

397   **Mouse infections:** For intraperitoneal infection the pooled barcode parasite library was expanded on

398   HFFs in a T175 flask. Once full parasite vacuoles were observed, parasites were scraped and syringe

399   lysed, counted on a haemocytometer and diluted to an inoculum of 37,000 viable parasites (data

400   represented in Figure 3) or 12,000 viable parasites (data represented in Figure 4) in 200 µL of PBS per

401   mouse. The numbers of viable parasites in the IP infection inoculums were determined by plaque assay.

402   At the time of inoculation 2×10^6 parasites were frozen as an initial population control. In addition, three

403   inoculum control samples were expanded immediately on HFF T25 flasks. After 48 or 72 hours three

404   to five mice were euthanized to isolate parasites in the peritoneal exudate. Specifically, 10 mL of PBS

405   was injected by 25G needle into the peritoneal cavity, mice were rocked vigorously, and peritoneal fluid

406   removed by syringe. Parasites and exudate cells were washed twice in 10 mL of media containing

407   penicillin/streptomycin, pelleted at 1,500 rpm and plated on HFFs T25 flasks. Parasites were harvested

408   when they approached full lysis of the monolayer pelleted and frozen for genomic DNA isolation. After

409   28 days or three months, the remaining mice were euthanized. Carcasses were incubated in 20% bleach

410   for 10 minutes and the brain was excised in the biosafety cabinet under sterile conditions. To isolate

411   parasites the brains were mashed though a 70 µm filter using 25 mL PBS with 5% FBS and

412   penicillin/streptomycin. Brain mash was pelleted for 10 minutes at 1,500 rpm, washed twice with PBS

413   and penicillin/streptomycin then plated on HFF monolayers in T75 flasks. After 36 hours, media was

414   changed to remove debris. Parasites were harvested by syringe lysis when the HFF monolayer was

415   nearly lysed out (approximately two weeks), pelleted and frozen for genomic DNA isolation. To

416   confirm cyst formation in the brain at one month (28 days) or three months post infection, 1/50th of the

417   mash was reserved, fixed in 4% paraformaldehyde for 15 minutes then stained with a 1:500 dilution of

418   dolichos bifluorus agglutinin conjugated to FITC in PBS (Vector Labs). FITC-positive cysts were

419   confirmed by fluorescence and morphology under 20x magnification and the total cyst burden per brain

420   was back-calculated.

421

422   **Bottleneck Analysis and Chord Distance Calculations:** Genetic selection bottlenecks experienced

423   within the murine host were estimated by calculating changes in the relative frequencies of barcodes

424   within dynamic *T. gondii* populations in relation to the starting population in the inoculum. The

425   following equations [26] were used to calculate chord distance:

426

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

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

428

429   $$\cos\theta = \sum_{i=1}^{k}\sqrt{f_{P1,i}f_{P2,i}}$$

430

431   **References**

432   1   Ribet, D. & Cossart, P. How bacterial pathogens colonize their hosts and invade deeper tissues.

433   *Microbes Infect* **17**, 173-183, doi:10.1016/j.micinf.2015.01.004 (2015).

434   2   Luft, B. J. *et al.* Toxoplasmic encephalitis in patients with the acquired immunodeficiency

435   syndrome. Members of the ACTG 077p/ANRS 009 Study Team. *The New England journal of*

436   *medicine* **329**, 995-1000, doi:10.1056/NEJM199309303291403 (1993).

437   3   Rodgers, J. Trypanosomiasis and the brain. *Parasitology* **137**, 1995-2006,

438   doi:10.1017/S0031182009991806 (2010).

439   4   Wohlfert, E. A., Blader, I. J. & Wilson, E. H. Brains and Brawn: Toxoplasma Infections of the

440   Central Nervous System and Skeletal Muscle. *Trends Parasitol* **33**, 519-531,

441   doi:10.1016/j.pt.2017.04.001 (2017).

442   5   Barrett, M. P., Kyle, D. E., Sibley, L. D., Radke, J. B. & Tarleton, R. L. Protozoan persister-

443   like cells and drug treatment failure. *Nat Rev Microbiol* **17**, 607-620, doi:10.1038/s41579-019-

444   0238-x (2019).

445   6   Waldman, B. S. *et al.* Identification of a Master Regulator of Differentiation in Toxoplasma.

446   *Cell* **180**, 359-372 e316, doi:10.1016/j.cell.2019.12.013 (2020).

447   7   Kim, K. S. Mechanisms of microbial traversal of the blood-brain barrier. *Nat Rev Microbiol* **6**,

448   625-634, doi:10.1038/nrmicro1952 (2008).

8   Elsheikha, H. M. & Khan, N. A. Protozoa traversal of the blood-brain barrier to invade the central nervous system. *FEMS microbiology reviews* **34**, 532-553, doi:10.1111/j.1574-6976.2010.00215.x (2010).

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

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

11   Pfeiffer, J. K. & Kirkegaard, K. Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. *Proc Natl Acad Sci U S A* **103**, 5520-5525, doi:10.1073/pnas.0600834103 (2006).

12   Grant, A. J. *et al.* Modelling within-host spatiotemporal dynamics of invasive bacterial disease. *PLoS Biol* **6**, e74, doi:10.1371/journal.pbio.0060074 (2008).

13   Lim, C. H. *et al.* Independent bottlenecks characterize colonization of systemic compartments and gut lymphoid tissue by salmonella. *PLoS Pathog* **10**, e1004270, doi:10.1371/journal.ppat.1004270 (2014).

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

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

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

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

476   18   Donald, R. G. & Roos, D. S. Insertional mutagenesis and marker rescue in a protozoan parasite:

477   cloning of the uracil phosphoribosyltransferase locus from Toxoplasma gondii. *Proc Natl Acad*

478   *Sci U S A* **92**, 5749-5753 (1995).

479   19   Rico, E., Jeacock, L., Kovarova, J. & Horn, D. Inducible high-efficiency CRISPR-Cas9-

480   targeted gene editing and precision base editing in African trypanosomes. *Scientific reports* **8**,

481   7960, doi:10.1038/s41598-018-26303-w (2018).

482   20   Vincent, I. M. *et al.* A molecular mechanism for eflornithine resistance in African

483   trypanosomes. *PLoS Pathog* **6**, e1001204, doi:10.1371/journal.ppat.1001204 (2010).

484   21   Smith, A. M. *et al.* Quantitative phenotyping via deep barcode sequencing. *Genome Res* **19**,

485   1836-1842, doi:10.1101/gr.093955.109 (2009).

486   22   Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical

487   analyses: 2018 update. *Nucleic Acids Res* **46**, W537-W544, doi:10.1093/nar/gky379 (2018).

488   23   Profaci, C. P., Munji, R. N., Pulido, R. S. & Daneman, R. The blood-brain barrier in health and

489   disease: Important unanswered questions. *J Exp Med* **217**, doi:10.1084/jem.20190062 (2020).

490   24   Watson, G. F. & Davis, P. H. Systematic review and meta-analysis of variation in Toxoplasma

491   gondii   cyst   burden   in   the   murine   model.   *Exp   Parasitol*   **196**,   55-62,

492   doi:10.1016/j.exppara.2018.12.003 (2019).

493   25   McGovern, K. E., Cabral, C. M., Morrison, H. W. & Koshy, A. A. Aging with Toxoplasma

494   gondii results in pathogen clearance, resolution of inflammation, and minimal consequences to

495   learning and memory. *Scientific reports* **10**, 7979, doi:10.1038/s41598-020-64823-6 (2020).

496   26   Cavalli-Sforza, L. L. & Edwards, A. W. Phylogenetic analysis. Models and estimation

497   procedures. *American journal of human genetics* **19**, 233-257 (1967).

498   27   Plantinga, E. A., Bosch, G. & Hendriks, W. H. Estimation of the dietary nutrient profile of free-

499   roaming feral cats: possible implications for nutrition of domestic cats. *The British journal of*

500   *nutrition* **106 Suppl 1**, S35-48, doi:10.1017/S0007114511002285 (2011).

501   28   Dubey, J. P. Tissue cyst tropism in Toxoplasma gondii: a comparison of tissue cyst formation

502   in organs of cats, and rodents fed oocysts. *Parasitology* **115 (Pt 1)**, 15-20 (1997).

503   29   Kuss, S. K., Etheredge, C. A. & Pfeiffer, J. K. Multiple host barriers restrict poliovirus

504   trafficking in mice. *PLoS Pathog* **4**, e1000082, doi:10.1371/journal.ppat.1000082 (2008).

505   30   Kaiser, P., Slack, E., Grant, A. J., Hardt, W. D. & Regoes, R. R. Lymph node colonization

506   dynamics after oral Salmonella Typhimurium infection in mice. *PLoS Pathog* **9**, e1003532,

507   doi:10.1371/journal.ppat.1003532 (2013).

508   31   Hullahalli, K. & Waldor, M. K. Pathogen clonal expansion underlies multiorgan dissemination

509   and   organ-specific   outcomes   during   murine   systemic   infection.   *Elife*   **10**,

510   doi:10.7554/eLife.70910 (2021).

511   32   Alsford, S. *et al.* High-throughput phenotyping using parallel sequencing of RNA interference

512   targets in the African trypanosome. *Genome Res* **21**, 915-924, doi:10.1101/gr.115089.110

513   (2011).

514   33   Bushell, E. *et al.* Functional Profiling of a Plasmodium Genome Reveals an Abundance of

515   Essential Genes. *Cell* **170**, 260-272 e268, doi:10.1016/j.cell.2017.06.030 (2017).

516   34   Sidik, S. M. *et al.* A Genome-wide CRISPR Screen in Toxoplasma Identifies Essential

517   Apicomplexan Genes. *Cell* **166**, 1423-1435 e1412, doi:10.1016/j.cell.2016.08.019 (2016).

518   35   Abel, S. *et al.* Sequence tag-based analysis of microbial population dynamics. *Nat Methods* **12**,

519   223-226, 223 p following 226, doi:10.1038/nmeth.3253 (2015).

520   36   Hullahalli, K., Pritchard, J. R. & Waldor, M. K. Refined Quantification of Infection Bottlenecks

521   and   Pathogen   Dissemination   with   STAMPR.   *mSystems*   **6**,   e0088721,

522   doi:10.1128/mSystems.00887-21 (2021).

523   37   Dean, S. *et al.* A toolkit enabling efficient, scalable and reproducible gene tagging in

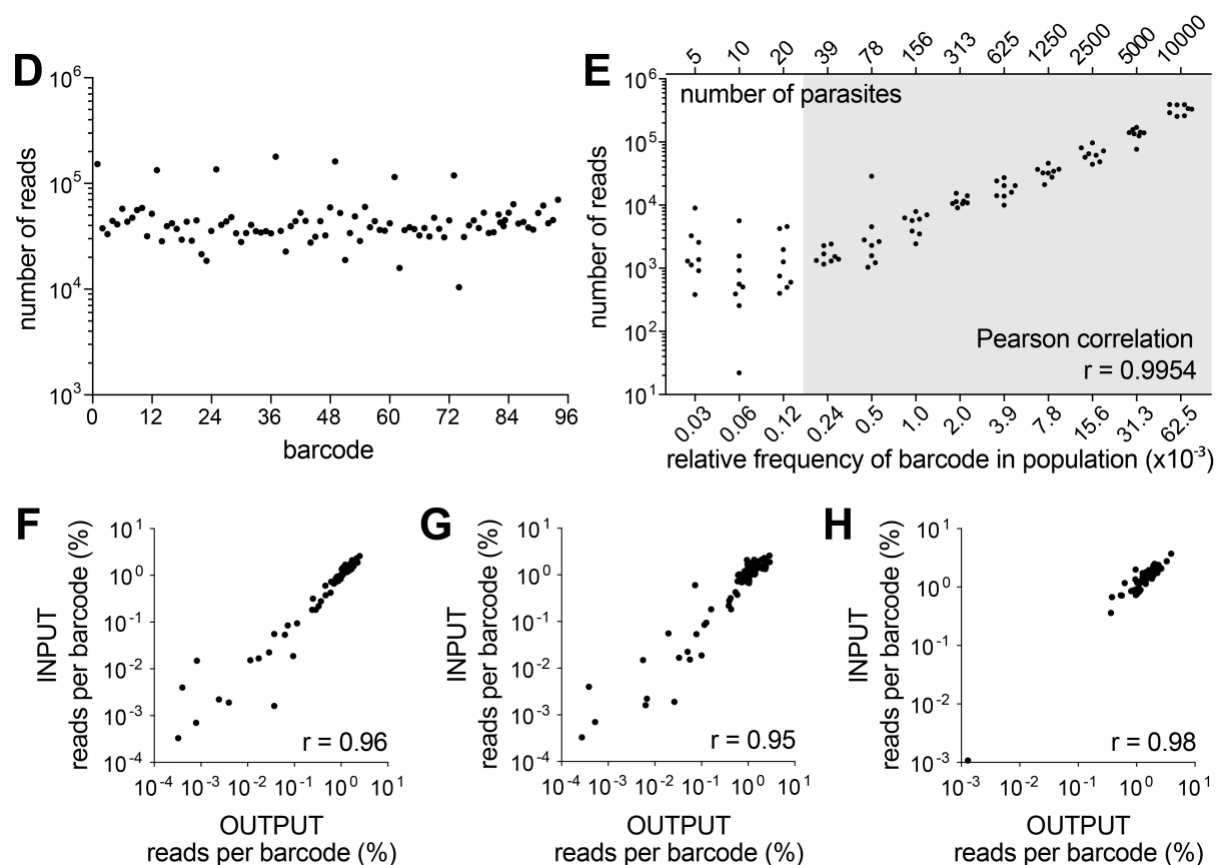524   trypanosomatids. *Open Biol* **5**, 140197, doi:10.1098/rsob.140197 (2015).

525

526   **Acknowledgements:**

537

538   **Author Contributions:** Conception/design of the work: EMF, CT, SEE, MAC.

539   Acquisition/analysis/interpretation of data: CJW, GS, CGC, HJB, FBL, CGC, MB, ARF, EA, IA, LG,

540   SEE, MAC. Manuscript drafting and revision: CJW, GS, HJB, EA, EMF, CT, SEE, MAC.

541

542   **Competing interests:** The authors declare no competing interests.

543

544   **Materials & Correspondence**: All correspondence and material requests should be addressed to MAC

545   (m.child@imperial.ac.uk) and SEE (se2s@virginia.edu).

Figure 1: Protozoan pathogens 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 in *T. gondii* or the *AAT6* locus in *T. b. brucei*, where Cas9 endonuclease activity introduces

550   a double-strand break. This break is repaired by homologous DNA repair using the co-transfected

551   60mer barcode oligo donor template. **B**) Sanger sequencing confirmation that barcode integration

552   results in disruption of the *UPRT* in *T. gondii* or *AAT6* in *T. b. brucei* coding sequence and destruction

553   of the protospacer DNA sequence and PAM. **C**) Multiplexed transfection strategy and NGS pipeline to

554   generate barcoded libraries of *T. gondii* tachyzoites for *in vitro* and *in vivo* studies. **D**) Complex libraries

555   containing 96 uniquely barcoded strains can be produced and the relative frequency of each barcode

556   quantified by NGS. Scatter plot presents read counts for individual barcodes. **E**) Individual barcodes

557   can be detected at low frequencies, and read depth is a sensitive proxy for parasite number. Scatter plot

558   presents the number of reads for a serial two-fold dilution series of known numbers of parasites and the

559   relative frequency of individual barcodes in the population. The shaded area indicates the data used to

560   calculate the PCC provided: $r = 0.9954$, $n = 9$, P (two-tailed) $= <0.0001$. **F-G**) The population structure

561   of multiplexed barcode libraries is stable *in vitro*. Scatter plot comparing 96 individual barcode

562   frequencies within the INPUT pooled library population and the OUTPUT sample harvested after one

563   ~36-hour lytic growth cycle (**F**, PCC $r = 0.96$, $n = 96$, P (two-tailed) $= <0.0001$), or six lytic growth

564   cycles (**G**, PCC $r = 0.95$, $n = 96$, P (two-tailed) $= <0.0001$). **H**) The population structure of multiplexed

565   barcode libraries is stable during the early stage of *in vivo* infection. Scatter plot comparing 63

566   individual barcode frequencies within a pooled library population for an intraperitoneal inoculum of

567    $2x10^6$ (INPUT) and the OUTPUT sample harvested from the peritoneal cavity of C57BL6 mice 36

568   hours post-infection. PCC $r = 0.98$, $n = 63$, P (two-tailed) $= <0.0001$.

569

570

571

572

573

574

575

576

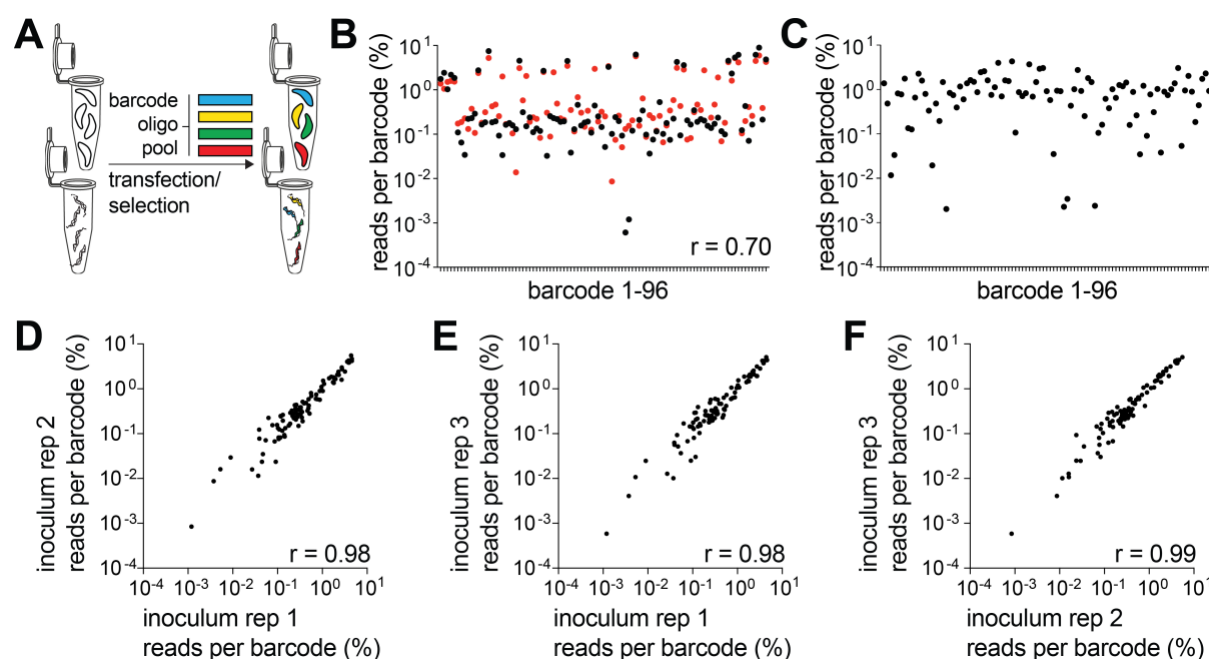**Figure 2: Barcoded libraries can be generated via one-pot transfections. A**) Schematic of the one-pot transfection method to generate barcoded parasite libraries. **B**) Overlay of the relative percent frequency of each barcode in an RHΔ*ku80* population (black dots) or a PruΔ*ku80* population (red 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 (two-tailed) = <0.0001. **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 percent frequency of barcodes was highly correlated in pairwise comparison each inoculum sample: **D**) inoculum 1 vs. 2, PCC r = 0.98, n = 96, P (two-tailed) = <0.0001; **E**) inoculum 1 vs. 3, PCC r = 0.98, n = 96, P (two-tailed) = <0.0001; **F**) inoculum 2 vs. 3, PCC r = 0.99, n = 96, P (two-tailed) = <0.0001. PCC values provided on scatter plots indicate degree of correlation between populations being compared.
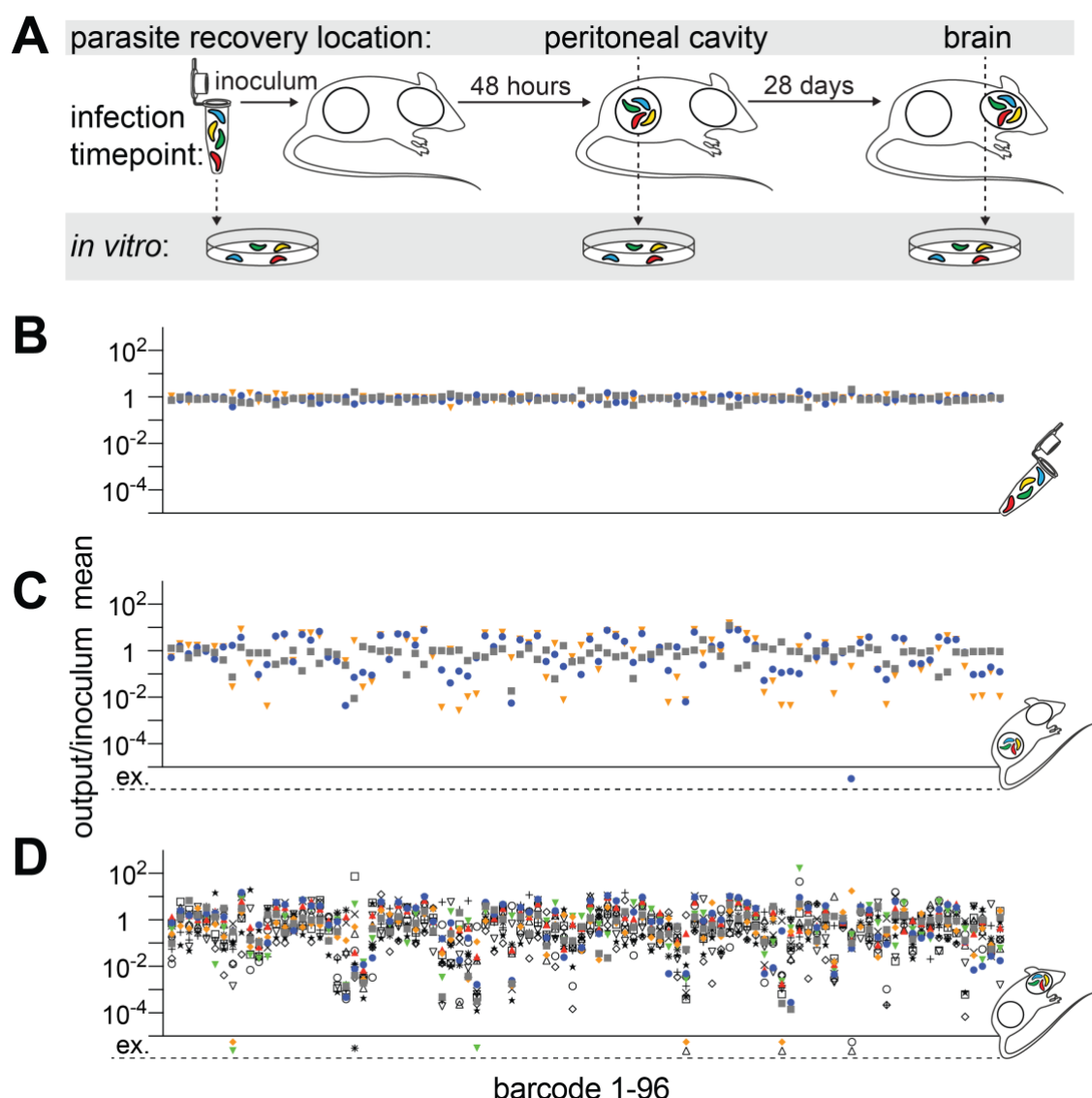
23

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 hours 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)** Scatter plot 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 hours (**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.
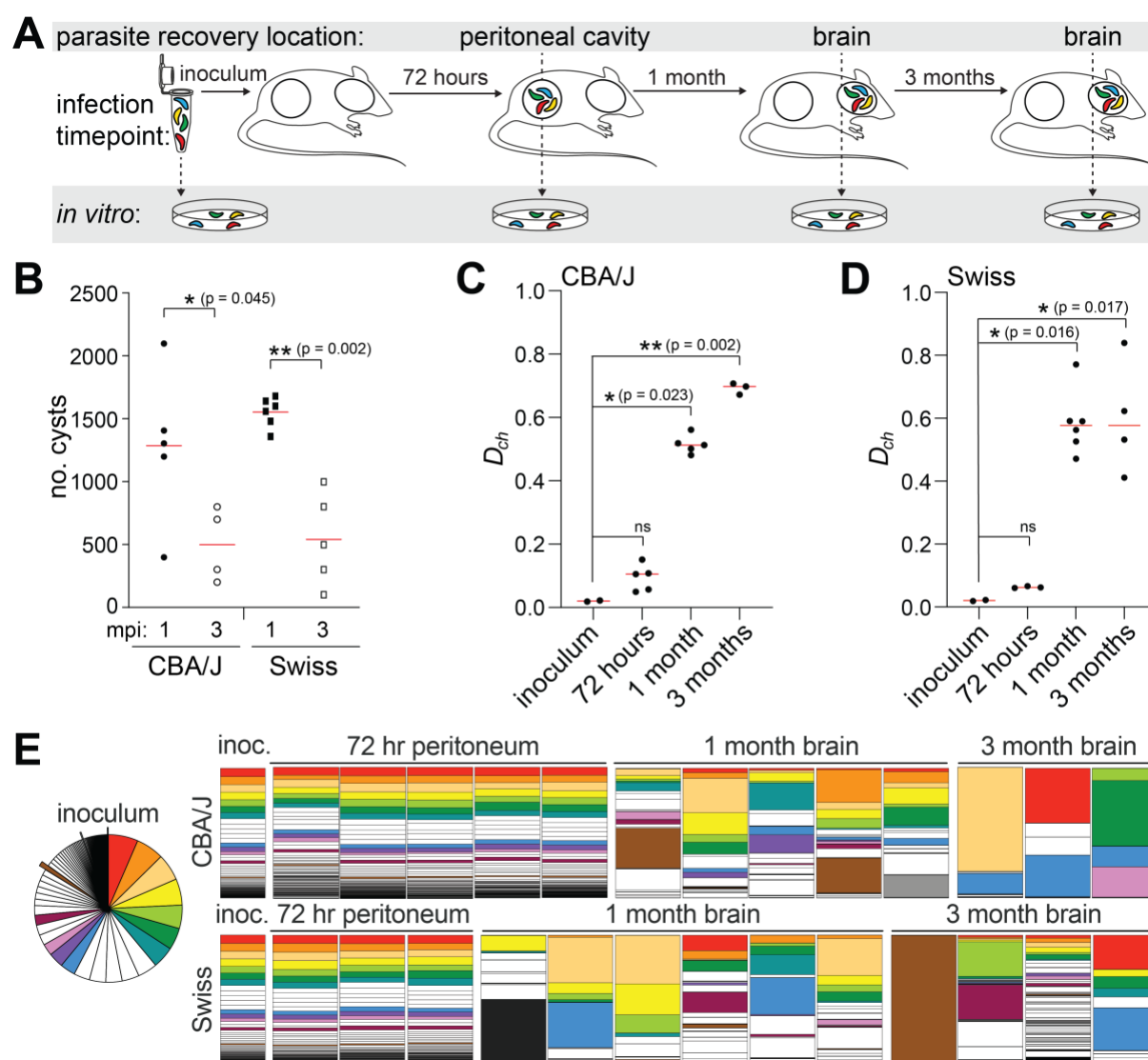
24

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 hours 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 with one-way ANOVA. **C-D)** Scatter plot of chord distances (D_{ch}) calculated from barcode representation in the inoculum relative to 48-hour 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 by Kruskal-Wallis one-way ANOVA, with

615 Mann-Whitney used for pairwise comparisons. **E**) Parts-of-whole charts representing the relative

616 frequency of each barcode within the mouse samples. Barcodes are ranked in descending order of

617 abundance in the inoculum mean (inoculum mean N=3, represented as a pie chart and bar charts labeled

618 'inoc.'). A colour was assigned to a barcode if it was the dominant barcode in any brain sample or if it

619 represented greater than 20% of all reads in any brain sample. Each colour is unique to one barcode.

620 Any barcode that dominated a mouse brain and was represented in the bottom 25% of inoculum reads

621 is outset in the inoculum pie chart for clarity and shaded brown, grey, or black in parts-of-whole charts.

622

623

624

625 **Supplemental Material**

626

627 **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

628

629 **Table S1:** Details of barcode extinctions occurring during the different phases of infection. Barcode

630 numbers and the mouse in which the extinction was observed are noted as shown in Figure 2C-D.

631 Extinctions are defined by an absence of the barcode sequence within the processed NGS read data.

632

633 **Table S2:** Sequences of barcoding oligo nucleotides.

634

635 **Table S3:** Raw data from all NGS experiments.
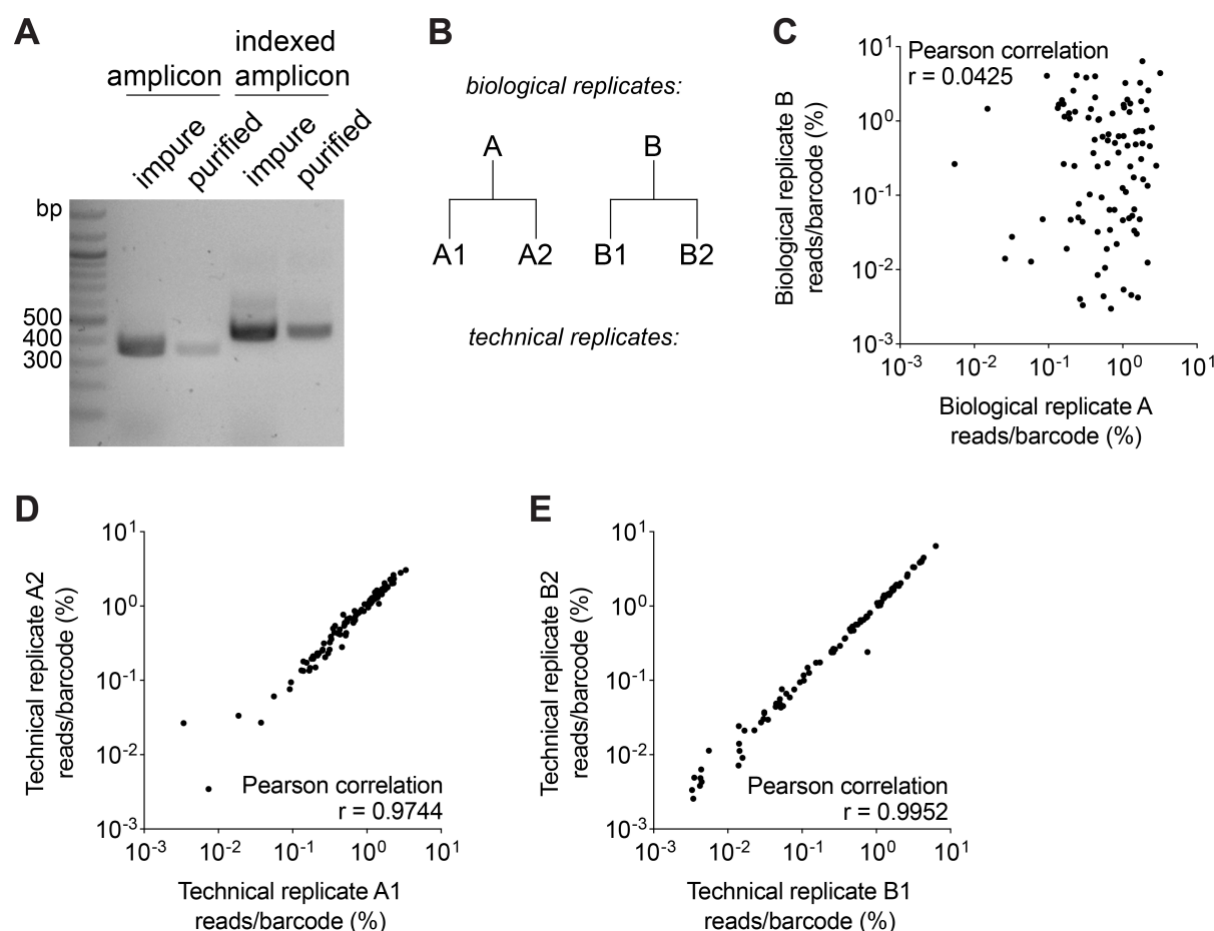
636

637

638

Figure S1

Figure S1: A robust pipeline for analysis of barcoded _T. gondii_ libraries by NGS. A) A single ~300 bp region of the UPRT locus 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). **B)** Strategy to isolate the primary source of variation to the multiplexed transfection barcoding strategy or the NGS pipeline. **C-E)** Scatter plots show the percent representation of individual barcodes within library pools for biological replicate transfections, (**C**, PCC r = 0.0425, n = 96, P (two-tailed) = 0.6807), or technical replicates of amplicon indexing (**D**, PCC r = 0.9744, n = 96, P (two-tailed) = <0.0001) and (**E**, PCC r = 0.9952, n = 96, P (two-tailed) = <0.0001). PCC values represent comparison between samples indicated on each x and y axis.

653   **Figure S2**
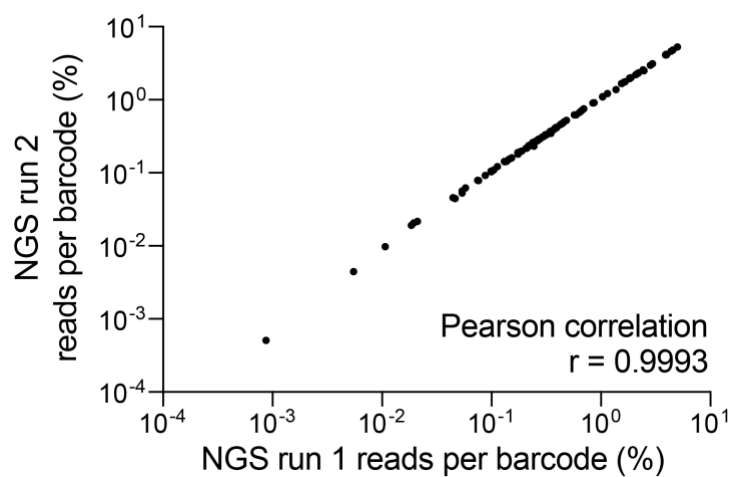
654

655   **Figure S2: Independent NGS runs assessing barcoded strain frequency distributions are highly**

656   **reproducible.** Scatter plot of relative percentage frequency of barcodes within a single genomic sample

657   processed on independent NGS runs. PCC r = 0.99, n = 96, P (two-tailed) = <0.0001.