

**Title:** *Giardia* Secretome Highlights Secreted Tenascins as a Key Component of Pathogenesis

**Running Title:** *Giardia* Secretome

**Authors:** Audrey Dubourg<sup>a</sup>, Dong Xia<sup>b,e</sup>, John P. Winpenny<sup>a</sup>, Suha Al Naimi<sup>a,c</sup>, Maha Bouzid<sup>a</sup>, Darren W. Sexton<sup>a,d</sup>, Jonathan M. Wastling<sup>b,f</sup>, Paul R. Hunter<sup>a</sup> and Kevin M. Tyler<sup>\*a</sup>

**Authors' affiliations:**

- a. NIHR Health Protection Research Unit in Gastrointestinal Infections, Norwich Medical School, University of East Anglia, Norwich, UK.
- b. Department of Infection Biology, Institute of Infection and Global Health, Faculty of Health & Life Sciences, University of Liverpool, UK.
- c. Department of Science and Technology, Faculty of Health and Science, University of Suffolk, Ipswich, UK
- d. School of Pharmacy and Biomolecular Sciences Liverpool John Moores University, Liverpool, UK.
- e. Research Support Office, Royal Veterinary College, University of London, UK.
- f. Faculty of Natural Sciences, Keele University, Staffordshire, UK.

Authors email:

AD: adubourg@g.ucla.edu

DX: dxia@rvc.ac.uk

JPW: john.winpenny@uea.ac.uk

SAN: s.al-naimi@ucs.ac.uk

MB: M.Bouzid@uea.ac.uk

DWS: D.W.Sexton@ljmu.ac.uk

JMW: j.wastling@keele.ac.uk

PRH: Paul.Hunter@uea.ac.uk

KT: k.tyler@uea.ac.uk

\*Correspondence: k.tyler@uea.ac.uk,

Tel +44 (0)1603-591225,

Fax: +44 (0) 1603 591750.

## Abstract

**Background:** *Giardia* is a protozoan parasite of public health relevance that causes gastroenteritis in a wide range of hosts. Two genetically distinct lineages (assemblages A and B) are responsible for the human disease. Although it is clear that differences in virulence occur, pathogenesis and virulence of *Giardia* remains poorly understood.

**Findings:** The genome of *Giardia* is believed to contain ORFs that could encode as many as 6,000 proteins. By successfully applying quantitative proteomic analyses to the whole parasite and to the supernatants derived from parasite culture of assemblages A and B, we confirm expression of ~1,600 proteins from each assemblage, the vast majority of which being common to both lineages. To look for signature enrichment of secreted proteins, we considered the ratio of proteins in the supernatant compared with the pellet, which defined a small group of enriched proteins, putatively secreted at a steady state by cultured growing trophozoites of both assemblages. This secretome is enriched with proteins annotated to have N-terminal signal peptide. The most abundant secreted proteins include known virulence factors such as cathepsin B cysteine proteases and members of a *Giardia* superfamily of cysteine rich

proteins that comprises VSPs, HCMPs and a new class of virulence factors, the *Giardia* tenascins. We demonstrate that physiological function of human enteric epithelial cells is disrupted by such soluble factors even in the absence of the trophozoites.

**Conclusions:** We are able to propose a straightforward model of *Giardia* pathogenesis incorporating key roles for the major *Giardia* derived soluble mediators.

**Keywords:** *Giardia*, Secretion, Proteomics, Quantitative Proteomics, Tenascin, Cysteine protease, Enteric Pathogen.

## Background

With some 280 million symptomatic cases, giardiasis causes more bouts of human illness than any other parasitic disease [1]. The mechanism and mediators of pathogenesis by *Giardia*, however, remain largely unknown. Thanks to human volunteer studies, the association of *Giardia* infection itself and the significance of the virulence of the infecting *Giardia* strain, is experimentally unambiguous [2]. The molecular definition associated with strain virulence is though largely unexplored. It is clear that the majority of *Giardia* infections are asymptomatic. It is also clear, that infection is primarily localized to the duodenum and that some localized damage, close to the sites of colonization, causes villus atrophy and apoptosis of surrounding cells. However, this localized damage cannot be the sole cause of the profound diarrhoea which is often characteristic of the disease and which appears to affect absorption over a much wider area of the digestive tract than the site of infection alone.

One of the secreted mediators of damage to the duodenum is believed to be cathepsin B protease [3]. Cathepsin B-like proteases compose one of the superfamilies belonging to the CA clan of cysteine peptidases [4]. Compared to other cathepsins, cathepsin B proteases possess an additional 20 amino acid insertion named the occluding loop that enables their function as an endo- or exopeptidase [5]. Although twenty-seven genes encoding cathepsin proteases have been identified in *Giardia*, for the majority of these proteases, function remains elusive [6]. While some parasites may secrete cathepsin B proteases to either evade or modulate their hosts immune responses [7], a recent study has demonstrated that *Giardia* trophozoites secrete cathepsin B-like proteases, degrading intestinal IL-8 and thereby reducing the inflammation reaction by the host [3]. Secreted *Giardia* cathepsin B protease (GCATB) may also contribute to degradation of intestinal mucin and facilitate trophozoite attachment to intestinal epithelia [8, 9].

Most of the proteomic studies so far reported for *Giardia* were undertaken in trophozoites undergoing encystation [10-12]. Only a few studies have focused on proteins secreted by *Giardia* and their role in the host-pathogen interaction [3, 13-15]. These studies were focused on parasite interaction with intestinal cell lines. No studies have yet attempted to quantify proteins that are the product of steady state secretion by healthy, growing *Giardia* trophozoites and

which we hypothesize as the primary mediators of giardiasis pathology. In this study, we have identified, to the limit of existing technology, the proteins expressed by populations of healthy, growing human infective *Giardia* trophozoites. We have provided quantitation of the relative abundance of retained and released trophozoite proteins from two human infective assemblages, affording calculation of the specific enrichment of released proteins and thereby the description of which proteins are most likely to be secreted by trophozoites of each assemblage. Thereafter, we compared the profile of enrichment between the two assemblages in order to identify conserved as well as assemblage-specific secreted proteins. We provide electrophysiological analysis which confirms that trophozoite secreted molecules adversely affect the homeostasis of enteric epithelia and our analysis of the heterogeneity of encoding genes between lineages demonstrates the direct selective pressure on these virulence factors and affords their use in discriminating clinically important strains and outbreaks. Finally, the discovery of tenascins as a highly represented and variable group of proteins secreted by trophozoites strongly implicates this new class of virulence factors in a novel model for the mechanism of *Giardia* pathogenesis. We propose that tenascin action follows degradation of the protective mucous afforded by the action of a secreted nuclease and GCATB, and damage to cellular junctions by GCATB. Tenascins acting by means of EGF receptor ligation, to prevent repair to those damaged junctions.

## Data description

Soluble and cytosolic fractions from *in vitro* grown assemblage A and B trophozoites, the aetiologic agents of human giardiasis, were extracted in order to establish which proteins are secreted in the steady state by healthy, growing trophozoite populations. We reasoned that secreted proteins would be overrepresented in the medium in which parasites were incubated compared with the trophozoites that produced them. This ostensibly straightforward assessment being reliant on the sensitive, specific and quantitative detection of the proteins expressed by *Giardia* trophozoites in whole cells and in the medium in which the trophozoites were incubated.

The WB (assemblage A – ATCC\_50803) and GS (assemblage B – ATCC\_50581) reference strains were utilized to facilitate ease of comparison between genetically divergent human infective isolates with the available reference genomes. For each experiment trophozoites were harvested from mid log growth and incubated in non-supplemented Dulbecco's Modified Eagle medium (DMEM) for 45 minutes at 37°C before supernatants and pellets were collected for proteomic and other analyses including validation of their viability by flow cytometry (Additional file 1: Fig S1). Proteomic analyses were based on samples from 3 distinct biological replicates. Each sample was analysed using two quantitative proteomic platforms the Orbitrap MS and the Q-Exactive MS. Thus, in total the results from 24 ( $2 \times 2 \times 2 \times 3$ ) proteomic analyses are reported.

The identification of abundant, secreted, *Giardia* virulence factors led us to consider whether the secretions from *Giardia* alone could effect changes in the behaviour of enteric epithelia - even in the absence of the trophozoites themselves. In order to determine the effect of *Giardia* trophozoite secreted factors

on the intestinal epithelia, chopstick type electrodes connected to a voltmeter were used to measure the trans-epithelial electrical resistance (TEER) of polarised CaCo-2 epithelial cells grown on permeable supports. CaCo-2 cells were cultured over 6 days until confluent. TEER across the developing CaCo-2 monolayer was measured on a daily basis as shown in Figure 2A. Once confluence was established, *Giardia* trophozoites were added to the apical side of the confluent epithelium and after 24 hours incubation the trophozoites were washed from the apical surface. In order to determine whether co-cultures of *Giardia* trophozoites or diluted *Giardia* supernatants affected the ion channels responsible for secretory movement across the epithelium, an Ussing chamber system was utilised with different chloride secretion inhibitor and activators.

Further details about sample collection, secretome analysis and electrophysiology can be found in the methods section and protocols provided.

## Analyses

### Protein expression in *Giardia* trophozoites

To describe definitive *Giardia* secretomes under a standard set of conditions with high confidence and based on a robust data set and to reduce the potential for technical artefact the two MS techniques: Q-Exactive and Orbitrap MS were used with similar settings on the same three independent replicates to increase coverage. Only proteins identified by both techniques within the three replicates datasets were included in the analysis to increase the robustness of the data. The protein quantification was performed using a label-free method: iBAQ (intensity based absolute quantification) which calculates the sum of parent ion intensities of identified peptides per proteins [16]. The average normalised abundance was divided by the iBAQ values giving the "Abundance-iBAQ". The quantitative datasets from both MS techniques and for each independent replicate were shown to be strongly correlated by a Spearman correlation test (data not shown) and therefore exploitable for proteomic analysis.

The Q-Exactive MS identified almost all of the proteins identified by use of the Orbitrap MS, and in total the two techniques identified 1,587 GS proteins and 1,690 WB proteins (Additional File 1: Fig S2). This represents over a quarter of the open reading frames (ORFs) predicted by the respective genomes in this single life-cycle stage under this steady state set of *in vitro* culture conditions and compares favourably with other recent proteomic analyses of *Giardia* [17, 18]. Lists of proteins detected in only one of the two assemblages are provided (Additional file 2: Table S1 and S2). Protein from two of the eight predicted assemblage-specific genes previously identified by comparative genomics was detected [19].

Overall, both assemblages gave comparable and consistent results using both platforms with the sensitivity of detection being greater for Q-Exactive MS; which provided a range of detection spanning 5 logs. In total, Q-Exactive MS identified 1,542 GS proteins and 1,641 WB proteins (Fig S3). Of these, 946 GS

proteins were present in both pellet and supernatant, 27 in the supernatant only and 569 GS proteins in pellet only. By comparison, 490 WB proteins were identified in supernatant and pellet and 24 in the supernatant only with 1,127 WB proteins in pellet only.

### ***Giardia secretome***

To evaluate supernatant enrichment, proteins identified in the supernatant (SP) datasets were gathered and compared to their concentration in the pellet (P) to provide a ratio using the following formula:  $\frac{SP\ abundance-iBAQ}{P\ abundance-iBAQ}$ . These proteins were then ranked from highest to lowest by ratiometric value and an arbitrary cut-off invoked such that the top 50 were considered as the most likely to be secreted. Proteins identified only in SP were also included in the analysis as most likely to be secreted. All the proteins selected as “of interest” were ranked according to their SP expression from most to least abundant to obtain a quantitative enrichment profile for each isolate and this was performed for each platform. Orbitrap and Q-Exactive enrichment profiles were compared and proteins were considered as most likely to be enriched in the supernatant when identified as such by Q-Exactive MS and confirmed by Orbitrap MS. The different enrichment profiles were then also compared between assemblages.

The results yielded a set of 15 orthologous proteins that were identified in both isolates by both techniques (Table 1). Eleven of these were predicted to possess an N-terminal signal sequence. Just two of these were of unknown function and two groups dominated the annotated genes encoding the rest of these proteins, five were annotated as tenascins and three as cathepsin B cysteine proteases. The most abundant enriched protein was found to be pyridoxamine 5'-phosphate oxidase (PNPO), an FMN dependent enzyme capable of fixing molecular oxygen that lacks a signal peptide and which was also recently identified as a secreted *Giardia* trophozoite protein upregulated during interaction with epithelial cells [15]. An extracellular nuclease was also present, along with a high cysteine membrane protein; as well as a protein product of a gene misannotated as a VSP (since it was well conserved between assemblages).

We considered that where proteins were shown to be enriched in the supernatant by both platforms and in both assemblages and possessed an N-terminal signal sequence that they were truly secreted proteins. Secreted proteins involved in adapting *Giardia* to the host environment of the human gut might be expected to be engaged in Red Queen evolution and have dN/dS indicative of positive selection. While amino-acid divergence between orthologs of secreted proteins varied considerably from 67% for the HCMP to 83% for (e.g. for the extracellular nuclease), only three proteins showed evidence of positive selections, two tenascins and one of the cathepsins. One cathepsin and one tenascin in particular showed evidence of evolution under a very high

degree of selective pressure (Table 1). Interestingly, some cathepsins and some tenascins with similar levels of amino-acid identity between the assemblages to those under high selective pressure showed little or no evidence of positive selection.

We considered whether lineage specific soluble mediators might also be present and identified by this method. Comparing those proteins identified by both methods as having the highest relative expression in the supernatant (Tables S3 and S4). The five most abundant conserved secreted proteins from Table 1 were also present in the top 10 secreted proteins from each assemblage amongst other VSPs, tenascins, and cathepsin B, and this regardless of the MS technique or the isolate. Unsurprisingly, VSPs were the primary proteins enriched in supernatants that were lineage-specific. Amongst the multigene families, however, there were also differences in the cathepsin B and tenascins/HCMP repertoires. No other proteins with N-terminal peptides were encoded in either assemblage except for one CxC-rich protein. Interestingly, none of the eight proteins encoded by assemblage-specific genes and identified by comparative genomics were found to be enriched in the supernatants.

When comparing secretion profiles between the two assemblages, seven proteins were over-represented in the supernatants by only one assemblage or only identified by Q-Exactive MS or present at very low abundance in one of the two (Table 2). Only two proteins, sentrin and A-type flavoprotein lateral transfer candidate, were present in the top 50 proteins of assemblage B (GS strain) trophozoites secretome. Whereas, the other five, one elongation factor 1- $\alpha$  (EF-1 $\alpha$ ), one ATP-binding cassette protein 5, one CxC rich protein, one translation initiation inhibitor and a peptide methionine sulfoxide reductase MsrB, were present in the top 55 proteins of assemblage A (WB strain) trophozoites secretomes. Interestingly, A-type flavoprotein lateral transfer candidate was also present in the top50 supernatant proteins by assemblage A trophozoites; however, its low supernatant enrichment ratio (< 0.2) suggests that this protein is unlikely to be secreted by assemblage A trophozoites.

### ***Giardia* soluble mediators disrupt intestinal cell functions**

Soluble and diffusible agents, able to disrupt gut function, could potentially mediate more diffuse and profound pathology for giardiasis than close range interactions between the trophozoites and the gastrointestinal epithelium alone. To determine whether *Giardia* secreted virulence factors could induce changes in the behaviour of intestinal epithelium, short-circuit current (Isc) was continuously measured across polarised CaCo-2 epithelial cells that had either been cultured without any additions, co-cultured with *Giardia* trophozoites or co-cultured with diluted (1:1000) *Giardia* supernatants (Figure 2B). Further experiments demonstrated that either after 24 hour co-culture with *Giardia* (Fig 2C) or 24 hour co-culture with diluted *Giardia* supernatants (Fig 2D) both experimental conditions dramatically inhibit both the cAMP-stimulated Isc (basolateral application of 10  $\mu$ M Forskolin) and the calcium-activated Isc (basolateral application of 100  $\mu$ M UTP). In order to identify what ion channels were being affected, the CFTR chloride ion channel inhibitor, GlyH101 (50  $\mu$ M), and the calcium-activated chloride ion channel inhibitor, DIDS (100  $\mu$ M), were added to the apical side of the Ussing chamber. The cAMP-

stimulated Isc is predominantly due to activation of CFTR chloride channels as it is inhibited by GlyH101 (Figure 2B-D). The calcium-activated Isc is predominantly due to activation of calcium-activated chloride channels as it is inhibited by DIDS (Figure 2B-D).

## Discussion

In this study, we have identified proteins secreted by trophozoites of both human infecting assemblages. Contaminating host serum proteins (mainly bovine albumin) in the supernatant samples were a concern, as previously described by others [20]. Such serum proteins bind to the parasite's surface and are continuously released which interfere with the characterisation of *Giardia* secretome. To overcome this issue, parasites were cleansed from the serum proteins and incubated in serum-free DMEM before collecting supernatants and pellets. To increase coverage and robustness of the analysis, two mass spectrometers (Orbitrap and Q-Exactive MS) were used on the same replicates and proteins identified by both MS were included in the analysis.

Previous studies have focused on protein secretion during *Giardia* trophozoite encystation; or protein secretion upon interaction with (or attachment to) host cells. Here instead, we chose to provide a detailed baseline from cultured *Giardia* trophozoites secreting proteins under a steady state *in vitro*. Nevertheless, our results are strongly supportive of a recent proteomic study looking at the effect of host attachment on the profile of *Giardia* secreted proteins [15]. Prior to that study, several metabolic enzymes had been proposed to be released by *Giardia* trophozoites upon interaction with intestinal epithelial cells (IEC) [13]: such as arginine deiminase (ADI), enolase, and ornithine carbamoyltransferase (OCT) which we were also able to identify from the culture supernatants of both assemblages.

Our study does confirm the previously observed enrichment of EF-1 $\alpha$ , in assemblage A culture supernatants [20] (Table 2 and Table S4). EF-1 $\alpha$  is a key enzyme in the protein synthesis process in eukaryotic cells [21] but many organisms have been shown to express EF-1 $\alpha$  in excess which suggests that this protein may have some other functions [21]. In the context of pathogenicity and virulence, the secreted *Leishmania* EF-1 $\alpha$  has been shown to down-regulate host inflammatory cell signalling [22]. In *Giardia*, EF-1 $\alpha$  has been shown to be an immunoreactive protein recognised by antibodies from patients who have previously had giardiasis [20]. Yet, its role as putatively secreted virulence factor in *Giardia* pathogenesis remains elusive. That this protein is only released by assemblage A trophozoites raises the possibility of associating its function with observable differences in pathogenesis or host range between the two human infective assemblages.

Our study shows some other differences in secretions between assemblage A and B trophozoites (Table 2). A-type flavoprotein lateral transfer candidate and sentrin were present in assemblage B (GS strain) trophozoites secretome; and ATP-binding cassette (ABC) protein 5, CxC rich protein, translation initiation inhibitor and peptide methionine sulfoxide reductase (MsrB) were present in assemblage A (WB strain) secretome.

A-type flavoprotein lateral transfer candidate has a high oxygen reductase activity during *Giardia* infection suggesting an O<sub>2</sub> scavenging function upon release in the host intestinal environment [23]; thus, potentially affording increased resilience to *Giardia* trophozoites in the small intestine and manipulating the parasites immediate microenvironment. Whether assemblage B trophozoites require A-type flavoprotein lateral transfer candidate throughout the infection or just in its early stage remains unclear. Sentrin is involved in the ubiquitination of proteins to render them resistant to degradation [24]. Sentrin is evolutionarily conserved and has been identified in prokaryotic and eukaryotic organisms such as *S. cerevisiae*, *A. thaliana* and *Homo sapiens*, which suggests a conserved specialised function in cell metabolism [24]. With its ubiquitination function, sentrin was expected to be only present in *Giardia* proteome but not in its secretome. Why this protein would be secreted or released by *Giardia* trophozoites remains unclear and raises the question of the advantages, for the parasite, of releasing sentrin into the host environment upon infection.

ABC proteins are a large and diverse canonical group of membrane proteins typically resident in the plasma membrane and associated, in eukaryotes, with the ATP dependent egress of metabolites and toxins; they can be determinants of virulence and drug resistance [25]. Here one *Giardia* ABC protein shows enrichment in the supernatant of WB but not of GS and it will be interesting to see if a functional correlation can be found. The CxC rich protein belongs to the HCMP superfamily that also includes VSPs, tenascins and HCMPs. The presence of orthologs in both strains is consistent with it not being a VSP protein. As with several other HCMPs, this CxC rich protein had a very high signal and only one TM domain suggesting that it may be a labile surface protein in WB, but its specific role and why it is much more abundant in the WB supernatant than the GS supernatant is not clear. Translation initiation inhibitors are proteins inhibiting the initiation of the translation of messenger RNA (mRNA) into proteins and are mainly located in the cell cytosol [26]. Yet, one translation initiation inhibitor is over-represented in the assemblage A trophozoite secretome (Top20 secreted proteins) probably due to its high solubility and stability. Peptide methionine sulfoxide reductase (MsrB) catalyses the reduction of free- and protein-bound methionine sulfoxides to corresponding methionines, which constitutes a mechanism for the scavenging of reactive oxygen species (ROS) responsible for a fundamental innate defence against pathogens in various host organisms [27]. MsrB is an antioxidant protein protecting organisms from the cytotoxic effects of ROS and therefore from cell death. This protein is crucial for the virulence of *S. typhimurium* and the immune evasion of *Schistosoma mansoni* [28, 29]. Whether msrB has a similar role in *Giardia* assemblage A pathogenicity remains unclear.

The difference in secretion between the two human infective assemblages observed in this study may also go some way to explaining the differences in pathogenesis, symptoms and host range previously observed between assemblage A and B.

The most abundant proteins, in both human isolates, primarily belong to four families of proteins: *Giardia* cathepsin B family (GCATB), high cysteine membrane proteins (HCMPs), variant surface proteins (VSPs) and tenascins.

The cathepsin B family of *Giardia* are confirmed virulence factors involved in many of the parasite's processes such as encystation and excystation [6]; secreted GCATBs degrade host IL-8 and inhibit neutrophil chemotaxis [3]. GCATB contains secreted and non-secreted trophozoite expressed proteins; the orthologues of which are predominantly common to GS (B) and WB (A) assemblages (Fig 1). Expression of sixteen GCATBs was proteomically confirmed, of which eleven were shown by our proteomic analysis to be secreted. These eleven fell into six orthologous groups and for three of these groups all group members were shown to be secreted. Secreted GCTAB GL50803\_15564 (WB) and its ortholog GL50581\_2036 (GS) show a dN/dS values of >26 indicative of strong positive selective pressure. Interestingly when GS was resequenced, GL50803\_15564 was found to comprise three recently diverged orthologs (GSB\_153537, GSB\_155477, GSB\_150353) and it may be that the positive selection pressure observed has been generated as a result of recent gene duplications in the assemblage B strain. GL50803\_16779, an assemblage A (WB) GCATB, has previously been shown to be up-regulated and involved in trophozoite motility in early pathogenesis of *Giardia* [15]. In this study, this protein was found to be in WB top 5 secreted proteins (Table S4); its GS ortholog (GL50581\_78) was also present but at a considerably lower level suggesting that for this GCATB may play a more significant role in assemblage A than assemblage B.

HCMPs are an enigmatic group of proteins with few associated functional studies. They may protect trophozoites against proteolysis [30, 31] and oxidative damage [32]. In *Giardia*, it appears that one lineage of HCMPs has given rise to the VSPs, whilst another has given rise to a group with high homology to mammalian tenascins. Tenascin, VSPs and HCMPs are then related multi-gene families that together form the largest group of proteins enriched in the *Giardia* supernatants. Interestingly, when aligned and analysed phylogenetically the secreted tenascins segregate into a monophyletic group (Figure S4). Both WB and GS orthologs of five tenascin gene products were secreted and in WB two other secreted tenascins were also detected that were not detected for the GS strain (Figure 1B).

VSPs are well-characterised surface glycoproteins with transmembrane domains, which are expressed one at a time by *Giardia* trophozoites through an RNAi regulated mechanism. They are quintessential virulence factors, responsible for antigenic variation. VSPs are hypervariable by nature and thus it is to be expected that they do not form orthologous pairs. This was the case for most we observed, intriguingly though, a few proteins annotated as VSPs were conserved between isolates suggesting that they are not actually VSPs and would not be subject to "one at a time" controlled expression - but are actually misannotated HCMPs which may have a conserved function in both GS and WB isolates. This study was not able to resolve whether the enrichment of such proteins in the supernatant observed is due to clipping or shedding from the parasite surface or whether the proteins are also secreted.

Tenascins are characterised by the presence of epidermal growth factor (EGF) repeats and are able to act as ligands for EGF receptors. Mammalian tenascins are extracellular matrix proteins, which modulate cell adhesion and migration [33]. They appear to have evolved from a group of proteins specific to vertebrates, presumably co-evolving with the EGF receptor and so the presence of homologous proteins in *Giardia* evolving independently from HCMPs

is a clear example of the kind of convergent evolution best described as molecular mimicry. Interestingly, a *Giardia* tenascin (WB-GL50803\_8687/GS-GL50581\_4316), secreted by both strains and *Giardia* tenascin (WB-GL50803\_14573/GS-GL50581\_1475) secreted only by WB strain (Table 1 and S4) were found to be induced by host soluble factors and implicated in regulation of trophozoite attachment [15], supporting the case for secreted tenascins acting as virulence factors in *Giardia* pathogenesis.

Most published studies concerning host cell-*Giardia* interactions have focused on the effects on the host intestinal epithelia upon attachment of the trophozoites to the cells. In this study, we have shown that diluted supernatant obtained from the steady growth of *Giardia* trophozoites *in vitro* has an effect on the intestinal cell function. The effect observed on chloride secretion by *Giardia* supernatants indicates that *Giardia* secretes a soluble factor, which is likely affecting secretion across the intestinal epithelial cells. Physiologically, cultured intestinal cells show sensitivity to *Giardia* proteins released by the parasite even at high dilution. Fig2D demonstrates that intestinal epithelial cells when acutely exposed to such *Giardia* proteins lose the ability to stimulate CFTR and calcium-activated chloride channels. The clear implication being that virulence determinants released from *Giardia* trophozoites interact with epithelial cell receptors and ion channels.

In this analysis, we have identified the proteins that are secreted by human infective *Giardia* trophozoites. Just two groups form the majority of these proteins: GCATBs and the HCMP superfamily encoding known virulence factors in addition to an abundant extracellular nuclease and an oxygen-fixing enzyme. The elucidation of this group of proteins dramatically increases our understanding of the pathogenic mechanisms underlying giardiasis at a molecular level. The genes encoding GCATBs and HCMP superfamily proteins are among the most heterogeneous of all genes between assemblages. Their probable role in interaction with the host and luminal environment is supported by the very high dN/dS values of some family members. Correlation of variation within assemblages at these loci with strain virulence is the essential next step for their use in diagnosis of virulent strains, risk assessment and disease prognosis.

Our results indicate that *Giardia* secretions are sufficient to disable normal function in enteric epithelial making cells less able to extract fluids from the lumen. In particular, they implicate PNPO, an extracellular nuclease, GCATBs and tenascins. The fact that both extracellular nuclease and GCATBs can be involved in the degradation of the intestinal mucus layer and that both GCATBs and tenascins can be associated with intestinal intracellular junction disruption suggests collaboration between these proteins. Therefore, we propose a pathogenic mechanism (Fig 3) whereby PNPO produces a reducing environment favouring growth of trophozoites, the extracellular nuclease degrades the outer layer of the intestinal mucus improving access for GCATBs for further degradation of the protective mucous barrier and subsequent disruption of intestinal intracellular junctions. Lastly, tenascins are involved in maintaining intestinal cell separation by ligation of EGF receptors present at the surface of intestinal cells and exacerbation of epithelial damage via increased levels of apoptosis amongst these more detached cells. Once the intestinal barrier is breached by the actions of giardia secreted virulence factors,

the sites of damage become prone to secondary infection by other opportunist microbes resident in the intestinal lumen and sensitive to irritation by allergens in foodstuffs leading to further inflammation and to the characteristic symptoms of the disease. Further investigations are necessary to verify this proposed mechanism of pathogenesis of giardiasis.

## METHODS

### Proteomic Analysis.

#### Samples preparation

*Giardia* trophozoites from the genome reference strains WB (assemblage A, ATCC\_50803) and GS (assemblage B, ATCC\_50581), were cultured in TYI-S-33 under standard conditions (5% CO<sub>2</sub>) [34] and harvested during the midlog phase of their *in vitro* growth curves. The total trophozoites (adhered and non-adhered) were washed 3 x in phosphate buffer saline (PBS) and then incubated in non-supplemented DMEM, with antibiotics to conserve an axenic milieu, for 45 minutes at 37°C (Figure S5 A) [38]. After incubation, an aliquot was analysed by flow cytometry to evaluate the viability of the *Giardia* samples. Trophozoites and supernatant were separated by centrifugation and both trophozoite pellet and supernatant were harvested. Proteins contained in supernatant were concentrated in Vivaspin columns (3,000 MWCO) with 25 mM ammonium bicarbonate (Ambic) (Figure S5 B) [39]. Supernatants were analysed by SDS PAGE [40] and were tested on cultured epithelial cells (Caco-2) to ensure the presence of proteins and biological activity (see below). Supernatants and pellets were sent to the Institute of Infection and Global Health at the University of Liverpool for mass spectrometry analysis (Figure S3) [41].

Protein samples were dispensed into low protein-binding microcentrifuge tubes (Sarstedt, Leicester, UK) and made up to 160 µl by addition of 25 mM Ambic. The proteins were denatured using 10 µl of 1% (w/v) RapiGest™ (Waters MS Technologies, Manchester, UK) in 25 mM Ambic followed by three cycles of freeze-thaw, and two cycles of 10 min sonication in water bath. Sample was then incubated at 80 °C for 10 min and reduced (addition of 10 µl of 60 mM DTT and incubation at 65 °C for 10 min) and alkylated (addition of 10 µl of 180 mM iodoacetamide and incubation at room temperature for 30 min in the dark). Trypsin (Sigma-Aldrich, Dorset, UK) was reconstituted in 50 mM acetic acid to a concentration of 0.2 µg/µl. Digestion was performed by the addition of 10 µl of trypsin to the sample followed by incubation at 37 °C overnight. The RapiGest™ was removed from the sample by acidification (1 µl of trifluoroacetic acid and incubation at 37 °C for 45 min) and centrifugation (15,000 × *g* for 15 min) [41]. After protein digestion, 1 µg of digest were injected into both the Orbitrap Velos and the Q-Exactive MS, for all samples.

#### Orbitrap Velos

Peptide mixtures were analysed by on-line nanoflow liquid chromatography using the nanoACQUITY-nLC system (Waters MS technologies, Manchester, UK) coupled to an LTQ-Orbitrap Velos (ThermoFisher Scientific, Bremen, Germany) mass spectrometer equipped with the manufacturer's nanospray ion source. The analytical column (nanoACQUITY UPLC™ BEH130 C18 15cm x 75µm, 1.7µm capillary column) was maintained at 35°C and a flow-rate of 300nl/min. The gradient consisted of 3-40% acetonitrile in 0.1% formic acid for 90min then a ramp of 40-85% acetonitrile in 0.1% formic acid for 3 min. Full scan MS spectra (m/z range 300-2000) were acquired by the Orbitrap at a resolution of 30,000. Analysis was performed in data dependant mode. The top 20 most intense ions from MS1 scan (full MS) were selected for tandem MS by collision induced dissociation (CID) and all product spectra were acquired in the LTQ ion trap. Ion trap and orbitrap maximal injection times were set to 50ms and 500ms, respectively.

#### Q-Exactive MS

Digests (2 µl) were analysed on a 50cm Easy-Spray column with an internal diameter of 75µm, packed with 2µm C18 particles, fused to a silica nano-electrospray emitter (Thermo Fisher Scientific). Reversed phase liquid chromatography was performed using the Ultimate 3000 nano system with a binary buffer system consisting of 0.1% formic acid (buffer A) and 80% acetonitrile in 0.1% formic acid (buffer B). The peptides were separated by a linear gradient of 5-40% buffer B over 110 min at a flow rate of 300nl/min. The column was operated at a constant temperature of 35°C and the LC system coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific). The Q-Exactive was operated in data-dependent mode with survey scans acquired at a resolution of 70,000 at m/z 200. Up to the top 10 most abundant isotope patterns with charge states +2, +3 and/or +4 from the survey scan were selected with an isolation window of 2.0Th and fragmented by higher energy collisional dissociation with normalized collision energies of 30. The maximum ion injection times for the survey scan and the MS/MS scans were 250 and 100ms, respectively, and the ion target value was set to 1E6 for survey scans and 1E4 for the MS/MS scans. Repetitive sequencing of peptides was minimized through dynamic exclusion of the sequenced peptides for 20s.

#### Data analysis

Thermo RAW files were imported into Progenesis LC-MS (version 4.1, Nonlinear Dynamics). Replicate runs were time-aligned using default settings and an auto-selected run as a reference. Peaks were picked by the software using default settings and filtered to include only peaks with a charge state of between +2 and +6. Peptide intensities of replicates were normalised against the reference run by Progenesis LC-MS. Spectral data were transformed to .mgf files with Progenesis LC-MS and exported for peptide identification using the PEAKS Studio 7 (Bioinformatics Solutions Inc.) search engine. Multiple search engine platform provided by PEAKS Studio named inChorus was used, which combines searching results from PEAKS DB (Bioinformatics Solutions Inc.), Mascot (Matrix Science), OMSSA (National Center for Biotechnology Information) and X!Tandem (Global Proteome Machine Organization). Tandem MS data were searched against a custom database that contained the common contamination and internal standards, GiardiaDB-3.1\_GintestinalisAssemblageA\_AnnotatedProteins or GiardiaDB-3.1\_GintestinalisAssemblageB\_AnnotatedProteins. The search parameters for Orbitrap-

Velos were as follows; precursor mass tolerance was set to 10ppm and fragment mass tolerance was set to 0.5 Da. One missed tryptic cleavage was permitted. Carbamidomethylation was set as a fixed modification and oxidation (M) set as a variable modification. The search parameters for Q Exactive were as follows; precursor mass tolerance was set to 10ppm and fragment mass tolerance was set to 0.01 Da. One missed tryptic cleavage was permitted. Carbamidomethylation was set as a fixed modification and oxidation (M) set as a variable modification. The false discovery rates (FDR) were set at 1% and at least two unique peptides were required for reporting protein identifications. Protein abundance (iBAQ) was calculated as the sum of all the peak intensities (from Proteomics output) divided by the number of theoretically observable tryptic peptides [16]. Protein abundance was normalised by dividing the protein iBAQ (intensity based absolute quantification) value by the summed iBAQ values for that sample. The reported abundance is the mean of the biological replicates.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository [24] with the dataset identifier PXD004398 and 10.6019/PXD004398.

### **Electrophysiology.**

#### *Giardia* trophozoites culture

*Giardia lamblia* WB and GS strain as well as the patients' strains (obtained from 3 patients with giardiasis from the NNUH) were grown in filter sterilized, modified TYI-S-33 medium with 10% adult bovine serum and 0.05% bovine bile [28] at 37°C in microaerophilic conditions and sub-cultured when confluent. To collect parasites for experiments, the medium was removed from the culture to eliminate unattached or dead parasites. The tube was refilled with cold, sterile medium and trophozoites detached by chilling on ice for 15 minutes.

Parasites were collected by centrifugation (1500 x g for 5 minutes at 4 °C) and washed once with the plating medium of 90% complete DMEM/10% *Giardia* medium. Parasites were then counted using a haemocytometer and diluted to the appropriate number.

To collect *Giardia* supernatant for experiments, the *Giardia* culture bottle was placed on ice for 15 minutes. The bottle then underwent centrifugation (1500 x g for 5 minutes at 4 °C). The supernatant was then collected and filtered 3 times using a 15mm diameter syringe filters (0.2µm pore size). Subsequently the post-filtered *Giardia* supernatant was diluted 1:1000 and saved in -20°C freezer until required.

#### Mammalian cell line (CaCo-2) preparation

CaCo-2 cells (passages 20-25) were grown in DMEM supplemented with nonessential amino acids, penicillin (12 IU/ml), streptomycin (12 µg/ml), gentamycin (47 µg/ml) and 20% (vol/vol) heat inactivated fetal calf serum (all from AMIMED, Bioconcept). The cells were seeded at a density of  $6 \times 10^4$  cells/cm<sup>2</sup> in 6-well Transwell filters (0.4 µm pore size) and cultured for 7-15 days until confluent. Confluent monolayers were then used for electrophysiological experiments, for co-culture experiments with *Giardia* parasites or for culture with *Giardia* supernatants [42].

#### CaCo-2 co-culture experiments with *Giardia* or *Giardia* supernatant

Confluent CaCo-2 monolayers were taken and the CaCo-2 cell media was removed and replenished with a combination of 90% complete DMEM/10% *Giardia* medium plus or minus *Giardia* trophozoites (100,000 total parasites per insert). Control cultures were maintained in a separate plate to prevent parasite contamination. Control inserts were inspected under the microscope to ensure there was no *Giardia* cross contamination. The co-cultures were incubated at 37°C and 5% CO<sub>2</sub> for 24 hours, after which the *Giardia* parasites were removed [42].

Confluent Caco-2 monolayers were also cultured with diluted (1:1000) *Giardia* supernatants for 24 hours. Briefly, the culture media was removed from the insert and Caco-2 cell media was replaced with a combination of 99.9% complete DMEM/ 0.1% *Giardia* medium plus or minus *Giardia* supernatant [42].

#### Transepithelial electrical resistance (TEER) Assay

Monolayers of CaCo-2 cells were grown on 6-well Transwell filters (0.4 µm pore size) for 7-15 days until confluent. The development of the polarised monolayer was assessed by measuring the TEER over a 7-15 day period. Once confluent, *Giardia* were added to the apical side of the Transwell filter and incubated for 24 hours. The integrity of the confluent polarised monolayer was assessed by measuring the TEER before and/or after apical infection by *Giardia* [42].

#### Electrophysiology Assay

Monolayers of CaCo-2 cells on Transwell filters were mounted into a Physiological Instruments EM-CSYS-2 Ussing chamber set-up, after establishment of a confluent monolayer and the short circuit current ( $I_{sc}$ ) across the monolayer was continuously measured [42].

Both sides of the epithelium were bathed in 5ml of Krebs Henseleit solution that was continuously circulated through the half chambers, maintained at 37°C and continuously bubbled with 95% O<sub>2</sub> / 5% CO<sub>2</sub>. The composition of the Krebs Henseleit bath solution used was similar to that used by Cuthbert [35] and had the following composition (in mM): NaCl 118, KCl 4.7, CaCl<sub>2</sub> 2.5, MgCl<sub>2</sub> 1.2, NaHCO<sub>3</sub> 25, KH<sub>2</sub>PO<sub>4</sub> 1.2 and glucose 11.1 (pH 7.4). The permeable supports

were left for 30 mins to equilibrate before experiments were started. All filters were treated with 10 $\mu$ M amiloride apically to eliminate electrogenic sodium absorption through epithelial sodium channels (ENaC) [42].

#### Data analysis

$I_{sc}$  was continuously monitored across the monolayers by a Physiological Instruments Multichannel Voltage/Current Clamp (VCC MC6) through 3M KCl/agar, Ag/AgCl<sub>2</sub> cartridge electrodes (Physiologic Instruments), and the raw data for  $I_{sc}$ , transepithelial resistance and transepithelial voltage were recorded using Acquire and Analyse version 1.3 software (Physiological Instruments). Data were exported to Microsoft Excel initially and then into GraphPad Prism version 5.0 for Windows package for data representation and statistical analysis.

#### Chemicals and Inhibitors

Forskolin (10 $\mu$ M), UTP (100 $\mu$ M), Amiloride (10 $\mu$ M), and DIDS (100 $\mu$ M) were obtained from Sigma Aldrich, and GlyH-101 (50  $\mu$ M) was obtained from Merck Chemicals. Stock solutions of Amiloride (10mM), GlyH-101 (50mM) were made by dissolving in DMSO. Final concentrations of drugs are as indicated in the text or figures and were produced by adding the appropriate volume of stock concentration to 5ml of either the basolateral or apical bathing solution.

#### Phylogeny

To look for sequence similarities between proteins of interest from a same protein family, the coding sequences of these proteins were retrieved from GiardiaDB (v 3.1, 4.0 and 5.0), aligned and compared using ClustalW.

Phylogenetic trees were built for these proteins, via Maximum likelihood approach using MEGA software (v. 6.06).

#### Availability of Supporting Data

All proteomic datasets are held by and can be accessed for free at the European Bioinformatics PRoteomics IDentifications (PRIDE) database (accession number PXD004398). Free Integrated functionality with other *Giardia* large datasets hosted at EupathDB [36]. Supporting data, including raw data in .csv format, alignments and phylogenetic analyses, are also available via the *GigaScience* repository GigaDB [37]. All protocols used in this study are available and can be accessed at protocols.io database [38-43].

**Abbreviations:** ABC (ATP-binding cassette); ADI (Arginine Deiminase); Ambic (Ammonium bicarbonate); ATP (Adenosine triphosphate); CaCo-2 (Human colonic adenocarcinoma derived epithelial cell line-2); DMEM (Dulbecco's Modified Eagle Medium); DIDS (4,4'-disothiocyanatostibene-2,2'-sulfonic acid); EF-1 $\alpha$  (Elongation Factor 1-alpha); EGF (Epidermal growth factor); ENaC (Epithelial Sodium Channel); FDR (False discovery rate); FMN (Flavin mononucleotide); GCATB (*Giardia* cathepsin B); GlyH101; HCMP (High cysteine membrane protein); iBAQ (Intensity based absolute quantification); IEC (Intestinal Epithelial Cells); IL (Interleukine); Isc (Short-circuit current); mRNA (messenger RNA); msrB (peptide methionine sulfoxide reductase B); OCT (Ornithine Carbamoyltransferase); ORF (Open reading frame); P (Pellet); PNPO (Pyridoxamine 5'-phosphate oxidase); rcf (Relative centrifugal force); RNA (Ribonucleic acid); ROS (Reductive oxygen species); rpm (Rotations per minute); SP (Supernatant); PRIDE (PRoteomics IDentifications); TEER (Transepithelial electrical resistance), VSP (Variant surface protein)

### Conflicts of Interest

The authors declare that they have no competing interests

### Authors' Contributions

K.T., J.M.W., J.P.W and P.H. conceived and designed the studies. K. T. and A. D. co-ordinated the experiments. A.D. and S.A.N. performed the electrophysiology with J.P.W. A.D. performed the Flow Cytometry with D.S. A.D. prepared the proteomic samples. D.X. performed the proteomic experiments. A.D. and M.B. performed the phylogenetic analysis. All authors contributed to the analysis of the data sets obtained and preparation of Figures and Tables. The manuscript was drafted by A.D. and K.T. and improved and approved prior to submission by all co-authors.

### Acknowledgements

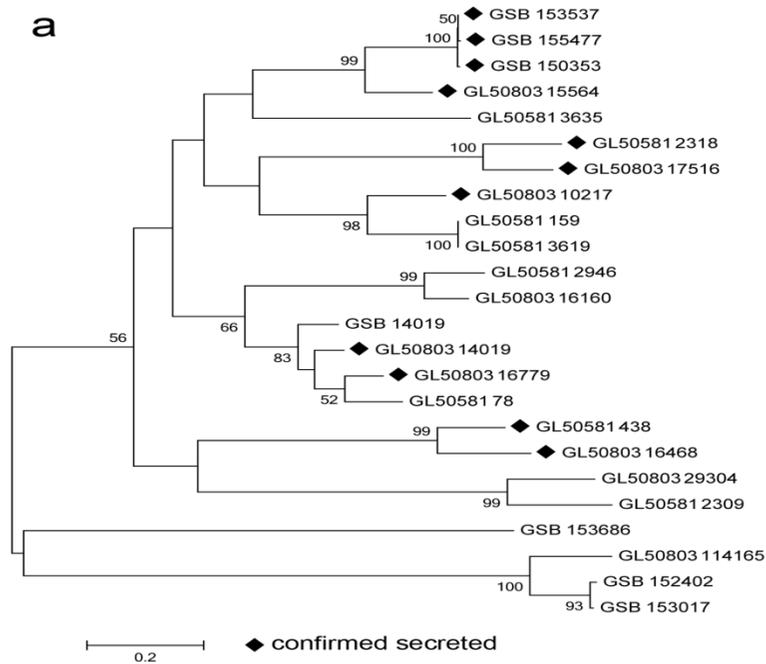
The research leading to these results was primarily funded from the European Union Seventh Framework Programme ([FP7/2007-2013] [FP7/2007-2011]) under Grant agreement no: 311846. PRH is supported by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at the University of Liverpool in partnership with Public Health England (PHE), and in collaboration with University of East Anglia, University of Oxford and the Institute of Food Research. Professor Hunter is based at University of East Anglia. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England. We thank Susanne Warrenfeltz and the EuPathDB team for invaluable assistance in making the integrated datasets available.

### REFERENCES

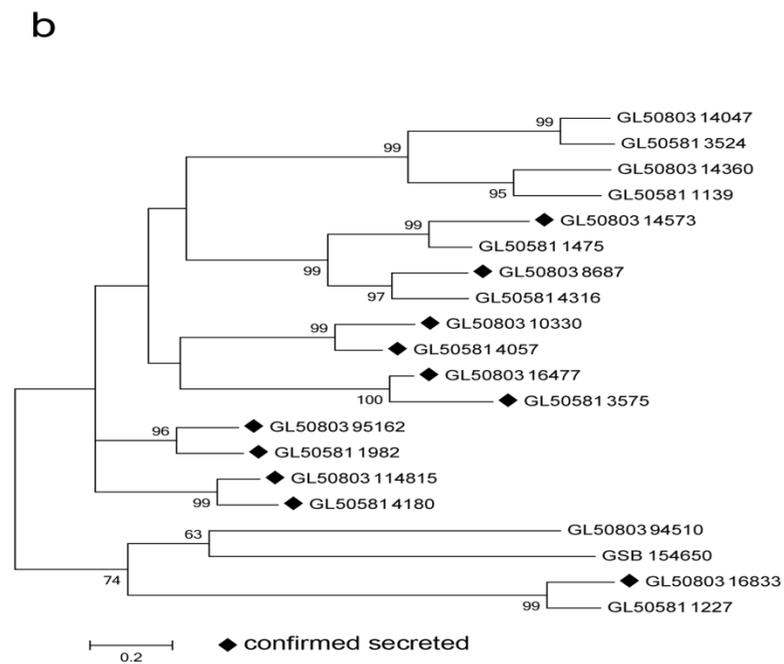
1. Esch KJ and Petersen CA. Transmission and epidemiology of zoonotic protozoal diseases of companion animals. *Clinical microbiology reviews*. 2013;26 1:58-85. doi:10.1128/CMR.00067-12.
2. Nash TE, Herrington DA, Losonsky GA and Levine MM. Experimental human infections with *Giardia lamblia*. *The Journal of infectious diseases*. 1987;156 6:974-84.
3. Cotton JA, Bhargava A, Ferraz JG, Yates RM, Beck PL and Buret AG. *Giardia duodenalis* cathepsin B proteases degrade intestinal epithelial interleukin-8 and attenuate interleukin-8-induced neutrophil chemotaxis. *Infect Immun*. 2014; doi:10.1128/IAI.01771-14.
4. Turk V, Stoka V, Vasiljeva O, Renko M, Sun T, Turk B, et al. Cysteine cathepsins: from structure, function and regulation to new frontiers. *Biochim Biophys Acta*. 2012;1824 1:68-88. doi:10.1016/j.bbapap.2011.10.002.
5. Musil D, Zucic D, Turk D, Engh RA, Mayr I, Huber R, et al. The refined 2.15 Å X-ray crystal structure of human liver cathepsin B: the structural basis for its specificity. *EMBO J*. 1991;10 9:2321-30.
6. DuBois KN, Abodeely M, Sakanari J, Craik CS, Lee M, McKerrow JH, et al. Identification of the major cysteine protease of *Giardia* and its role in encystation. *The Journal of biological chemistry*. 2008;283 26:18024-31. doi:M802133200 [pii]10.1074/jbc.M802133200.
7. Sajid M and McKerrow JH. Cysteine proteases of parasitic organisms. *Molecular and biochemical parasitology*. 2002;120 1:1-21. doi:S0166685101004388 [pii].
8. Rodriguez-Fuentes GB, Cedillo-Rivera R, Fonseca-Linan R, Arguello-Garcia R, Munoz O, Ortega-Pierres G, et al. *Giardia duodenalis*: analysis of secreted proteases upon trophozoite-epithelial cell interaction in vitro. *Mem Inst Oswaldo Cruz*. 2006;101 6:693-6.
9. Paget TA and James SL. The mucolytic activity of polyamines and mucosal invasion. *Biochem Soc Trans*. 1994;22 4:394S.
10. Wampfler PB, Tosevski V, Nanni P, Spycher C and Hehl AB. Proteomics of Secretory and Endocytic Organelles in *Giardia lamblia*. *PLoS One*. 2014;9 4:e94089. doi:10.1371/journal.pone.0094089.
11. Faso C, Bischof S and Hehl AB. The proteome landscape of *Giardia lamblia* encystation. *PLoS One*. 2013;8 12:e83207. doi:10.1371/journal.pone.0083207.
12. Lingdan L, Pengtao G, Wenchao L, Jianhua L, Ju Y, Chengwu L, et al. Differential dissolved protein expression throughout the life cycle of *Giardia lamblia*. *Exp Parasitol*. 2012;132 4:465-9. doi:10.1016/j.exppara.2012.09.014.
13. Ringqvist E, Palm JE, Skarin H, Hehl AB, Weiland M, Davids BJ, et al. Release of metabolic enzymes by *Giardia* in response to interaction with intestinal epithelial cells. *Molecular and biochemical parasitology*. 2008;159 2:85-91. doi:S0166-6851(08)00056-X [pii]10.1016/j.molbiopara.2008.02.005.
14. Roxstrom-Lindquist K, Palm D, Reiner D, Ringqvist E and Svard SG. *Giardia* immunity--an update. *Trends Parasitol*. 2006;22 1:26-31. doi:S1471-4922(05)00311-9 [pii]10.1016/j.pt.2005.11.005.
15. Emery SJ, Mirzaei M, Vuong D, Pascovici D, Chick JM, Lacey E, et al. Induction of virulence factors in *Giardia duodenalis* independent of host attachment. *Scientific reports*. 2016;6:20765. doi:10.1038/srep20765.
16. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. *Nature*. 2011;473 7347:337-42. doi:10.1038/nature10098.

17. Emery SJ, Lacey E and Haynes PA. Data from a proteomic baseline study of Assemblage A in *Giardia duodenalis*. Data in brief. 2015;5:23-7. doi:10.1016/j.dib.2015.08.003.
18. Emery SJ, Lacey E and Haynes PA. Quantitative proteomic analysis of *Giardia duodenalis* assemblage A: A baseline for host, assemblage, and isolate variation. *Proteomics*. 2015;15 13:2281-5. doi:10.1002/pmic.201400434.
19. Jerlstrom-Hultqvist J, Ankarklev J and Svard SG. Is human giardiasis caused by two different *Giardia* species? *Gut Microbes*. 2010;1 6:379-82. doi:10.4161/gmic.1.6.13608.
20. Skarin H, Ringqvist E, Hellman U and Svard SG. Elongation factor 1-alpha is released into the culture medium during growth of *Giardia intestinalis* trophozoites. *Exp Parasitol*. 2011;127 4:804-10. doi:S0014-4894(11)00017-8 [pii]10.1016/j.exppara.2011.01.006.
21. Condeelis J. Elongation factor 1 alpha, translation and the cytoskeleton. *Trends Biochem Sci*. 1995;20 5:169-70. doi:S0968000400889987 [pii].
22. Nandan D and Reiner NE. *Leishmania donovani* engages in regulatory interference by targeting macrophage protein tyrosine phosphatase SHP-1. *Clinical immunology*. 2005;114 3:266-77. doi:10.1016/j.clim.2004.07.017.
23. Di Matteo A, Scandurra FM, Testa F, Forte E, Sarti P, Brunori M, et al. The O<sub>2</sub>-scavenging flavodiiron protein in the human parasite *Giardia intestinalis*. *The Journal of biological chemistry*. 2008;283 7:4061-8. doi:10.1074/jbc.M705605200.
24. Kamitani T, Kito K, Nguyen HP, Fukuda-Kamitani T and Yeh ET. Characterization of a second member of the sentrin family of ubiquitin-like proteins. *The Journal of biological chemistry*. 1998;273 18:11349-53.
25. Sauvage V, Aubert D, Escotte-Binet S and Villena I. The role of ATP-binding cassette (ABC) proteins in protozoan parasites. *Molecular and biochemical parasitology*. 2009;167 2:81-94. doi:10.1016/j.molbiopara.2009.05.005.
26. Malys N and McCarthy JE. Translation initiation: variations in the mechanism can be anticipated. *Cellular and molecular life sciences : CMLS*. 2011;68 6:991-1003. doi:10.1007/s00018-010-0588-z.
27. Kim HY and Gladyshev VN. Methionine sulfoxide reduction in mammals: characterization of methionine-R-sulfoxide reductases. *Molecular biology of the cell*. 2004;15 3:1055-64. doi:10.1091/mbc.E03-08-0629.
28. Denkel LA, Horst SA, Rouf SF, Kitowski V, Bohm OM, Rhen M, et al. Methionine sulfoxide reductases are essential for virulence of *Salmonella typhimurium*. *PloS one*. 2011;6 11:e26974. doi:10.1371/journal.pone.0026974.
29. Oke TT, Moskovitz J and Williams DL. Characterization of the methionine sulfoxide reductases of *Schistosoma mansoni*. *The Journal of parasitology*. 2009;95 6:1421-8. doi:10.1645/GE-2062.1.
30. Davids BJ, Reiner DS, Birkeland SR, Preheim SP, Cipriano MJ, McArthur AG, et al. A new family of giardial cysteine-rich non-VSP protein genes and a novel cyst protein. *PLoS One*. 2006;1:e44. doi:10.1371/journal.pone.0000044.
31. Nash TE. Surface antigenic variation in *Giardia lamblia*. *Mol Microbiol*. 2002;45 3:585-90. doi:3029 [pii].
32. Requejo R, Hurd TR, Costa NJ and Murphy MP. Cysteine residues exposed on protein surfaces are the dominant intramitochondrial thiol and may protect against oxidative damage. *The FEBS journal*. 2010;277 6:1465-80. doi:10.1111/j.1742-4658.2010.07576.x.
33. Chiquet-Ehrismann R and Chiquet M. Tenascins: regulation and putative functions during pathological stress. *The Journal of pathology*. 2003;200 4:488-99. doi:10.1002/path.1415.

34. Keister DB. Axenic culture of *Giardia lamblia* in TYI-S-33 medium supplemented with bile. *Trans R Soc Trop Med Hyg.* 1983;77 4:487-8.
35. Cuthbert AW. Assessment of CFTR chloride channel openers in intact normal and cystic fibrosis murine epithelia. *Br J Pharmacol.* 2001;132 3:659-68. doi:10.1038/sj.bjp.0703859.
36. EuPathDB: Eukaryotic Pathogen Database Resources : *Giardia* DB 35  
[http://giardiadb.org/giardiadb/showXmlDataContent.do?name=XmlQuestions.News#giardiadb12\\_17\\_release](http://giardiadb.org/giardiadb/showXmlDataContent.do?name=XmlQuestions.News#giardiadb12_17_release) accessed 11 Jan 2018
37. Dubourg A, Xia D, Winpenny JP, Al-Naimi S, Bouzid M, Sexton DW, et al. Supporting data for "Giardia Secretome Highlights Secreted Tenascins as a Key Component of Pathogenesis" *GigaScience* Database. 2018. <http://dx.doi.org/10.5524/100381>
38. Dubourg A, Xia D, Winpenny JP, Al Naimi S, Bouzid M, Sexton DW, et al. *Giardia* supernatant cleansing protocol. *Protocols.io*.2018.  
[dx.doi.org/10.17504/protocols.io.k96cz9e](http://dx.doi.org/10.17504/protocols.io.k96cz9e)
39. Dubourg A, Xia D, Winpenny JP, Al Naimi S, Bouzid M, Sexton DW, et al. Concentration of *Giardia* supernatant proteins. *Protocols.io*.2018.  
[dx.doi.org/10.17504/protocols.io.mccc2sw](http://dx.doi.org/10.17504/protocols.io.mccc2sw)
40. Dubourg A, Xia D, Winpenny JP, Al Naimi S, Bouzid M, Sexton DW, et al. Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE) with Sypro straining. *Protocols.io*.2018.  
[dx.doi.org/10.17504/protocols.io.mcdc2s6](http://dx.doi.org/10.17504/protocols.io.mcdc2s6)
41. Dubourg A, Xia D, Winpenny JP, Al Naimi S, Bouzid M, Sexton DW, et al. Proteomics Analysis. *Protocols.io*.2018.  
[dx.doi.org/10.17504/protocols.io.mcec2te](http://dx.doi.org/10.17504/protocols.io.mcec2te)
42. Dubourg A, Xia D, Winpenny JP, Al Naimi S, Bouzid M, Sexton DW, et al. *Giardia* electrophysiological assays. *Protocols.io*.2018.  
[dx.doi.org/10.17504/protocols.io.mcfc2tn](http://dx.doi.org/10.17504/protocols.io.mcfc2tn)
43. **Dubourg A, Xia D, Winpenny JP, Al Naimi S, Bouzid M, Sexton DW, et al. *Giardia* Secretome Highlights Secreted Tenascins as a Key Component of Pathogenesis *Protocols.io*.2018.**  
[dx.doi.org/10.17504/protocols.io.k97cz9n](http://dx.doi.org/10.17504/protocols.io.k97cz9n)



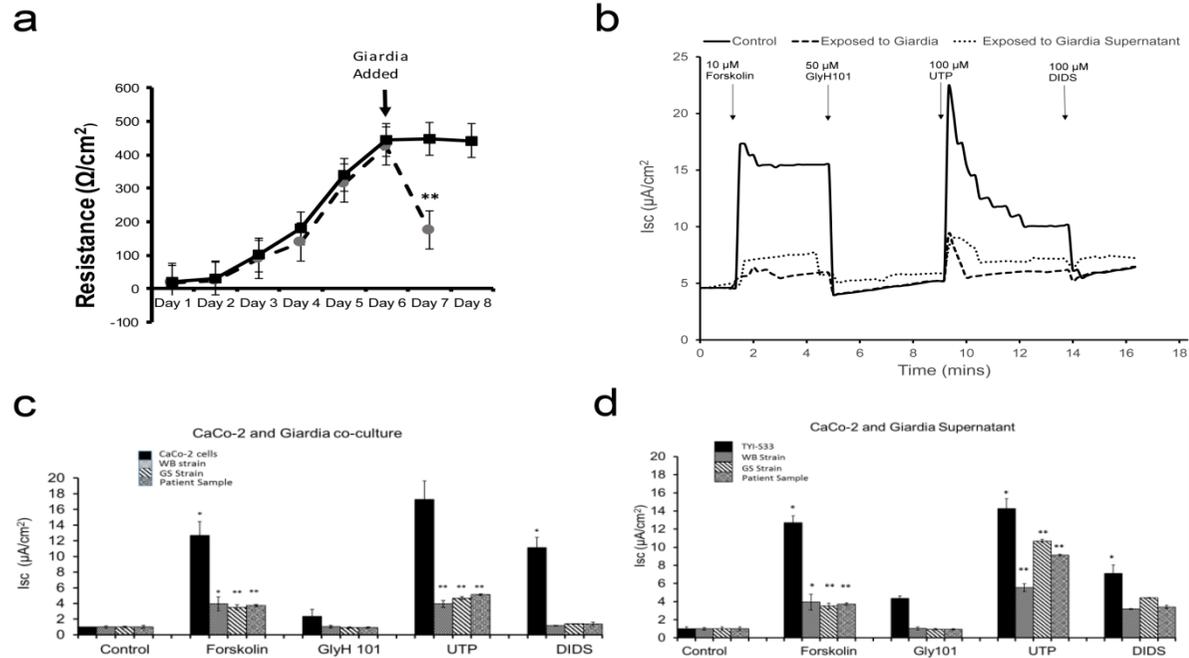
*Giardia Cathepsin B*



*Giardia Tenascin*

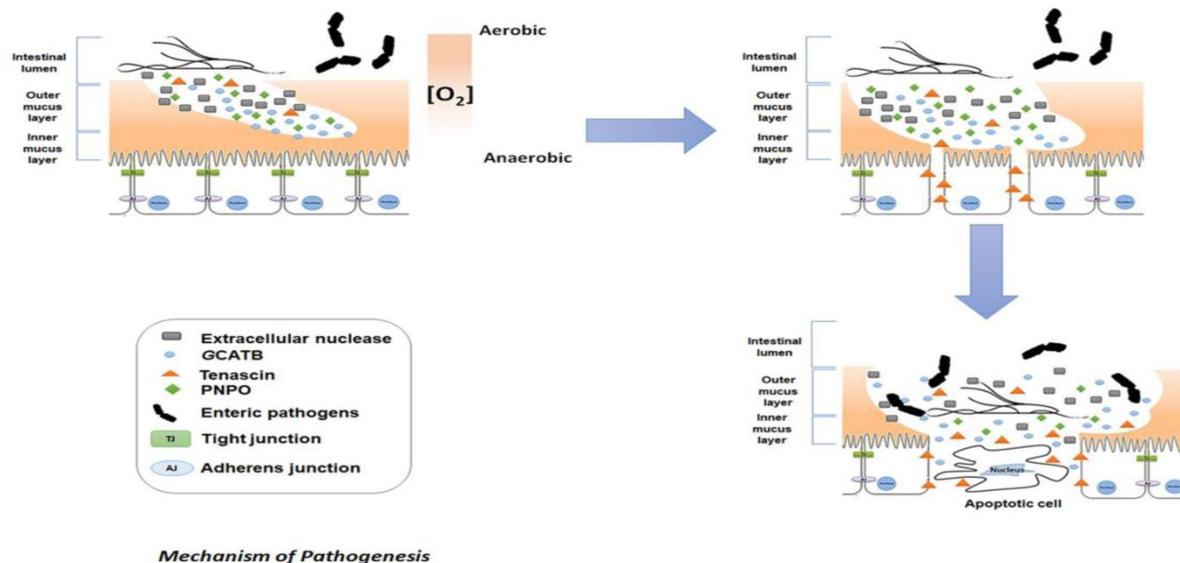
**Figure 1: Neighbour joining tree showing clustering of A) Cathepsin B and B) Tenascin gene families.** Genes were retrieved by gene name search on GiardiaDB. Gene sequences were downloaded and aligned using ClustalW generated with MEGA 6 software package. Maximum composite likelihood method was used, with 2000 bootstrap replicates. Bootstrap values greater than 50% are shown above the branches. ◆ proteins confirmed to be secreted using our proteomic analysis.

Preprint



**Figure 2: The effect of co-culture with *Giardia* or *Giardia* supernatants on the electrophysiological properties of CaCo-2 monolayers. A)** Transepithelial electrical resistance (TEER) in CaCo-2 monolayers following seeding on permeable supports. Data shows increase in TEER as monolayer develops. Confluence occurred around Day 6. *Giardia* were added on Day 6 after confluent monolayer formed and co-cultured with the CaCo-2 monolayer for 24 hours. TEER was measured after 24 hours and compared to TEER in monolayers that had not been exposed to *Giardia* (n=6). **B)** A representative short circuit current (Isc) against time recording from single monolayers of CaCo-2 cells in an Ussing chamber. The trace shows the activation of CFTR chloride channels (basolateral application of 10 μM Forskolin) and calcium-activated chloride channels (basolateral application of 100 μM UTP). Specificity of activation is confirmed by inhibition of Isc by the specific CFTR channel blocker, GlyH101; and specific calcium-activated chloride channel blocker, DIDS. The effect on Isc of 24 hour co-incubation of CaCo-2 monolayers with *Giardia* or with *Giardia* supernatant (1:1000 dilution) is also shown. **C)** Effect of 24 hour

co-incubation of CaCo-2 monolayers with different strains of *Giardia* (WB, GS and patient samples) on forskolin-stimulated and UTP-stimulated Isc (n=3). **D)** Effect of supernatant co-incubation from different strains of *Giardia* (WB, GS and patient samples) on forskolin-stimulated and UTP-stimulated Isc (n=3) from Caco-2 monolayers. The results were analysed by student's t-test and expressed as mean values  $\pm$  standard error mean (SEM). Significant difference expressed as \*P<0.05, \*\*P<0.01 compared to control.



**Figure 3: Proposed novel mechanism of pathogenicity for *Giardia* involving PNPO, extracellular nuclease, GCATB, Tenascin.** PNPO ( ) renders the intestinal environment more favourable to trophozoite's growth. Once a new *Giardia* colony is established, trophozoites release extracellular nuclease ( ), GCATB ( ) and Tenascin ( ). Extracellular nuclease may contribute to reducing the viscosity of the intestinal outer mucus layer, while GCATB may degrade mucins and disrupt intracellular junction. Finally, Tenascins may maintain intestinal cells apart by attaching to the EGF receptors present at the surface of intestinal cells that could over time lead to the apoptosis of these isolated intestinal cells.

**Table 1: The secretome of human infective *Giardia* trophozoites of assemblage A and B have a conserved repertoire of abundant secreted factors identified by both Orbitrap MS and Q-Exactive MS.** 15 proteins were identified as most likely to be secreted by both GS and WB isolates. 12 are annotated proteins and 3 are hypothetical proteins. Proteins are ranked according to GS Q-Exactive Supernatant (SP) protein abundance, from most to least abundant. Of the 12 annotated proteins, 5 are tenascins and 3 are related high cysteine membrane proteins or VSP and three are cathepsin Bs. The other annotated abundant secreted protein is an extracellular nuclease. Protein ranking represents the proteins rank within this table, from most to least abundant. Detailed breakdown of the secretome for each assemblage by each method are provided in Supplemental tables 1-4.

Protein description	GI Number Assemblage A	GI Number Assemblage B	A:B Identity	dN/dS	Signal Peptide <sup>b</sup>	Protein Abundance		SP/P Ratio	Abundance ranking
						Pellet (P) iBAQ	Supernatant (SP) iBAQ		
PNPO	GL50803_5810	GL50581_4133	99.2	0.038	NP <sup>c</sup>	5.71E+07	1.18E+08	2.063091	1
Tenascin	GL50803_95162	GL50581_1982	76.2	<b>1.597<sup>a</sup></b>	0.99	2.97E+07	4.77E+07	1.607024	2
Tenascin	GL50803_10330	GL50581_4057	73.5	0.347	0.99	4.66E+06	2.02E+07	4.342293	3
Cathepsin B	GL50803_16468	GL50581_438	83.6	0.1072	0.78	9.63E+06	1.79E+07	1.861309	4
Tenascin	GL50803_8687	GL50581_4316	77.6	<b>44.176<sup>a</sup></b>	0.98	6.22E+06	1.10E+07	1.770526	5
Uncharacterised	GL50803_5258	GL50581_2767	91.2	0.029	NP <sup>c</sup>	3.93E+06	1.09E+07	2.780918	6
Extracellular nuclease	GL50803_8742	GL50581_3607	83.1	0.234	1	1.03E+06	4.57E+06	4.436193	7
Tenascin-37	GL50803_16477	GL50581_3575	79.8	0.1256	0.99	8.35E+05	4.03E+06	4.830956	8
Cathepsin B	GL50803_15564	GL50581_2036	79.1	<b>26.5782<sup>a</sup></b>	1	1.10E+06	3.95E+06	3.589608	9
CKS1	GL50803_2661	GL50581_3484	100	0.001	NP <sup>c</sup>	1.14E+06	3.20E+06	2.803062	10
Tenascin	GL50803_113038	GL50581_4180	79	0.0949	1	1.20E+06	3.14E+06	2.620931	11
HCMP Group 1	GL50803_7715	GL50581_727	67	0.1821	0.99	ND <sup>d</sup>	2.54E+06	∞	12
Uncharacterised	GL50803_16522	GL50581_352	76	0.1591	NP <sup>c</sup>	1.15E+06	2.21E+06	1.928833	13
HCMP	GL50803_12063	GL50581_2622	83	0.246	1	3.62E+05	1.94E+06	5.354665	14
Cathepsin B	GL50803_17516	GL50581_2318	72.8	0.2056	1	ND <sup>d</sup>	7.81E+05	∞	15

<sup>a</sup> dN/dS in bold indicate protein show evidence of positive selective pressure during divergence from a common ancestor

<sup>b</sup> Probability of N-terminal signal peptide using SignalP

<sup>c</sup> Not predicted

<sup>d</sup> Not detected

**Table 2: Human infective *Giardia* trophozoites of assemblage A and B secrete a small set of different proteins.** Seven proteins were identified as most likely to be secreted by either GS or WB isolates. Two are most likely to be secreted by GS isolate (shown in *italic*) and five are most likely to be secreted by WB isolate (shown in **bold**). One GS isolate- and two WB isolate-secreted were identified only via Q-Exactive MS in the other assemblage's dataset (shown in red). The abundance ranking represents the protein ranking within the secretome of both assemblages according to their abundances in the supernatant.

Protein description	GI Number Assemblage A	GI Number Assemblage B	Signal Peptide <sup>a</sup>	Protein Abundance						Abundance ranking in secretome	
				Assemblage A			Assemblage B			Assemblage A	Assemblage B
				Pellet (P) iBAQ	Supernatant (SP) iBAQ	SP/P Ratio	Pellet (P) iBAQ	Supernatant (SP) iBAQ	SP/P Ratio		
<i>A-type flavoprotein lateral candidate</i>	<i>GL50803_10358</i>	<i>GL50581_1626</i>	<i>NP<sup>c</sup></i>	2.11E+07	3.70E+06	0.17546	4.35E+06	8.70E+06	2.001741	47	10
<i>Sentrin</i>	<i>GL50803_7760</i>	<i>GL50581_3210</i>	<i>NP<sup>c</sup></i>	<i>1.44E+05</i>	<i>ND<sup>c</sup></i>	<i>ND<sup>c</sup></i>	<i>ND<sup>c</sup></i>	<i>3.67E+04</i>	$\infty$	-	31
<b>EF-1<math>\alpha</math></b>	<b>GL50803_112312</b>	-	<b>NP<sup>c</sup></b>	<b>ND<sup>c</sup></b>	<b>2.22E+06</b>	$\infty$	-	-	-	<b>17</b>	-
<b>ATP-binding cassette protein 5</b>	<b>GL50803_8227</b>	<b>GL50581_3399</b>	<b>NP<sup>c</sup></b>	<b>2.93E+06</b>	<b>2.89E+06</b>	<b>0.98546</b>	<b>2.01E+05</b>	<b>1.43E+04</b>	<b>0.071052</b>	<b>15</b>	<b>897</b>
<b>CxC rich protein</b>	<b>GL50803_17476</b>	<b>GL50581_4509</b>	<b>1</b>	<b>4.80E+05</b>	<b>2.83E+05</b>	<b>0.58945</b>	<b>4.35E+04</b>	<b>3.59E+04</b>	<b>0.823376</b>	<b>43</b>	<b>819</b>
<b>Peptide methionine sulfoxide reductase MsrB</b>	<b>GL50803_5180</b>	<b>GL50581_3084</b>	<b>NP<sup>c</sup></b>	<b>1.24E+05</b>	<b>8.80E+04</b>	<b>0.70952</b>	<b>1.09E+06</b>	<b>1.30E+06</b>	<b>1.19568</b>	<b>53</b>	<b>331</b>
<b>Translation initiation inhibitor</b>	<b>GL50803_480</b>	<b>GL50581_4017</b>	<b>NP<sup>c</sup></b>	<b>9.47E+06</b>	<b>4.08E+06</b>	<b>0.43038</b>	<b>1.06E+07</b>	<b>9.76E+06</b>	<b>0.920148</b>	<b>15</b>	<b>95</b>

<sup>a</sup> Probability of N-terminal signal peptide using SignalP

<sup>b</sup> Not predicted

<sup>c</sup> Not detected

Preliminary