



**HAL**  
open science

# Extensive transcriptome analysis correlates the plasticity of *Entamoeba histolytica* pathogenesis to rapid phenotype changes depending on the environment

Christian Weber, Mikael Koutero, Marie-Agnes Dillies, Hugo Varet, Cesar Lopez-Camarillo, Jean Yves Coppee, Chung-Chau Hon, Nancy Guillén

## ► To cite this version:

Christian Weber, Mikael Koutero, Marie-Agnes Dillies, Hugo Varet, Cesar Lopez-Camarillo, et al.. Extensive transcriptome analysis correlates the plasticity of *Entamoeba histolytica* pathogenesis to rapid phenotype changes depending on the environment. *Scientific Reports*, 2016, 6 (1), pp.35852. 10.1038/srep35852 . pasteur-02557494

**HAL Id: pasteur-02557494**

**<https://pasteur.hal.science/pasteur-02557494v1>**

Submitted on 28 Apr 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# SCIENTIFIC REPORTS



OPEN

## Extensive transcriptome analysis correlates the plasticity of *Entamoeba histolytica* pathogenesis to rapid phenotype changes depending on the environment

Christian Weber<sup>1,2</sup>, Mikael Koutero<sup>3</sup>, Marie-Agnes Dillies<sup>3,4</sup>, Hugo Varet<sup>3,4</sup>, Cesar Lopez-Camarillo<sup>5</sup>, Jean Yves Coppée<sup>3</sup>, Chung-Chau Hon<sup>1,2,\*</sup> & Nancy Guillén<sup>1,2</sup>

Amoebiasis is a human infectious disease due to the amoeba parasite *Entamoeba histolytica*. The disease appears in only 20% of the infections. Diversity in phenotypes may occur within the same infectious strain in the gut; for instance, parasites can be commensal (in the intestinal lumen) or pathogenic (inside the tissue). The degree of pathogenesis of clinical isolates varies greatly. These findings raise the hypothesis that genetic derivation may account for amoebic diverse phenotypes. The main goal of this study was to analyse gene expression changes of a single virulent amoebic strain in different environmental contexts where it exhibit different degrees of virulence, namely isolated from humans and maintained through animal liver passages, in contact with the human colon and short or prolonged *in vitro* culture. The study reveals major transcriptome changes in virulent parasites upon contact with human colon explants, including genes related to sugar metabolism, cytoskeleton rearrangement, stress responses and DNA repair. Furthermore, in long-term cultured parasites, drastic changes in gene expression for proteins with functions for proteasome and tRNA activities were found. Globally we conclude that rapid changes in gene expression rather than genetic derivation can sustain the invasive phenotype of a single virulent isolate of *E. histolytica*.

The parasite *Entamoeba histolytica* is the etiological agent of amoebiasis, an infectious disease that affects the intestine and the liver of humans. Amoebiasis occurs at high incidence in large populations with limited modern sanitation systems. The infestation arises after ingestion of cysts contaminating water and food. Upon de-cystation, a vegetative cell, the trophozoite, is formed that colonizes the intestine or becomes invasive. During intestinal invasion, *E. histolytica* degrades the colonic mucosa and then interacts with the intestinal epithelium, crosses the basal lamina and disrupts the extracellular matrix (ECM). Trophozoites colonize the human gut mainly as commensals that feed on bacteria, these are able to invade the mucosa then initiating the acute phase of the disease. Intestinal amoebiasis has multiple clinical manifestations ranging from frequent mild diarrhoea to dysentery with blood and mucus in the stool. A major research interest on *E. histolytica* is driven by the observation that only one out of five infected persons develops the disease<sup>1</sup> while most reported infections are asymptomatic. The development of the disease has been associated with human factors<sup>2</sup>, environmental conditions<sup>3,4</sup> and

<sup>1</sup>Institut Pasteur, Cell Biology of Parasitism Unit, F-75015 Paris, France. <sup>2</sup>Inserm, U786, F-75015 Paris, France.

<sup>3</sup>Institut Pasteur, Transcriptome and EpiGenome, BioMics, Center for Innovation and Technological Research, F-75015, Paris, France. <sup>4</sup>Institut Pasteur, Hub Bioinformatique et Biostatistique – Centre de Bioinformatique, Biostatistique et Biologie Intégrative (C3BI, USR 3756 IP CNRS) – F-75015 Paris, France. <sup>5</sup>Universidad Autonoma de la Ciudad de Mexico, Genomics Sciences Program, Mexico City, Mexico. \*Present address: RIKEN Center for Life Science Technologies, Yokohama Institute, Division of Genomic Technology, Yokohama, 230-0045, Japan. Correspondence and requests for materials should be addressed to N.G. (email: nancy.guillen@pasteur.fr)

especially with trophozoite virulence<sup>5</sup>. The amoebic invasive process can then induce an inflammatory response accompanied by disease-relevant cell death (for a review<sup>6</sup>). In some cases, subsequent to intestinal infection, or by direct acquisition through the blood, the parasite disseminates via the circulatory system, reaches the liver and induces hepatic abscesses that can be fatal for the infected person. Circulating trophozoites are confronted with the host's innate immune response, in particular with cytotoxic compounds, as well as with oxygen tensions higher than those usually found in the gut. In addition, trophozoites are exposed to other host defence mechanisms – they are notably attacked by reactive oxygen and nitrogen intermediates produced by phagocyte cells, these compounds are highly concentrated in the infected tissue. For example, it has been shown experimentally that trophozoites invading the liver are highly sensitive to blood complement and tissue responses<sup>7,8</sup>. The diversity of human compartments that *E. histolytica* crosses during the infection combined with the variability of symptoms induced, suggest that this parasite expresses distinct transcriptional programs although the key components regulating gene expression remain to be described.

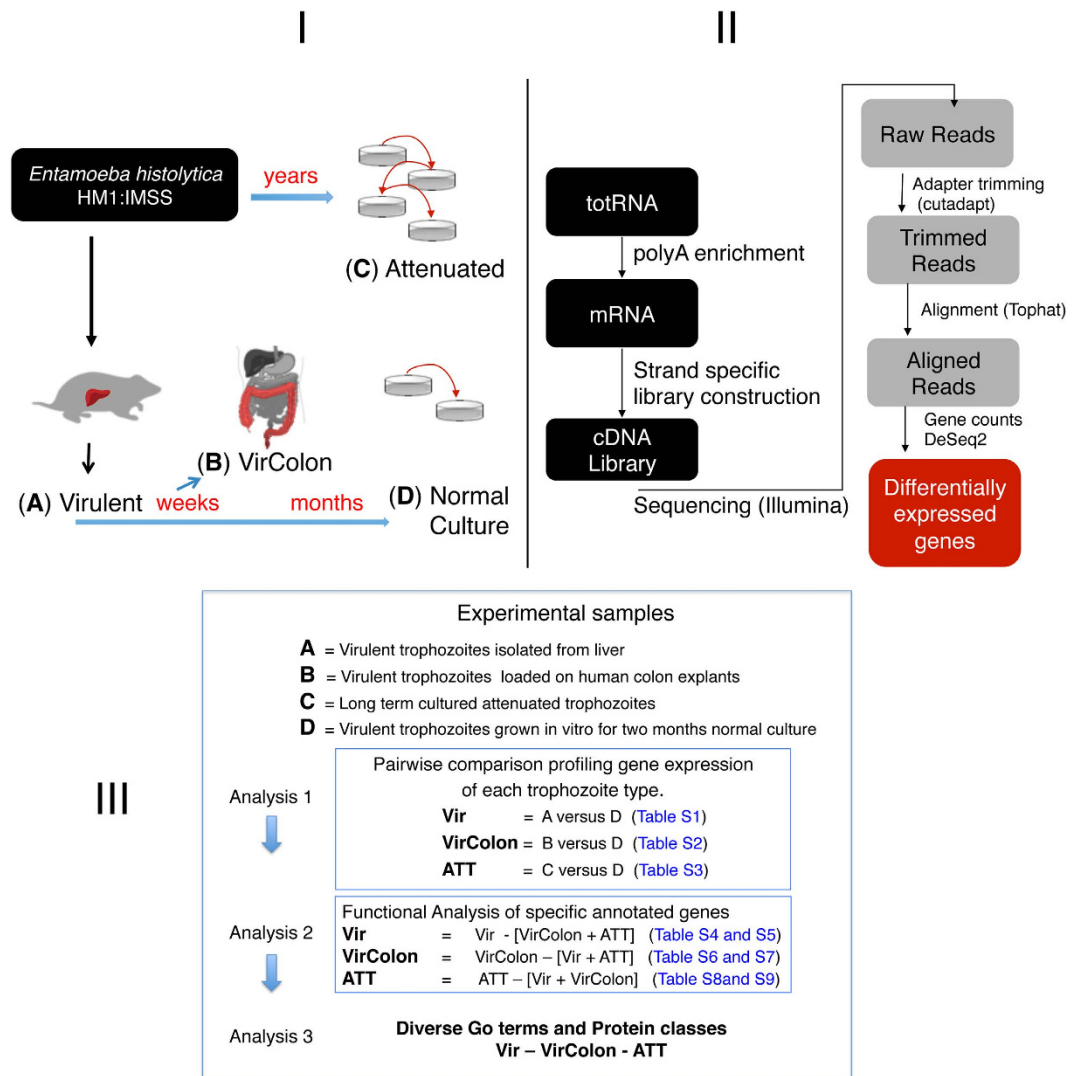
Following the sequencing and annotation of the *E. histolytica* genome<sup>9</sup>, microarray approaches have been developed to analyse gene expression in this organism. The data obtained revealed important regulatory networks involved in strain phenotype differences<sup>10,11,12</sup>, colonic and hepatic invasion<sup>12–15</sup>, responses to stress<sup>16,17</sup>, drug treatments<sup>18,19</sup>, metabolic changes<sup>20–23</sup> and the encystation process<sup>24</sup>. Overall, these studies have identified diverse pathogenic factors, new pathways for drug treatment and some elements of transcriptional gene regulation. Thus, transcriptome approaches opened solid avenues to investigate the mechanism supporting invasion and pathogenicity of *E. histolytica*. Since this parasite only targets humans and the gut is the main reservoir for the cyst form, it is difficult to experimentally study infections in the context of amoebic human natural host aiming to understand the molecular basis of amoebiasis. Human colon explants have recently been used as an alternative system, and experiments have provided new knowledge to understand the pathophysiological bases of intestinal invasion<sup>14,25–27</sup>. For instance, the role of amoebic proteinases and human metalloproteinases<sup>27</sup> in the degradation of the ECM and activation of immune responses has been outlined. The comparison of the gene expression profiles of parasites isolated from an asymptomatic carrier (i.e. the non-virulent Rahman strain) and from a symptomatic patient (i.e. the virulent HM1:IMSS strain) in contact with human colon explants has provided a global view of gene expression changes<sup>11</sup>. One remarkable feature of the virulent phenotype resides in the up-regulation of genes implicated in carbohydrate metabolism and processing of glycosylated residues of compounds building the intestinal barrier. However, it can be argued that trophozoites from the isolates used in these experiments have been in culture for decades and may harbour some genetic deviations influencing their virulence. In conjunction with the expression of genes encoding the so-called virulence factors, the infectious process also requires the expression of genes whose products are important for adaptation to different host responses and to counteract environmental factors of the invaded tissues.

In this work we hypothesized that for a given *E. histolytica* strain, changes in the environment (in particular those occurring during the interaction with the human host) trigger the major gene expression marks related to different degrees of virulence. To test this hypothesis, we applied high-throughput RNA sequencing to profile, for the first time, samples from the same isolate (HM1:IMSS) in contact with different environments. We harvested trophozoites from liver abscess experimentally induced in the hamster animal model, a condition enhancing amoebic virulence, and from these virulent parasites loaded in contact with human colon explants, reflecting the intestinal invasive process. Searching for factors that might be implicated in the aggravation of the disease, we first defined a set of genes specifically expressed in virulent trophozoites isolated from the liver that we further refined by studying trophozoites interacting with the human colon. In addition to stress responses and DNA repair, the expression of important cytoskeleton-related factors was the mark of cultured virulent parasites. Remarkably, virulent parasites change their profile upon interaction with the human colon by increasing the expression of genes related to metabolism of carbohydrates and glycosylated residues. The profiles established for virulent trophozoites were compared to those of long-term *in vitro* cultured parasites of the same strain that have acquired a virulence-attenuated phenotype<sup>7</sup>. The marks associated with virulence and intestinal invasion are lost in virulence-attenuated amoebae, which appeared highly active in proteasome activities and repressed in tRNA transfer enzymes. Based on our results we concluded that rapid and highly specific changes in the amoebic transcriptome, rather than genetic variation of the amoeba isolate, can account for the invasion of human colon by virulent *E. histolytica*.

## Results and Discussion

### Characterization of the transcriptome of *Entamoeba histolytica* with different degrees of virulence.

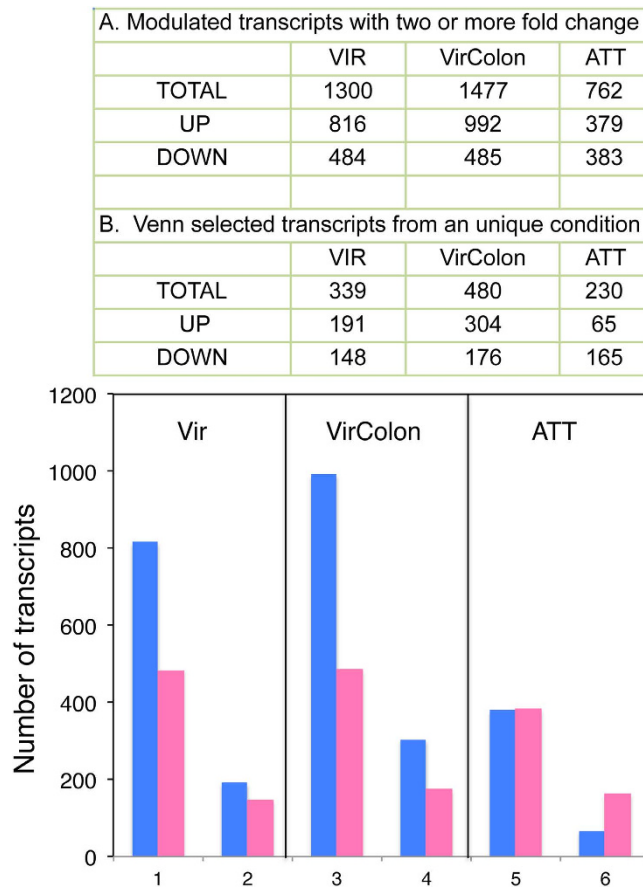
To analyse transcriptional changes in trophozoites of the same strain undergoing changes in virulence we used RNAseq. The RNA libraries were constructed from polyA + mRNA from the HM1:IMSS strain grown in four different conditions (Fig. 1A) Virulent parasites extracted from liver abscesses of infected hamsters (Vir), B) Virulent parasites exposed to explants of human colon (VirColon), C) long-term cultured virulence-attenuated amoeba (ATT) that have lost the ability to form liver abscesses and D) trophozoites short-time cultured (Normal Culture) for roughly less than three months (24 passages). Three replicates of each condition were performed resulting in 12 RNA libraries in total that were sequenced on an Illumina HiSeq 2000 platform (see Material and Methods). Data were processed, using bioinformatical and statistical workflows as already published<sup>28</sup>. Coding gene models were based on the bona fide gene models defined in previous work ( $n = 7312$ )<sup>28</sup>. We determined that in all the diverse conditions almost 95% of the annotated coding regions had at least one read indicating that our dataset is deep enough to analyse the majority of annotated transcripts. Differentially expressed genes were defined as genes with  $\leq 5\%$  false discovery rate and  $\geq 2$ -fold or  $\leq 0.5$ -fold changed expression rates following the statistical analysis performed with DESeq2 as described in the Materials and Methods section (Supplementary Tables S1 to S3).



**Figure 1. Workflow for the experimental procedures.** I. The amoeba strain HM1:IMSS was isolated from the liver of hamster and then cultivated in several conditions: (A) *in vitro* overnight (Vir condition), (B) *in vitro* for two weeks and then deposited on the top of human colon explants (Vir Colon condition) or (D) *in vitro* for several weeks (Normal condition). Long-term cultured parasites ATT were also tested (C). II. RNAs were organically isolated, polyA-enriched and directional libraries were constructed and sequenced. The raw reads were filtered for quality and sequence contaminants and aligned to the reference genome (TopHat) to finally compute differential expression levels (DeSeq2). III. Paths followed for the bioinformatics analysis of the data obtained by RNASeq.

**Pairwise comparison of gene expression profiles.** To characterize the differences in gene expression of the HM1:IMSS strain in the four conditions tested; we used the virulent parasites grown for a limited period in culture medium (referred as D) as the reference. Pairwise comparisons were performed with the three growth conditions, A, B and C. (Fig. 1 and Supplementary Table S2). The comparative analysis identified the differentially expressed genes (Fig. 2), with 1300 genes in the condition A (Vir) (816 up- and 484 down-regulated); 1477 genes (992 up- and 485 down-regulated) in condition B (VirColon) and 762 genes (379 up- and 383 down-regulated) in condition C (ATT) with respect to the reference condition (D).

**Expression profile of virulent parasites isolated from the liver.** The A versus D comparison highlights genes that are modulated upon extraction of parasite from the liver tissue and overnight culture. Among the 816 up-regulated genes, 381 had been previously annotated whereas 435 were of unknown function. The data (Supplementary Tables S4 to S9) showed important changes in genes related to virulence, and in particular genes involved in stress responses, as previously suggested from microarray data<sup>14–16</sup>. Genes encoding several so-called heat shock proteins were up-regulated, as well as genes for proteasome and protein degradation pathway components described as over expressed in stress conditions. We observed for the first time important changes affecting genes related to the cytoskeleton. Among the 214 annotated down-regulated genes (Supplementary Table S5),

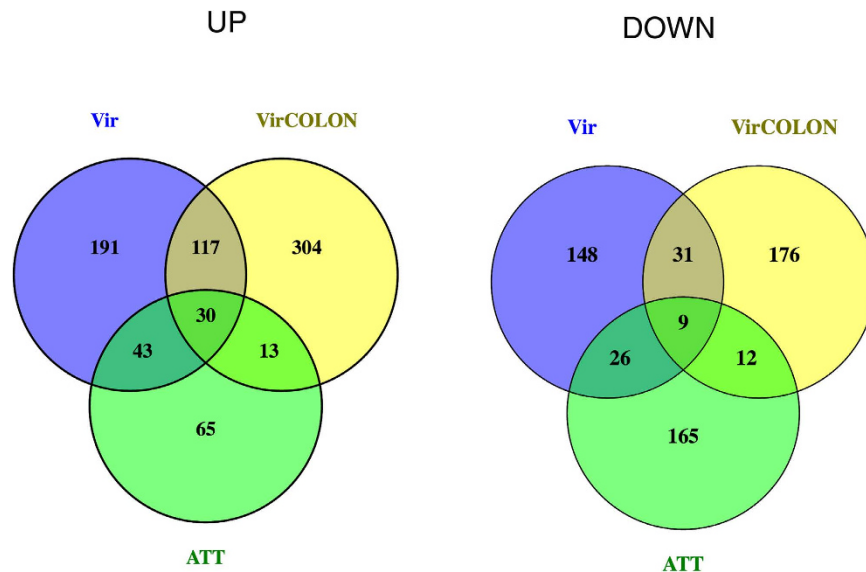


**Figure 2. Analysis of the differentially expressed genes.** After mapping of short reads on the amoebic genome, the number of modulated genes was determined for each condition (A) in comparison to genes from the normal culture condition. Comparing only the annotated genes in a Venn diagram the number of gene modulated unique to each condition (B) was determined. The figure in the bottom summarizes the different values: those represented in blue indicating the UP regulated genes and those in pink correspond to down regulated one. Two bars appears per condition and these correspond firstly (left bar) to all transcript modulated by condition and secondly (right bar), to unique transcripts modulated in each condition.

several members of the AIG family (GTPase immunity-associated protein family) encoding genes are the most differentially expressed.

**Expression profile of virulent parasites interacting with human colon explants.** For virulent trophozoites in contact with human colon explants, a well-defined gene expression pattern indicated the ability of *E. histolytica* to trigger rapid expression changes during early stages of colon invasion (comparison B versus D). Among the 994, up-regulated genes, 464 were already annotated and included genes involved in carbohydrate metabolism and protein glycosylation, AIG factors, the Gal-GalNAc lectin (heavy and intermediate subunits) and competence EC (Supplementary Table S6). We compared the file corresponding to annotated up-regulated genes in contact with the human colon with our previously published data (Supplementary Table S10) obtained by microarray experiments<sup>12</sup> and confirmed that incubation with the colon triggers a unique gene expression profile that is qualitatively different from the one of parasites from the same isolate freshly isolated from the liver (B/D versus A/D), or from long-term cultured parasites (B/D versus C/D). In particular, regulators of nonsense transcripts and alpha- or beta-amylase encoding genes were found, the latter being an essential protein in the pathogenic process of *E. histolytica*<sup>12</sup>. Compared to microarrays, RNASeq appears highly sensitive allowing for a deeper analysis of the stress response (see below). Among the 485 down-regulated genes in the B vs. D comparison (Supplementary Table S7), the 228 previously annotated genes provided evidence that the transcription of some genes involved in amoeba pathogenicity was diminished when amoebae were in contact with the human colon, notably the genes coding for the Gal-GalNAc lectin light chains and the lysine- and glutamic acid-rich protein 1 (KERP1).

**Expression profile of virulence-attenuated parasites.** Attenuated parasites (ATT) were derived from the virulent HM1:IMSS strain by prolonged axenic culture<sup>7</sup>. Among the 379 up-regulated genes (151 annotated) we identified (Supplementary Table S8) several genes belonging to the *ariel* multicopy gene family which is specific to *E. histolytica* species among the *Entamoeba* genus. Although this family was identified 15 years



**Figure 3. Venn diagram of the differentially expressed genes.** The number in each circle represents the amount of differentially expressed genes between the different comparisons (test versus control). Only the annotated genes were considered. The overlapping number stands for the mutual differentially expressed genes between the different comparisons and the non overlapping numbers specify the genes unique to each condition: virulent (Vir), Virulent on the colon (Vir Colon) and long term cultured ATT (attenuated). Up and Down regulated genes were determined.

ago, no function has been attributed to the ARIEL protein so far<sup>29</sup>. Genes encoding Hsp20 -Hsp70 or START domain-containing proteins were up-regulated. All these gene categories have been previously identified by microarray experiments as being up regulated in strains with low virulence, such as the Rahman strain<sup>12</sup>. Among the 212 annotated and down-regulated genes of ATT trophozoites (Supplementary Table S9), we identified genes encoding proteins associated with the parasite surface, including the immuno-dominant variable antigen M17<sup>30</sup>, the BspA gene family<sup>31</sup> or the serine-threonine-isoleucine-rich protein<sup>32</sup> as well as genes encoding numerous signalling molecules such as small GTPases and protein kinases<sup>33</sup>.

**Functional analysis of modulated genes.** The PANTHER classification system, searching for GO terms, allowed the functional interpretation of RNASeq datasets (Up or Down genes)<sup>34</sup>. The *E. histolytica* genome was only recently added to the PANTHER database and it is important to note that not all genome data entries are associated with a hit in PANTHER protein class, family or pathway classifications, due to the incomplete genome annotation.

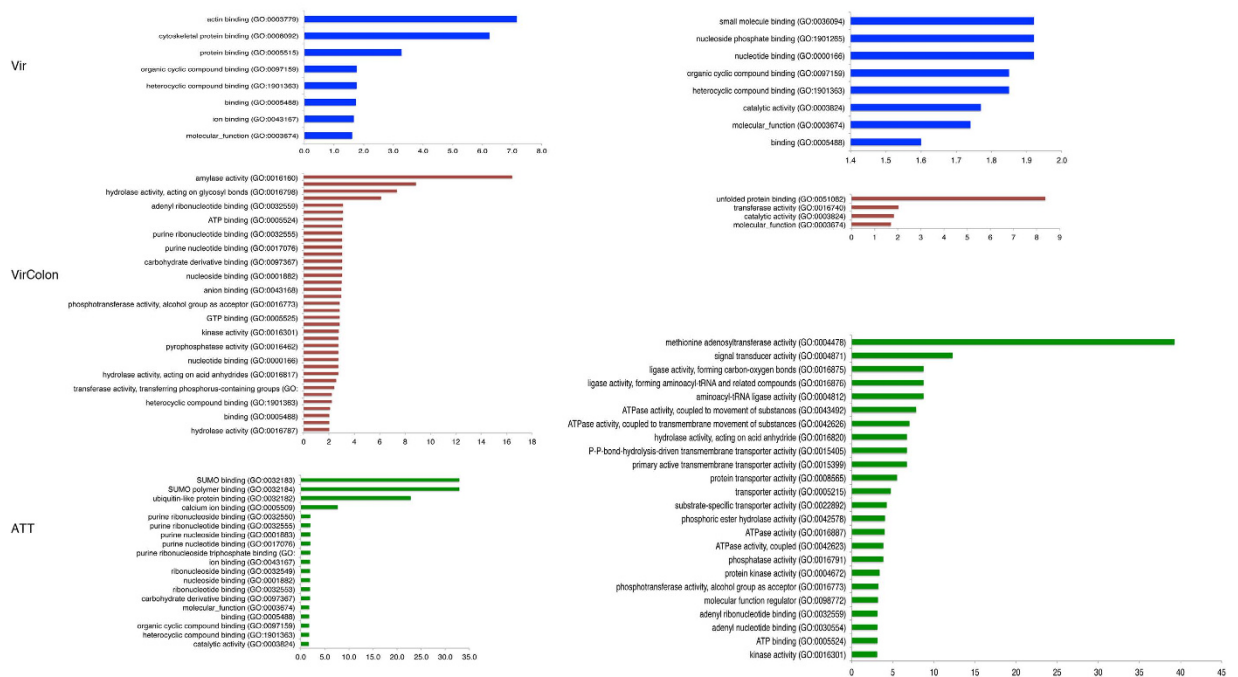
For functional predictions, we first used Venn diagrams to identify genes whose expression regulation was specific to one of the three conditions, as well as genes whose expression regulation was observed in two or three conditions (Fig. 3). The latter were removed and were not further analysed. We found 304, 191 and 65 genes that were up-regulated in conditions B (VirColon, amoebae in contact with colon), A (Vir, virulent amoebae from liver) and C (ATT, long-term cultured, virulence-attenuated amoebae) respectively. Down-regulated genes were analysed similarly revealing 176, 148 and 165 genes specifically repressed in VirColon, Vir and ATT respectively.

Following the Venn selections, an enrichment analysis was performed to identify GO categories that were over- or under-represented among the list of the specific differentially expressed genes. GO terms were sorted into the different subcategories for molecular function (MF) (Fig. 4) and biological processes (BP) (Fig. 5). Diverse levels of enrichment were found in the two main GO aspects (MF, BP) for the three datasets (A/D, B/D and C/D). For the induced genes (Table 1), the Vir condition (191 genes) showed a significant enrichment in MF for actin binding (GO:0003779) and cytoskeletal protein binding (GO:0008092) categories with 7 and 6 fold enrichments respectively. In BP we found the ribonucleoprotein complex biogenesis (GO:0022613) and with 4 fold enrichment, RNA metabolic process (GO:0016070) with 2.82 fold, while 65 genes were not classified.

In MF the GO terms analysis for up-regulated genes in the VirColon condition (304 genes) highlighted amylase activity (GO:0016160) with 16 fold enrichment, hydrolase activity hydrolysing O-glycosyl compounds (GO:0004553) with 9 fold, and hydrolase activity acting on glycosyl bonds (GO:0016798) with 7 fold enrichment, while 65 genes were not classified. For the biological processes, carbohydrate metabolic process (GO:0005975) was enriched 3 fold, while 83 genes were not classified.

Finally, the GO terms analysis for MF in the ATT condition revealed the SUMO binding (GO:0032183), SUMO polymer binding (GO:0032184) with both 33 fold enrichment and the ubiquitin-like protein binding (GO:0032182) with 23 fold enrichment. The BP found as the most significant were the cellular protein catabolic process (GO:0044257), proteolysis involved in cellular protein catabolic process (GO:0051603) and protein catabolic process (GO:0030163) all with 4 fold enrichment. SUMO is the Small Ubiquitin-related Modifier protein

## MOLECULAR FUNCTION



**Figure 4. Distribution of GO functional classifications.** Upon analysis with Panther tools<sup>34</sup>, the distribution of the fold enrichment levels of terms for diverse molecular functions categories for Vir, VirColon and ATT (Up and Down modulated genes) is represented.

that is important for gene transcription regulation<sup>35</sup>. SUMO modifies many proteins including promoter-specific transcription factors, cofactors and chromatin-modifying enzymes. A total of 14 genes were unclassified.

Concerning the GO term analysis of down-regulated genes (Figs 4 and 5) in the Vir condition, lipid metabolism (GO:0006629) was the most relevant with 6 fold enrichment whereas 47 genes were unclassified.

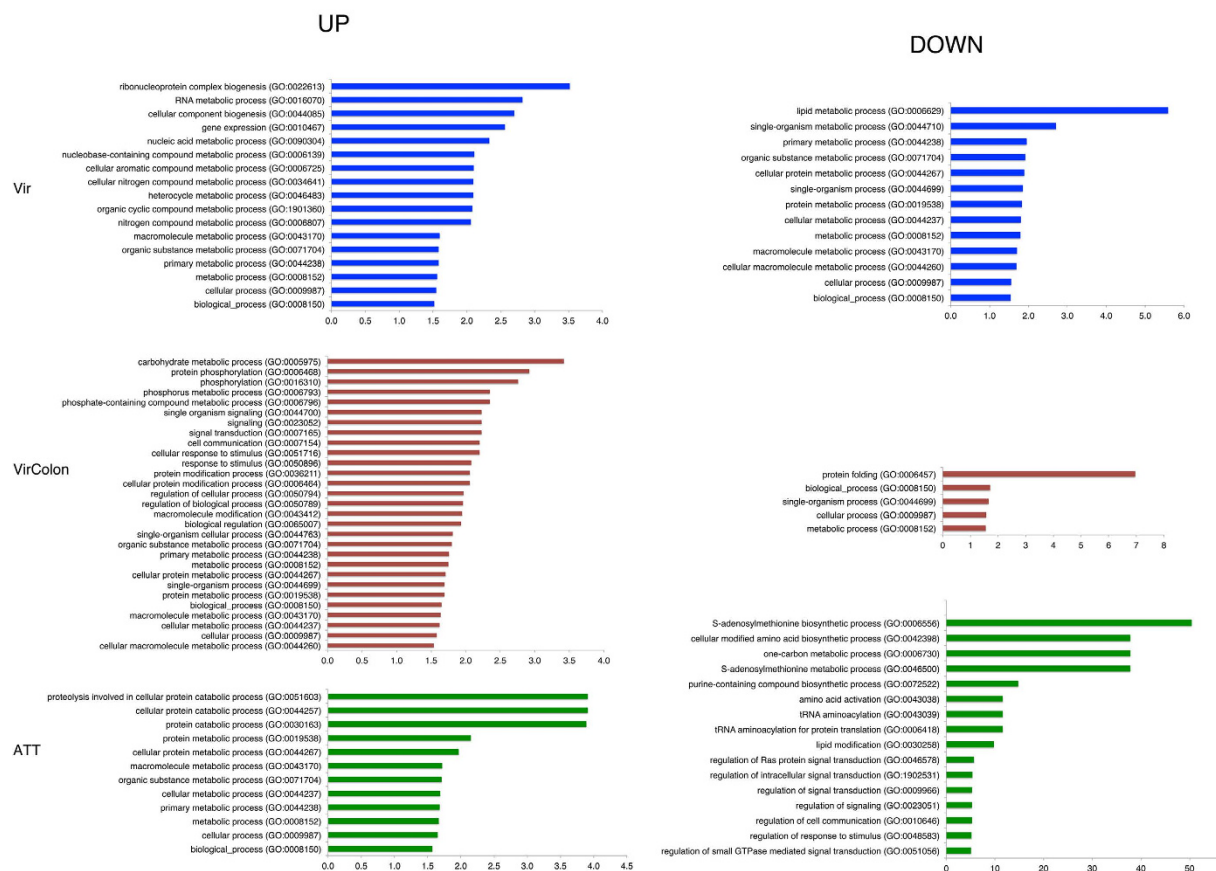
The most relevant GO term found for genes down regulated in the VirColon condition in BP was the protein-folding category (GO:0006457) with 7 fold. In MF the unfolded protein binding (GO:0051082) was enriched 8 fold. 43 genes were unclassified.

Finally numerous GO term enrichments were identified in the ATT condition indicating a high impact in down-regulation of genes for this amoebic strain. In BP the following categories were significantly enriched: S-adenosylmethionine biosynthetic process (GO:0006556) with 35 fold, purine-containing compound biosynthetic process (GO:0072522) with 14 fold enrichment, tRNA aminoacylation for protein translation (GO:0006418) with 9 fold enrichment, tRNA aminoacylation (GO:0043039) with 9 fold enrichment, amino acid activation (GO:0043038) with 9, fold enrichment lipid modification (GO:0030258) with 9 fold enrichment and regulation of Ras protein signal transduction (GO:0046578) with 5 fold enrichment and several others with changes less than 5-fold. Only 38 genes were unclassified. Many molecular functions were associated with these GO terms (Figs 4 and 5): methionine adenosyltransferase activity (GO:0004478) with 39 fold, signal transducer activity (GO:0004871) with 12 fold and aminoacyl-tRNA ligase activity (GO:0004812), ligase activity, forming aminoacyl-tRNA and related compounds (GO:0016876), ligase activity, forming carbon-oxygen bonds (GO:0016875), all with 9 fold enrichment.

Overall the GO term analysis determined the high specificity of transcriptome patterns for each of the three conditions tested.

**Protein classes highly represented in the different growth conditions.** To generate a more focused view of the GO terms, we looked at protein class enrichments between the PANTHER categories. The software plots the Log of fractional difference (observed versus expected), which is calculated for each testing list as: (number of genes for the category - number of genes expected)/number of genes expected (Fig. 6). For the genes up regulated in the Vir condition, we observed a significant enrichment for Heat Shock proteins where Hsp70A, Hsp70 and Hsp90 appearing as the most overrepresented (Fig. 6). For the first time we revealed the association of genes involved in RNA metabolism and amoebic virulence. This includes the DNA-RNA polymerase (EHI\_122780, EHI\_125350), mRNA splicing factors (EHI\_049370, EHI\_183900, EHI\_060350, polyadenylation factors (EHI\_067580)) as well as diverse components of the mRNA transduction machinery (Supplementary Table S11).

## BIOLOGICAL PROCESS



**Figure 5. Distribution of GO functional classifications.** Upon analysis with Panther tools<sup>34</sup>, the distribution of the fold enrichment levels of terms for diverse biological process categories for Vir, VirColon and ATT (Up and Down modulated genes) is represented.

In the up-regulated genes the actin-rich cytoskeleton is clearly a mark for virulence in the context of liver isolated parasites, (Supplementary Table S12) showing over 32 proteins identified including actin, actin binders and cytoskeleton constituents as the most significant. These proteins belong to the actin-binding proteins category with filament crosslinking activities able to organize microfilaments in orthogonal and/or parallel networks<sup>36</sup>, including: filamin (EHI\_104630, EHI\_167130), villin (EHI\_007480), gelsolin (EHI\_009570), filopodin (EHI\_080740), cortexillin (EHI\_103430), Diaphanous-Formin (EHI\_190990, EHI\_192460, EHI\_118260) Paxillin (EHI\_069060, EHI\_022960) as well as actin binders including coronin (EHI\_083590, EHI\_122800), Myosin II (EHI\_110180), cofilin (EHI\_186840) and a recently identified Arp related protein Arpv2 (EHI\_048630) unique to *E. histolytica*<sup>33</sup>.

Amylases, DNA repair and primary metabolism are the mark for VirColon. Several alpha-amylases (EHI\_153100, EHI\_152880, EHI\_023360, EHI\_129810) and beta-amylase (EHI\_153590, EHI\_009020, EHI\_049700, EHI\_058340, EHI\_148800, EHI\_035700) appeared represented. Previously<sup>12</sup> the beta-amylase EHI\_192590 allele showed (in microarray experiments) a fold change up to 25 in the comparison between *E. histolytica* HM1:IMSS strain (virulent) and *E. histolytica* Rhaman strain (non-virulent) cultured *in vitro* or during their interaction with human colon explants. Here expression of the EHI\_192590 allele didn't show any change probably because we are comparing only an unique virulent strain in different conditions; instead significant overexpression of four copies of beta-amylase encoding gene was observed (Supplementary Table S13) strongly sustaining our conclusion on the role of beta-amylase in the infectious process.

Several genes encoding proteins with helicase activities involved in DNA repair and genome stability maintenance were up-regulated and overrepresented in trophozoites in contact with host colon (Supplementary Table S13). The DNA damage response represents a survival mechanism that confers cellular protection against stresses induced by DNA-damaging agents, several classes of antibiotics, and oxidative agents that may attack amoebae during colon invasion. Two genes potentially involved in nucleotide excision repair (NER) of DNA, namely ERCC2 (EHI\_054240) and ERCC3 (EHI\_088430), were significantly overexpressed. Both ERCC2 and ERCC3 are DNA helicase component of the core-TFIIH basal transcription factor involved in NER and RNA transcription by RNA polymerase II in human cells<sup>37</sup>. We also found an increased expression of a gene (EHI\_187240) with high homology to WRNIP1, a Werner DNA helicase interacting protein 1, which functions as a modulator for initiation events during DNA polymerase delta-mediated DNA synthesis and sensor of DNA damage<sup>38</sup>. In addition,



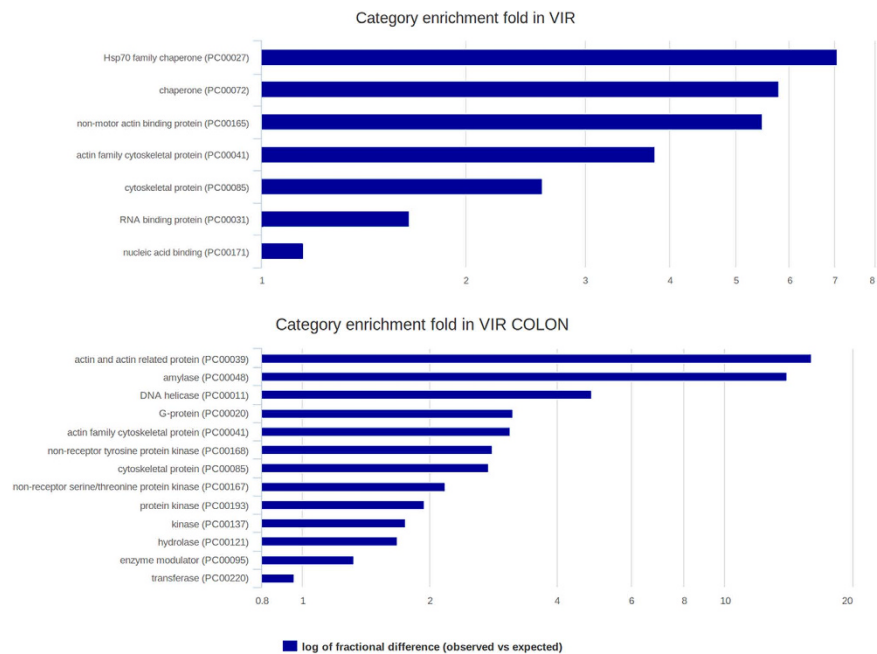
	Fold	P-Value
<b>Molecular function</b>		
Vir		
actin binding (GO:0003779)	7	7.81E-04
cytoskeletal protein binding (GO:0008092)	6	2.59E-03
protein binding (GO:0005515)	3	5.26E-04
Vir Colon		
hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553)	9	3.18E-05
hydrolase activity, acting on glycosyl bonds (GO:0016798)	7	1.99E-04
GTPase activity (GO:0003924)	6	2.75E-02
ATP binding (GO:0005524)	3	9.05E-18
adenyl nucleotide binding (GO:0030554)	3	9.05E-18
adenyl ribonucleotide binding (GO:0032559)	3	9.05E-18
carbohydrate derivative binding (GO:0097367)	3	3.88E-26
purine ribonucleoside triphosphate binding (GO:0035639)	3	1.54E-25
purine nucleotide binding (GO:0017076)	3	1.54E-25
purine nucleoside binding (GO:0001883)	3	1.54E-25
purine ribonucleotide binding (GO:0032555)	3	1.54E-25
purine ribonucleoside binding (GO:0032550)	3	1.54E-25
nucleoside binding (GO:0001882)	3	1.98E-25
ribonucleoside binding (GO:0032549)	3	1.98E-25
ribonucleotide binding (GO:0032553)	3	3.88E-25
ATT		
SUMO polymer binding (GO:0032184)	33	4.94E-02
SUMO binding (GO:0032183)	33	4.94E-02
ubiquitin-like protein binding (GO:0032182)	23	1.40E-03
calcium ion binding (GO:0005509)	8	1.81E-02
<b>Biological process</b>		
Vir		
ribonucleoprotein complex biogenesis (GO:0022613)	4	2.29E-03
Vir colon		
carbohydrate metabolic process (GO:0005975)	3	2.27E-02
ATT		
cellular protein catabolic process (GO:0044257)	4	3.48E-02
proteolysis involved in protein catabolic process (GO:0051603)	4	3.48E-02
protein catabolic process (GO:0030163)	4	3.68E-02

**Table 1. GO terms for up regulated genes.** Were only analysed the annotated genes: Vir condition 191 genes, Vir Colon 304 genes and ATT 65 genes. The table present only the categories enriched 3 fold or more.

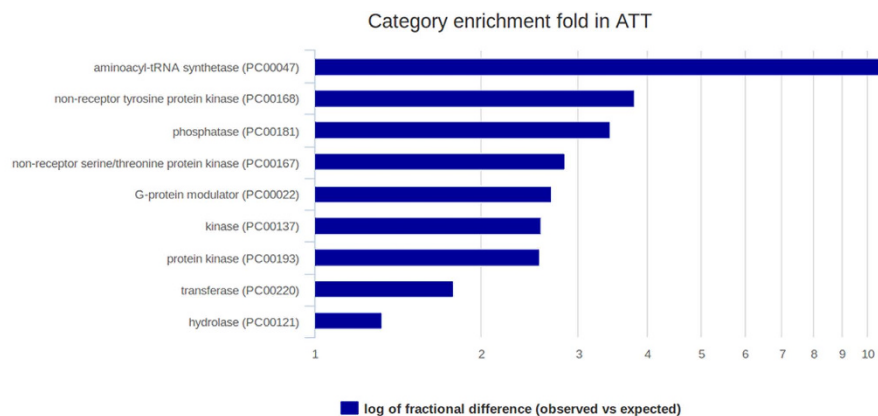
two genes homologous to genes coding for DNA replication licensing factors, MCM2 (EHI\_117970) and MCM7 (EHI\_158110), involved in DNA replication and cell division in higher eukaryotes<sup>39</sup> were also up-regulated. Intriguingly, two genes (EHI\_178520, EHI\_035550) sharing homologies with regulator of nonsense transcripts 1 (UPF1) were up-regulated in virulent trophozoites. In human cells, UPF proteins have RNA-dependent helicase and ATPase activities required for nonsense-mediated decay (NMD) of mRNAs containing premature stop codons<sup>40</sup>. Notably, NMD triggered by mRNA-mRNA interactions have been linked to cell migration and invasion of cancer cells<sup>41</sup>.

The strikingly strong enrichment found among the up-regulated categories for ATT trophozoites concerned genes encoding E3 ubiquitin ligase (EHI\_142060, EHI\_158200, EHI\_098590), a protein transferring ubiquitin to lysine residues of a protein to target it to the proteasome for degradation (Supplementary Table S14). The U-ligases found here contain a Zinc finger domain (RING domain). Ubiquitin hydrolases (EHI\_064460), Ubiquitin conjugating enzyme (EHI\_100110), several cysteine proteases and proteasome subunits were identified (Supplementary Table S14). Among the down-regulated genes for the conditions Vir and VirColon, no significant overrepresentation of protein classes was found. In contrast, for the ATT condition numerous signalling molecules including small GTPases (PC 00022), serine/tyrosine protein kinase (PC00167) and aminoacyl-tRNA synthetase (PC00047) with glycyl, methionyl, isoleucyl, tryptophanyl and histidyl tRNA synthetases were identified (Fig. 7). These are enzymes that link the appropriate amino acid onto its cognate tRNA to allow mRNA translation into proteins.

Finally, in order to try to identify common potential regulatory elements enhancing gene expression we performed a gene clustering analysis for the UP regulated genes. This confirmed the major findings above described for the diverse treatments applied to *E. histolytica* HM1:IMSS (Supplementary Fig. 2). The data provide gene



**Figure 6. Enrichment of protein classes for Vir and VirColon UP regulated genes.** The log of fractional difference between observed versus expected is plotted upon identification of protein classes representatives of Vir and Vir colon whose genes are up regulated.



Attenuated amoeba. Protein class overrepresented among the DOWN regulated genes						
UniProt	Gene ID	Gene name	PANTHER FAMILY		Protein class	
C4LXL0	EHI_073460	Glycyl-tRNA synthetase, putative;EHI_073460;ortholog	GLYCINE--TRNA LIGASE (PTHR10745:SF0)		aminoacyl-tRNA synthetase(PC00142)	
C4M2K7	EHI_079860	Methionyl-tRNA synthetase, putative;EHI_079860;ortholog	METHIONINE--TRNA LIGASE, CYTOPLASMIC (PTHR11946:SF86)		aminoacyl-tRNA synthetase(PC00142)	
C4LY19	EHI_177660	Isoleucyl-tRNA synthetase, putative;EHI_177660;ortholog	ISOLEUCINE--TRNA LIGASE, CYTOPLASMIC (PTHR11946:SF73)		aminoacyl-tRNA synthetase(PC00142)	
C4M7U9	EHI_020240	Tryptophanyl-tRNA synthetase, putative;EHI_020240;ortholog	TRYPTOPHAN--TRNA LIGASE, CYTOPLASMIC (PTHR10055:SF1)		aminoacyl-tRNA synthetase(PC00142)	
C4LU94	EHI_110420	Tryptophanyl-tRNA synthetase, putative;EHI_110420;ortholog	TRYPTOPHAN--TRNA LIGASE, CYTOPLASMIC (PTHR10055:SF1)		aminoacyl-tRNA synthetase(PC00142)	
C4M6S3	EHI_180020	Histidyl-tRNA synthetase, putative;EHI_180020;ortholog	HISTIDYL--TRNA SYNTHETASE, ISOFORM D (PTHR11476:SF7)		RNA binding protein(PC00171);aminoacyl-tRNA synthetase(PC00031)	

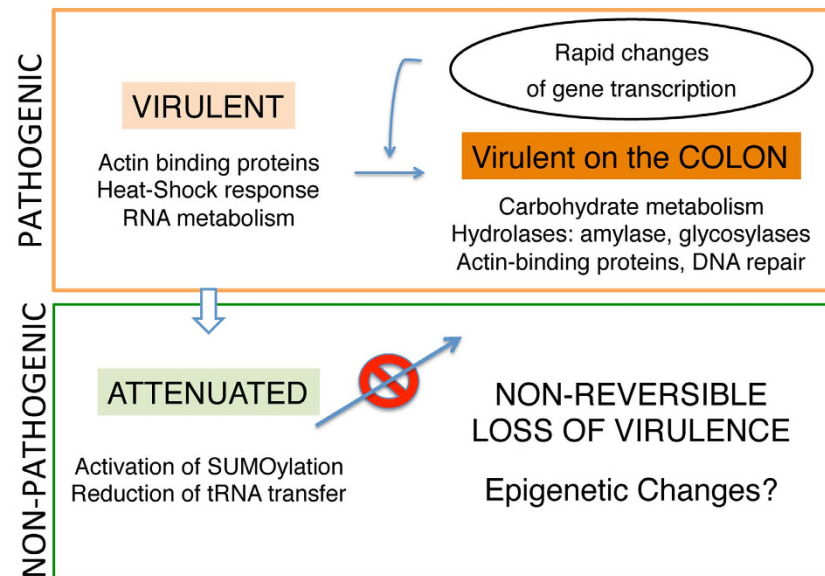
**Figure 7. Enrichment of protein classes for ATT down regulated genes.** The log of fractional difference between observed versus expected is plotted upon identification of protein classes representatives of ATT condition whose genes are down regulated.

identities useful for further bioinformatics and experimental analysis of 5' or 3' elements important for gene expression regulation in each condition.

### General Conclusion

In the present study, we used extensive comparative transcriptome analysis to determine the gene expression profiles of *E. histolytica* associated with virulence in the context of intestinal and liver infection. According to the read coverage obtained in the RNASeq approach we conclude that globally the entire genome appears to be transcribed in all tested conditions. The complete analysis of gene expression of highly virulent trophozoites freshly isolated from liver abscess and grown for only 16 hours *in vitro* identifies for the first time important features

## Specific gene expression changes of *E. histolytica* (HM1:IMSS) associated with different degrees of virulence



**Figure 8. Summary and conclusions.** *E. histolytica* HM1:IMSS virulent strain rapidly adapts its transcriptome according to changes of the environment. To trigger the pathogenic process, the parasite nourishes upon degradation of the mucus layer, overcomes the various tissue stresses and moves into the deep layers of the intestine, all these functions are rapidly activated upon contact with the intestine. Trophozoites adapted to long-term culture are not able to trigger these reactions and instead, protein degradation is highly active as well as the reduction of amino acids supply to the ribosome for protein synthesis.

implicated in this phenotype including actin cytoskeleton, heat shock response and RNA metabolism. Changes in stress responses have already been observed in precedent work but the analysis presented here is clearly more exhaustive. The analysis of one amoebic strain in diverse environmental conditions influencing pathogenicity revealed that modulation of the amoebic transcriptome is capital for parasite adaptation allowing survival, growth and invasive behaviour. We thus discovered the genes that rapidly responded to the contact of this parasite with human colon. The patterns of differentially expressed genes of amoebae originating from the same strain and challenged with different environmental conditions closely correlates to the gene responses identified during the comparison of two particular isolates<sup>12</sup>: one from an asymptomatic carrier (*E. histolytica* Rhaman strain) and the other from a symptomatic patient (*E. histolytica* HM1:IMSS) and allows to conclude that only the contact of virulent trophozoites with the human colon triggers an unique gene transcription profile. The genes identified as significantly modulated in the process of human colon invasion confirm that metabolic processes including the activity of glycosylases, together with cytoskeleton and DNA repair activities, are the most important mechanisms underlying amoebic intestine invasion (Fig. 8). For the first time, we also highlighted two important phenotypic features of long-term cultured virulence-attenuated trophozoites, namely the activation of SUMO activity and the down-regulation of tRNA synthetases, suggesting that in these trophozoites the increase of proteasome activities and the down-regulation of the translational machinery may be involved in gene regulation, which is different from the virulent parasites.

The major conclusion of this work is that a rapid shift in gene expression is the most important phenomenon driving amoebic pathogenicity during its interaction with the human colon. In our working model, we consider several factors that may explain the shift of the intestinal amoebic phenotype. First, a change in the composition of the bacterial flora which should take place at any time. For instance, it is known that co-infection in the presence of some Enterobacteriaceae increases the virulence of the parasite<sup>42</sup>. Another plausible factor is the human immune response because the amoebae are chemoattracted by pro-inflammatory molecules such as tumor necrosis factor<sup>43</sup> and some intestinal resident cells behaving like a TNF reservoir (eg. mast cells) potentially activated at an early stage of infection. In a general view, thank to the impact of this work, the hypothesis that genetic variability sustains differences in pathogenicity<sup>44</sup> gain in precision and seems highly related to the lost of invasive ability of commensal parasites, a fact opening new avenues to investigate whether epigenetic mechanisms may sustain such commensal behaviour. Our findings provide the first insight into the molecular basis of gene expression variation of pathogenic *E. histolytica* according to the environmental conditions encountered during invasive infection.

### Material and Methods

**In vitro culture of *E. histolytica*.** *Entamoeba histolytica* HM1:IMSS is a virulent strain isolated in 1967 from a colonic biopsy of rectal ulcer from adult human male with amoebic dysentery, Mexico City, Mexico. The

HM1-IMSS was deposited in the American strain collection (ATCCH 30459TM) and it is a gift of Professor Ruy Perez Tamayo and Dr Alfonso Olivos (UNAM, Mexico). To maintain virulence, the HM1-IMSS strain has been passed periodically through the liver of hamsters (roughly 180 passages since isolation). The axenic *Entamoeba histolytica* strain HM1:IMSS was cultured in TYI-S-33 medium at 37 °C<sup>45</sup>. The fresh liver isolate was maintained in culture for 16 hours, for two weeks or maximally three months. Cultures at two months were used as the reference condition for the transcriptome comparisons. Attenuated trophozoites were derived from the HM1-IMSS virulent strain by prolonged axenic culture (more than 10 years) and have lost the ability to form liver abscesses in the hamster animal model<sup>7</sup>.

**Hepatic inoculation procedure.** Four-weeks-old male Syrian golden hamsters (*Mesocricetus auratus*), with a weight ranging from 90 to 100 g, were inoculated by the intraportal route with  $1 \times 10^6$  virulent HM1:IMSS trophozoites. Treatment of animals and surgical procedures were carried out in accordance with relevant guidelines and regulation. The experimental protocol (N° ND/JLA-06.254) has been published<sup>8</sup> and was approved by “Service Hygiène, sécurité et protection de l’environnement” of Institut Pasteur, France. At 7 days post-infection, hamsters were sacrificed and livers removed. For RNASeq experiments, pieces of infected liver were incubated in TYI-S-33 medium and parasites were allowed to adhere to the culture tube for 16 hours. Liver tissue was then discarded and the pool of trophozoites used to purify RNA or for further growth.

**Interaction of *E. histolytica* with human colon explants.** Previous experimental published conditions were used for handling human colon pieces<sup>25</sup> and RNA purification<sup>12</sup>. Briefly,  $1.6 \times 10^5$  trophozoites were added to the luminal face of the colon and incubated in Krebs buffer at 37 °C for 1 and 7 hours. After 1 hour of incubation, trophozoites interacting with the mucus were collected by recovering the mucus layer for RNA purification. Tissues (3 independent samples) were processed in accordance with the French government’s guidelines for research on human tissues and the French Bioethics Act. Patient-written informed consent was obtained at Foch Hospital and the data were analysed anonymously at the Pasteur Institute. The applied protocol was approved by a regional investigational review board (Comité de Protection des Personnes, Ile de France VII, Paris, France) and an institutional review board (Institut Pasteur Recherche Biomedicale, Paris, France; reference RBM/2009.50).

**RNA extraction, library construction and sequencing.** Total RNA was extracted from approximately  $1 \times 10^6$  *E. histolytica* trophozoites (with each culture performed in triplicates) of using Trizol (Invitrogen). The polyA fraction was purified from 10 µg of total RNA using Dynabeads according to the manufacturer’s instructions (ThermoFisher). Libraries were constructed using ScriptSeq mRNA-Seq Library Preparation Kit (Illumina) following manufacturer’s recommendations and were quality controlled using Agilent Bioanalyzer. Sequencing was performed on a HiSeq 2000 (Illumina) to obtain 58 base single-end reads. Adapter sequences were trimmed using Cutadapt<sup>46</sup> and reads shorter than 15 nucleotides removed.

These processed reads data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-4882.

**Analysis of transcriptome sequencing data.** Sequence reads were mapped to the *E. histolytica* genome assembly (AmoebaDB v1.7, <http://amoebadb.org/amoeba/>) using Tophat version 2.0.6 with default parameters. Count data for coding genes (based on the bona fide gene models defined in previous work<sup>28</sup> were analysed using R version 3.2.2<sup>47</sup> and the DESeq2 Bioconductor package version 1.10.0<sup>48</sup>. Data were normalized with DESeq2 and the default parameter. The dispersion estimation and statistical test for differential expression were performed with default parameters (including outlier detection and independent filtering). Raw p-values for each comparison (each growth condition versus control) were adjusted for multiple testing according to the Benjamini and Hochberg (BH) procedure<sup>49</sup> and genes with an adjusted p-value lower than 0.05 were considered differentially expressed. A Principal Component Analysis (PCA) was performed to explore the structure of the data (Supplementary Fig. 1). The input data for the PCA was a subset of the VST-transformed counts matrix containing the 500 most variant genes. GO terms analysis and Protein classes identification was performed with the tools of PANTHER<sup>34</sup>. Only gene expression rates with a fold-change greater than 2 or lower than 0.5 were considered. To draw Venn diagrams we used Venny (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>). We used the lists of differentially expressed genes resulting from the comparison of virulent HM1:IMSS extracted from the liver, virulent HM1:IMSS incubated on human colon, attenuated HM1:IMSS versus HM1:IMSS in culture.

## References

1. Watanabe, K. & Petri, W. A. Molecular biology research to benefit patients with *Entamoeba histolytica* infection. *Mol Microbiol.* **98**(2), 208–217 (2015).
2. Duggal, P. *et al.* A mutation in the leptin receptor is associated with *Entamoeba histolytica* infection in children. *J Clin Invest.* **121**(3), 1191–1198 (2011).
3. Mondal, D. *et al.* Contribution of enteric infection, altered intestinal barrier function, and maternal malnutrition to infant malnutrition in Bangladesh. *Clin Infect Dis.* **54**(2), 185–192 (2012).
4. Petri, W. A. Jr, Mondal, D., Peterson, K. M., Duggal, P. & Haque, R. Association of malnutrition with amebiasis. *Nutr Rev.* **67** Suppl 2, S207–S215 (2009).
5. Marie, C. & Petri, W. A. Jr. Regulation of virulence of *Entamoeba histolytica*. *Annu Rev Microbiol.* **68**, 493–520 (2014).
6. Faust, D. M. & Guillen, N. Virulence and virulence factors in *Entamoeba histolytica*, the agent of human amoebiasis. *Microbes Infect.* **14**(15), 1428–1441 (2012).
7. Olivos, A. *et al.* *Entamoeba histolytica*: mechanism of decrease of virulence of axenic cultures maintained for prolonged periods. *Exp Parasitol.* **110**(3), 309–312 (2005).
8. Rigother, M. C. *et al.* Fate of *Entamoeba histolytica* during establishment of amoebic liver abscess analyzed by quantitative radioimaging and histology. *Infect Immun.* **70**(6), 3208–3215 (2002).
9. Loftus, B. *et al.* The genome of the protist parasite *Entamoeba histolytica*. *Nature.* **433**(7028), 865–868 (2005).

10. MacFarlane, R. C. & Singh, U. Identification of differentially expressed genes in virulent and nonvirulent *Entamoeba* species: potential implications for amebic pathogenesis. *Infect Immun.* **74**(1), 340–351 (2006).
11. Davis, P. H., Schulze, J. & Stanley, S. L. Jr. Transcriptomic comparison of two *Entamoeba histolytica* strains with defined virulence phenotypes identifies new virulence factor candidates and key differences in the expression patterns of cysteine proteases, lectin light chains, and calmodulin. *Mol Biochem Parasitol.* **151**(1), 118–128 (2007).
12. Thibeaux, R. *et al.* Identification of the virulence landscape essential for *Entamoeba histolytica* invasion of the human colon. *PLoS Pathog.* **9**(12), e1003824 (2013).
13. Gilchrist, C. A. *et al.* Impact of intestinal colonization and invasion on the *Entamoeba histolytica* transcriptome. *Mol Biochem Parasitol.* **147**(2), 163–176 (2006).
14. Santi-Rocca, J. *et al.* Endoplasmic reticulum stress-sensing mechanism is activated in *Entamoeba histolytica* upon treatment with nitric oxide. *PLoS One.* **7**(2), e31777 (2012).
15. Santos, F. *et al.* Maintenance of intracellular hypoxia and adequate heat shock response are essential requirements for pathogenicity and virulence of *Entamoeba histolytica*. *Cell Microbiol.* **17**(7), 1037–1051 (2015).
16. Weber, C. *et al.* Stress by heat shock induces massive down regulation of genes and allows differential allelic expression of the Gal/GalNAc lectin in *Entamoeba histolytica*. *Eukaryot Cell.* **5**(5), 871–875 (2006).
17. Ali, I. K., Ehrenkaufer, G. M., Hackney, J. A. & Singh, U. Growth of the protozoan parasite *Entamoeba histolytica* in 5-azacytidine has limited effects on parasite gene expression. *BMC Genomics.* **8**, 7–26 (2007).
18. Debnath, A. *et al.* A high-throughput drug screen for *Entamoeba histolytica* identifies a new lead and target. *Nat Med.* **18**(6), 956–960 (2012).
19. Mfotie Njoya, E. *et al.* Bioassay-guided fractionation of extracts from *Codiaeum variegatum* against *Entamoeba histolytica* discovers compounds that modify expression of ceramide biosynthesis related genes. *PLoS Negl Trop Dis.* **8**(1), e2607 (2014).
20. Hernández-Cuevas, N. A., Weber, C., Hon, C. C. & Guillen, N. Gene expression profiling in *Entamoeba histolytica* identifies key components in iron uptake and metabolism. *PLoS One.* **9**(9), e107102 (2014).
21. Baumel-Alterzon, S., Weber, C., Guillén, N. & Ankri, S. Identification of dihydropyrimidine dehydrogenase as a virulence factor essential for the survival of *Entamoeba histolytica* in glucose-poor environments. *Cell Microbiol.* **15**(1), 130–144 (2013).
22. Husain, A., Jeelani, G., Sato, D. & Nozaki, T. Global analysis of gene expression in response to L-Cysteine deprivation in the anaerobic protozoan parasite *Entamoeba histolytica*. *BMC Genomics.* **12**, 275 (2011).
23. Penuliar, G. M. *et al.* Transcriptional and functional analysis of trifluoromethionine resistance in *Entamoeba histolytica*. *Antimicrob Chemother.* **67**(2), 375–386 (2012).
24. Ehrenkaufer, G. M., Haque, R., Hackney, J. A., Eichinger, D. J. & Singh, U. Identification of developmentally regulated genes in *Entamoeba histolytica*: insights into mechanisms of stage conversion in a protozoan parasite. *Cell Microbiol.* **9**(6), 1426–1444 (2007).
25. Bansal, D. *et al.* An *ex-vivo* human intestinal model to study *Entamoeba histolytica* pathogenesis. *PLoS Negl Trop Dis.* **3**(11), e551 (2009).
26. Thibeaux, R. *et al.* Newly visualized fibrillar collagen scaffolds dictate *Entamoeba histolytica* invasion route in the human colon. *Cell Microbiol.* **14**(5), 609–621 (2012).
27. Thibeaux, R. *et al.* The parasite *Entamoeba histolytica* exploits the activities of human matrix metalloproteinases to invade colonic tissue. *Nat Commun.* **5**, 5142 (2014).
28. Hon, C. C. *et al.* Quantification of stochastic noise of splicing and polyadenylation in *Entamoeba histolytica*. *Nucleic Acids Res.* **41**(3), 1936–1952 (2013).
29. Willhoeft, U., Buss, H. & Tannich, E. DNA sequences corresponding to the ariel gene family of *Entamoeba histolytica* are not present in *E. dispar*. *Parasitol Res.* **85**(8–9), 787–789 (1999).
30. Edman, U., Meraz, M. A., Rausser, S., Agabian, N. & Meza, I. Characterization of an immuno-dominant variable surface antigen from pathogenic and nonpathogenic *Entamoeba histolytica*. *J Exp Med.* **172**(3), 879–888 (1990).
31. Silvestre, A. *et al.* In *Entamoeba histolytica*, a BspA family protein is required for chemotaxis toward tumour necrosis factor. *Microbial Cell.* **2**(7), 235–246 (2015).
32. MacFarlane, R. C. & Singh, U. Identification of an *Entamoeba histolytica* serine-, threonine-, and isoleucine-rich protein with roles in adhesion and cytotoxicity. *Eukaryot Cell.* **6**(11), 2139–2146 (2007).
33. Hon, C. C., Nakada-Tsukui, K., Nozaki, T. & Guillen, N. Dissecting the actin cytoskeleton of *Entamoeba histolytica* from a genomic perspective. *Molecular and Cell biology of Anaerobic Protists*. Chapter 4. (Horizon Press, 2009).
34. Mi, H. *et al.* PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucl. Acids Res.* **44**(D1), D336–D342 (2016).
35. Bosch, D. E. & Siderovski, D. P. Structural determinants of ubiquitin conjugation in *Entamoeba histolytica*. *J Biol Chem.* **288**(4), 2290–2302 (2013).
36. Pollard, T. D., Actin and Actin-Binding Proteins. *Cold Spring Harb Perspect Biol.* **8**(8), pii: a018226 (2016).
37. Fukuda, A., Nogi, Y. & Hisatake, K. The regulatory role for the ERCC3 helicase of general transcription factor TFIIF during promoter escape in transcriptional activation. *Proc Natl Acad Sci USA* **99**(3), 1206–1211 (2002).
38. Tsurimoto, T., Shinozaki, A., Yano, M., Seki, M. & Enomoto, T. Human Werner helicase interacting protein 1 (WRNIP1) functions as a novel modulator for DNA polymerase delta. *Genes Cells.* **10**(1), 13–22 (2005).
39. Vijayraghavan, S. & Schwacha, A. The eukaryotic Mcm2-7 replicative helicase. *Subcell Biochem.* **62**, 113–134 (2012).
40. Franks, T. M., Singh, G. & Lykke-Andersen, J. Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense-mediated mRNA decay. *Cell.* **143**(6), 938–950 (2010).
41. Gong, C., Tang, Y. & Maquat, L. E. mRNA-mRNA duplexes that autoelicit Staufen1-mediated mRNA decay. *Nat Struct Mol Biol.* **20**(10), 1214–1220 (2013).
42. Galván-Moroyoqui, J. M., Del Carmen Domínguez-Robles, M., Franco, E. & Meza, I. The interplay between *Entamoeba* and enteropathogenic bacteria modulates epithelial cell damage. *PLoS Negl Trop Dis.* **2**(7), e266 (2008).
43. Blazquez, S. *et al.* Human tumor necrosis factor is a chemoattractant for the parasite *Entamoeba histolytica*. *Infect Immun.* **74**(2), 1407–1411 (2006).
44. Ali, I. K. *et al.* Evidence for a link between locus R-R sequence type and outcome of infection with *Entamoeba histolytica*. *Clin Microbiol Infect.* **18**(7), E235–E237 (2012).
45. Diamond, L. S., Harlow, D. R. & Cunnick, C. C. A new medium for the axenic cultivation of *Entamoeba histolytica* and other *Entamoeba*. *Trans R Soc Trop Med Hyg.* **72**, 431–432 (1978).
46. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* **17**, 10–12 (2011).
47. R Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna, Austria, 2015.
48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology.* **15**(12), 550 (2014).
49. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society.* **57**(1), 289–300 (1995).

## Acknowledgements

The authors gratefully acknowledge the help of Sylvie Syan, Elisabeth Labruyère and Roman Thibeaux (BCP Unit) for animal and colon infection experiments. Thanks to Odile Sismeiro and Rachel Legendre (Transcriptome and EpiGenome Platform) for their advice with RNA seq and for data submission. Special thanks to Daniela Faust (BCP Unit) for critical reading this manuscript. The Transcriptome and EpiGenome Platform is a member of the France Génomique consortium (ANR10-NBS-09-08). The work was funded by the National French Research Agency (ANR-2010-GENM-0011-01, GENAMIBE), the French Parasitology consortium “LabexParaFrap” (ANR-11-LABX0024) and the Mexico–France programs ECOS NORD grant M08-S02. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author Contributions

C.W. and M.K. performed the experiments. C.-C.H., M.-A.D. and H.V. performed the bioinformatics analysis. C.L.-C., J.Y.C. and N.G. made the biological analysis and wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

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

**How to cite this article:** Weber, C. *et al.* Extensive transcriptome analysis correlates the plasticity of *Entamoeba histolytica* pathogenesis to rapid phenotype changes depending on the environment. *Sci. Rep.* **6**, 35852; doi: 10.1038/srep35852 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2016