Title

The commensal protist *Tritrichomonas musculus* exhibits a dynamic life cycle that induces B cell-modulated remodeling of the gut microbiota

Authors

Ana Popovic[1,2], Eric Yixiao Cao[3], Joanna Han[3], Nirvana Nursimulu[1,4], Eliza V.C. Alves-Ferreira[5], Andrea Kennard[5], Noor Alsmadi[3], Michael E. Grigg[5], Arthur Mortha[3], John Parkinson[1,2,6,*]

Affiliations

[1]Program in Molecular Medicine, Hospital for Sick Children, Toronto, Ontario, Canada

[2]Department of Biochemistry, University of Toronto, Toronto, Ontario, Canada

[3]Department of Immunology, University of Toronto, Toronto, Ontario, Canada

[4]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

[5]Molecular Parasitology Section, Laboratory of Parasitic Diseases, NIAID, National Institutes of Health, Bethesda, Maryland, USA

[6]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

Corresponding authors

*To whom correspondence should be addressed:

john.parkinson@utoronto.ca

Keywords: Eukaryotic microbiota; Parasites; Tritrichomonas; Microbiome; Gut pathogens

Summary

Commensal protists and gut bacterial communities exhibit complex relationships, mediated at least in part through host immunity. To improve our understanding of this tripartite interplay, we investigated community and functional dynamics between the murine protist *Tritrichomonas musculus* (*T. mu*) and intestinal bacteria in healthy and B cell-deficient mice. We identified dramatic, *T. mu*-dependent remodeling of resident bacterial composition and metabolism in parallel to increases in *T. mu* metabolism and the expression of putative virulence factors, responses accelerated in the absence of B cells. Upregulation of iron acquisition and tryptophan consumption revealed competition between bacteria and the protist. Single cell transcriptomics identified distinct *T. mu* life stages, providing evidence for active growth, sexual replication and the formation of pseudocysts. Unique cell states were validated *in situ*, and appeared modulated by constituent bacteria, suggesting their role in protist development. Our results reveal complex microbial dynamics during the establishment of a commensal protist.

Introduction

Intestinal microbiota are key regulators of host health and the development of disease[1,2]. The majority of studies currently address roles of bacteria and fungi, however emerging data supports an appreciable presence of protists in the gut[3-5]. These include species considered commensals, diarrheagenic parasites and pathobionts that exert influential roles on resident bacterial communities. Conversely, bacterial communities have been shown to impact the ability of protists to successfully colonize[6,7]. In addition to direct interactions such as competition for resources, protist-bacterial cross-talk appears largely mediated through the modulation of host immune pathways[8,9]. Currently little is known concerning the molecular basis which drive these complex protist:host:bacterial interactions.

Among protists capable of both asymptomatic gut colonization and virulence are members of the trichomonads, microbes prevalent in the human gut with links to colitis, irritable bowel syndrome and colon cancer[5,10-13]. In mice, trichomonads *Tritrichomonas musculus* (*T. mu*) and *Tritrichomonas muris* establish chronic asymptomatic infections, which result in elevated baseline intestinal immune activation, dependent on increases in IFN-gamma producing Th1 and interleukin-17 producing Th17 cells. Although sustained inflammation from prolonged *T. mu* colonization can increase susceptibility to T-cell-driven colitis, it also confers protection to mice against enteric infections with *Salmonella*[9,14-16]. The stimulation of intestinal tuft cells by *Tritrichomonas*-produced succinate furthermore leads to a Th2-based immune response, triggering structural remodelling of the intestinal epithelium, and ultimately restricts colonization by parasitic helminths[14,19,20]. Colonization by *Tritrichomonas* spp. has also been shown to shift commensal bacterial diversity and the abundance of *Helicobacter* and Lactobacillaceae taxa[17,18]. Similarly, bacteria can modulate *T. mu* engraftment. For example, *Bifidobacterium* spp. together with inulin supplements were shown to influence *T. mu* colonization in the context of a high fat diet, suggesting complex microbial interactions[17]. These interactions between protist and bacteria may also be driven through immunoglobulins (Ig) A and M, antibodies secreted by B cells and elevated in response to *T. mu* colonization[9,14]. Although both bacteria

64 and protists are known to impact B cell antibody responses, it is not well understood what role B cells play
65 in modulating the physiology of a complex community comprised of members from diverse microbial
66 kingdoms[9,21,22].
67
68 Here, we begin to unravel the tripartite interactions between *T. mu*, resident gut bacteria and the host through
69 the application of 16S rRNA sequence surveys and transcriptomics to microbial communities in healthy
70 (WT) and B cell-deficient (*muMT*$^{-/-}$) mice. We show that introduction of *T. mu* induces dramatic
71 diversification of gut bacterial populations in both host genotypes during the first 28 days of colonization.
72 Shifts in bacterial growth and metabolism, in response to protist colonization, track with upregulation of *T.*
73 *mu* core metabolic processes and the expression of putative virulence factors, changes which are accelerated
74 in the absence of B cells. To dissect the activities of *T. mu* in the mouse intestinal tract at high resolution,
75 we performed single cell transcriptomics. This exceptional dataset reveals actively growing and replicating
76 trophozoites, and evidence of *T. mu* sexual replication and a pseudocyst stage. Using fluorescent RNA
77 labelling of uniquely expressed genes, we validate these protist cell states *in situ*. Finally, we describe
78 potential cross-kingdom competition for resources between protist and resident bacteria, specifically in the
79 acquisition and utilization iron and key amino acids tryptophan and arginine. Our data demonstrate the co-
80 adaptation of resident bacteria and a new protist commensal, which together reshape the gut environment
81 in a manner dependent on the host immune response.
82
83 **Results**
84
85 ***T. mu* colonization drives diversification of the intestinal microbiome**
86 Colonization with *T. mu* induces profound changes in the mouse intestinal immune landscape, including
87 increased levels of host IgA, an abundant antibody known to maintain gut microbial homeostasis[9,22]. To
88 explore the downstream impact on resident gut microbiota, and how this may be influenced by antibody-

89 producing B cells, we measured microbial composition and gene expression during the first 28 days of *T.*
90 *mu* colonization in WT and B cell-deficient (*muMt*$^{-/-}$) C57Bl/6 mice, through 16S and metatranscriptomics
91 sequencing (Figure 1A). Protists expanded rapidly in mouse caeca, in both the presence and absence of B
92 cells, reaching 54 million and 68 million cells respectively (Figure 1B). Host IgA levels in WT mice tracked
93 *T. mu* growth, and were significantly higher relative to naïve mice at both days 2 and 28 ($p < 0.05$)
94 confirming previous reports (Figure 1C)[23].
95
96 The introduction of *T. mu* dramatically altered gut bacterial composition in both WT and *muMt*$^{-/-}$ mice
97 (Figure 2A). Sequencing of fecal 16S revealed 470 bacterial taxa (OTUs with ≥5 reads) in mouse fecal
98 samples, with median 176 OTUs per sample (Figure 2A and Tables S1 and S2). Naïve mice were dominated
99 by Bacteroidetes and Firmicutes, including members of Muribaculaceae, Bacteroidaceae, Lactobacillaceae
100 and Lachnospiraceae (Figure 2A). Upon colonization with *T. mu*, bacterial richness (observed OTUs)
101 increased, with a delayed increase in *muMt*$^{-/-}$ mice, before receding to near day 0 levels (LME, day 7 vs day
102 0, $\beta = 94.7$, 95% CI [49.0, 140.5], $p < 0.001$; day 14 vs day 0 in *muMt*$^{-/-}$, $\beta = 66.4$, 95% CI [1.8, 131.1], $p <$
103 0.05; two-way ANOVA, day 28 colonized vs naïve, *ns*) (Figure 2B). Increases in evenness (Shannon
104 diversity), however, persisted to day 28 across both host genotypes, reflecting expansions in multiple taxa
105 from Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria clades (LME, day 2 vs day 0 x host
106 genotype, $\beta = -0.52$, 95% CI [-1.02, -0.01], $p < 0.05$; day 7 vs day 0 x host genotype, $\beta = -1.00$, 95% CI [-
107 1.50, -0.49], $p < 0.001$; days 2 to 28 vs day 0, $p < 0.001$; two-way ANOVA, day 28 colonized vs naïve,
108 $F(1,12) = 7.06$, $p < 0.05$) (Figures 2A and 2B). A spike in evenness during the first week in WT hosts is the
109 result of rapid initial expansions of multiple taxa compared to a single Lachnospiraceae taxon in *muMt*$^{-/-}$
110 mice (Figure 2A). While bacterial communities in WT and *muMt*$^{-/-}$ mice differed prior to infection, their
111 compositions changed progressively after day 2 of protist colonization and trended toward a convergent
112 community profile (Adonis, time x host, $F(5,55) = 6.2$, $p < 0.001$; pre-infected vs day 2, *ns*) (Figure 2C).
113 That compositions in day 28 control mice remained most similar to those at day 0, prior to *T. mu*

114 colonization, underlines a protist-driven shift (Adonis, day 28 ctl vs day 0, protist x host genotype, $F(1,15)$
115 $= 5.7$, $p < 0.05$; *ns* in WT mice). We identified 55 taxa that changed significantly as a result of *T. mu* ($p <$
116 0.05), clustered into six distinct dynamics patterns (Figure 2D). The majority of bacteria increased in
117 abundance in both WT and *muMt*$^{-/-}$ hosts: 26 taxa continued expanding until the termination of the
118 experiment (including *Bifidobacterium*, *Parabacteroides*, *Alistipes* and two lactobacilli) and 10 taxa peaked
119 at day 21 after which their abundances fell (e.g. the Proteobacteria *Desulfovibrio* and *Helicobacter*). Taxa
120 whose abundances decreased as a result of *T. mu* colonization (11 taxa) included *Blautia* and several other
121 clostridia, taxa from the Muribaculaceae family and *Enterorhabdus*. Five taxa exhibited discordant patterns
122 between host genotypes, including *Parasutterella*, *Bacteroides*, and a species of Lachnospiraceae which
123 experienced temporary reductions in WT mice but expanded significantly in *muMt*$^{-/-}$ mice; and two species
124 of Muribaculaceae which displayed the opposite response. We identified 82 taxa with host genotype-
125 dependent patterns, independent of the protist, however as WT and *muMt*$^{-/-}$ mice were not co-housed, we
126 cannot rule out caging effects. Changes in bacterial abundances were validated through qPCR using clade-
127 specific primers for four groups: Firmicutes, Bacteroidetes, Actinobacteria and Gammaproteobacteria (day
128 21 vs day 0, $p < 0.05$, except Firmicutes in *muMT*$^{-/-}$, $p = 0.057$, and Bacteroidetes in WT and *muMT*$^{-/-}$, $p =$
129 0.057) (Figure 2E).
130

131 ***T. mu* induces changes in bacterial metabolism**

132 To evaluate how *T. mu*-induced shifts in gut bacterial composition translate to changes in community
133 function, we performed bulk RNA sequencing (metatranscriptomics) of mouse caecal contents at days 2
134 and 28 of the experiment, timepoints reflecting early and chronic stages of infection (Figure 3A).
135 Sequencing produced a total of 2.4 billion paired reads across 32 samples (median 76.8 million) (Table S3).
136 After filtering for low quality, host and bacterial rRNA, 835 million reads were identified as bacterial (741
137 million in naïve and 94 million in colonized mice), 157 million as *T. mu* and 349 million reads remained
138 unannotated (293 million in naïve and 56 million in colonized mice) (Figure S1 and Table S3). The majority

139 of reads mapping to bacterial transcripts belonged to Firmicutes taxa (597 million), followed by
140 Bacteroidetes (75.4 million) and Proteobacteria (21.5 million) (Figure 3A). Colonization with *T. mu*
141 associated with blooms in Bacteroidetes and particularly Proteobacteria activity. Bacteroidetes transcripts
142 increased in proportion from 14.9% to 26.6% in WT colonized mice and from 12.4% to 24.2% in *muMT*$^{-/-}$
143 colonized mice, whereas naïve mice displayed relatively modest increases (9.1% to 11.5% in WT; 5.9% to
144 7.6% in *muMT*$^{-/-}$) (Figure 3A). Changes in proteobacterial activity were more striking, varying between 1%
145 and 3.4% in naïve mice, but increasing from 4% to 15.6% of reads in WT mice by day 28, and already
146 accounting for 13.7% of reads by day 2 in *muMT*$^{-/-}$ mice, suggesting accelerated increases and a possible
147 role for B cells in regulating these taxa. The majority of these reads belonged the *Helicobacteraceae* clade
148 (9.1% WT and 6.2% *muMT*$^{-/-}$ at day 28), bacteria previously associated with enteric inflammation and
149 exacerbated disease in T cell-driven colitis [9,14,24].
150
151 The presence of *T. mu* RNA, rRNA in particular, substantially reduced transcriptional information from
152 bacteria in colonized mice, from a median of 22.1 million reads per naïve mouse to 2.1 million (Table S3).
153 Protozoan rRNA is not captured by commercial rRNA depletion kits used in sequencing library
154 preparation. To account for the variation in sequencing depth, we evaluated gene expression in data rarefied
155 to 600,000 bacterial reads per sample. Principal component analysis of bacterial gene expression re-
156 capitulated patterns observed in 16S compositional data (Figures 2C and 3B). Bacterial expression differed
157 between host genotypes, and progressively changed after introduction of *T. mu*. The protist induced overall
158 upregulation of bacterial genes, with strong responses in *muMt*$^{-/-}$ mice already by day 2 (Figure 3C).
159 Interestingly, host genotype-related differences decreased over time in both colonized and naïve mice
160 (Figure 3C).
161
162 The analysis of bacterial metabolic pathways revealed an association between *T. mu* colonization and
163 increased bacterial glycolysis, carbohydrate metabolism, peptidoglycan, polysaccharide and O-Antigen

164 sugar biosynthesis, and the biosynthesis and metabolism of numerous amino acids (Figure 3D). Amino acid
165 biosynthesis and metabolism, in particular, showed earlier upregulation in the absence of B cells. At day 2,
166 we detected metabolism primarily in Firmicutes taxa, but noted shifts toward Bacteroidetes and
167 Proteobacteria by day 28 (Figures S2 and S3). The expansion of Proteobacteria (facultative anaerobes) and
168 increased cycling through succinyl-CoA synthetase (SucC) in the TCA cycle, an enzyme suppressed under
169 anaerobic conditions[25], may suggest oxygen exposure, characteristic of gut inflammation and dysbiosis
170 (Figure S2). Increasing metabolism, and particularly the expression of enzymes required for biosynthesis
171 of surface lipopolysaccharides (LPS) and O-antigen sugars, the variable polysaccharide coating the outer
172 LPS layer, suggests that the establishment of a chronic infection by *T. mu* may induce bacterial growth in
173 the caecum, and specifically the growth of particular taxa.

174

175 To further explore this notion, we probed for changes in the expression of bacterial cell division and cell
176 wall biosynthetic machinery. We first identified homologues to 44 proteins involved in these processes, as
177 previously identified in the model *E. coli* system, many of which are conserved across bacteria and form a
178 network of functionally associated protein modules[26] (Figure 3E). Homologues to 40 of these proteins were
179 detected in the experiment, and 27 were present in the rarefied dataset (represented by 667 unique genes).
180 Patterns of gene expression were broadly consistent with those described in other pathways—genes were
181 upregulated during *T. mu* colonization, with earlier responses in B cell-deficient mice (12 bacterial genes
182 upregulated in *muMT*$^{-/-}$ mice at day 2, compared to three in WT mice), and a shift from predominantly
183 Firmicutes gene expression at day 2 to Bacteroidetes (primarily *Bacteroides* and *Parabacteroides*) and
184 Proteobacteria (primarily *Helicobacter*) at day 28, suggesting an environment favouring gram-negative
185 taxa. At day 2, gene *ftsZ*, the initiator of bacterial cell division, was upregulated nearly 160 and 1600-fold
186 in colonized WT and *muMT*$^{-/-}$ mice, respectively. At this stage, several additional genes were upregulated
187 in *muMT*$^{-/-}$ mice, including *ftsK* (~95-fold) which functions in chromosome segregation, *mrdA* and *mrdB*
188 (40 and 114-fold) which have roles in bacterial cell shape and elongation, and peptidoglycan biogenesis

genes *murCDE* (84 to 206-fold). By day 28, these genes were upregulated in bacteria from both groups of mice, along with cell division protein *ftsA* (112 and 274-fold). Other key cell division genes (e.g., ftsBLNQ) were not detected in the rarefied data. Genes active in protein secretion (*secA*, *yjjK*, *prlA*) followed similar patterns, and the porin *ompA*, which is known to modulate infection and host immunity in gram-negative bacteria, was upregulated at day 28 (62 and 247-fold in WT and *muMT*$^{-/-}$ mice). A number of genes showed host genotype-specific expression at day 28, including peptidoglycan biosynthesis genes *murG* and *mraY* (upregulated in the WT host) and nutrient stress response gene *relA* (upregulated in *muMt*$^{-/-}$ microbiota). Collectively, these data demonstrate substantial changes in bacterial functions in response to the establishment of a new protist resident.

Profiling of *T. mu* gene expression during colonization required assembly of a draft metagenome

Concurrent with changes in bacterial gene expression during *T. mu* colonization, we tracked the activities of the protist itself in the mouse gut. To allow mapping and interpretation of gene expression in the protist, a draft metagenome assembly of purified *T. mu* was produced using one million PacBio sequences (median length 11,924 bp), polished with 8.6 million Illumina MiSeq reads. The assembly consisted of 1,780 contigs with a total length of 225 Mbp. Sequences were AT-rich (71.4%) with repetitive regions across 52% of the length (see Tables S4 and S5 for metagenome statistics). A total of 26,723 genes were predicted (mean length 1,878 bp), of which 18,405 were identified as protein coding. We also identified two distinct rRNA sequences, inclusive of the 18S, 5.8S, 28S and internal transcribed spacers (ITS), with 97% sequence identity, suggesting the presence of one additional tritrichomonad organism (Figure 4A). Phylogenetic comparison of the ITS sequences with other parabasalids confirmed that both belong to the Tritrichomonadidae family (Figure 4B). The major sequence (81% of all reads) was synonymous with *T. mu*, whereas the minor sequence (19% of reads) belonged to a novel species distinct from all parabasalids catalogued in the GenBank database, with 85.4% identity at the ITS locus to *T. muris*.

214 Functional annotation of protist genes identified 1,932 unique Pfam domains, 999 unique Gene Ontology
215 (GO) terms and 471 unique enzymes (EC numbers). Among the top 20 most frequent Pfam domains in *T.*
216 *mu* is an interesting BspA-type Leucine rich repeat region (PF13306), identified in 317 protist genes and
217 present also in large numbers in *T. vaginalis*, *E. histolytica* and *E. dispar*[27,28]. Seven of these genes were
218 closely related to *T. vaginalis* TvBspA625, a cell surface protein with known antigenic properties (Tables
219 S6 and S7)[27]. In bacteria, these domains were reported in virulence factors that mediate interactions with
220 the host extracellular matrix and promote invasion at mucosal sites[29,30]. We also identified adhesins, lectins
221 and cysteine proteases, other protein families implicated in facilitating host interactions and parasite
222 infection (Tables S6 and S7). Such proteins localize to the cell surface or are secreted, and may represent
223 key factors in *T. mu*'s ability to establish a chronic colonization[27,31-33]. We furthermore report homologues
224 to enzymes of the arginine dihydrolase pathway, a pathway linked to immune escape in *G. duodenalis*
225 through the depletion of local L-arginine, ultimately inhibiting host production of toxic nitrous oxide[34-36].
226 Enzymes in this pathway (arginine deiminase (ADI), TMU_00005093 and TMU_00022804; ornithine
227 carbamoyltransferase (OCT), TMU_00009104; and carbamate kinase (CK), TMU_00010220 and
228 TMU_00019703) have also been identified in *T. vaginalis* and *T. foetus*[37], and may contribute to chronicity
229 of colonization. Future research is set to validate the role of these enzymes in host interactions and
230 permanent colonization.
231

232 ***T. mu* gene expression changes during colonization**

233 The assembly of the *T. mu* metagenome allowed us to map 157 million metatranscriptomic reads to the
234 protist. Of these, 20% mapped uniquely across 17,212 genes (between 247,166 and 3,810,677 reads per
235 sample), and were used to explore *T. mu* activity during colonization of a healthy and B cell-deficient gut.
236 Gene expression changed over time in both hosts (666 DE genes in WT and 627 in $muMt^{-/-}$ mice, $p < 0.05$),
237 and the presence of host B cells appeared to have a greater impact later in the infection, at day 28 (Figures
238 4C and 4D). Pathway analysis showed upregulation of *T. mu* carbohydrate metabolism, including galactose

239 (ec00052) and starch and sucrose (ec00500) during colonization, increased degradation of glycans
240 including N-glycans and/or gangliosides (ec00511), as well as upregulation of the tricarboxylic acid cycle
241 (TCA; ec00020) and glycolysis (ec00010), consistent with growth (Figure 4E). The expression of 41 *T. mu*
242 genes changed in a B cell-dependent manner, including four BspA-like proteins (with predicted
243 extracellular immunogenic regions), which increase in expression from day 2 to 28 in WT mice (Figure 4F,
244 Table S8). Interestingly, in *muMT*$^{-/-}$ mice, these genes showed either no change or decreasing expression
245 over time (TMU_00005602 and TMU_00010916; $p < 0.05$). Other genes with potential cytoadherence and
246 virulence roles (BspA-like, adhesins, lectins and cysteine proteases) exhibited a diverse array of expression
247 changes in response to colonization time and/or the presence of B cells (Figures 4G and S4). For example,
248 homologues of adhesins AP33 and AP120 (TMU_00013231 and TMU_00017025) and TvBspA625
249 (TMU_00015121), genes with documented roles in *T. vaginalis* infection, were upregulated at day 28, with
250 intriguing increases already at day 2 in *muMT*$^{-/-}$ mice (Figure 4G; day 28 vs day 2 in WT, $p < 0.05$ and $p <$
251 0.01; *muMT*$^{-/-}$, *ns*; WT vs *muMT*$^{-/-}$ at day 2, *ns*). Conversely two of three identified cysteine proteases
252 (TMU_00012588 and TMU_00013765) were downregulated at day 28 (Figure S4B). These data
253 demonstrate that dynamic metabolic and accessory functions are required by the protist as it adapts to the
254 host environment.
255

256 **Single cell profiling of *T. mu* reveals distinct life cycle stages**

257 Since trichomonads display a complex life cycle featuring a variety of cell states (e.g. trophozoite,
258 amoeboid, pseudocyst) we applied single cell RNA sequencing (scRNA-Seq) to explore whether the
259 dramatic changes in protist gene expression, defined in the previous section, reflects a shift in the
260 distribution of cell states. Here, we were also interested in seeing how the absence of the bacterial
261 microbiome might impact the colonization and life cycle propagation of *T. mu*, and therefore performed
262 scRNA-Seq on protists isolated from both conventionalized and germ-free (GF) mice. GF mice are also
263 known to possess an immature B cell compartment with low levels of IgA prior to colonization[38].

264 Sequencing produced 232 and 224 million reads in GF and conventionalized mice, respectively, with 70%
265 and 76% uniquely mapped to the protist genome assembly (Table S9). Gene expression was analysed in
266 3,000 *T. mu* cells per host, with medians of 3,776 and 3,140 transcripts and 823 and 589 genes per cell. We
267 identified 15 distinct protist subpopulations, subsequently assigned to three 'superclusters' (Figure 5A).
268 Clusters 1 through 4 (designated supercluster 'A') and clusters 10 and 11 were notably enriched in protists
269 isolated from the GF mouse, while clusters 5 through 7 (within supercluster 'B') and 15 were represented
270 primarily by protists from the conventionalized mouse (Figure 5B).
271
272 Functional enrichment of GO terms in genes upregulated in cell clusters suggested the presence of actively
273 metabolizing cells, those at various stages of cell cycle, and protists potentially undergoing pseudocyst
274 formation (Figure 5C). Clusters 3 and 5-7, for example, were enriched in carbohydrate metabolism (7 GO
275 terms), including galactose and monosaccharide metabolism (GO:0006012; GO:0005996). Cytoskeleton
276 remodeling and actin polymerization/depolymerization (19 GO terms) were upregulated in clusters 1-3 and
277 6-8, and biosynthetic processes (14 GO terms), including peptide and macromolecule biosynthesis
278 (GO:0043043; GO:0009059) were enriched in clusters 1-4 and cluster 8, as well as in clusters 12-14
279 (supercluster 'C'). We therefore hypothesize that clusters 1-4 ('A') and 5-11 ('B') may represent actively
280 feeding and growing *T. mu* trophozoites. Sphingolipid metabolism (GO:0006665), which contributes to
281 regulation of *Giardia* encystation[39], was enriched only in cluster 8. Interestingly oxidative stress response
282 (GO:0006979) was upregulated in the neighbouring clusters 12 and 13, and the metabolism and
283 biosynthesis of biogenic amines (GO:0042401, GO:0006576) were upregulated in cluster 14. As biogenic
284 amines increase in *Entamoeba invadens* during encystation[40], we suggest that the oxygen stress response
285 may represent stress-induced activation of the encystation transcriptional program in clusters 12 and 13,
286 and that cluster 14 may represent a pseudocyst form of *T. mu*, previously identified in *T. foetus* and *T.*
287 *vaginalis*[41,42]. Cluster 14, incidentally, has the lowest transcriptional complexity, defined by the ratio of
288 detected genes per transcripts, and the fewest detected genes (Figure 5D).

289

290 Enrichment of chromosome and DNA reorganization GO terms in clusters 1, 5 and 11 (eight GO terms),
291 and DNA replication and expression of cell cycle checkpoint genes (10 GO terms) in cluster 10, suggests
292 these protists may be undergoing cell cycle. We note that cluster 15, which shares some of these functions,
293 also features increased DNA recombination (GO:0006310) and may be undergoing sexual or parasexual
294 replication. To lend support to these findings, we carried out a comprehensive analysis of G1/S (*hisH3*,
295 *hisH4*, *tk*) and G2 (*pcna*, *cycA* and *cycB*) cell cycle marker genes, as well as those specific to and/or required
296 for meiosis (Tables S6 and S7)[43,44]. Functional enrichment scores were calculated based on the average
297 expression of each gene set relative to randomly chosen control genes. Scores for expression of G1/S marker
298 genes were highest in Clusters 1 and 11, and cluster 10 scored highest for the G2 phase (G1/S cluster 1 or
299 cluster 11 vs remaining clusters, $p < 0.001$; cluster 1 vs cluster 11, $p < 0.01$; G2 cluster 10 vs remaining
300 clusters, $p < 0.001$; Figures 5E and 5F). We confirmed increased expressions of histones H3 and H4 (highest
301 during G1/S) and thymidine kinase (*tk*; expressed during S phase) in clusters 1 and 11 (Figures S5A and
302 S5B). Genes which achieve highest expression during G2 phase (*pcna* and cyclins *cycA* and *cycB*), and
303 minichromosome maintenance (*mcm*) genes, expressed during G1/S and G2 phases, were upregulated in
304 cluster 10.

305

306 To probe for evidence of meiosis, we searched for 26 essential proteins—previously identified in *T.*
307 *vaginalis*[44] and involved in recombination, and synaptonemal and meiotic sister chromatid cohesion
308 complexes—in the *T. mu* proteome. Homologues to 24 of these proteins were present, including all eight
309 with meiosis-specific activity: Dmc1, Spo11-1, Hop1, Hop2A, Hop2B, Mnd1, Msh4, Msh5 and Mer3
310 (Table S6). Spo11-1 mapped to two highly similar *T. mu* proteins, however the expression of only one was
311 detected in our experiment. As above, we calculated an enrichment score for meiosis-specific gene
312 expression across all clusters. Cluster 15, representing a small group of 60 cells (52 from the
313 conventionalized and eight from the GF mouse), scored highest ($p < 0.001$; Figures 5E and 5F). In addition

314 to meiosis-specific genes *dmc1*, *hop2A* and *mnd1*, required genes *rad1*, *mre11*, *smc2*, *smc3* and *smc5* were
315 also significantly upregulated in cluster 15 ($p < 0.001$; Figure S5C), consistent with protists undergoing
316 sexual or parasexual replication. *Smc2* and *smc5* were upregulated also in cluster 10, supporting DNA
317 replication ($p < 0.001$). Given the differences in cluster distributions between conventionalized and GF
318 mice (Figure 5B), including the proportion of cells undergoing meiosis, our results suggest that the
319 constituent microbiome plays a role in modulating the protist's life cycle.
320

321 **Protist subpopulations express distinct sets of virulence-related genes**

322 Since the distribution of protist life cycle stages are altered in the absence of an intact constituent
323 microbiome, we were interested in how these differences might impact the ability of *T. mu* to colonize. We
324 therefore examined the expression of genes suspected to facilitate colonization (BspA-like proteins,
325 adhesins, lectins and cysteine proteases; Figure 5G). All four groups of genes were expressed in *T. mu*
326 isolated from both GF and conventionalized hosts, and their expression patterns were broadly defined by
327 the three major superclusters (A, B and C). Of the 317 identified BspA-like genes, we detected 176 in the
328 scRNA-Seq data. Remarkably, the expression of many of these genes was restricted to specific
329 subpopulations and, with the exception of clusters 12-14 (putative encysting stages), they differed between
330 protists from the GF and conventionalized mice (Table S10). A number of genes, for example, was enriched
331 in clusters 1-4 across both hosts (e.g. TMU_00019122, TMU_00026100), while the exclusive expression
332 of others in the conventionalized mouse (e.g. TMU_00025647, TMU_00025644) led to separation of these
333 clusters based on the presence of a constituent microbiome. Gene expression patterns in the two hosts were
334 similar in each of clusters 12-14, however a distinct shift in cluster 14 may reflect the terminal stage of
335 encystation. Notably, cells associated with cluster 15 (the putative sexual stage) exhibited distinct profiles
336 of BspA-like genes between the GF and conventionalized hosts, supporting the idea that the constituent
337 microbiome impacts multiple stages of the protist life cycle.
338

339 Eight of the nine *T. mu* adhesin homologues, which in *T. vaginalis* mediate binding to host epithelial cells
340 (AP33, AP51, AP65 and AP120)[31], were detected and notably depleted in clusters 4, 13 and 14 (Figure 5G).
341 The latter is perhaps indicative of the switch to pseudocyst forms that may facilitate transmission. Lectin
342 homologues, likewise displayed distinct patterns of expression across clusters, with similar depletion in
343 clusters 12-14, as well as in the putative sexual stage, cluster 15. Interestingly, cells from clusters 7-10
344 expressed more lectins in the GF mouse relative to their conventionalized counterparts. Similar to the other
345 groups of putative virulence factors, cysteine proteases exhibited distinct expression profiles across clusters,
346 with clear differences between the GF and conventionalized mice and limited expression in cluster 14. Of
347 the four cysteine proteases expressed in putative pseudocysts, two are members of the C54 peptidase family
348 with similarity to yeast ATG4 which is involved in autophagy. C54 cysteine proteinases, along with other
349 autophagy related proteins, have in *T. vaginalis* been implicated in cell starvation and differentiation[45,46].
350
351 This dynamic arsenal of putative virulence factors appears to have lifecycle stage-specific roles in *T. mu*,
352 that are further modulated by the presence of a constituent microbiota. Their distinct patterns of expression
353 further support the transition to a pseudocyst form in clusters 12-14, while clusters 1-4 and 5-11 represent
354 two distinct, actively growing forms of the protist, perhaps the trophozoite and amoeboid states.
355
356 **Distinct *T. mu* cell stages are detected *in vivo***
357 We validated the presence of distinct *T. mu* subpopulations in the intestinal lumen through *in situ*
358 fluorescent labelling of protozoan mRNAs that were uniquely expressed or distinctively upregulated across
359 specific clusters (Figure 6A). Microscopic imaging of caecal sections from WT mice revealed protists as
360 large, nucleated cells in the gut lumen, with no apparent adherence to the host epithelium (Figure 6B).
361 Transcript TMU_00005724, which was expressed in a majority of cells of supercluster 'B', was labelled in
362 31% of caecal protists (Figures 6A and 6C). TMU_00016742, expressed in a subset of 'B' (clusters 5, 6,
363 and 9), was detected in 9% of cells also positive for TMU_00005724, and was uniquely expressed in another

364 19%. Fluorescent *in situ* hybridization flow cytometry (FiSH-Flow) revealed expression of
365 TMU_00005724 across all protists freshly isolated from caeca (Figure 6D), with reproducible detection of
366 cluster-specific transcripts TMU_00016742 (clusters 5, 6 and 9), TMU_00009244 (cluster 12) and
367 TMU_00001185 (cluster 14) and in 2.6%, 0.08% and 0.19% of cells, respectively (Figure 6E). Greater
368 proportions of protists expressed TMU_00016742 in *muMT*$^{-/-}$ mice ($p < 0.05$) and trended towards more
369 TMU_00001185-positive but fewer TMU_00009244-expressing cells (*ns*), suggesting possible B cell
370 influence on the *T. mu* transcriptional program (Figure 6E). Since TMU_00001185 is uniquely expressed
371 in cells of cluster 14, those we hypothesize to represent *T. mu* pseudocysts, and pseudocyst formation occurs
372 during egress from the gastrointestinal tract, we tracked TMU_00001185 expression in protists freshly
373 isolated from the caecum and colon, and for up to 3 days of *in vitro* culturing (Figure 6F). Caecal isolates
374 contained the fewest TMU_00001185-expressing cells, consistent with their identity as pseudocysts
375 undergoing egress through the colon. In line with our speculations, proportions of TMU_00001185-
376 expressing cells increased during *in vitro* culturing. Since chitin is a critical component of cysts or
377 pseudocysts in other trichomonads and protozoan parasites, we measured levels of chitin in freshly isolated
378 or cultured *T. mu* by staining with fluorophore-labelled wheat-germ agglutinin (WGA). WGA fluorescence
379 charted a pattern similar to that of TMU_00001185: lowest detection in fresh caecal isolates, with a steady
380 increase during *in vitro* culturing (Figures 6G and 6H)[42,47,48]. Studies are ongoing to characterize the
381 pseudocyst stage at the ultrastructural level. These *in situ* results validate the presence of transcriptionally
382 distinct groups of protists, and provide an effective means to track *T. mu* population dynamics and
383 functional adaptations during colonization.

384

385 ***T. mu* and gut bacteria compete for nutrients**

386 We surmise that the establishment of a chronic *T. mu* infection depends on its ability to compete with
387 resident bacteria, as well as the host, for nutrients in the intestinal tract. Recently we showed a link between
388 iron supplementation (an essential and limiting micronutrient required for microbial growth), protist

389 colonization and microbiome structure[5]. Since iron is essential for trichomonad growth, and pathogenesis
390 in urogenital tract infection[49,50], we hypothesized that its consumption by intestinal *T. mu* would exert
391 pressure on bacterial iron acquisition and storage systems. Our metatranscriptomic data confirmed
392 increasing *T. mu* iron consumption through the upregulation of the pyruvate:ferredoxin oxidoreductase
393 (TMU_00017025, $p < 0.01$) and predicted ferredoxin-1 (TMU_00007489, *ns*), key iron-sulfur cluster
394 containing proteins required for trichomonad energy production (Figure 7A)[51,52]. The enzymes were
395 expressed across protist populations, with reductions in cluster 4, and in clusters 13 and 14, which aligns
396 with reduced energy requirements in the pseudocyst stage (Figure 7B). The upregulation instead of
397 ferredoxins TMU_00018447 and TMU_00005891 in clusters 4, and particularly 12 and 14 may be linked
398 to encystation-related stress responses. Similar expression of these genes in *T. mu* from conventionalized
399 and GF mice may suggest that the protist is agnostic in terms of iron acquisition to the resident microbiome.
400 In response to the increasing demand for iron from the protist, bacteria showed upregulated iron acquisition
401 and storage systems (Figures 7C and 7D). Gene expression in *T. mu*-colonized mice was enriched for
402 siderophore transport genes at day 28, predominantly in *Helicobacter*, *Bacteroides* and *Parabacteroides* (p
403 < 0.001), and for genes with identified siderophore transport *potential* already at day 2 (WT, $p < 0.05$;
404 $muMt^{-/-}$, $p < 0.001$; Figures 7C and S5C). Numerous siderophore synthesis genes were also upregulated at
405 days 2 and 28 (Figure 7D), in addition to iron gene regulation (WT, $p < 0.01$; $muMt^{-/-}$, $p < 0.05$; primarily
406 Firmicutes), iron storage (WT, *ns*; $muMt^{-/-}$, $p < 0.05$; *Mucispirillum schaedleri* and unidentified bacteria)
407 and magnetosome formation (WT, $p < 0.05$; $muMt^{-/-}$, $p < 0.01$; primarily *Bacteroides*) (Figures 7C and
408 S5C). Host B cell status appeared to have little impact on bacterial iron-related machinery in the context of
409 the protist (Figure S6B), which may suggest a more direct competition between resident microbiota and the
410 new colonizer for an increasingly limited nutrient.

411

412 Another nutrient at the interface of gut microbiota and host is the essential amino acid tryptophan.
413 Tryptophan plays a central role in the formation of bioactive metabolites related to host physiology e.g.

414 including indole, tryptamine, kynuneurine and serotonin. The gut microbiome has emerged as a potent
415 driver for modulating its metabolism[53], and we therefore examined how *T. mu* and resident bacteria might
416 compete for tryptophan and route its conversion. In *T. mu*, a putative tryptophanase (EC 4.1.99.1;
417 TMU_0006921), which converts tryptophan to indole, was significantly upregulated at day 28 relative to
418 day 2 (WT, $p < 0.05$; *muMt*$^{-/-}$, $p < 0.01$; Figure 7E). Enzymes associated with alternative pathways, in the
419 conversion of tryptophan to kynurenate (EC 3.5.1.9; TMU_00003049) or to tryptamine (EC 4.1.1.28;
420 TMU_00000947), were expressed at relatively low levels and exhibited no significant differences over
421 time, suggesting that *T. mu* preferentially drives production of indole (Figure 7E). This metabolite, also
422 previously shown to be an end product of tryptophan metabolism in *T. foetus* and *T. vaginalis*, is known to
423 impact host immune response, mucosal homeostasis and barrier function, and serves in interspecies
424 signaling among bacteria[54-56]. Select bacteria, in the presence of *T. mu*, appear to upregulate their own
425 tryptophanases (Figure 7F). Six of 19 bacterial tryptophanases were differentially expressed in *T. mu*
426 colonized-mice: the two most highly expressed increased in expression between days 2 and 28, and mapped
427 to *Bacteroides* and *Prevotella* (CDE07436; colonized vs naïve at day 2 in WT, $p < 0.05$, and *muMT*$^{-/-}$, $p <$
428 0.01; at day 28 in both hosts, $p < 0.0001$), and to *Oscillibacter* (R9MHE9; colonized vs naïve at day 28 in
429 both hosts, $p < 0.001$; WT vs *muMt*$^{-/-}$ at day 28,; day 28 vs day 2 in WT colonized, $p < 0.01$). Four identified
430 arylformamidases (EC 3.5.1.9) showed no significant *T. mu*-dependent differences, and two other enzymes
431 with possible activity in the kynurenine pathway (tryptophan 2,3-dioxygenase, EC 1.13.11.11; kynurenine
432 aminotransferase, EC 2.6.1.7) were absent after data rarefaction. This may suggest that tryptophan
433 metabolism is further driven toward the production of indole by gut bacteria in the presence of the protist.
434 In line with these observations, our single cell data showed upregulation of *T. mu* tryptophanase in the
435 presence of a constituent microbiota, and reduced expression during transition to the pseudocyst form in
436 clusters 13 and 14 (Figure 7E), consistent with reduced dependence on host mucosal homeostasis.
437

438 The consumption of arginine in the gut, an important source of energy for both trichomonads and bacteria,
439 also has an immunomodulatory role by limiting host production of antimicrobial NO[36,57]. Enzymes of the
440 associated arginine dihydrolase pathway—arginine deiminase (ADI; EC 3.5.3.6), ornithine
441 transcarbamylase (OTC; EC 2.1.3.3) and carbamate kinase (CK; EC 2.7.2.2)—are encoded by both the
442 protist and bacteria. We explored competition for the nutrient by tracking the expression of these enzymes
443 among *T. mu* and the resident microbiota. None of the enzymes expressed by *T. mu* were significantly
444 associated with either colonization time or B cell status, suggesting continual arginine consumption (Figure
445 S4D). Interestingly, the two isoforms of ADI, the only enzyme specific to this pathway, were selectively
446 upregulated in protist subpopulations, either clusters 1-4 (TMU_00005093) or clusters 5-11 and 15
447 (TMU_00022804) (Figure 7G). The same pattern was observed for the two isoforms of CK
448 (TMU_00010220 associated with clusters 1-3 and TMU_00019703 with clusters 6-11). The single OTC
449 showed greater expression in clusters 5-15, and consistently so in conventionalized mice (GF vs
450 conventionalized, clusters 5-14, $p < 0.001$; cluster 15, $p < 0.05$), while the constituent microbiota showed
451 no significant effects on ADI and CK genes. The presence of *T. mu*, however, may impact arginine
452 metabolism in select bacteria. Numerous enzymes with potential activity in the arginine dihydrolase
453 pathway were detected in bacteria (104 ADI, 740 OTC and 396 CK). In the rarefied data used to evaluate
454 gene expression, a single ADI (belonging to a member of the Lactobacillales order) was detected in *T. mu*-
455 colonized mice only, with highest expression at day 28 (Figure 7H). Of the two most highly expressed OTC
456 and CK genes each, one OTC (belonging to *Helicobacter*) increased in expression in both WT and *muMT*$^{-}$
457 $^{/-}$ colonized mice, whereas one CK (from an unclassified bacterium) increased in WT colonized mice only
458 (day 28, $p < 0.05$). These data show complex changes in microbial arginine consumption during protist
459 colonization.
460

461 The dynamics of the acquisition and metabolism machinery for these three key gut nutrients suggests
462 competitive (iron) and even cooperative relationships (tryptophan) between the protist and the resident
463 microbiota, and provide a high resolution view of complex tripartite relationships in the gut ecosystem.

464

465 **Discussion**

466 Gut colonization with *T. mu* and its close relative *T. muris* is known to stimulate the mouse immune
467 system[9,14]. Such immune changes induced by other parasitic protozoa were shown to mediate antagonistic
468 and mutualistic interactions with intestinal bacteria, suggesting the intriguing possibility that the resident
469 microbiome may mediate protist colonization or virulence, and vice versa[7,8,58]. Here, we profiled the
470 activities of *T. mu* during a new colonization event, and report its ability to reshape both resident bacterial
471 composition and gene expression in healthy hosts, and more rapidly in those with a compromised immune-
472 mediated ability to regulate the microbiota. Protist-induced shifts in bacterial composition have previously
473 been documented for other protozoa, including *G. duodenalis*, *T. foetus* and *T. gondii,* as well as in a recent
474 study of *T. mu*[8,17,18,59]. Some of the compositional changes that we noted also occurred in the study by Wei
475 *et al.*, including a reduction of *Lactobacillaceae* and increase in Bacteroidetes and *Parabacteroides*
476 abundances[17]. However, while the previous study found reduced bacterial diversity and *Helicobacter*
477 abundance after colonization, we observed the opposite in both healthy and B cell-deficient mice (possibly
478 due to distinct mouse lines, animal facilities and variations in *Tritrichomonas* strains), and indeed increased
479 *Helicobacter* RNA expression in the presence of *T. mu*. Increased abundance of *Helicobacter* has also been
480 reported after *T. foetus* colonization of cats[18].

481

482 Increased bacterial metabolism and peptidoglycan biogenesis, and particularly the shift from Firmicutes to
483 gram negative taxa (Bacteroidetes and Proteobacteria) at day 28, is a particularly interesting consequence
484 of colonization, and may signal environmental pressure favouring the growth of particular taxa. As gram
485 negative bacteria harbor surface LPS capable of boosting immunity, their increasing abundance and activity

486 may contribute to the intestinal immune activation by *T. mu*[9,60]. The earlier occurrence of these changes (at
487 day 2) in B cell-deficient mice suggests a failure to control commensal bacteria, protist expansion or both.
488 Absence of B cells is known to result in aberrant control of gut bacterial communities, and we recently
489 showed altered IgA reactivity against commensal bacteria in the presence of *T. mu*[22,61,62]. We speculate this
490 reactivity might serve to limit adverse responses by the resident bacteria as the protist attempts to colonize.
491 The reactivity of IgA against trichomonad cell surface proteins has also previously been demonstrated in
492 *T. vaginalis*, *T. foetus* and *Trichomonas tenax*[63]. Here, we show earlier upregulation of several putative
493 surface proteins in *T. mu* in the absence of B cells and IgA, including TMU_00015121, a homologue of the
494 antigenic *T. vaginalis* BspA625 protein previously detected in clinical isolates, and homologues of *T.*
495 *vaginalis* AP33 and AP120 adhesins[27]. These may signal changes in protist immunogenicity modulated by
496 B cells, or a more rapid expansion of TMU_00015121-expressing *T. mu* in their absence. This would be
497 consistent with the central role of antibodies (including IgA) in parasite clearance. While IgA-deficient
498 mice are unable to clear *Giardia* infections, for example, there is no evidence that antibodies against *T. mu*
499 will facilitate protozoan clearance from the intestine[64].
500
501 We noted distinct patterns of expression among *T. mu* subpopulations of BspA-like proteins, adhesins,
502 lectins and cysteine proteases, protein families implicated in host cell binding or parasite virulence[27,31-33].
503 This may suggest a change in host adhesive properties or antigenicity as the protist adopts different cell
504 states during its life cycle. Related trichomonads *T. vaginalis* and *T. foetus* are known to exist in one of
505 three forms—actively growing trophozoites, host-adherent amoeboid cells or environmentally resistant
506 pseudocysts[41,42,65-67]. The capture of actively metabolizing *T. mu* undergoing cell cycle in the scRNA-Seq
507 data, and cells with stress-activated pathways neighbouring those expressing relatively fewer genes (with
508 little to no expression of adhesins) suggests we have captured at least the trophozoite and pseudocyst states.
509 The upregulation of biogenic amine biosynthesis in the putative pseudocyst is in line with the increased
510 levels of biogenic amines observed, for example, during *Entamoeba invadens* encystation[40]. Although

511 adhesins were depleted in these *T. mu* cells, the binding of *T. foetus* cysts to vaginal epithelia was previously
512 demonstrated through electron microscopy, a possible survival strategy against expulsion[68]. If *T. mu*
513 pseudocyst forms are capable of cytoadherence, the distinct set of expressed BspA-like genes may
514 contribute to such a strategy. To gain access to host epithelia, particularly a trophozoite or amoeboid *T. mu*
515 would need to penetrate the intestinal mucous layer. *T. vaginalis* utilizes cysteine proteases to degrade host
516 mucins along the vaginal wall[69]. In addition to the large repertoire of cysteine proteases in the *T. mu* genome,
517 enzymes of the N-glycan degradation pathway, encoded by both *T. mu* and *T. vaginalis*, are also capable of
518 degrading O-glycans, the dominant glycans found in mucins[70].
519
520 Since encystation occurs prior to exit from the intestinal tract, or under conditions of stress, we expected
521 few pseudocysts in the caecum. In fresh caecal isolates, we indeed detected transcripts of the pseudocyst
522 marker gene in fewer than 1% of protists. This proportion increased during incubation *in vitro*, accompanied
523 by a reduction in motile cells, lending support to cluster 14 as pseudocysts. A similar transcript-labelling
524 approach could be used to capture cells in the amoeboid state. Exposure to host epithelia and cell surface
525 glycoproteins in other trichomonads triggers a transformation of trophozoites to host-adherent amoeboid
526 cells[65,66]. Identification of uniquely expressed transcripts in such cells of *T. mu* could help to track the
527 appearance of amoeboid cells, particularly in contexts of induced colitis and models of colorectal cancer
528 which are exacerbated by *T. mu* coloinzation[9]. One possibility is that we have already captured the
529 amoeboid state in our scRNA-Seq data, in supercluster 'A'. The majority of protists in these populations
530 were recovered from the GF mouse, which are known to have a more penetrable intestinal mucus[71]. These
531 cells express different suites of adhesins, cysteine proteases and enzymes involved in the degradation of
532 arginine relative to supercluster 'B', which we suggest represents the trophozoite form. Our attempts to
533 capture the cells of state 'A' in mice with conventional microbiomes using fluorescent labelling have so far
534 failed. As microscopic analysis both in our study and that of Wei *et al.* found no evidence of host adherence,
535 it is possible that the more accessible host epithelium in GF mice promotes transformation to an amoeboid

536 state[14]. This could have important implications for protist colonizations in a diseased or inflamed gut where
537 the mucus layer is compromised.
538
539 One intriguing finding was the expression of meiosis-specific genes in a small subpopulation of *T. mu* cells,
540 suggesting the presence of sexual or parasexual replication. Sexual recombination has previously been
541 hypothesized in *T. vaginalis*, based on the presence and expression of core meiotic genes in its genome (27
542 of 29)[44,72]. Synapsed chromosomes observed in small subpopulations of cultured *T. vaginalis* and linkage
543 disequilibrium in its genome, suggesting the non-random segregation of genetic alleles, have also been
544 presented as evidence of a sexual cycle[72-74]. Meiotic genes have similarly been identified in *G. duodenalis*,
545 known to undergo parasexual recombination[75]. The expression of meiosis-specific gene *hop2B* in other *T.*
546 *mu* clusters suggests this gene may have alternative roles in cellular processes, seen also in the ciliate
547 *Tetrahymenoa thermophila* which requires *hop2B* for vegetative growth[76]. While our data suggests that the
548 presence of a microbiome may promote meiosis (as noted by the higher proportion of meiotic cells in the
549 conventionalized relative to GF mice), we nonetheless note that despite extensive efforts to purify *T. mu*
550 from mouse caeca prior to infection of new mice, we detected bacterial DNA in the GF mouse. Unlike the
551 GF mouse, however, the conventionalized mouse was exposed to commensal bacteria for four weeks prior
552 to the introduction of *T. mu*. This would have allowed ample time for microbiome-induced changes in
553 intestinal morphology and immune cell differentiation, which reach a new homeostasis within eight to 30
554 days[77]. Consequently, the protist would have encountered two very different intestinal environments.
555
556 The upregulation of bacterial iron acquisition machinery, consequent to *T. mu* colonization, could be a
557 competitive response to nutrient depletion by *T. mu*, and may function as a complementary means to
558 maintain control over protist growth. Iron is an essential micronutrient. To limit the growth and virulence
559 of invading pathogens, the host restricts access to luminal iron through sequestration by proteins such as
560 lactoferrin[49,50,78]. Parasitic trichomonads require particularly high quantities to sustain their growth. In *T.*

561 *vaginalis*, enzymes of the hydrogenosome, an energy-producing organelle resembling mitochondria,
562 achieve peak activity at over 200 μM iron, whereas at 100 μM, growth is restricted and the protist
563 upregulates iron sequestration[50]. Low iron conditions have also been shown to hinder trichomonad adhesion
564 to host epithelia[79]. Iron uptake by resident bacteria has the potential, therefore, to further restrict protist
565 growth and virulence, an effect which may be lost in a context of surplus iron provided for example by
566 nutritional supplements. Previous studies have associated iron supplements with increased growth of
567 enteropathogens, diarrhea and intestinal inflammation in infants[5]. In a cohort of young children from
568 Pakistan supplemented with iron and vitamins, we recently showed increased carriage of numerous
569 intestinal protozoa and fungi, with possible carry-over effects to bacterial composition[5]. In the present
570 study, we noted blooms of siderophore transport in *Helicobacter* and iron storage in *Mucispirillum*
571 *schaedleri*, a pathobiont causally associated with Crohn's-like colitis in immunodeficient mice[80].
572 Interestingly, changes in iron levels have also been shown to alter expression of *T. vaginalis* BspA-like
573 genes *in vitro*, potentially influencing protist virulence[27]. The question of how iron supplementation or
574 restriction influences the balance and/or virulence of gut protozoa and bacteria remains an important
575 question. In addition to iron, we also noted a potential for competition over the catabolism of tryptophan,
576 with the expression of a putative tryptophanase by the protist. Recently there has been much interest in the
577 production of neuroactive kynurenines from tryptophan by the gut microbiome, and their potential role in
578 the gut brain access[81,82]. Through shifting the production of indole at the expense of alternative products,
579 tryptamine and kynurenine, our findings suggest that colonization by the protist may similarly modulate
580 the gut-brain access.
581
582 We note also several limitations in the metatranscriptomic study. Although we used littermate controls to
583 compare the microbiota of protist-colonized and naïve mice, host genotype-dependent differences may have
584 been confounded by caging effects. Differences in bacterial composition and gene expression diminished
585 over time in *T. mu* colonized mice, possibly a due in part to transfer of bacteria adherent to the protist. In

586 the microbial RNA-Seq data, we were limited in our ability to predict bacterial gene function and taxonomy
587 by the current available reference microbial data. We note that the large number of unclassified RNA
588 sequences may belong to taxa identified by 16S rRNA data but for which genome related information is
589 sparse, or to non-coding regulatory RNAs not examined in this study. Metatranscriptomics also revealed
590 the unexpected presence of murine astrovirus in mouse caeca, represented by 5.5 million reads and
591 assembled into full-length genomes belonging to at least two viral strains. As this virus is capable of altering
592 the host intestinal barrier, its presence represents a potential confounding factor in delineating protist-
593 specific effects on the microbiome[83]. It nevertheless showcases the potential of this method to identify
594 additional agents contributing to inflammatory disease. We note that we found no association between virus
595 detection, protozoan colonization and host genotype in this study.
596
597 Through a series of sequencing studies, we showed that introduction of a new protist leads to significant
598 remodeling of resident microbiota in the gut. As these changes are accelerated in the absence of B cells, we
599 suggest that antibodies moderate the expansion of the protist during the early days of colonization. We
600 furthermore demonstrated powerful means to track protist activities and cell states through combined
601 sequence-based and *in situ* methodologies, which will contribute substantially to our ability to decode the
602 strategies employed by gut parasites during colonization.
603
604 **Author contributions**
605 J.P., A.M., M.E.G., and A.P. conceived and designed the study. A.P. and A.K. isolated nucleic material for
606 high throughput sequencing, and A.P. performed sequence data analyses. E.Y.C. performed mouse
607 colonization experiments, performed flow cytometry and microscopic analyses. N.N. performed protozoan
608 enzyme annotation. N.A. and J.H. participated in sample collection and *in vitro* culturing experiments.
609 E.C.A.F. generated the phylogenetic tree. A.P., J.P., A.M. and M.E.G. wrote the manuscript. All authors
610 reviewed and/or edited the paper.

Acknowledgements

This work was funded by the Canadian Institutes for Health Research (MRT-168043) to J.P., M.E.G. and A.M; the Natural Sciences and Engineering Research Council (RGPIN-2019-06852) to J.P.; the Intramural Research Program of the National Institute of Allergy and Infectious Diseases at the National Institutes of Health to M.E.G.; a Restracomp scholarship administered by the Research Training Centre (Hospital for Sick Children) and a graduate scholarship from the Government of Ontario to A.P. A.M. is supported by the Canadian Foundation for Innovation John R. Evans Leaders Fund, the Canadian Institutes for Health Research (PJT-388337, PJT-480765) and a Natural Sciences and Engineering Research Council (RGPIN-2019-04521). A.M. is the Tier 2 Canadian Research Chair in Mucosal Immunology and supported by the Tier 2 CRC-CIHR program (CRC-2021-00511). Computing resources were provided by the SciNet High Performance Computing (HPC) Consortium; SciNet is funded by the Canada Foundation for Innovation under the auspices of Compute Canada, the Government of Ontario, Ontario Research Fund - Research Excellence, and the University of Toronto.

Declaration of interests

The authors declare no competing interests.

629 **Main figure titles and legends**

630

631

632 **Figure 1. Study design to test the impact of *T. mu* colonization on resident microbiota.**

633 (A) Schematic of microbial interactions explored in the present study using 16S sequencing,
634 metatranscriptomics (RNA-Seq) and/or single-cell transcriptomics (scRNA-Seq). (B) Experimental design.
635 Groups of four WT and *muMT*$^{-/-}$ mice were colonized with *T. mu*. Fecal pellets and caecal samples were
636 collected at indicated timepoints for 16S and metatranscriptomic profiling. (C) Expansion of *T. mu* cells in
637 WT mouse caeca and (D) mucosal IgA levels over 28 days of colonization.

638

639

640 **Figure 2. *T. mu* colonization drives bacterial diversification.**

641 (A) Relative abundance of bacterial composition as determined by 16S amplicon sequencing of mouse fecal
642 pellets. Plots represent profiles from four WT and *muMt*$^{-/-}$ mice sampled over a four-week colonization time
643 course, as well as two groups of uninfected control mice sampled at the end of the experiment. Colours
644 represent genus-level abundances within indicated phylum colour groups. (B) Bacterial richness (observed
645 OTUs) and evenness (Shannon diversity) over the course of *T. mu* colonization. (C) Principal coordinate
646 analysis showing compositional differences among samples, as determined Bray-Curtis dissimilarities. (D)
647 Patterns of bacterial changes during *T. mu* colonization in WT (grey) and *muMt*$^{-/-}$ (red) mice. Significantly
648 differentially abundant OTUs were clustered into groups based on patterns of abundance changes. Pie charts
649 show taxonomies of OTUs included in each cluster. (E) Phylum-level abundance determined through qPCR
650 using clade-specific primers, at before (day 0) or 21 days after colonization with *T. mu* in WT *(top)* and
651 *muMt*$^{-/-}$ mice *(bottom)*. Abundances were normalized to Ct values obtained with universal 16S rRNA
652 probes. Error bars represent SEM. Mann-Whitney U test for significance was performed, and significance
653 is indicated as follows: $*p < 0.05$, $** < 0.01$, *ns* not significant.

654

655

656 **Figure 3. Bacterial gene expression is altered in the presence of the protist.**

657 (A) Taxonomic profiles of putative bacterial mRNA reads. (B) Principal component analysis of bacterial
658 gene expression. Each circle denotes a sample isolated from a single mouse, coloured by host genotype,
659 presence of *T. mu* and timepoint, as shown. (C) Numbers of significantly upregulated or downregulated
660 bacterial genes between protist-colonized and naïve mice, or due to host genotype (*muMt*$^{-/-}$ vs WT), grouped
661 as indicated (n = 3 or 4). (D) Bacterial metabolic pathway enrichment associated with *T. mu* colonization.
662 Enrichment was determined by the overrepresentation of EC terms in differentially expressed genes of
663 KEGG-defined pathways. Total ECs detected in the experiment and in the rarefied gene expression matrix
664 used in the analysis are indicated in the bar charts to the right. Significantly enriched pathways are
665 represented as coloured dots: grey indicates *ns*, and the gradient from yellow to red represents decreasing
666 p values, beginning from $p<0.05$. Sizes of dots represent odds ratios. Significance was calculated using
667 Fisher's exact tests. Asterisks indicate infinite odds ratios. (E) Interaction network of cell division, cell wall
668 biogenesis and secretory genes, based on a previously determined network of *E. coli* proteins[26]. Coloured
669 nodes represent protein homologues identified in the metatranscriptomic data, and edges are previously
670 described protein-protein or functional interactions. Genes significantly up- and downregulated are denoted
671 with blue and red borders, respectively. Node sizes correlate with log2-fold changes between *T. mu*-
672 colonized and naïve mice. Colours within node pie charts represent proportions of gene expression assigned
673 to various taxa, as shown: yellows represent Proteobacteria, greens are Bacteroidetes, pinks and purples are
674 Firmicutes, and black denotes unclassified bacteria. Colours encircling groups of proteins represent
675 functional modules as previously described[26].

676

677

678 **Figure 4. Protist gene expression changes during colonization of WT and *muMt*$^{-/-}$ mice.**

679 (A) Graphical depiction of two trichomonad rRNA sequences identified in the metagenome assembly. The
680 major sequence is identical to *T. mu* and maps to 81% of reads, and the minor sequence maps to 19% of

681 reads. The sequences were aligned using MUSCLE. Mismatches in the minor sequence to *T. mu* are
682 indicated in black. (B) Phylogeny of the two distinct ribosomal ITS sequences recovered from the
683 trichomonad metagenome. An unrooted tree was constructed from aligned ITS regions using the Maximum-
684 likelihood method. The major sequence refers to *T. mu*, showing 99.4% identity to the reference, and the
685 minor sequence shares 85.4% identity with *T. muris*. Scale = number of SNPs per site. (C) Principal
686 component analysis of protist gene expression patterns over time, in WT or *muMt*$^{-/-}$ mouse caeca. (D)
687 Numbers of differentially expressed genes associated with colonization time (day 28 vs day 2) or host
688 genotype (*muMT*$^{-/-}$ vs WT). (E) Metabolic pathway enrichment in *T. mu* during colonization. The dotplot
689 represents enrichment of EC terms in KEGG-defined pathways, within the set of genes differentially
690 expressed over time or based on B cell status. Coloured dots represent those significantly enriched: the
691 gradient from yellow to red marks decreasing *p* values, beginning from $p<0.05$. Grey dots are *ns*.
692 Significance was evaluated using Fisher's exact tests, and dot sizes correlate with odds ratios. Total ECs
693 detected in the experiment are shown in the bar charts to the right. (F) Normalized expression of BspA-like
694 genes showing interaction with host B cell status and (G) differentially expressed homologous to genes
695 implicated in *T. vaginalis* virulence. Differences were evaluated using the Wilcoxon rank-sum test, and
696 adjusted for multiple testing using the Benjamini-Hochberg approach. *$p < 0.05$, ** < 0.01, *** < 0.001,
697 *ns*, non-significant.
698
699
700 **Figure 5. Single cell transcriptional profiling of caecal *T. mu* populations.**
701 (A) Protist subpopulations isolated from colonized GF and conventionalized mouse caeca. Each point
702 represents one of 6000 protists cells, clustered by gene expression profiles using UMAP dimensionality
703 reduction, and coloured by cluster designation (3000 cells from each of two hosts). Groups of clusters,
704 designated 'superclusters' A, B or C, are indicated. (B) Numbers of cells assigned to each cluster by host
705 are shown. Total cluster sizes are shown on the right. (C) Heatmap of GO term enrichment within *T. mu*
706 clusters. Rows represent significantly enriched GO terms in one or more clusters (columns). Cells are

707 coloured and clustered by (log-transformed) numbers of upregulated proteins per GO term. (D)
708 Transcriptional complexity of protist clusters, defined as numbers of genes per transcript. Boxplots indicate
709 medians and interquartile ranges. (E) Violin plots and (F) feature plots depicting G1/S, G2 and meiotic
710 scores in cells across each *T. mu* cluster. Scores were determined based on the average expression of pre-
711 defined gene sets relative to randomly chosen control genes. Significance was tested using the two-sided
712 Wilcoxon-rank sum test, *p<0.05, **<0.01, ***<0.001. (G) Heatmap showing expression of (*top* to *bottom*)
713 BspA-like proteins, putative adhesin proteins, lectins and Cysteine proteases across protist clusters. Colours
714 to the right of Cysteine proteases indicate the protein family (Pfam).
715
716
717 **Figure 6. *T. mu* subpopulations are detected *in situ*.**
718 (A) Feature plots showing the expression of select cluster-specific genes. (B) Immunofluorescence
719 micrograph of mouse cecal tissue after colonization with *T. mu* for 21 days. Nuclear staining (blue) shows
720 host and protozoan DNA, and white dotted lines denote the mucus layer. (scale bar = 10μm). (C) RNAscope
721 immunofluorescence of *T. mu* transcripts TMU_00005724 (green) and TMU_00016742 (red) on sections
722 of cecal tissue. Nuclear DNA is stained blue. Panels on either side (scale bars = 10μm), and centre images
723 (scale bars = 1μm). Frequencies of protists expressing cluster-specific genes are shown on the right. (D)
724 FISH-Flow analysis of freshly isolated *T. mu* at 21 days post colonization. Histogram shows control
725 unstained cells (grey) and cells stained with fluorescent probes for TMU_00005724 (purple) (E) FISH-
726 Flow analysis of *T. mu* hybridized with fluorescent RNA probes for the TMU_00005724 transcript,
727 multiplexed with probes for TMU_00016742, TMU_00009244 or TMU_00001185 transcripts. Contour
728 plots indicate fluorescence within isolated protists from a representative experiment from WT mice.
729 Percentages of cells positive for hybridized probes are indicated within gates. Adjacent bar plots show
730 frequencies of *T. mu* positive for transcripts 28 days post colonization in WT (grey) and *muMt*$^{-/-}$ (red) mice
731 (n=4 per group). Significance was tested using the Wilcoxon rank sum test. (F) Percentages of protists
732 freshly isolated from WT mouse caeca and colons, or cultured *in vitro* for one and three days, which were

733 positive for the TMU_00001185 transcript or (G) stained with WGA-FITC. Representative contour plots
734 of WGA-stained protists are shown (n=3 animals or culture plates per group). Significance was tested using
735 one-way ANOVA, and adjusted for multiple comparisons using Tukey's test. (H) Cytospins of WGA-
736 stained protists. Representative images are shown at 63x magnification. Error bars represent SEM. *$p <$
737 0.05, **$p < 0.01$, *ns* not significant.
738
739
740 **Figure 7. *T. mu* and resident bacteria compete for resources in the gut lumen**
741 Shown are changes in expression of bacterial and *T. mu* genes related to iron, A-D, generation of tryptophan
742 catabolites, E and F, and the arginine dihydrolase pathway, G and H. (A) Expression of putative *T. mu*
743 pyruvate:ferredoxin oxidoreductase (PFOR) and ferredoxin (Fdx) genes in metatranscriptomic data, at days
744 2 and 28 of the experiment. (B) PFOR and Fdx gene expression among protist subpopulations isolated from
745 GF and conventionalized mice. The colour intensity correlates with average expression across cells and the
746 sizes of the dots indicate the percentage of cells in each group with detected expression. (C) Enrichment of
747 iron-related gene expression in bacteria, in the context of *T. mu* colonization. Shown are numbers of DE
748 genes per iron-related gene family in *T. mu*-colonized versus naïve mice at days 2 and 28. Total numbers
749 of genes identified in each family in the metatranscriptomic data are indicated in brackets. Significance of
750 enrichment was tested using the hypergeometric test. (D) Median log2 fold-change of bacterial siderophore
751 transport and synthesis gene expression between colonized and naïve mice. Significantly DE genes are
752 shown as coloured points: the gradient from yellow to red marks decreasing p values, beginning from
753 $p<0.05$. (E) Expression of *T. mu* and (F) the most highly expressed bacterial (read count >100) enzymes
754 with predicted function in the production of tryptophan metabolites. Dotplots in E show gene expression
755 among protist subpopulations, as above. Pie charts in F represent the total numbers of tested and DE
756 bacterial enzymes. Blue and green dots in the graphic show enzyme activities identified in *T. mu* and/or
757 bacterial genes, respectively. (G) Expression of *T. mu* enzymes associated with the arginine dihydrolase
758 pathway among protist subpopulations isolated from GF conventionalized mice, as above. (H) Expression

759 of bacterial enzymes associated with the arginine dihydrolase pathway. Shown are the only ADI present in
760 the rarefied gene count matrix, and the two differentially expressed OTC and CK genes. Gene expression
761 in metatranscriptomics data was evaluated using DESeq2, and in scRNA-Seq data using Seurat. All
762 expression values are derived from normalized gene counts. Boxplots and violin plots show medians and
763 interquartile ranges. *P* values were corrected for multiple testing using the Benjamini-Hochberg approach.
764 *$p<0.05$, **<0.01, ***<0.001
765

766 **STAR Methods**

767

768 **Mice**

769 C57Bl/6J and B6.129S2-*Ighm*^{tm1Cgn}/J (*muMt*$^{-/-}$) female mice were purchased from Jackson Laboratory (Bar
770 Harbor, ME, USA) and maintained in specific pathogen free (SPF) conditions at the University of Toronto
771 Division of Comparative Medicine animal facility. Animals were housed in a closed caging system and
772 provided with an irradiated chow diet (Envigo Teklad 2918), non-acidified water (reverse-osmosis and UV-
773 sterilized) with a 12hr light/dark cycle. Experiments were conducted at 6-8W of age, using age and sex
774 matched controls. Following colonization with *T. mu*, animals were housed in separated cages to avoid
775 contamination of non-colonized mice. Animal experiments were approved by the Local Animal Care
776 Committee (LACC) at the Faculty of Medicine, University of Toronto (animal use protocol 20012400 to
777 AM).

778

779 For the single-cell RNA sequencing experiment, 12-week-old germ-free (GF), female C57Bl/6 mice were
780 purchased from Taconic Biosciences (Germantown, NY, USA) and maintained in the GF facility until the
781 day of analysis, when they were transported to specific pathogen free (SPF) conditions. Those
782 conventionalized with a microbiome received a 200 µL bacterial suspension through oral gavage, derived
783 from two C56Bl/6J fecal pellets resuspended in 1ml sterile PBS, followed by centrifugation at 1000rpm for
784 3min to remove large debris. Conventionalized mice were then maintained under SPF conditions for 4
785 weeks prior to *T. mu* colonization.

786

787 ***T. mu* colonization**

788 *T. mu* used in this study was originally isolated at the Icahn School of Medicine at the Mount Sinai animal
789 facility in New York[9]. Purification of *T. mu* from mice was performed as previously described[9]. The cecal
790 content of a colonized C57Bl/6J mouse was harvested into 20ml sterile cold PBS and filtered through a

791 70µm cell strainer. Cecal contents were centrifuged at 500g for 7min at 4°C. The supernatant was discarded,
792 and the pellet was washed twice with 40ml sterile PBS, and resuspended in 5ml 40% Percoll overlaid on
793 5ml of 80% Percoll. Samples were centrifuged at 1300g for 20min at 20°C with the breaks off. The
794 interphase containing the protist was collected, washed with sterile PBS, pelleted at 4°C, and resuspended
795 in 5ml sterile PBS and filtered as above. Protist cells were sorted into sterile PBS on a BD Influx Cell Sorter
796 (BD Biosciences, Franklin Lakes, NJ, USA) using the 100µm nozzle at 27psi at 4°C, with >99% purity.
797 Sorted cells were once more pelleted as above and resuspended in sterile PBS. Two million collected *T. mu*
798 cells were orally gavaged immediately after into each C57Bl/6J wildtype (WT) or *muMt*$^{-/-}$ mouse, or for the
799 scRNA-Seq experiment into 16-week-old female GF or previously conventionalized C57Bl/6 mice.
800
801 **Trichomonad genome sequencing and annotation**
802 High molecular weight genomic DNA was extracted from 100 million sorted tritrichomonad cells using the
803 MagAttract HMW DNA Kit (QIAGEN, Hilden, Germany). Library generation and sequencing using
804 PacBio Sequel technology on two SMRT cells were carried out at the McMaster University Farncombe
805 Metagenomics Facility (Hamilton, Canada). Reads were error corrected using CANU v1.8[84], and
806 subsequently assembled into contigs using Flye v.2.4.1[85]. Contig sequences were polished through
807 alignment with eight million 300 bp paired-end Illumina MiSeq reads (sequenced at the National Institute
808 of Allergy and Infectious Diseases, NIH, Bethesda, Maryland) using BWA[86] and Pilon v.1.23[87]. The
809 genome assembly is available at: https://github.com/ParkinsonLab/Tritrichomonas-murine-microbiome-
810 interactions/draft-genome-assembly
811
812 Genome annotation was carried out using the Maker v2.31 pipeline[88] as follows. Repetitive regions were
813 identified *de novo* and masked using RepeatModeler v.1.0.11[89] and RepeatMasker v.4.0.7[90]. Transfer RNA
814 genes were predicted with tRNAscan-SE v2.0.6[91]. *Ab initio* prediction of gene models was carried out using
815 SNAP version 2013-11-29[92] and Augustus version 3.3.1[93], with an initial two rounds of training using

816 SNAP followed by one using Augustus with BUSCO v.2.0.1[94]. Reference information for gene prediction
817 included data from related trichomonad organisms: the *T. foetus* expressed sequence tag (EST) library
818 (CX159216.1), *T. foetus* strain K (3AUP000179807) and *Trichomonas vaginalis* G3 (3AUP000001542)
819 proteomes, and available *Pentatrichomonas hominis* and *Dientamoeba fragilis* protein sequences within
820 the NCBI nr database (accessed August 21, 2019). Functional annotation including Gene Ontology (GO)
821 term and PFAM prediction was performed with InterProScan v.5.30-69.0[95] and the hmmscan algorithm,
822 from HmmerWeb version 2.41.2[96], with minimum 1e-05 E-value cut-offs. Enzyme Commission (EC)
823 numbers were assigned to proteins using the enzyme prediction program ARCHITECT[97] with minimum
824 0.5 confidence scores. Genes encoding adhesins, meiosis and cell cycle-related proteins were identified
825 based on protein-level homology with relevant *T. vaginalis* proteins[31,44] as follows. Sequences of *T.*
826 *vaginalis* proteins were retrieved from the TrichDB database[98], and submitted for sequence similarity
827 searches against the *T. mu* proteome using BLAST[99], with minimal scoring criteria of E-value 1e-5, 30%
828 sequence identity and 40% query coverage. Where proteins matched multiple candidates in *T. mu* with a
829 10% or greater difference in sequence identity, only the highest scoring match was retained.
830
831 For functional comparison of *T. mu* genes with other protozoa and a model organism, GO terms, PFAM
832 domains and EC numbers were downloaded for *T. vaginalis* G3 (release 2.0), *Giardia intestinalis*
833 assemblage A isolate WB (release 1.0) and *Entamoeba dispar* SAW760 (release 49) from VEuPathDB[100],
834 and the *Tritrichomonas foetus* (UP000179807), *Blastocystis* sp. ST1 (UP000078348) and *Saccharomyces*
835 *cerevisiae* (UP000002311) proteomes in UniProt[101] (accessed Oct 27, 2020). GO terms were summarized
836 into GO Slim categories using the R package GSEABase[102].
837
838 **Protist phylogeny analysis**
839 Ribosomal internal transcribed spacer (ITS) sequences were extracted from PacBio whole genome
840 sequence data. Parabasalid sequences downloaded from the GenBank database

841 (https://www.ncbi.nlm.nih.gov/genbank/) and used in the analysis of phylogeny are listed in Table S11.
842 Multiple sequence alignments were generated using the MUSCLE algorithm[103], and an unrooted phylogeny
843 was constructed using the Maximum-likelihood method with a Tamura-Nei model[104] in the MEGA-X
844 software[105]. Bootstrap values were generated using 1000 replicates, with only values >65% shown.
845

846 **scRNA sequencing and analysis of gut-colonizing protist**

847 One GF and one conventionalized C56Bl/6J female mouse (Taconic Biosciences, Germantown, NY, USA)
848 were colonized with *T. mu* through oral gavage, as described above. Four weeks post colonization, the GF
849 mouse was transferred to the SPF facility, where the conventionalized mouse was being housed, and protist
850 cells were isolated from the caecal contents as described above. Purified *T. mu* was immediately transferred
851 on ice to the Princess Margaret Genomics Centre (Toronto, Canada) for STAMP library preparation using
852 the DropSeq approach and sequencing on the NextSeq 500 platform (Illumina, San Diego, CA, USA)[106].
853 Reads were processed using Drop-seq Tools v.1.13[107] as per the authors' instructions and aligned to the *T.*
854 *mu* genome assembly using STAR v. 2.5.3a[108]. A total of 3,000 cells, with minimum 200 detected genes
855 and 500 transcripts, were selected from each mouse for gene expression analysis in Seurat v4[109,110]. Only
856 genes detected in a minimum of 3 cells were included. Anchor cells were identified using
857 FindVariableFeatures and FindIntegrationAnchors functions based on 2000 most variable genes, and used
858 to align the transcriptome datasets. Cells were grouped into clusters based on gene expression using graph-
859 based clustering with the following parameters: resolution=0.8, principal component dimensions=18;
860 clusters were visualized using UMAP nonlinear dimensionality reduction[111].
861

862 Differences in metabolic pathway regulation among protist cell clusters were identified as follows.
863 Differentially expressed genes were identified for each cluster using the FindAllMarkers function in Seurat,
864 with minimum *p*-value 0.05, adjusted for multiple testing using Bonferroni correction. All enzymes in the
865 *T. mu* gene set were identified by their assigned EC numbers. Differential regulation of metabolic pathways

866 (the set of enzymes associated with each pathway, as defined by the Kyoto Encyclopedia of Genes and
867 Genomes, KEGG[112]) between cell clusters was determined using the hypergeometric test, to test for
868 overrepresentation of pathway enzymes within the differentially expressed gene set. Only pathways having
869 a minimum of 5 and maximum of 500 enzymes in the *T. mu* genome assembly were considered. Functional
870 enrichment within differentially expressed genes was also performed using GO term overrepresentation
871 with the topGO package[113] and the Fisher's Exact test. The AddModuleScore function in Seurat was used
872 to score enrichment of custom-defined gene sets (e.g. meiosis, G1/S and G2 phase genes), and score
873 differences between clusters were tested for significance using the two-sided Wilcoxon-rank sum test.
874 Benjamini-Hochberg correction was applied for multiple testing[114]. Heatmaps of gene expression were
875 generated using the pheatmap 1.0.12[115] package in R and clustered using the Ward.D2 hierarchical
876 clustering algorithm.

877

878 **16S microbiome profiling and metatranscriptomics**

879 Groups of four WT or *muMt*$^{-/-}$ six-week-old female mice were infected with two million protists each by
880 oral gavage for profiling of 16S rRNA and microbial gene expression. A schematic of the experiment is
881 presented in Figure 1a. Protist burdens were determined in mouse caeca using a hemocytometer.

882

883 *16S microbiome profiling.* For 16S profiling, mouse fecal pellets were collected from individual mice and
884 stored at -80°C prior to *T. mu* colonization and at days 2, 7, 14, 21 and 28 post-colonization. Pellets from
885 two groups of uninfected control mice were collected at day 28 for comparison. Microbial DNA was
886 extracted using the DNeasy PowerSoil Kit (QIAGEN, Hilden, Germany) according to manufacturer
887 instructions, with an initial 5min bead beating step using the FastPrep-24 homogenizer (MP Biomedicals,
888 Santa Ana, CA, USA) at 5M/s. Generation of 16S rRNA amplicons and sequencing were carried out by
889 the Integrated Microbiome Resource facility at Dalhousie University (Halifax, Canada). Briefly, 16S
890 variable regions 6 to 8 (V6-V8) were amplified using B969F and BA1406R primers[116] and Phusion DNA

891 polymerase (New England Biolabs, Ipswich, MA, USA). The first amplification round was carried out
892 using primers fused to Illumina adapters and multiplexing barcodes, followed by 30 cycles as follows: 98°C
893 for 10s, 55°C for 30s, 72°C for 30s, and a final extension at 72°C for 5min. Libraries were generated and
894 250 bp paired-end sequencing was carried out with MiSeq v2 chemistry (Illumina, San Diego, CA, USA).
895 Data were quality filtered, denoised and clustered *de novo* to 97% operational taxonomic units (OTUs)
896 using Dada2[117] within QIIME2[118] (parameters: truncate f 280, r 240, trim 10 (f/r), min-parent-fold 5). OTU
897 taxonomies were assigned using QIIME2 with a custom classifier trained on V6-V8 regions extracted from
898 SILVA[119] v132 nr SSU sequences and clustered to 97% sequence identity.
899
900 *Bacterial diversity and differential abundance.* Diversity and multivariate analyses were performed with
901 R-packages Phyloseq 1.26.1[120] and vegan v.2.5[121]. Microbiota richness (number of OTUs) and evenness
902 (Shannon Diversity Index, H) values were averaged from 100 independently rarefied datasets to the
903 minimum read depth (19,427). To evaluate the effects of *T. mu* colonization time and host genotype on
904 bacterial diversity, we implemented linear mixed effects (LME) models using the package lme4[122], and
905 performed a two-way ANOVA on day 28 samples to compare diversities of infected and naïve mice. Beta
906 diversity analyses, using Bray Curtis and weighted Unifrac (wUnifrac) dissimilarity scores, were carried
907 out on rarefied data, using OTUs present at a minimum of 10 reads in 10% of samples, and significance
908 was tested using the adonis function in vegan. Differential taxon abundance was evaluated with DESeq2
909 v.1.22.2[123], and groups of taxa with similar changes in abundance were identified using DEGreport[124].
910
911 *Microbial metatranscriptomics.* Caecal contents were recovered from sacrificed *T. mu*-colonized and
912 uninfected mice at days 2 and 28 of the experiment and immediately frozen at -80°C. On the same day, 30-
913 90mg caecal samples were added to 1ml ice cold TRIzol (Invitrogen, Carlsbad, CA, USA) containing
914 200mg 0.1mm Zirconia beads (BioSpec Products, Bartlesville, OK, USA). Samples were homogenized
915 with three 2min cycles of bead-beating at 25 Hz using the TissueLyser (QIAGEN, Hilden, Germany), and

916 chilled on ice for 5min between cycles. Homogenized samples were incubated on ice for an additional
917 15min, followed by 5min at room temperature. Total RNA extraction was achieved using the
918 PureLink RNA Mini Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's TRIzol Plus
919 Total Transcriptome Isolation protocol. Prokaryotic and eukaryotic rRNA was depleted using the QIAseq
920 FastSelect rRNA HMR and 5S/16S/23S kits (QIAGEN, Hilden, Germany). Sequencing libraries were
921 prepared using NEBNext Ultra II Directional RNA Library Prep Kits for Illumina (New England Biolabs,
922 Ipswich, MA, USA) and paired-end sequenced (150 bp) on the NovaSeq 6000 platform (Illumina, San
923 Diego, CA, USA) at Genome Quebec (Montréal, Canada). Low quality reads, host and vector
924 contamination, rRNA and tRNA reads were filtered out using the MetaPro pipeline[125]. Remaining reads
925 were mapped to the *T. mu* genome assembly using STAR v. 2.5.3a[108], and all unmapped reads were returned
926 to MetaPro for contig assembly, gene model prediction and functional and taxonomic annotation. Proteins
927 with iron-related functions were identified using FeGenie[126] searches of the annotated gene and protein
928 sequences from the metatranscriptome as well as the *T. mu* proteome. Putative cell wall biogenesis
929 machinery was identified by conducting sequence similarity searches of translated genes and proteins in
930 the metatranscriptomics data against previously described proteins[26] in the *E. coli* MG1655 proteome using
931 DIAMOND v0.9.22[127] with a minimum E-value 1e-10.
932
933 *Differential gene expression and functional enrichment.* Differential gene expression analyses were
934 performed using DESeq2 v.1.22.2[123]. Analysis for *T. mu* was performed on genes prefiltered for a minimum
935 detection of two reads in two samples. Due to differences in read depth of bacterial genes between *T. mu*-
936 colonized and naïve samples, gene count matrices were rarefied to 600,000 reads per sample. Differential
937 gene expression was calculated for 15 independent rarefactions, and median log2-fold changes and average
938 false discovery rate-corrected *p* values were recorded. Principal component analyses of *T. mu* and bacterial
939 gene expression were performed on DESeq2 variance stabilized counts. Functional enrichment was
940 evaluated using EC or GO term overrepresentation in differentially expressed genes, as described for single

941 cell data above, where enrichment of up- or downregulated enzymes in any given pathway was used as a
942 proxy for pathway activity.
943
944 **RNA-FISH Flow Cytometry**
945 Purified *T. mu* were divided into staining aliquots of 4 million cells and samples were washed twice with
946 PBS by spinning at 500g for 7min at 4°C. Cells were stained with Zombie Aqua viability dye (BioLegend,
947 San Diego, CA, USA) in PBS in the dark (1:1000 – 15min – 4°C), washed twice with PBS, and the
948 remaining steps were performed as described in the Stellaris RNA-FISH protocol for cells in suspension.
949 After hybridization, cells were resuspended in Wash Buffer B and analyzed on a BD LSR Fortessa X-20
950 cell analyzer (BD Biosciences, Franklin Lakes, NJ, USA). For detection of chitin on putative pseudocysts,
951 *T. mu* was isolated from the mouse caecum and colon at day 21 post colonization and stained with wheat
952 germ agglutinin (WGA)-FITC (Sigma Aldrich, St. Louis, MO, USA) in FACS buffer (PBS – 2% FBS –
953 5mn EDTA) in the dark. To monitor encystation of protists upon isolation from the gastrointestinal tract,
954 isolated protists were either stained immediately or cultured anaerobically *in vitro* for one to three days
955 prior to staining. Cytospins were visualized using a Zeiss AXIO Observer microscope (Carl Zeiss AG, Jena,
956 Germany).
957
958 **RNAScope**
959 RNAScope was performed as indicated in the RNAScope Multiplex Fluorescent Reagent Kit v2 (Advanced
960 Cell Diagnostics, Newark, CA, USA) protocol. Approximately 0.5cm long cecum sections were excised
961 from *T. mu*-colonized C57Bl/6J WT or *muMt*$^{-/-}$ mice, placed in 10% neutral buffered formalin and fixed
962 overnight at RT with gentle agitation. The next day, samples were gently washed twice with PBS, placed
963 in 70% ethanol, and embedded in paraffin and sliced to 7µm sections at the Toronto Centre for
964 Phenogenomics. Paraffin sections were pretreated as follows. Samples were baked at 40°C for 30min in a
965 HybEZ Oven (Advanced Cell Diagnostics, Newark, CA, USA), and hydrogen peroxide was applied at RT

966 for 10min. Antigen target retrieval was conducted at 99°C in a water bath under the 15min standard antigen
967 retrieval procedure. A barrier was created around sections using an ImmEdge pen (Vector Laboratories,
968 Burlingame, CA, USA) and allowed to dry at RT for 15min. Samples were treated with protease at 40°C
969 for 30min in a HybEZ Oven, and stored in sterile saline sodium citrate solution overnight (175.3g NaCl –
970 88.2g sodium citrate – 800ml ddH2O – pH 7). Multiplex probe amplification was conducted as per the
971 manufacturer's instructions: TSA Plus Fluorophores Fluorescein and Cyanine 3 were hybridized against
972 protist probes TMU_00005724 and TMU_00016742 respectively, and samples were visualized using a
973 Zeiss AXIO Observer microscope (Carl Zeiss AG, Jena, Germany).
974
975 **qPCR**
976 *DNA Isolation for quantitative PCR.* DNA isolation was performed from frozen fecal pellets using the
977 QIAQuick qPCR purification kit (QIAGEN Hilden, Germany) with modifications. Fecal weights were
978 measured, and pellets were thawed on ice for 10min. SDS buffer (8g SDS – 30mL H_20 – 0.22µm) and DNA
979 buffer A (20mM Tris pH 8 – 2mM EDTA – 200mM NaCl) were combined in a 1:1.41 ratio to prepare SDS-
980 A buffer. Samples were transferred to cryogenic 2ml tubes and 400µl zirconium beads, 550µl of phenol-
981 chloroform, and 482µl of SDS-A buffer were mixed with each sample. Samples were homogenized using
982 the Mini-BeadBeater (Biospec Products, Bartlesville, OK, USA) at 4000rpm for 5min, centrifuged at 1800g
983 for 5min, and the supernatant was transferred into a PCR purification plate containing 650µl of PM buffer.
984 Samples were washed twice with PE buffer, centrifuged and eluted into 100µl H_2O. DNA concentrations
985 were quantified using the Quant-iT kit (Thermo Fisher Scientific, Waltham, MA, USA).
986
987 *RNA isolation and reverse transcription of mRNA for quantitative real time RT PCR.* Two million *T. mu*
988 cells were combined with 500µL TRIzol (Invitrogen, Waltham, MA, USA) and incubated at RT for 5min.
989 100µL chloroform was added to each sample, samples were mixed and incubated at RT for 3min and
990 centrifuged for 15min at 12,000g at 4°C. The aqueous phase containing RNA was mixed with 250µL

991 isopropanol, mixed and incubated at 4°C for 10min. Samples were centrifuged for 10min at 12,000g at 4°C,
992 the supernatant was discarded and 500µL of 75% ethanol was added. Samples were vortexed for 10s and
993 centrifuged for 5min at 7,500g at 4°C. The supernatant was discarded and samples were allowed to air dry
994 for 10min at RT. RNA was resuspended in 50µL of RNase-free water and stored at -80°C. *T. mu* mRNA
995 was converted to cDNA using the High-Capacity RNA-to-cDNA kit (Applied Biosystems, Waltham, MA,
996 USA). In brief, 2µg purified RNA per reaction was mixed with 2X Reverse Transcription Buffer Mix, 20X
997 Reverse Transcription Enzyme Mix and nuclease free H_2O (up to 20µL). The thermal cycling conditions
998 were conducted as follows: 37°C for 60min – 95°C for 5min – infinite hold at 4°C.
999
1000 *Quantitative Real-Time RT-PCR*. qPCR was performed following the PowerTrack SYBR Green Master
1001 Mix kit instructions (Applied Biosystems, Waltham, MA, USA). In brief, 10ng of DNA was mixed with
1002 2X PowerTrack Master Mix, and 400nM of forward/reverse primers in 10µL reactions. The thermal cycling
1003 conditions were as follows: enzyme inactivation at 95°C for 2 min, 40 cycles of denaturation at 95°C for
1004 15s and annealing/extension at 60°C for 60s. Relative expression of target genes was normalized to the *T.*
1005 *mu* glyceraldehyde 3-phosphate dehydrogenase (GAPDH) or universal 16S gene for *T. mu* and bacterial
1006 genes respectively, and calculated using the delta-delta Ct method. Primer sequences are included in Table
1007 S12.
1008
1009 **IgA ELISA**
1010 Between 100µl and 200µl of mouse blood was collected and serum was isolated using Microvette CV 300
1011 Z serum separation tubes (SARSTEDT, Newton, NC. USA) according to the manufacturer's protocol, and
1012 stored at -80°C. Immunoglobulins were absorbed onto high-bound Nunc MaxiSorp ELISA plates
1013 (BioLegend, San Diego, CA, USA) and anti-IgA ELISA was performed using anti-mouse IgA-HRP
1014 antibody (SouthernBiotech, Birmingham, AL, USA). In brief, 50µl of anti-mouse Ig master mix
1015 (SouthernBiotech, Birmingham, AL, USA) was diluted 1:2000 in PBS, and used to coat 96-well ELISA

1016 plates overnight at 4°C. The following day, 200µl of 1% bovine serum albumin (Sigma-Aldrich, St. Louis,
1017 MI, USA) in PBS was added to plates to block for 1hr at 37°C. 50µl of purified IgA (1:1000 in assay buffer)
1018 or serum (1:100 in assay buffer) was added to 100µl assay buffer (0.1M Tris-HCl, 0.1M NaCl, 1% FBS,
1019 0.05% Tween-20), and 6x 1 in 3 serial dilutions were prepared. Mixtures were incubated for 1 hour at 37°C
1020 and washed 3 times with 200µl 0.1% PBST. After adding 50µl of anti-mouse IgA-HRP (1:1000 in assay
1021 buffer) (SouthernBiotech, Birmingham, AL, USA), mixtures were incubated for 1 hour more at 37°C and
1022 washes repeated as before. 50µl of TMB substrate (BioLegend, San Diego, CA, USA) was added, incubated
1023 for 30s, and reactions were stopped with the addition of 50µl of H_2SO_4. Concentrations were measured by
1024 spectrophotometry (450 nm spectral absorbance) and values were calibrated against those of the purified
1025 IgA standard curve.
1026

1027 **Data availability**

1028 Sequence data generated in this study have been deposited to the NCBI Sequence Read Archive under the
1029 BioProject identifiers PRJNA913581 and PRJNA914770.
1030

1031 **Supplemental item titles**

1032

1033 Table S1. 16S sequencing read and OTU counts

1034 Table S2. 16S OTU table

1035 Table S3. Metatranscriptomic read processing and classification

1036 Table S4. Metagenome characteristics.

1037 Table S5. Frequency of predicted transfer RNA genes in the metagenome assembly.

1038 Table S6. Putative *T. mu* homologues of potential virulence and meiosis-specific genes.

1039 Table S7. *T. mu* HMMER protein domain predictions

1040 Table S8. Genes differentially expressed in *T. mu* during WT or B-cell deficient mouse colonization

1041 Table S9. Single cell RNA sequencing statistics of *T. mu*

1042 Table S10. Read counts for genes containing BspA-like domains across *T. mu* subpopulations

1043 Table S11. Parabasalid ribosomal ITS sequences used in the analysis of phylogeny

1044 Table S12. DNA primers used for quantitative RT-PCR

1045

1046 **References**

1047

1048 1 Fan, Y. & Pedersen, O. Gut microbiota in human metabolic health and disease. *Nat Rev*
1049 *Microbiol* **19**, 55-71, doi:10.1038/s41579-020-0433-9 (2021).
1050 2 Libertucci, J. & Young, V. B. The role of the microbiota in infectious diseases. *Nat Microbiol* **4**,
1051 35-45, doi:10.1038/s41564-018-0278-4 (2019).
1052 3 Sardinha-Silva, A., Alves-Ferreira, E. V. C. & Grigg, M. E. Intestinal immune responses to
1053 commensal and pathogenic protozoa. *Front Immunol* **13**, 963723,
1054 doi:10.3389/fimmu.2022.963723 (2022).
1055 4 Scanlan, P. D. *et al.* The microbial eukaryote Blastocystis is a prevalent and diverse member of
1056 the healthy human gut microbiota. *FEMS microbiology ecology* **90**, 326-330, doi:10.1111/1574-
1057 6941.12396 (2014).
1058 5 Popovic, A. *et al.* Micronutrient supplements can promote disruptive protozoan and fungal
1059 communities in the developing infant gut. *Nat Commun* **12**, 6729, doi:10.1038/s41467-021-
1060 27010-3 (2021).
1061 6 Benson, A., Pifer, R., Behrendt, C. L., Hooper, L. V. & Yarovinsky, F. Gut commensal bacteria
1062 direct a protective immune response against Toxoplasma gondii. *Cell Host Microbe* **6**, 187-196,
1063 doi:10.1016/j.chom.2009.06.005 (2009).
1064 7 Chappell, C. L. *et al.* Fecal Indole as a Biomarker of Susceptibility to Cryptosporidium Infection.
1065 *Infect Immun* **84**, 2299-2306, doi:10.1128/IAI.00336-16 (2016).
1066 8 Raetz, M. *et al.* Parasite-induced TH1 cells and intestinal dysbiosis cooperate in IFN-gamma-
1067 dependent elimination of Paneth cells. *Nat Immunol* **14**, 136-142, doi:10.1038/ni.2508 (2013).
1068 9 Chudnovskiy, A. *et al.* Host-Protozoan Interactions Protect from Mucosal Infections through
1069 Activation of the Inflammasome. *Cell* **167**, 444-456 e414, doi:10.1016/j.cell.2016.08.076 (2016).
1070 10 Jokelainen, P. *et al.* Dientamoeba fragilis, a Commensal in Children in Danish Day Care Centers.
1071 *J Clin Microbiol* **55**, 1707-1713, doi:10.1128/JCM.00037-17 (2017).
1072 11 Cuffari, C., Oligny, L. & Seidman, E. G. Dientamoeba fragilis masquerading as allergic colitis. *J*
1073 *Pediatr Gastroenterol Nutr* **26**, 16-20, doi:10.1097/00005176-199801000-00003 (1998).
1074 12 Yakoob, J. *et al.* Blastocystis hominis and Dientamoeba fragilis in patients fulfilling irritable
1075 bowel syndrome criteria. *Parasitol Res* **107**, 679-684, doi:10.1007/s00436-010-1918-7 (2010).
1076 13 Zhang, N. *et al.* High prevalence of Pentatrichomonas hominis infection in gastrointestinal cancer
1077 patients. *Parasit Vectors* **12**, 423, doi:10.1186/s13071-019-3684-4 (2019).
1078 14 Escalante, N. K. *et al.* The common mouse protozoa Tritrichomonas muris alters mucosal T cell
1079 homeostasis and colitis susceptibility. *The Journal of experimental medicine* **213**, 2841-2850,
1080 doi:10.1084/jem.20161776 (2016).
1081 15 Tuzlak, L. *et al.* Ultrastructural identification and molecular characterization of two new
1082 parabasalid species that naturally colonize laboratory mice, Tritrichomonas musculus and
1083 Tritrichomonas casperi. *BioRxiv*, doi:https://doi.org/10.1101/2023.01.20.524969 (2023).
1084 16 Chiaranunt, P. *et al.* NLRP1B and NLRP3 Control the Host Response following Colonization
1085 with the Commensal Protist Tritrichomonas musculis. *J Immunol* **208**, 1782-1789,
1086 doi:10.4049/jimmunol.2100802 (2022).
1087 17 Wei, Y. *et al.* Commensal Bacteria Impact a Protozoan's Integration into the Murine Gut
1088 Microbiota in a Dietary Nutrient-Dependent Manner. *Appl Environ Microbiol* **86**,
1089 doi:10.1128/AEM.00303-20 (2020).
1090 18 Bierlein, M., Hedgespeth, B. A., Azcarate-Peril, M. A., Stauffer, S. H. & Gookin, J. L. Dysbiosis
1091 of fecal microbiota in cats with naturally occurring and experimentally induced Tritrichomonas
1092 foetus infection. *PLoS One* **16**, e0246957, doi:10.1371/journal.pone.0246957 (2021).
1093 19 Schneider, C. *et al.* A Metabolite-Triggered Tuft Cell-ILC2 Circuit Drives Small Intestinal
1094 Remodeling. *Cell* **174**, 271-284 e214, doi:10.1016/j.cell.2018.05.014 (2018).

1095 20 Howitt, M. R. *et al.* Tuft cells, taste-chemosensory cells, orchestrate parasite type 2 immunity in
1096 the gut. *Science* **351**, 1329-1333, doi:10.1126/science.aaf1648 (2016).
1097 21 Ost, K. S. *et al.* Adaptive immunity induces mutualism between commensal eukaryotes. *Nature*
1098 **596**, 114-118, doi:10.1038/s41586-021-03722-w (2021).
1099 22 Pabst, O., Cerovic, V. & Hornef, M. Secretory IgA in the Coordination of Establishment and
1100 Maintenance of the Microbiota. *Trends Immunol* **37**, 287-296, doi:10.1016/j.it.2016.03.002
1101 (2016).
1102 23 Rojas, O. L. *et al.* Recirculating Intestinal IgA-Producing Cells Regulate Neuroinflammation via
1103 IL-10. *Cell* **176**, 610-624 e618, doi:10.1016/j.cell.2018.11.035 (2019).
1104 24 Polk, D. B. & Peek, R. M., Jr. Helicobacter pylori: gastric cancer and beyond. *Nat Rev Cancer*
1105 **10**, 403-414, doi:10.1038/nrc2857 (2010).
1106 25 Partridge, J. D., Scott, C., Tang, Y., Poole, R. K. & Green, J. Escherichia coli transcriptome
1107 dynamics during the transition from anaerobic to aerobic conditions. *J Biol Chem* **281**, 27806-
1108 27815, doi:10.1074/jbc.M603450200 (2006).
1109 26 Peregrin-Alvarez, J. M., Xiong, X., Su, C. & Parkinson, J. The Modular Organization of Protein
1110 Interactions in Escherichia coli. *PLoS Comput Biol* **5**, e1000523,
1111 doi:10.1371/journal.pcbi.1000523 (2009).
1112 27 Noel, C. J. *et al.* Trichomonas vaginalis vast BspA-like gene family: evidence for functional
1113 diversity from structural organisation and transcriptomics. *BMC Genomics* **11**, 99,
1114 doi:10.1186/1471-2164-11-99 (2010).
1115 28 Wilson, I. W. *et al.* Genetic Diversity and Gene Family Expansions in Members of the Genus
1116 Entamoeba. *Genome Biol Evol* **11**, 688-705, doi:10.1093/gbe/evz009 (2019).
1117 29 Inagaki, S., Onishi, S., Kuramitsu, H. K. & Sharma, A. Porphyromonas gingivalis vesicles
1118 enhance attachment, and the leucine-rich repeat BspA protein is required for invasion of epithelial
1119 cells by "Tannerella forsythia". *Infect Immun* **74**, 5023-5028, doi:10.1128/IAI.00062-06 (2006).
1120 30 Rego, S. *et al.* Structural and Functional Analysis of Cell Wall-anchored Polypeptide Adhesin
1121 BspA in Streptococcus agalactiae. *J Biol Chem* **291**, 15985-16000, doi:10.1074/jbc.M116.726562
1122 (2016).
1123 31 Garcia, A. F. & Alderete, J. Characterization of the Trichomonas vaginalis surface-associated
1124 AP65 and binding domain interacting with trichomonads and host cells. *BMC Microbiol* **7**, 116,
1125 doi:10.1186/1471-2180-7-116 (2007).
1126 32 Guzman-Tellez, P. *et al.* Lectins as virulence factors in Entamoeba histolytica and free-living
1127 amoebae. *Future Microbiol* **15**, 919-936, doi:10.2217/fmb-2019-0275 (2020).
1128 33 Siqueira-Neto, J. L. *et al.* Cysteine proteases in protozoan parasites. *PLoS Negl Trop Dis* **12**,
1129 e0006512, doi:10.1371/journal.pntd.0006512 (2018).
1130 34 Das, P., Lahiri, A., Lahiri, A. & Chakravortty, D. Modulation of the arginase pathway in the
1131 context of microbial pathogenesis: a metabolic enzyme moonlighting as an immune modulator.
1132 *PLoS Pathog* **6**, e1000899, doi:10.1371/journal.ppat.1000899 (2010).
1133 35 Rath, M., Muller, I., Kropf, P., Closs, E. I. & Munder, M. Metabolism via Arginase or Nitric
1134 Oxide Synthase: Two Competing Arginine Pathways in Macrophages. *Front Immunol* **5**, 532,
1135 doi:10.3389/fimmu.2014.00532 (2014).
1136 36 Stadelmann, B., Merino, M. C., Persson, L. & Svard, S. G. Arginine consumption by the
1137 intestinal parasite Giardia intestinalis reduces proliferation of intestinal epithelial cells. *PLoS One*
1138 **7**, e45325, doi:10.1371/journal.pone.0045325 (2012).
1139 37 Yarlett, N., Lindmark, D. G., Goldberg, B., Moharrami, M. A. & Bacchi, C. J. Subcellular
1140 localization of the enzymes of the arginine dihydrolase pathway in Trichomonas vaginalis and
1141 Tritrichomonas foetus. *J Eukaryot Microbiol* **41**, 554-559, doi:10.1111/j.1550-
1142 7408.1994.tb01516.x (1994).
1143 38 Hapfelmeier, S. *et al.* Reversible microbial colonization of germ-free mice reveals the dynamics
1144 of IgA immune responses. *Science* **328**, 1705-1709, doi:10.1126/science.1188454 (2010).

1145 39 Hernandez, Y. *et al.* Novel role of sphingolipid synthesis genes in regulating giardial encystation.
1146 *Infect Immun* **76**, 2939-2949, doi:10.1128/IAI.00116-08 (2008).
1147 40 Jeelani, G. *et al.* Metabolic profiling of the protozoan parasite Entamoeba invadens revealed
1148 activation of unpredicted pathway during encystation. *PLoS One* **7**, e37740,
1149 doi:10.1371/journal.pone.0037740 (2012).
1150 41 Pereira-Neves, A. & Benchimol, M. Tritrichomonas foetus: budding from multinucleated
1151 pseudocysts. *Protist* **160**, 536-551, doi:10.1016/j.protis.2009.05.001 (2009).
1152 42 Beri, D. *et al.* Demonstration and Characterization of Cyst-Like Structures in the Life Cycle of
1153 Trichomonas vaginalis. *Front Cell Infect Microbiol* **9**, 430, doi:10.3389/fcimb.2019.00430
1154 (2019).
1155 43 Horlock-Roberts, K. *et al.* Drug-Free Approach To Study the Unusual Cell Cycle of Giardia
1156 intestinalis. *mSphere* **2**, doi:10.1128/mSphere.00384-16 (2017).
1157 44 Malik, S. B., Pightling, A. W., Stefaniak, L. M., Schurko, A. M. & Logsdon, J. M., Jr. An
1158 expanded inventory of conserved meiotic genes provides evidence for sex in Trichomonas
1159 vaginalis. *PLoS One* **3**, e2879, doi:10.1371/journal.pone.0002879 (2007).
1160 45 Hirt, R. P. *et al.* Trichomonas vaginalis pathobiology new insights from the genome sequence.
1161 *Adv Parasitol* **77**, 87-140, doi:10.1016/B978-0-12-391429-3.00006-X (2011).
1162 46 Brennand, A. *et al.* Autophagy in parasitic protists: unique features and drug targets. *Mol*
1163 *Biochem Parasitol* **177**, 83-99, doi:10.1016/j.molbiopara.2011.02.003 (2011).
1164 47 Arroyo-Begovich, A., Carabez-Trejo, A. & Ruiz-Herrera, J. Identification of the structural
1165 component in the cyst wall of Entamoeba invadens. *J Parasitol* **66**, 735-741 (1980).
1166 48 Ward, H. D., Alroy, J., Lev, B. I., Keusch, G. T. & Pereira, M. E. Identification of chitin as a
1167 structural component of Giardia cysts. *Infect Immun* **49**, 629-634, doi:10.1128/iai.49.3.629-
1168 634.1985 (1985).
1169 49 Castro, C. *et al.* Iron-modulated pseudocyst formation in Tritrichomonas foetus. *Parasitology*
1170 **143**, 1034-1042, doi:10.1017/S0031182016000573 (2016).
1171 50 Lehker, M. W. & Alderete, J. F. Iron regulates growth of Trichomonas vaginalis and the
1172 expression of immunogenic trichomonad proteins. *Mol Microbiol* **6**, 123-132, doi:10.1111/j.1365-
1173 2958.1992.tb00844.x (1992).
1174 51 Vanacova, S. *et al.* Iron-induced changes in pyruvate metabolism of Tritrichomonas foetus and
1175 involvement of iron in expression of hydrogenosomal proteins. *Microbiology (Reading)* **147**, 53-
1176 62, doi:10.1099/00221287-147-1-53 (2001).
1177 52 Horvathova, L. *et al.* Transcriptomic identification of iron-regulated and iron-independent gene
1178 copies within the heavily duplicated Trichomonas vaginalis genome. *Genome Biol Evol* **4**, 1017-
1179 1029, doi:10.1093/gbe/evs078 (2012).
1180 53 Agus, A., Planchais, J. & Sokol, H. Gut Microbiota Regulation of Tryptophan Metabolism in
1181 Health and Disease. *Cell Host Microbe* **23**, 716-724, doi:10.1016/j.chom.2018.05.003 (2018).
1182 54 Lloyd, D., Lauritsen, F. R. & Degn, H. The parasitic flagellates Trichomonas vaginalis and
1183 Tritrichomonas foetus produce indole and dimethyl disulphide: direct characterization by
1184 membrane inlet tandem mass spectrometry. *J Gen Microbiol* **137**, 1743-1747,
1185 doi:10.1099/00221287-137-7-1743 (1991).
1186 55 Li, X., Zhang, B., Hu, Y. & Zhao, Y. New Insights Into Gut-Bacteria-Derived Indole and Its
1187 Derivatives in Intestinal and Liver Diseases. *Front Pharmacol* **12**, 769501,
1188 doi:10.3389/fphar.2021.769501 (2021).
1189 56 Zhang, L. S. & Davies, S. S. Microbial metabolism of dietary components to bioactive
1190 metabolites: opportunities for new therapeutic interventions. *Genome Med* **8**, 46,
1191 doi:10.1186/s13073-016-0296-x (2016).
1192 57 Novak, L. *et al.* Arginine deiminase pathway enzymes: evolutionary history in metamonads and
1193 other eukaryotes. *BMC Evol Biol* **16**, 197, doi:10.1186/s12862-016-0771-4 (2016).
1194 58 Yilmaz, B. *et al.* Gut microbiota elicits a protective immune response against malaria
1195 transmission. *Cell* **159**, 1277-1289, doi:10.1016/j.cell.2014.10.053 (2014).

1196 59 Barash, N. R., Maloney, J. G., Singer, S. M. & Dawson, S. C. Giardia Alters Commensal
1197 Microbial Diversity throughout the Murine Gut. *Infect Immun* **85**, doi:10.1128/iai.00948-16
1198 (2017).
1199 60 Wang, J., Chen, W. D. & Wang, Y. D. The Relationship Between Gut Microbiota and
1200 Inflammatory Diseases: The Role of Macrophages. *Front Microbiol* **11**, 1065,
1201 doi:10.3389/fmicb.2020.01065 (2020).
1202 61 Cao, E. *et al.* The protozoan commensal Tritrichomonas musculis is a natural adjuvant for
1203 mucosal IgA. *BioRxiv*, doi:https://doi.org/10.1101/2022.10.08.511442 (2022).
1204 62 Catanzaro, J. R. *et al.* IgA-deficient humans exhibit gut microbiota dysbiosis despite secretion of
1205 compensatory IgM. *Sci Rep* **9**, 13574, doi:10.1038/s41598-019-49923-2 (2019).
1206 63 Mundodi, V., Kucknoor, A. S., Chang, T. H. & Alderete, J. F. A novel surface protein of
1207 Trichomonas vaginalis is regulated independently by low iron and contact with vaginal epithelial
1208 cells. *BMC Microbiol* **6**, 6, doi:10.1186/1471-2180-6-6 (2006).
1209 64 Langford, T. D. *et al.* Central importance of immunoglobulin A in host defense against Giardia
1210 spp. *Infect Immun* **70**, 11-18, doi:10.1128/IAI.70.1.11-18.2002 (2002).
1211 65 Kusdian, G., Woehle, C., Martin, W. F. & Gould, S. B. The actin-based machinery of
1212 Trichomonas vaginalis mediates flagellate-amoeboid transition and migration across host tissue.
1213 *Cell Microbiol* **15**, 1707-1721, doi:10.1111/cmi.12144 (2013).
1214 66 Petropolis, D. B., Fernandes Rodrigues, J. C., da Rocha-Azevedo, B. & Silva-Filho, F. C. E. The
1215 binding of Tritrichomonas foetus to immobilized laminin-1 and its role in the cytotoxicity exerted
1216 by the parasite. *Microbiology (Reading)* **154**, 2283-2290, doi:10.1099/mic.0.2007/015941-0
1217 (2008).
1218 67 Dias-Lopes, G. *et al.* Morphologic study of the effect of iron on pseudocyst formation in
1219 Trichomonas vaginalis and its interaction with human epithelial cells. *Mem Inst Oswaldo Cruz*
1220 **112**, 664-673, doi:10.1590/0074-02760170032 (2017).
1221 68 Mariante, R. M., Lopes, L. C. & Benchimol, M. Tritrichomonas foetus pseudocysts adhere to
1222 vaginal epithelial cells in a contact-dependent manner. *Parasitol Res* **92**, 303-312,
1223 doi:10.1007/s00436-003-1026-z (2004).
1224 69 Lehker, M. W. & Sweeney, D. Trichomonad invasion of the mucous layer requires adhesins,
1225 mucinases, and motility. *Sex Transm Infect* **75**, 231-238, doi:10.1136/sti.75.4.231 (1999).
1226 70 Alsmark, C. *et al.* Patterns of prokaryotic lateral gene transfers affecting parasitic microbial
1227 eukaryotes. *Genome Biol* **14**, R19, doi:10.1186/gb-2013-14-2-r19 (2013).
1228 71 Johansson, M. E. *et al.* Normalization of Host Intestinal Mucus Layers Requires Long-Term
1229 Microbial Colonization. *Cell Host Microbe* **18**, 582-592, doi:10.1016/j.chom.2015.10.007 (2015).
1230 72 Bradic, M. *et al.* Genetic Indicators of Drug Resistance in the Highly Repetitive Genome of
1231 Trichomonas vaginalis. *Genome Biol Evol* **9**, 1658-1672, doi:10.1093/gbe/evx110 (2017).
1232 73 Drmota, T. & Kral, J. Karyotype of Trichomonas vaginalis. *European Journal of Protistology* **33**,
1233 131-135 (1997).
1234 74 Bradic, M. & Carlton, J. M. Does the common sexually transmitted parasite Trichomonas
1235 vaginalis have sex? *PLoS Pathog* **14**, e1006831, doi:10.1371/journal.ppat.1006831 (2018).
1236 75 Ramesh, M. A., Malik, S. B. & Logsdon, J. M., Jr. A phylogenomic inventory of meiotic genes;
1237 evidence for sex in Giardia and an early eukaryotic origin of meiosis. *Current biology : CB* **15**,
1238 185-191, doi:10.1016/j.cub.2005.01.003 (2005).
1239 76 Mochizuki, K., Novatchkova, M. & Loidl, J. DNA double-strand breaks, but not crossovers, are
1240 required for the reorganization of meiotic nuclei in Tetrahymena. *J Cell Sci* **121**, 2148-2158,
1241 doi:10.1242/jcs.031799 (2008).
1242 77 El Aidy, S. *et al.* Temporal and spatial interplay of microbiota and intestinal mucosa drive
1243 establishment of immune homeostasis in conventionalized mice. *Mucosal Immunol* **5**, 567-579,
1244 doi:10.1038/mi.2012.32 (2012).
1245 78 Gastelum-Martinez, A. *et al.* Iron-modulated virulence factors of Entamoeba histolytica. *Future*
1246 *Microbiol* **13**, 1329-1341, doi:10.2217/fmb-2018-0066 (2018).

1247 79 Alderete, J. F., Nguyen, J., Mundodi, V. & Lehker, M. W. Heme-iron increases levels of AP65-
1248 mediated adherence by Trichomonas vaginalis. *Microb Pathog* **36**, 263-271,
1249 doi:10.1016/j.micpath.2003.12.007 (2004).
1250 80 Herp, S., Durai Raj, A. C., Salvado Silva, M., Woelfel, S. & Stecher, B. The human symbiont
1251 Mucispirillum schaedleri: causality in health and disease. *Med Microbiol Immunol* **210**, 173-179,
1252 doi:10.1007/s00430-021-00702-9 (2021).
1253 81 Chen, L. M. *et al.* Tryptophan-kynurenine metabolism: a link between the gut and brain for
1254 depression in inflammatory bowel disease. *J Neuroinflammation* **18**, 135, doi:10.1186/s12974-
1255 021-02175-2 (2021).
1256 82 Ortega, M. A. *et al.* Microbiota-gut-brain axis mechanisms in the complex network of bipolar
1257 disorders: potential clinical implications and translational opportunities. *Mol Psychiatry*,
1258 doi:10.1038/s41380-023-01964-w (2023).
1259 83 Cortez, V. *et al.* Astrovirus infects actively secreting goblet cells and alters the gut mucus barrier.
1260 *Nat Commun* **11**, 2097, doi:10.1038/s41467-020-15999-y (2020).
1261 84 Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and
1262 repeat separation. *Genome Res* **27**, 722-736, doi:10.1101/gr.215087.116 (2017).
1263 85 Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using
1264 repeat graphs. *Nat Biotechnol* **37**, 540-546, doi:10.1038/s41587-019-0072-8 (2019).
1265 86 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1266 *arXiv:1303.3997v1 [q-bio.GN]* (2013).
1267 87 Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and
1268 genome assembly improvement. *PLoS One* **9**, e112963, doi:10.1371/journal.pone.0112963
1269 (2014).
1270 88 Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging model
1271 organism genomes. *Genome Res* **18**, 188-196, doi:10.1101/gr.6743907 (2008).
1272 89 Smit, A. & Hubley, R. *RepeatModeler Open-1.0* (2008-2015).
1273 90 Smit, A., Hubley, R. & Green, P. *RepeatMasker Open-4.0.*, <http://www.repeatmasker.org>
1274 (2013-2015).
1275 91 Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences.
1276 *Methods Mol Biol* **1962**, 1-14, doi:10.1007/978-1-4939-9173-0_1 (2019).
1277 92 Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, doi:10.1186/1471-2105-5-59
1278 (2004).
1279 93 Stanke, M., Schoffmann, O., Morgenstern, B. & Waack, S. Gene prediction in eukaryotes with a
1280 generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* **7**,
1281 62, doi:10.1186/1471-2105-7-62 (2006).
1282 94 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO:
1283 assessing genome assembly and annotation completeness with single-copy orthologs.
1284 *Bioinformatics* **31**, 3210-3212, doi:10.1093/bioinformatics/btv351 (2015).
1285 95 Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**,
1286 1236-1240, doi:10.1093/bioinformatics/btu031 (2014).
1287 96 Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity
1288 searching. *Nucleic Acids Res* **39**, W29-37, doi:10.1093/nar/gkr367 (2011).
1289 97 Nursimulu, N., Moses, A. M. & Parkinson, J. Architect: A tool for aiding the reconstruction of
1290 high-quality metabolic models through improved enzyme annotation. *PLoS Comput Biol* **18**,
1291 e1010452, doi:10.1371/journal.pcbi.1010452 (2022).
1292 98 Aurrecoechea, C. *et al.* GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic
1293 protist pathogens Giardia lamblia and Trichomonas vaginalis. *Nucleic Acids Res* **37**, D526-530,
1294 doi:10.1093/nar/gkn631 (2009).
1295 99 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search
1296 tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

100 Amos, B. *et al.* VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. *Nucleic Acids Res*, doi:10.1093/nar/gkab929 (2021).
101 UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515, doi:10.1093/nar/gky1049 (2019).
102 Morgan, M., Falcon S & R, G. *GSEABase: Gene set enrichment data structures and methods. R package version 1.52.*, <https://rdrr.io/bioc/GSEABase/man/GSEABase-package.html> (2020).
103 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).
104 Tamura, K. & Nei, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**, 512-526, doi:10.1093/oxfordjournals.molbev.a040023 (1993).
105 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547-1549, doi:10.1093/molbev/msy096 (2018).
106 Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214, doi:10.1016/j.cell.2015.05.002 (2015).
107 Nemesh, J. & Wysoker, A. *Drop-seq tools v1.13*, <https://github.com/broadinstitute/Drop-seq/releases> (
108 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
109 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).
110 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).
111 McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *Journal of Open Source Software* **3**, 861, doi:https://doi.org/10.21105/joss.00861 (2018).
112 Okuda, S. *et al.* KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res* **36**, W423-426, doi:10.1093/nar/gkn282 (2008).
113 Alexa, A. & Rahnenfuhrer, J. *topGO: Enrichment Analysis for Gene Ontology. R package version 2.42.0.*, <https://github.com/lyijin/topGO_pipeline> (2020).
114 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal statistical society: series B* **57**, 289-300, doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x (1995).
115 Kolde, R. *pheatmap: Pretty Heatmaps. R package version 1.0.12.*, <https://CRAN.R-project.org/package=pheatmap> (2019).
116 Comeau, A. M., Li, W. K., Tremblay, J. E., Carmack, E. C. & Lovejoy, C. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS One* **6**, e27492, doi:10.1371/journal.pone.0027492 (2011).
117 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581-583, doi:10.1038/nmeth.3869 (2016).
118 Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852-857, doi:10.1038/s41587-019-0209-9 (2019).
119 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).
120 McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217, doi:10.1371/journal.pone.0061217 (2013).
121 Oksanen, J. B., F. Guillaume; Friendly, Michael; Kindt, Roeland; Legendre, Pierre; McGlinn, Dan; Minchin, Peter R. ; O'Hara, R. B.; Simpson, Gavin L.; Solymos, Peter; Henry M.; Stevens,

1348 H.; Szoecs, Eduard; Wagner, Helene. *vegan: Community Ecology Package*, <https://CRAN.R-
1349 project.org/package=vegan> (2017).
1350 122 Bates, D., Maechler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using
1351 lme4. *Journal of Statistical Software* **67**, 1-48, doi:10.18637/jss.v067.i01 (2015).
1352 123 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
1353 RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
1354 124 Pantano, L. *DEGreport: Report of DEG analysis. R package version 1.26.0*,
1355 <http://lpantano.github.io/DEGreport/> (2020).
1356 125 Taj, B. *et al.* MetaPro: A scalable and reproducible data processing and analysis pipeline for
1357 metatranscriptomic investigation of microbial communities. *BioRxiv*,
1358 doi:https://doi.org/10.1101/2021.02.23.432558 (2021).
1359 126 Garber, A. I. *et al.* FeGenie: A Comprehensive Tool for the Identification of Iron Genes and Iron
1360 Gene Neighborhoods in Genome and Metagenome Assemblies. *Front Microbiol* **11**, 37,
1361 doi:10.3389/fmicb.2020.00037 (2020).
1362 127 Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat*
1363 *Methods* **12**, 59-60, doi:10.1038/nmeth.3176 (2015).

1364

1365

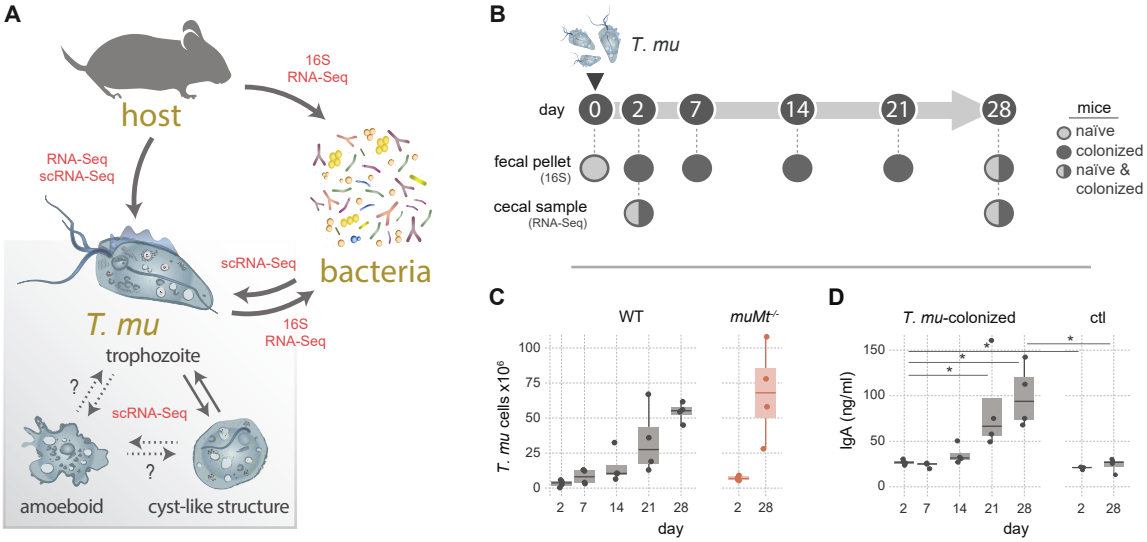

A
host
16S
RNA-Seq
RNA-Seq
scRNA-Seq
bacteria
scRNA-Seq
T. mu
16S
RNA-Seq
trophozoite
?
scRNA-Seq
?
amoeboid
cyst-like structure
B
T. mu
day
0
2
7
14
21
28
fecal pellet
(16S)
cecal sample
(RNA-Seq)
mice
naïve
colonized
naïve &
colonized
C
WT
muMt-/-
T. mu cells x10^6
100
75
50
25
0
2
7
14
21
28
2
28
day
D
T. mu-colonized
ctl
IgA (ng/ml)
150
100
50
0
2
7
14
21
28
2
28
day

Figure 1

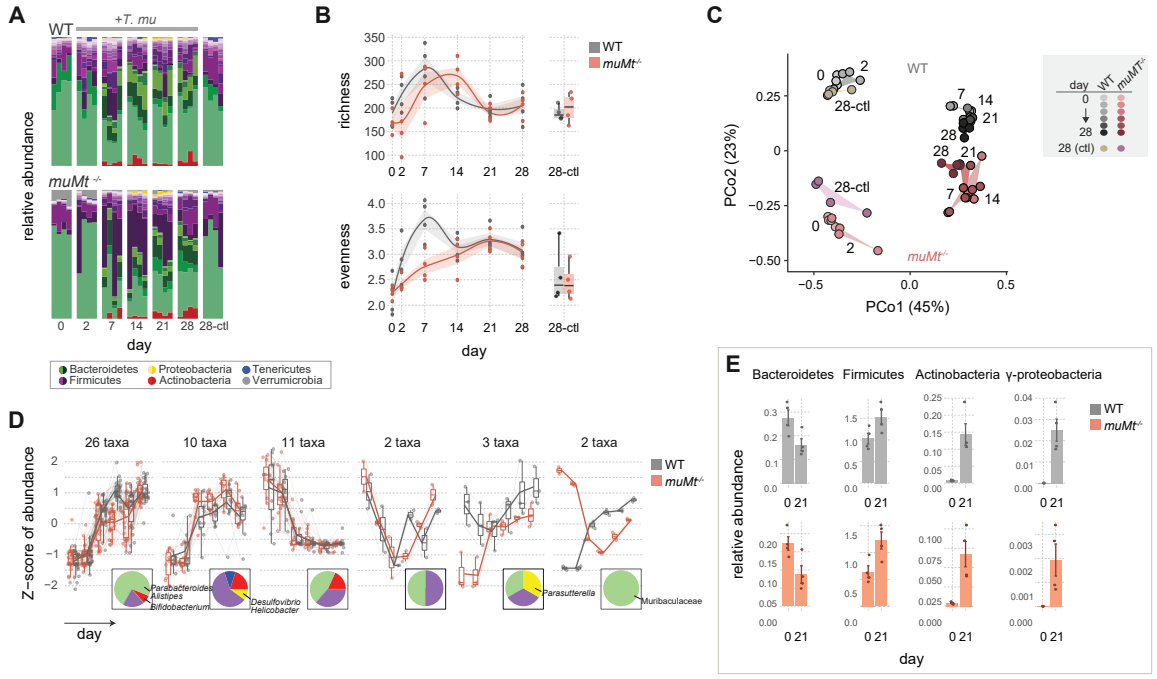

A
WT
+T. mu
muMt -/-
relative abundance
0 2 7 14 21 28 28-ctl
day
Bacteroidetes
Firmicutes
Proteobacteria
Actinobacteria
Tenericutes
Verrumicrobia
B
richness
evenness
WT
muMt-/-
day
C
PCo2 (23%)
PCo1 (45%)
WT
muMt-/-
28-ctl
day
28 (ctl)
D
26 taxa
10 taxa
11 taxa
2 taxa
3 taxa
2 taxa
Z−score of abundance
Parabacteroides
Alistipes
Bifidobacterium
Desulfovibrio
Helicobacter
Parasutterella
Muribaculaceae
day
E
Bacteroidetes
Firmicutes
Actinobacteria
γ-proteobacteria
relative abundance
day

Figure 2

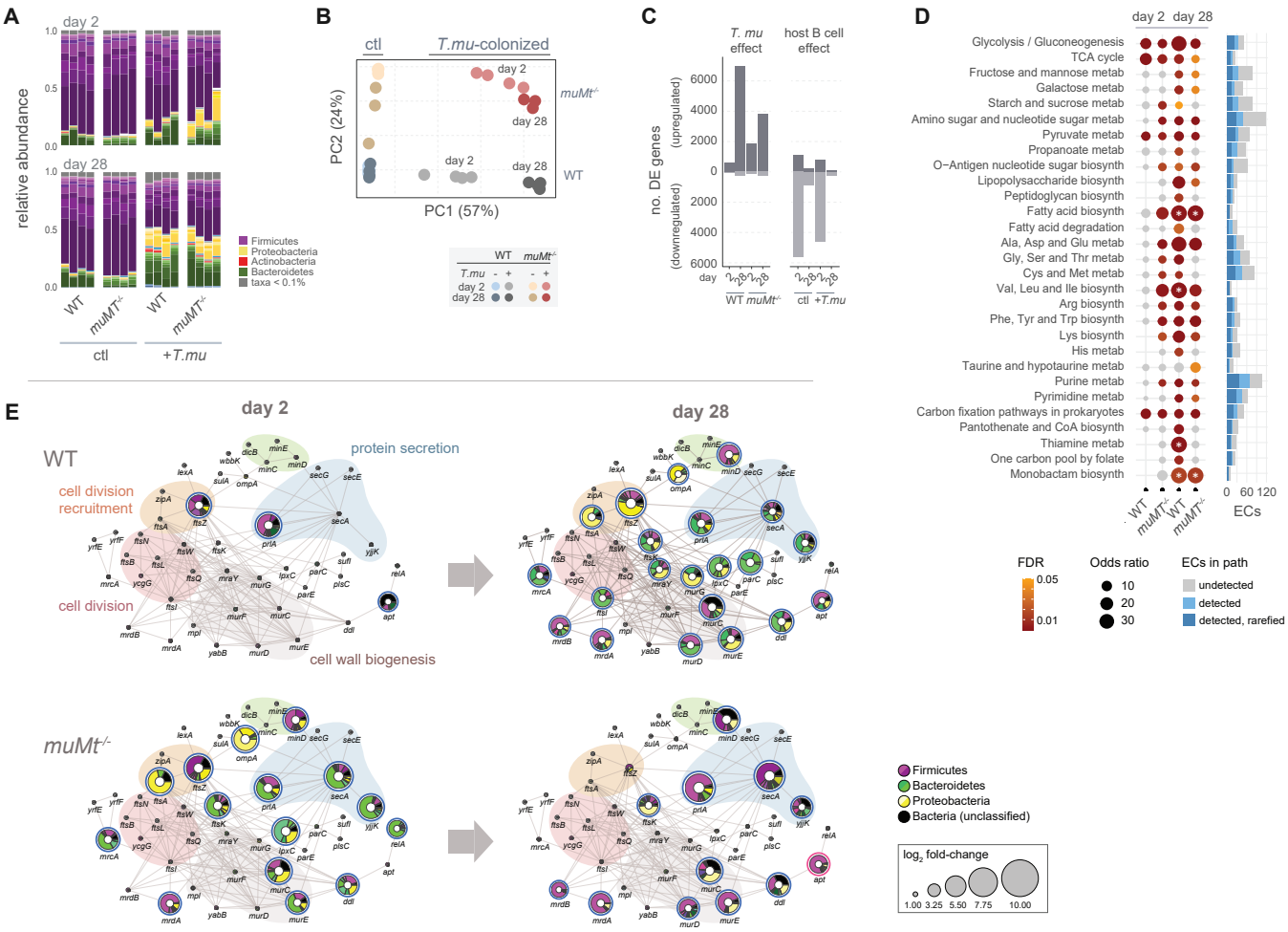

A
day 2
day 28
relative abundance
1.0
0.5
0.0
WT
muMt-/-
ctl
+T.mu
Firmicutes
Proteobacteria
Actinobacteria
Bacteroidetes
taxa < 0.1%
B
ctl
T.mu-colonized
PC2 (24%)
PC1 (57%)
day 2
day 28
muMt-/-
WT
T.mu
C
T. mu effect
host B cell effect
no. DE genes
(upregulated)
(downregulated)
6000
4000
2000
0
day
D
day 2
day 28
Glycolysis / Gluconeogenesis
TCA cycle
Fructose and mannose metab
Galactose metab
Starch and sucrose metab
Amino sugar and nucleotide sugar metab
Pyruvate metab
Propanoate metab
O−Antigen nucleotide sugar biosynth
Lipopolysaccharide biosynth
Peptidoglycan biosynth
Fatty acid biosynth
Fatty acid degradation
Ala, Asp and Glu metab
Gly, Ser and Thr metab
Cys and Met metab
Val, Leu and Ile biosynth
Arg biosynth
Phe, Tyr and Trp biosynth
Lys biosynth
His metab
Taurine and hypotaurine metab
Purine metab
Pyrimidine metab
Carbon fixation pathways in prokaryotes
Pantothenate and CoA biosynth
Thiamine metab
One carbon pool by folate
Monobactam biosynth
0 60 120
ECs
FDR
0.05
0.01
Odds ratio
10
20
30
ECs in path
undetected
detected
detected, rarefied
E
day 2
day 28
WT
muMt-/-
protein secretion
cell division recruitment
cell division
cell wall biogenesis
Firmicutes
Bacteroidetes
Proteobacteria
Bacteria (unclassified)
log2 fold-change
1.00 3.25 5.50 7.75 10.00

Figure 3

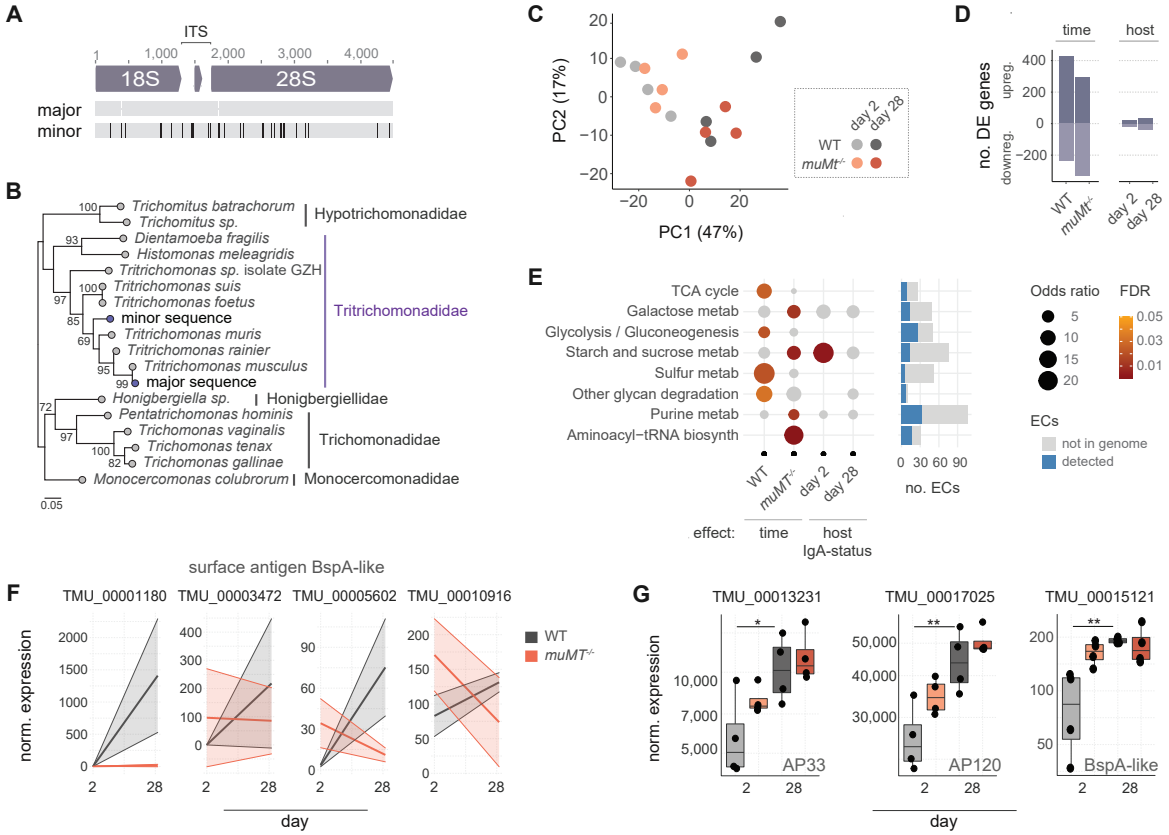

A
ITS
1
1,000
2,000
3,000
4,000
18S
28S
major
minor
B
Trichomitus batrachorum
Trichomitus sp.
Hypotrichomonadidae
Dientamoeba fragilis
Histomonas meleagridis
Tritrichomonas sp. isolate GZH
Tritrichomonas suis
Tritrichomonas foetus
minor sequence
Tritrichomonas muris
Tritrichomonas rainier
Tritrichomonas musculus
major sequence
Tritrichomonadidae
Honigbergiella sp.
Honigbergiellidae
Pentatrichomonas hominis
Trichomonas vaginalis
Trichomonas tenax
Trichomonas gallinae
Trichomonadidae
Monocercomonas colubrorum
Monocercomonadidae
0.05
C
PC2 (17%)
PC1 (47%)
day 2
day 28
WT
muMt-/-
D
time
host
no. DE genes
upreg.
downreg.
WT
muMt-/-
day 2
day 28
E
TCA cycle
Galactose metab
Glycolysis / Gluconeogenesis
Starch and sucrose metab
Sulfur metab
Other glycan degradation
Purine metab
Aminoacyl−tRNA biosynth
WT
muMT-/-
day 2
day 28
no. ECs
effect:
time
host
IgA-status
Odds ratio
FDR
ECs
not in genome
detected
F
surface antigen BspA-like
TMU_00001180
TMU_00003472
TMU_00005602
TMU_00010916
norm. expression
WT
muMT-/-
day
G
TMU_00013231
TMU_00017025
TMU_00015121
norm. expression
AP33
AP120
BspA-like
day

Figure 4

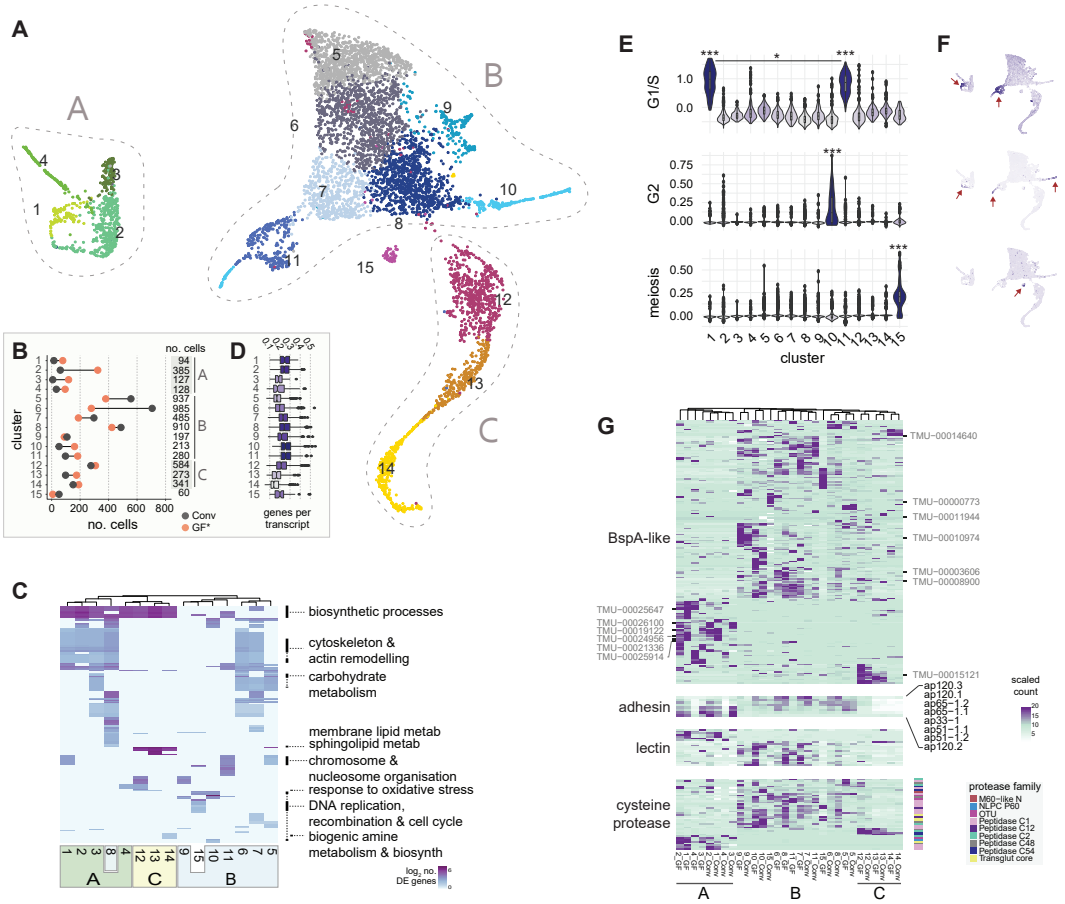
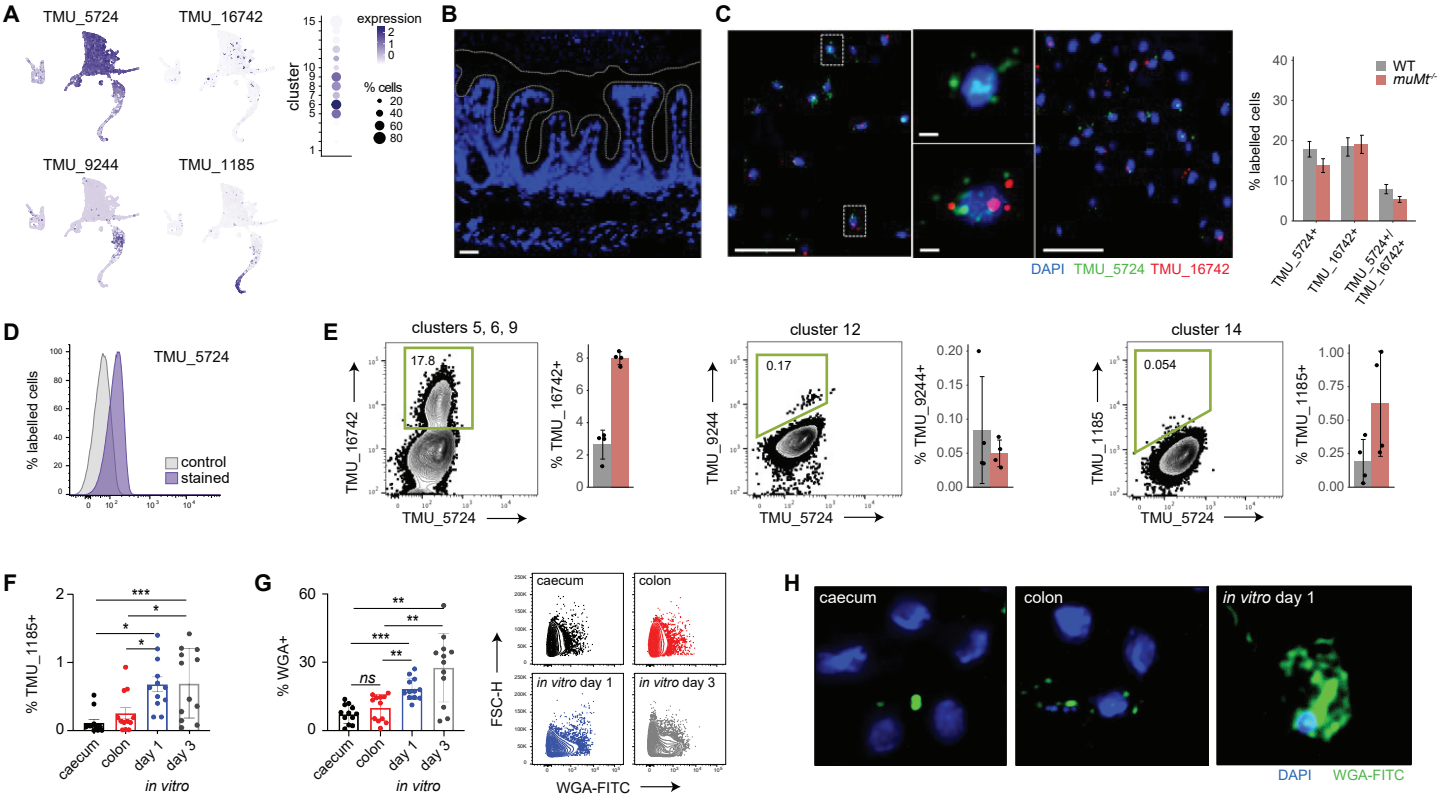

A
B
C
D
E
F
G
cluster
no. cells
Conv
GF*
genes per transcript
G1/S
G2
meiosis
cluster
biosynthetic processes
cytoskeleton & actin remodelling
carbohydrate metabolism
membrane lipid metab
sphingolipid metab
chromosome & nucleosome organisation
response to oxidative stress
DNA replication, recombination & cell cycle
biogenic amine metabolism & biosynth
log2 no. DE genes
BspA-like
adhesin
lectin
cysteine protease
scaled count
protease family
M60-like N
NLPC P60
OTU
Peptidase C1
Peptidase C12
Peptidase C2
Peptidase C48
Peptidase C54
Transglut core

Figure 5


A
TMU_5724
TMU_16742
TMU_9244
TMU_1185
expression
cluster
% cells
B
C
DAPI TMU_5724 TMU_16742
WT
muMt-/-
% labelled cells
TMU_5724+
TMU_16742+
TMU_5724+/ TMU_16742+
D
TMU_5724
% labelled cells
control
stained
E
clusters 5, 6, 9
17.8
TMU_16742
TMU_5724
% TMU_16742+
cluster 12
0.17
TMU_9244
% TMU_9244+
cluster 14
0.054
TMU_1185
% TMU_1185+

F
% TMU_1185+
caecum
colon
day 1
day 3
in vitro
G
% WGA+
ns
caecum
colon
day 1
day 3
in vitro
FSC-H
WGA-FITC
in vitro day 1
in vitro day 3
H
caecum
colon
in vitro day 1
DAPI WGA-FITC

Figure 6

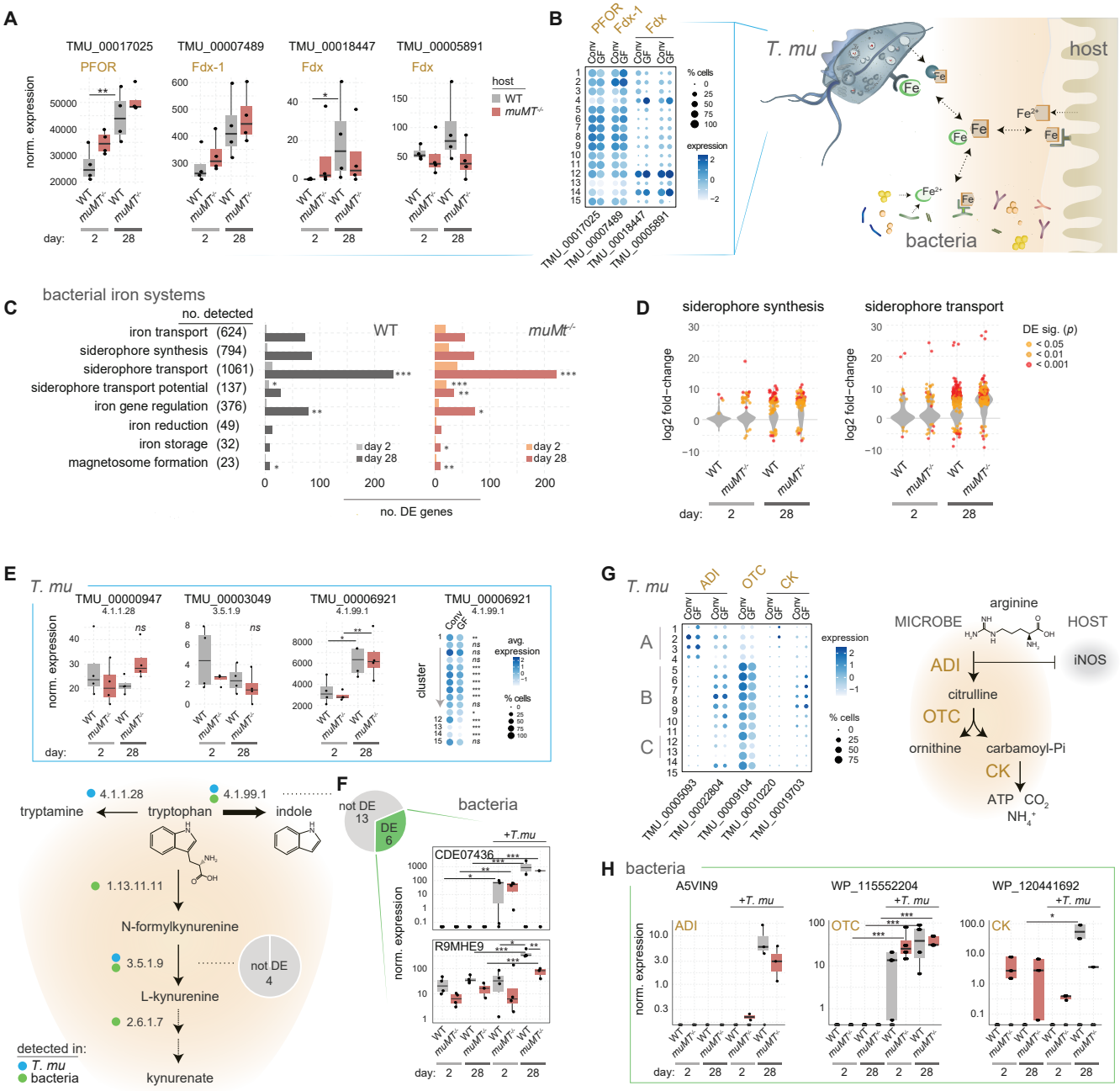

A
TMU_00017025
PFOR
TMU_00007489
Fdx-1
TMU_00018447
Fdx
TMU_00005891
Fdx
norm. expression
host
WT
muMT-/-
day:
B
PFOR
Fdx-1
Fdx
% cells
expression
T. mu
host
bacteria
C
bacterial iron systems
no. detected
iron transport (624)
siderophore synthesis (794)
siderophore transport (1061)
siderophore transport potential (137)
iron gene regulation (376)
iron reduction (49)
iron storage (32)
magnetosome formation (23)
WT
muMt-/-
day 2
day 28
no. DE genes
D
siderophore synthesis
siderophore transport
log2 fold-change
DE sig. (p)
< 0.05
< 0.01
< 0.001
day:
E
T. mu
TMU_00000947
4.1.1.28
TMU_00003049
3.5.1.9
TMU_00006921
4.1.99.1
TMU_00006921
4.1.99.1
norm. expression
cluster
avg. expression
% cells
tryptamine
tryptophan
indole
N-formylkynurenine
L-kynurenine
kynurenate
not DE 4
detected in:
T. mu
bacteria
F
not DE 13
DE 6
bacteria
+T.mu
CDE07436
R9MHE9
norm. expression
day:
G
T. mu
ADI
OTC
CK
expression
% cells
TMU_00005093
TMU_00022804
TMU_00009104
TMU_00010220
TMU_00019703
MICROBE
arginine
HOST
ADI
iNOS
citrulline
OTC
ornithine
carbamoyl-Pi
CK
ATP
CO2
NH4+
H
bacteria
A5VIN9
WP_115552204
WP_120441692
+T. mu
ADI
OTC
CK
norm. expression
day:

Figure 7