In-depth comparative analysis of *Tritrichomonas foetus* transcriptomics reveals genes linked to host-adaptation

Andrés M. Alonso (✉ amalonso@intech.gov.ar)

IIB-INTECH: Instituto de Investigaciones Biotecnológicas

Nicolás Schcolnicov

IIB-INTECH: Instituto de Investigaciones Biotecnológicas

Luis Diambra

CREG: Universidad Nacional de la Plata Centro Regional de Estudios Genomicos

Veronica Coceres (✉ coceres@intech.gov.ar)

IIB-INTECH: Instituto de Investigaciones Biotecnológicas

Research

Keywords: *Tritrichomonas foetus*, transcriptomics, adaptation

Posted Date: May 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-419937/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

BACKGROUND: *Tritrichomonas foetus* is a protozoa flagellated that reside as a parasite or commensal in organ cavities such as the gastrointestinal and reproductive tracts of their hosts. While this parasite is an important venereal pathogen in cattle and the causative agent of chronic diarrhea in the domestic cat, the mechanisms that define the host specificity of this parasite are still unknown.

METHODS: Here, we integrate the genomic and transcriptomic information of the parasites obtained from different hosts (bovine, feline and porcine), to determine the gene expression profiles of *T. foetus* associated with host adaptation.

RESULTS: We demonstrated the existence of patterns of co-expressed genes specific to each strain and related to known transcription factors (Myb DNA-binding proteins), proteases and genes related to protein phosphorylation process. Also, the genes related to Myb DNA-binding proteins and protein kinases were differentially expressed between strains being those genes different for each strain.

CONCLUSIONS: On the basis of the expression profile variability of genes involved in transcription, intracellular signaling and proteases between the strains (pathogenic and non-pathogenic), we propose that these genes have roles in *T. foetus* adaptation to different hosts. This integrated approach will serve as a useful resource for future studies about the host-parasite interaction and for the future identification of new targets for diagnosis, vaccines and therapeutic intervention to control the bovine and feline tritrichomonosis.

Background

Trichomonads are extracellular flagellated protozoa that reside as parasites or commensals of warm and anaerobic body cavities of their hosts. One of them, *Tritrichomonas foetus*, has been described as a commensal and facultative pathogen of cattle, cats and pigs (1–3). This parasite is an important venereal pathogen in cattle, which causes endometritis, infertility, early embryonic death and thus significant economic losses. In contrast, in cats, *T. foetus* infects the ileum, caecum and colon, causing chronic large bowel diarrhea. Moreover, it has been described as a commensal and facultative pathogen of pigs, being found in the nasal cavity, stomach, caecum, colon and occasionally in the small intestine of these animals (4). Although it is not unusual for a parasite to have a wide host range, the diverse routes of transmission, tissue tropism and pathogenicity of *T. foetus* isolates within different hosts is confounding. Previously, cross-infection experiments were developed to assay *T. foetus* tropism and pathogenicity. These experiments confirmed *T. foetus* cross-infection among bovine and porcine hosts (5, 6). On the other hand, it was proved that feline strain can establish infections in bovine hosts and vice versa, although the pathology is mild in comparison with their original hosts (7). These observations contributed to the current discussion about whether the *T. foetus* strains from bovine, feline, and porcine are or not the same species and generated a question about this parasite adaptation to different hosts.

A variety of molecular studies were conducted to find genetic differences among bovine, feline and porcine strains. Mostly, it was confirmed the closed relationship between porcine and bovine genotypes (6, 8, 9), while at feline genotype punctual differences were documented (9–12). In contrast, proteomics studies among bovine and feline strains revealed a similar proteomic profile, but showed differences in cysteine proteases expression profiles (12). Additionally, an RNA-seq experiment, by *de novo* assembly strategy, demonstrated similarities between bovine and feline transcriptomes, although it is important to highlight that the major similarities were found between the bovine and porcine genotypes. In this context, Morin-Adeline *et al* suggested that these similarities were due to the recent parasite adaptation to their respective hosts that colonization could be a product of regulation of gene transcription (13, 14). It is important to note that those approaches were documented when no genome assembled profile of genes related to *T. foetus* was available, which was a great limitation for functional genomics analysis and gene annotation. Nowadays, a genome assembly was documented for *T. foetus* K1 strain (15), but there are still no reports on the integration of genomics and transcriptomics datasets.

Here we performed a guided assembly of the previously reported transcriptomics data for bovine (BP-4), feline (G10/1) and porcine (PIG30/1) strains (13, 14), using as reference the K1 strain genome assembly, since we hypothesized that adaptation of *T. foetus* to different hosts could be product of differential gene expression. By integration of available genomic and transcriptomic data for *T. foetus*, we performed a clustering procedure over the expression data to reduce redundancy and propose hidden patterns of co-expressed genes (clusters). For this analysis, we took into account the genes related to pathogenicity and adaptation to the host, previously described for trichomonads, such as cysteine proteases (16–21). We could observe that cysteine proteases were grouped by our analysis with genes related to known transcription factors, the Myb DNA-binding proteins (22), and genes related to protein phosphorylation process. In addition, we observed that protein phosphorylation processes were the most differentially regulated between strains, and interestingly, the expression profile of genes related to those processes was different for each strain. The same occurred when we observed the genes related to Myb proteins, which allow us to speculate about the possible function of CPs, Myb proteins and protein phosphorylation in *T. foetus* strain adaptation to the host.

Methods

Data acquisition and analysis

Raw data of RNA sequencing from *Tritrichomonas foetus* strains: porcine (PIG30/1; SRX973684) (14), bovine (BP4; SRX540117) and feline (G10/I; SRX540971) (13) were obtained from sequence read archive database (23). Quality check was performed by FASTQC tool (24) and reads were filtered by trimmomatic software (25). Surviving reads were aligned to *Tritrichomonas foetus* K1 reference genome (ASM183968v1) employing HISAT2 alignment tool (26). For transcriptome assembly and the analysis of differential expressed genes we conduct a guided protocol using the Cufflinks pipeline (27) (*T. foetus* K1 as reference). List of differentially expressed genes for each comparison are listed in the Additional file 1:Table S1. Data quality check, plots and analysis were conducted by R scripts and packages.

Agglomerative procedure and analysis

To reduce the redundancy of the data set we performed a hierarchical clustering method (UPGMA) to aggregate a 26,268 gene data set in clusters by similar expression values. Genes with zero FPKM values in all samples were filtered and obtained a matrix of genes of dimension 24767x3. To estimate an adequate number of clusters we performed the agglomerative procedure for different numbers of clusters, and calculate a measure of the clustering merit (Davies-Bouldin index, DBI) (28). Low values of DBI indicated good cluster structure; as a result we group our data set from 24767 genes in 690 clusters (Additional file 2:Figure S1). The resulting cluster matrix (690x3) and cluster composition is listed in the Additional file 3:Table S2 and Additional file 4:Table S3 respectively. Heatmaps plots were constructed by R scripts and packages.

Proteases annotation

The complete transcriptome assembly was mapped to the MEROPS database (29) by a blastx protocol. To avoid false positive discovery, we searched in the library of selected peptidase sequences (27); only five matches per seq with $1e-10$ E-value of the cutoff was accepted. The resulting matches for each transcript were mapped back to the corresponding genes. List of mapped proteases, and putative cysteine proteases can be found in Additional file 5:Table S4.

Gene annotation and GO enrichment analysis

For the annotation of the differential expressed genes Hmmer2go suite was employed using the pfam HMM models database (30) to predict domains from transcripts that could codify for functional proteins. Transcripts with predicted ORFs of 300 nucleotides were accepted as the minimum limit. For domain prediction, results were filtered by best E-value (< 0.001). For gene ontology term analysis, the background *T. foetus* K1 reference proteome was downloaded from the UniProt database (UP000179807). The annotated differential expressed genes in the three samples were mapped to obtain the corresponding GO term, data can be found in the Additional file 6:Table S5.

Results

Bovine, Porcine and Feline *Tritrichomonas foetus* transcriptomics overview

As we have previously mentioned, genetic differences between *T. foetus* strains are consistent, but not sufficient to define as different species the bovine, porcine and feline strains. Moreover, those differences are not enough for understanding the *T. foetus* adaptive capacity to different hosts. In this work, we hypothesized that the ability of *T. foetus* to adapt to different hosts could be explained at transcriptional regulation level. In this sense, we conducted a mapping experiment employing three available transcriptomics datasets for *T. foetus* strains: BP-4 (bovine), PIG30/1 (porcine), G10/1 (feline) (13, 14); against the available public genome for *T. foetus* (bovine K1 strain (15)). Our results demonstrated a high mappability rate of the three strains against *T. foetus* K1 strain (Table 1) and thus, a close relationship among the different strains at transcriptomic level with the bovine K1 reference strain.

Next, we performed a guided transcriptome assembly (K1 strain as reference) and we reconstructed a common transcriptome for each one of *T. foetus* strains analyzed (BP-4, G10/1 and PIG30/1). We obtained an assembly of 29,361 isoforms (contigs), which represents a total of 26,284 genes for the three strains. Taking into account that we used as reference the K1 assembly, differences between previously de novo assemblies are clear (Table 1). Finally, we highlight that the *T. foetus* K1 assembly showed 25,336 genes, and we were able to obtain 948 sequences assemblies that could be new genes sequences. Afterwards, by a Principal Component Analysis (PCA), we demonstrated that BP-4 and PIG30/1 strains were grouped and distant from G10/1 strain; concluding that similarities between bovine and porcine strains, at the gene expression level, are high and not related to feline strain gene expression (Fig. 1). In this context, expression data for each gene were plotted for each strain against the other, which exhibited clear differences between expressing genes for the different strains (Additional file 2:Figure S2A). Once again a close relationship between BP-4 and PIG30/1 strains arises, while a marked difference in gene expression level between G10/1 and their counterparts is evident.

Table 1
Statistics from the *Tritrichomonas foetus* transcriptomes

	K1(ref.)	PIG30/1a	BP-4b	G10/1b
Assembly size (nt)	51862250	47094268	37882427	29525551
Contigs (n°)	29361	43308	42363	36559
Largest	21237	17203	14314	17195
Shortest	71	201	201	201
Average	1744.38	1087	895.25	806.61
N50	2454	1503	1259	1178
Map vs K1(%)	-	95.68	96.65	91.36
K1(ref.): guided assembly performed in this work using <i>Tritrichomonas foetus</i> K1 genome as reference. ^a Summary of transcriptome statics from Morin-Adeline et al.; 2015 (14), ^b Summary of transcriptome statics from Morin-Adeline et al.; 2014 (13).				

Expression patterns in *T. foetus* strains transcriptomics

Considering the difficulty of analysis of the great amount of data obtained, we hypothesized that reducing the dimension of the data could help us to explore differences in a search of gene expression patterns that could be related with each strain in a different context. Since high throughput technologies have generated large quantities of data available in recent years, determining expression patterns from the resulting datasets could be problematic without a previous dimension reduction procedure. In this sense, computational clustering methods are used to improve exploration strategies by reducing the complexity of the data sets (31). Moreover, gene expression clusters are composed of similar function

categories, a particularity that could be exploited to infer the functionality of uncharacterized genes that are part of the same cluster (32).

In order to reveal the patterns of gene expression for each *T. foetus* strains, we performed a hierarchical clustering procedure that let us reduce our data from 26,284 genes to 690 clusters of genes for each strain. A heatmap plot splitted by dendrogram in 5 groups allowed us to demonstrate clear differential patterns. In concordance with previous data, PIG30/1 and BP-4 strains appear grouped and separate from G10/1 strain (Fig. 2, cluster composition can be found in Additional file 4:Table S3). As can be seen, characteristic patterns of highly expressed clusters of genes arise from G10/1 strain at clusters B, C and D. On the other hand, cluster patterns for PIG30/1 and BP-4 strains were part of section A and B. Is possible to see that patterns in groups B, C and D appeared to be the most particular. In fact, while gene clusters in section B appeared to be greatly expressed in the three strains (with some particular differences), gene clusters in section C and D were more expressed in G10/1 strain (Fig. 2 and Additional file 2:Figure S3). In concordance with those observations, we explored the three sections with the aim of describing the gene composition of each one.

A closer inspection over section B revealed the presence of a gene product related to cathepsin L-like cysteine peptidase in cluster 498 (Fig. 2). This gene product is homologous to TfCP8 protein of *T. foetus* F2 strain (gene bank: X87781.1) (33). Additionally, other genes products related to proteases were present, such as TRFO_22235 at cluster 520, TRFO_09351 in cluster 189, TRFO_43126, and TRFO_29624 (both at cluster 389; see Additional file 4:Table S3). In cluster 316 we found the TRFO_05369, another putative cysteine protease; and in this cluster, also we observed two putative malate dehydrogenase enzymes, a 40S ribosomal protein, and a hydrogenosomes membrane protein precursor. In addition, we have been able to identify a gene TRFO_27838 that codifies a putative precursor of adhesin AP65-1 in cluster 66 and another putative adhesin (TRFO_07867) in cluster 619. This type of adhesins possesses a role as moonlighting proteins (a subclass of multifunctional proteins previously documented in trichomonads parasites with a role in pathogenesis and adaptation to host cells) (34). Interestingly, at cluster 140, a gene that encodes for a tetraspanin protein (TRFO_34204) was found grouped with ribosomal proteins and rubrerythrin gene, a protein related to oxidative stress protection (35).

Section C showed clusters formed of genes related to pathogenesis and moonlighting proteins. We identified genes associated with homeostasis regulation (thioredoxin, HSP90), amino acid metabolic process (aminotransferases, aminopeptidases, PDXDC1, protein phosphatase 2C1, protein serine/threonine phosphatase, cysteine synthase), and nitroreductase family genes (related to nucleotide metabolism). As well as other genes involved in Ca²⁺ signaling machinery (EF-hand family protein and CBL-interacting serine/threonine-protein kinase 8), synthesis of phospholipids (myosin cross-reactive antigen, inositol 3 phosphate synthase) and the GDP-L-fucose biosynthesis via de novo pathway (GDP mannose 4,6 dehydratase). Finally, we demonstrated the presence in cluster 592 (Fig. 2) of a gene that codify for a Clan CA, family C1, cathepsin L-like cysteine peptidase homologous to the documented cysteine protease gene from F2 strain (CP7, genbank:X87780.1) (33).

In section D, as in section B, we detected genes associated with oxidative stress defense such as superoxide dismutase, rubrerythrin, thioredoxin, peroxiredoxin, pirin, OsmC (osmotically inducible protein C), and quercetin 2,3-dioxygenase. As well as, NifU-like domains containing genes and genes related to homeostasis regulation (HSP90-2 and HSP71). Also, we identified genes related to energy metabolism (malate dehydrogenase, glucose 6 phosphate 1 dehydrogenase, NADP dependent isopropanol dehydrogenase, pyruvate decarboxylase isozyme 3/, glyceraldehyde 3 phosphate dehydrogenase, alcohol dehydrogenase iron-containing family protein, glutamate decarboxylase, glucokinase 1, transketolase family protein and fumarate hydratase class II) and lipid transport (Lipid A export ATP binding/permease protein MsbA). In these sections, genes related to signalling (CaMK, ABC transporter family protein, MFS or major facilitator superfamily transporter, small GTP-binding protein; V type proton ATPase subunit B and V-type proton ATPase 16 kDa proteolipid subunit, Ubiquitin-conjugating enzyme E2 and a saposin-like gene) were also identified.

We were also able to detect genes associated with regulation of enzyme activities and/or gene expression (Adenylate and Guanylate cyclase), regulation of transcription (Myb like DNA-binding domain-containing gene), and translation (initiation factor eIF-5A gene family). Finally, ribosomal genes (40S ribosomal S17-B, 40S ribosomal S27, 60S ribosomal L19, 60S ribosomal export gene NMD3), cell division cycle gene 48; and proteases such as a cysteine proteinases (CPs): Clan CD, family C14, metacaspase like cysteine peptidase, Clan CD, family C13, asparaginyl endopeptidase-like cysteine peptidase were also grouped at this section. In conclusion, our clustering analysis was able to reduce the great dimensionality of the transcriptomic data and let us find clusters of genes with potential roles in adaptation and pathogenesis, in addition was possible to observe characteristic patterns of gene cluster expression for each strain.

Cysteine Proteases (CPs) expression patterns in *T. foetus* strains

In trichomonads, cysteine proteases have been implicated in the adherence to host cells, cytotoxicity, nutrient acquisition and the evasion of the host immune response (36). Additionally, it has been reported that such proteases are differentially expressed among *T. foetus* strains. Morin-Adelin *et al* (14) documented that cysteine protease 7 (CP7) was more expressed in the G10/1 strain, in contrast to cysteine protease 8 (CP8), which is more expressed in BP4 and PIG30/1 strains. Taking into account this observation, we performed a homology search in our transcriptome assembly with the aim of finding all the possible genes related to the curated protease database MEROPS (29). A blastx analysis gave us a total of 520 proteases (Additional file 5: Table S4), then we mapped these results against our clustering analysis in order to find clusters that contained those proteases, and thereby confirm their expression in the *T. foetus* strains.

The gene proteases were distributed in 194 clusters of the 690 in our agglomerative analysis (28%), which tells us that a great percent of gene clusters that contain almost one gene that codifies for a putative protease; demonstrating the great distribution of these proteins in the *T. foetus* genome. Particularly, we searched for new proteases that could belong to the CPs family, and our blastx results

demonstrated the existence of a total of 265 putative CPs (~45%); of them, 132 correspond to hypothetical proteins (~50%) in the K1 reference genome. Additionally, we obtained five new sequences related to CPs as a result of our assembly method applied that were not present in K1 genome reference (Additional file 5: Table S4, genes with only Assembly id). This family of proteases (CPs) was represented in 130 clusters of the 192 clusters where proteases were found (~68%). This observation showed the great representation of CPs in the total of proteases expressed by *T. foetus*, in the three strains. In addition, it is important to mention the existence of a great variety of genes in most clusters where the CPs were present, and as has been largely documented, genes in the same cluster could be co-expressed and participate in similar processes (31).

In this context, studying particular differences between strains at clusters composed by CPs could give us more information about adaptation to the host and the pathogenesis process. As shown in Fig. 3A patterns of CPs clusters arise, revealing differences between strains. One of those differences has been documented and involves a cathepsin L-like cysteine peptidase homologous to TfCP7 from F2 strain (genbank: X87780.1; TRFO_22235; cluster 592) and TfCP8 (TRFO_20864, cluster 498), both previously mentioned in this work.

Then, we searched for clusters with similar behavior to TfCP7 (592) and TfCP8 (498) at each strain to propose new gene targets related to these pathogenic factors and the processes where those could participate. We highlighted five clusters related to TfCP7(592) behavior (Fig.3B): cluster 228 that contains a cathepsin L-like endopeptidase gene termed “crustapain” (TRFO_20787) in addition to six more genes, two of them termed as hypothetical proteins and one new gene product as result of our assembly. It highlights the presence of a gene that encodes the alpha amylase, a CAMK kinase protein and GTPase activator protein (Additional file 4:Table S3). Cluster 315 contains 21 genes including a pro-cathepsin H (TRFO_30800) and a Clan CD, family C14, metacaspase-like cysteine peptidase (TRFO_39208) with a CAMK family protein. A numerous genes were termed as hypothetical, while the rest were related to GTPase activity (Additional file 4:Table S3). Cluster 345 contained 4 genes, including a ubiquitin-specific peptidase 21 orthologue (TRFO_42709), the cluster 471 was composed by 4 genes including a Clan CD, family C14, metacaspase-like cysteine peptidase (TRFO_19395), and finally, the cluster 545 included seven genes including an asparagine endopeptidase of the family C13 (Legumain;TRFO_32118). Once again a CAMK family protein stands out sharing the cluster with CPs.

On the other hand, five clusters are highlighted as related to CP8 (498) behavior (Fig.3C), those were: cluster 144, composed of six genes including cathepsin K (TRFO_42629) and crustapain (TRFO_02020) that was confirmed by us as the homologous to CP5 from *T. foetus* F2 strain (Genbank: X87778.1) (33). Cluster 280 included 14 genes such as cathepsin L-like cysteine peptidase annotated as “cysteine protease 8” (TRFO_16156) homologous to TfCP6 from F2 strain (X87779.1), in addition to an EF-hand protein and CAMK family protein. Cluster 521 with three genes, including a papain homolog annotated as “Oryzain alpha chain” (TRFO_07604), cluster 602 was composed only by a Clan CA, family C1, cathepsin L-like cysteine peptidase (TRFO_30793) as the cluster 625 that only included a cathepsin L1 (TRFO_08265), homologous to the TfCP9 from F2 strain (X87782.1). As can be seen CPs, shared clusters

with a great variety of genes that are co-expressed. While a great proportion was termed in reference as hypothetical protein, others are related to CAMK family and EF-hand proteins, those gene products were documented as important factors for pathogeny and migration in other parasite models (37). In this sense our in-depth analysis is useful for studying large families of genes and proposing new gene targets related to pathogenesis.

Functional analysis of differentially expressed genes (DEGs)

Since our exploratory analysis revealed characteristic patterns of gene expression for each strain, we decided to perform a more strict analysis to investigate differences at the gene expression level between strains to map those differences to biological processes or molecular functions that could explain strains adaptability.

To achieve the above objective we performed a differential expression (DE) analysis between strains. The analysis was conducted as follows: G10/1 vs PIG30/1; BP-4 vs PIG30/1 and BP4 vs G10/1, significant DEGs were listed in Additional file 1:Table S1. Volcano plots (Additional file:Figure S2B) summarize the results of the analysis, as can see for G10/1 vs PIG30/1 we observed 157 genes upregulated for G10/1 strain (downregulated for PIG30/1 strain) while another 445 genes were downregulated (upregulated for PIG30/1 strain). For BP-4 vs PIG30/1 analysis, we observed 52 genes upregulated for BP-4 strain (downregulated for PIG30/1) and 79 genes downregulated (upregulated for PIG30/1). Finally for BP-4 vs G10/1 367 were upregulated for BP-4 strain (downregulated for G10/1) and 203 were downregulated (upregulated for G10/1 strain). Initially, we observed that the major differential expressed genes (DEGs) arise when G10/1 strain was compared against the PIG30/1 or BP-4 counterpart, while in BP-4 vs PIG30/1 analysis the number of DEGs was much lower (Additional file:Figure S2B).

Since we conducted a guided transcriptome assembly using as reference *Tritrichomonas foetus* K1 draft genome, gene annotations were taken from the reference. In this context, genes that were not present in *T. foetus* K1 annotation were documented in our assembly (see Additional file 4:Table S3, genes with only Assembly_id). Taking into account that *T. foetus* K1 strain genome remains poorly annotated, we implemented a strategy for to annotate most quantity of genes present in our assembly (and not present in *T. foetus* K1 assembly) and assigned them putative functions by hidden Markov models (see material and methods). As a result, possible functions to genes that were part of the DEGs between strains were assigned (Additional file 6: Table S5). Next, we map the gene ontology terms related to the annotated DEGs. As can be seen in Fig. 4, at the three comparisons, biological processes were common between up and downregulated genes. A major percent of DEGs were associated to protein phosphorylation processes (GO: 0006468) and we observed that these proteins were related to kinase domains (PF00069) and tyrosine and serine/threonine kinase domains (PF07714). It is important to highlight that even though this process is the most represented in the three comparisons, the genes that are related to this process are not the same, suggesting that each strain uses different groups of kinases at the different contexts (Fig. 5A).

An inspection over G10/1 vs PIG30/1 comparison revealed that chromatin remodeling process (GO:0006338) was presented in upregulated and downregulated genes, particularly by those which share sequences related with a c-Myb DNA binding domain. Moreover, carbohydrate metabolic (GO: 0005975) and glucose metabolic (GO: 0006006) processes were more represented in upregulated genes. In the downregulated genes process were more represented those involved in the regulation of transcription (GO: 0006355), regulation of DNA replication (GO: 0006275) and DNA replication (GO: 0006260) processes. Molecular functions shared between upregulated and downregulated genes were related to kinase activity (GO: 0004672), GTP binding (GO: 0005525), and GTPase activity (GO: 0003924). Only in upregulated group, we observed genes related to calcium ion binding (GO:0005509), and those related to EF-hand domain-containing proteins (PF00036), a Ca²⁺ ion binding domain essential for cell signaling (10.1042/BJ20070255). Finally, protein binding function (GO:0005515) appears represented only in downregulated genes due to the presence of genes that codify for Leucine rich repeat proteins (PF13855), that are known to participate in protein-protein interaction (38).

For BP-4 vs PIG30/1 DE analysis results showed that genes that are up and down regulated were involved in the carbohydrate metabolic process, this is due to the differential expression of genes that codify for Glycosyl hydrolases (PF04616, PF01055) related to molecular function of hydrolase activity, hydrolyzing O-glycosyl compounds (GO:0004553), and a glucose-6-phosphate 1-dehydrogenase family protein (PF02781) that participate in addition at glucose metabolic process (GO:0006006) and oxidation-reduction process (GO:0055114). Transmembrane transport (GO: 0055085) was an upregulated process, since the differential expression of a gene that codifies for a major facilitator superfamily protein or MFS (PF07690). Interestingly, MFS proteins facilitate the transport across the cytoplasmic or internal membranes of a variety of substrates including ions, sugar phosphates, drugs, neurotransmitters, nucleosides, amino acids, and peptides (39). On the other hand, translational initiation (GO:0006413), chromatin remodeling and DNA replication were downregulated processes showing that those processes, related to growth and cell replication, were more active in the PIG30/1 strain. Molecular functions like GTP binding and GTPase activity were upregulated in contrast to nucleic acid binding (GO: 0003676) and hydrolase activity (GO: 0004553) that were downregulated processes.

Finally, BP-4 vs G10/1 analysis revealed common processes between up and downregulated genes such as chromatin remodeling and regulation of transcription DNA-templated (GO: 0006355). Cell redox-homeostasis (GO: 0045454) was an upregulated process since thioredoxin genes were found upregulated. Also, regulation of DNA replication was upregulated, since proliferating cell nuclear antigen (PCNA) genes were differentially expressed. Moreover, the signal transduction (GO: 0035556) and carbohydrate metabolic process were downregulated processes. An interesting observation, as was shown above for DE kinases, was that the genes related to the chromatin remodeling process (GO: 0006338) were not the same in each comparison, opening questions about gene regulation at different contexts for *T. foetus* strains (Fig. 5B).

Next, all the annotated DEGs were mapped in our clustering analysis with the aim of extending our annotation procedure. We observed that a great proportion of DEGs (at the three comparisons) were

included in clusters of section E in our heatmap representation, with poor representation of clusters that are part of sections A,B and D, and none of the genes from clusters of section C were differentially expressed (no significant). Cluster 134 (section E), with 43 genes, was the most represented cluster in G10/1 vs PIG30/1 comparison when upregulated genes are explored. The majority of the DEGs were involved in processes such as chromatin remodeling and phosphorylation, and also a papain family cysteine protease (TRFO_23170) was highlighted. This was the only CP that was significantly expressed between strains in our analysis and is the homologue to the CP1 from *T. foetus* D1 strain (U13153.1) (40). It is important to highlight that, in cluster 134, our analysis grouped three genes related to CPs, six genes related to Myb proteins and eleven genes related to proteins that participate in phosphorylation processes (Additional file 4:Table S3). This observation is consistent with our hypothesis that these three groups of genes participate in a common process for adaptation in a parasite's environment: protein kinases integrating signals from context, Myb proteins regulating the gene expression for host colonization and CPs that participates in virulence and adherence to host.

On downregulated genes, cluster 21 (Section E) was the most represented with 89 genes, and were included particular process as cell redox homeostasis (GO: 0045454), intraciliary transport (GO: 0042073), vacuolar transport and regulation of autophagy (GO: 0010506), although phosphorylation process (GO: 0006468) and GTPase activity (GO: 0003924) function were the most represented. Once again, genes related to Myb proteins, protein phosphorylation and CPs are highlighted.

For BP-4 vs PIG30/1 upregulated genes analysis, cluster 124 (Section E), contributed with 18 genes related to phosphorylation processes (GO: 0006468) and with kinase activity (GO: 0004672), while cluster 494 (Section E), in downregulated genes, was the most represented with 16 genes. Finally, cluster 21 was the most represented (59 genes) at upregulated genes for BP-4 vs G10/1 comparison, while for downregulated genes cluster 154 (section E,37 genes) was highlighted with different process such as calcium binding (GO:0005509), regulation of transcription (GO:0006355), transmembrane transport (GO:0055085), vesicle-mediated transport (GO:0016192) and phosphorylation (GO:0006468). Interestingly, the clusters described above are composed of a great proportion of genes that codify for hypothetical proteins (i.e. for cluster 134, ~50%), in this sense our analysis could contribute to functional annotation of genes.

Discussion

The present study analyzed the transcriptomics data from bovine (BP-4), feline (G10/1) and porcine (PIG30/1) strains using as reference the K1 strain genome assembly, in order to integrate the genomic and transcriptomic information of the parasites obtained from different hosts, contributing to functional genetics study of *Tritrichomonas foetus*. In this sense, transcriptomic analysis performed from a virulent and an attenuated strain of *Histomonas meleagridis* (order Trichomonadida) identified a list of specific transcripts for each one, in addition to the common transcripts (41). Studies in *Entamoeba histolytica* and *Trichomonas vaginalis*, organisms related to *T. foetus*, also suggested that differences in pathogenicity among various isolates of the same protozoan parasite were due to changes in virulence

genes expression (42, 43). Moreover, it has been demonstrated through the analysis of transcriptomes of *Entamoeba histolytica* that the changes in environmental conditions trigger the expression of virulence genes (44), which could be suggesting an environmental influence in transcription regulation and pathogenesis.

Previously, different studies were focused on genetic differences among *T. foetus* strains but do not finally explain the capability of *T. foetus* to host adaptation and their eventual capacity to trigger the infection in different hosts (10, 11). Here, our exploratory analysis over the transcriptomics of the strains (BP-4, G10/1 and PIG30/1) revealed differences in expression levels, thereby we analyzed the processes where those differentially expressed genes were included. First, we used clustering methodologies to predict gene functions and cellular processes; and mining datasets in order to find new targets of study, which is relevant considering that the annotation of the *T. foetus* K1 genome is bounded (~ 70% of gene products are documented as hypothetical proteins). We reduced redundancy of the transcriptomic data set of ~ 26000 genes to 690 clusters by agglomerative protocol and our representation of the gene cluster showed a clear characteristic patterns of co-expressed genes for each strain.

Next, taking account that cysteine proteases (CP) play key roles in the biology and pathogenicity of different parasites (45), we analyzed those clusters where genes that codify for the CPs were present so that identify other possible genes, related to pathogenesis, included in each section. About the relevance of these proteases in parasites, has been reported that the key virulent cysteine protease (CP5) is present in *Entamoeba histolytica*, and is absent in the closely related but non-pathogenic *Entamoeba dispar* (46). Besides, the increase in CPs abundance was related to higher virulence in *T. vaginalis* (47) and specifically, an increased secretion of CP2, CP3 and CP4 in *Trichomonas vaginalis* demonstrated to favor the parasite's ability to induce host cell apoptosis (48). Also, the existence of difference in CPs expression patterns has been described for *T. foetus* strains. At this sense, as were documented by Morin Adelin *et al.*; 2014, a gene product related to the cysteine protease 8, unique gene in cluster 498, appears to be more expressed in BP-4 and PIG30/1 strains; and cysteine protease 7, unique gene in cluster 592 is part of a pattern of clusters clearly more expressed in G10/1 strain.

Our homology search of genes that codify for cysteine proteases revealed that they are widely distributed in the transcriptome and integrate numerous clusters, suggesting a relevant role of these proteases in pathogenicity and adaptation to host cells. We could observe that clusters contain CPs genes strains-specific, in fact, our heatmap shows that bovine and porcine strain are more related each other than with feline strain. An exploration over these clusters, particularly the cluster 592 (CP7), more expressed in feline strain, revealed the presence of potential virulence factors, like Pro-cathepsin H (cluster 315) and Crustapain (cluster 228), that are homologous genes to proteases of *Trichomonas vaginalis* CP65, CP39 y CP4 (MER0002336). These CPs are pathogenic factors that have been related to cytotoxicity (20, 49, 50). Additionally, there is great evidence that this type of proteases, that belong to proteases family C1, are important for evasion of the immune system and nutrition in other parasites (51). Also, we revealed the presence of legumain peptidase gene (cluster 545), a documented virulence factor in *T. vaginalis* (52). Finally, we reported the existence of five new putative CP genes that could be future targets of study.

In this context, we could speculate that the changes in transcript levels of specific CP genes suggest non redundant functions for the individual proteases. While our analysis contributes to the discovery of factors important for host adaptation and pathogenicity of *T. foetus*, further studies are needed to establish the specific roles of these proteins in *T. foetus* strains.

On the other hand, our differential gene expression analysis allowed us to study the different up and down regulated processes between strains, by a gene annotation procedure. Protein phosphorylation is the major mechanism by which external stimuli are transformed into intracellular signals to which cells respond and interestingly, we could observe that the phosphorylation process was modulated in the three *T. foetus* strains. In *Trichomonas vaginalis*, has been reported that the protein phosphorylation is the most frequent post translational modification (PTM) since it possesses numerous kinase genes distributed in the genome, suggesting that it could perform protein phosphorylation reactions under different environmental conditions (53, 54). Studies in *Entamoeba histolytica* revealed that protein kinases accounting for about 3% of the total proteome size which is slightly more than the size of kinome of most other eukaryotes, suggesting that protein phosphorylation could be a key mechanism of regulation of this parasite by sensing changes and integrating signals from environment (55, 56). Additionally, in *T. vaginalis* has been proposed that cysteine proteases are functionally regulated by PTMs as protein phosphorylation (57). In fact, it was documented that a *T. vaginalis* legumain peptidase (TvLEGU-1) is phosphorylated and this PTM could have a relevant role in activation and immunogenicity in the host during infection (52) (19). On the basis of the above mentioned, we speculate that phosphorylation process is the most represented since is probably integrates external signals from context, with the aim to regulate genes that are relevant for *T. foetus* adaptability such as the peptidases (pathogenic factors related with adherence and colonization of the host).

In addition, our analysis revealed that Myb proteins were differentially expressed in *T. foetus* strains, and were products of different genes. These proteins were reported as transcription factors that regulate the proliferation and differentiation in eukaryotic cells (22). Surprisingly and as were observed for protein kinases, the differentially expressed Myb genes are not the same in the different *T. foetus* strains. Whence, we hypothesized that the expression of different Myb genes could be influenced by the environment and those Myb genes could be regulating genes related to *T. foetus* adaptation to context. In fact, it is known that Myb proteins in *E. histolytica*, *Giardia lamblia* and *Toxoplasma gondii* regulates the transcription of genes related to stage conversion to cyst (58–60). Also, in *E. histolytica*, the expression profiles of the Myb genes were different and it has been shown that this Myb transcription factor regulates the expression of a subset of stage-specific genes (58).

On the other hand, in *T. vaginalis* was described that an iron-inducible gene that participates in adherence to host, known as *ap65-1*, is transcriptionally regulated by Myb transcription factors (61). The promoter of *ap65-1* contains recognition elements (MRE) that are binding sites for three Myb-like transcription factors that regulate the gene transcription by interaction with MRE sequences. Furthermore, by *in silico* analysis, MRE elements were identified in genes corresponding to virulence factors and that could be recognized by Myb proteins in response to parasite context (19). Finally, is important to highlight that Myb protein

function is influenced by protein phosphorylation in higher eukaryotes (62) and in *T. vaginalis* it is known that TvMyb3 is phosphorylated to finally translocate to the nucleus in response to iron concentration (63). Then, our observation that the protein phosphorylation process was one of the most modulated processes in our analysis, becomes even more important.

Conclusion

In the present study, we have been using an extensive comparative analysis of the transcriptomes to determine the gene expression profiles of *T. foetus* associated with virulence in the context of different hosts. We revealed the existence of patterns of co-expressed genes (clusters) and the expression profile variability of genes involved in transcription, intracellular signaling and proteases between the strains (pathogenic and non-pathogenic); which makes us speculate that host-context could influence the transcriptional machinery of the parasite, thus in the expression of genes involved in signaling and proteolysis. While further studies are needed to confirm the relation of those processes with the *T. foetus* pathogenicity, our study makes a valuable contribution to the functional genomics of this Trichomonadidae.

Declarations

ETHICS DECLARATIONS

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

CONSENT FOR PUBLICATION

Not applicable.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

This research was supported with a grant from the ANPCyT grant BID PICT 2018-1013 (VMC). The funders had no role in decision to publish and preparation of the manuscript.

AUTHORS' CONTRIBUTIONS

VMC and AMA designed the study. AMA and NS conducted the experiments, AMA, VMC and LD analyzed the data. AMA and VMC wrote the manuscript and LD revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

VMC, AMA and LD are researchers from the National Council of Research (CONICET). NS is an undergraduate student at UNSAM.

References

1. Rae DO, Crews JE. *Tritrichomonas foetus*. The Veterinary clinics of North America Food animal practice. 2006;22(3):595–611.
2. Dabrowska J, Karamon J, Kochanowski M, Sroka J, Zdybel J, Cencek T. *Tritrichomonas Foetus* as a Causative Agent of Tritrichomonosis in Different Animal Hosts. Journal of veterinary research. 2019;63(4):533–41.
3. Yao C, Koster LS. *Tritrichomonas foetus* infection, a cause of chronic diarrhea in the domestic cat. Veterinary research. 2015;46:35.
4. Mostegl MM, Richter B, Nedorost N, Maderner A, Dinhopl N, Weissenbock H. Investigations on the prevalence and potential pathogenicity of intestinal trichomonads in pigs using in situ hybridization. Veterinary parasitology. 2011;178(1–2):58–63.
5. Fitzgerald PR. Bovine trichomoniasis. The Veterinary clinics of North America Food animal practice. 1986;2(2):277–82.
6. Tachezy J, Tachezy R, Hampl V, Sedinova M, Vanacova S, Vrlik M, et al. Cattle pathogen *trichomonas foetus* (Riedmuller, 1928) and pig commensal *Tritrichomonas suis* (Gruby & Delafond, 1843) belong to the same species. J Eukaryot Microbiol. 2002;49(2):154–63.
7. Stockdale HD, Givens MD, Dykstra CC, Blagburn BL. *Tritrichomonas foetus* infections in surveyed pet cats. Veterinary parasitology. 2009;160(1–2):13–7.
8. Kleina P, Bettim-Bandinelli J, Bonatto SL, Benchimol M, Bogo MR. Molecular phylogeny of Trichomonadidae family inferred from ITS-1, 5.8S rRNA and ITS-2 sequences. International journal for parasitology. 2004;34(8):963–70.
9. Dabrowska J, Keller I, Karamon J, Kochanowski M, Gottstein B, Cencek T, et al. Whole genome sequencing of a feline strain of *Tritrichomonas foetus* reveals massive genetic differences to bovine and porcine isolates. International journal for parasitology. 2020;50(3):227–33.
10. Slapeta J, Craig S, McDonell D, Emery D. *Tritrichomonas foetus* from domestic cats and cattle are genetically distinct. Experimental parasitology. 2010;126(2):209–13.
11. Slapeta J, Muller N, Stack CM, Walker G, Lew-Tabor A, Tachezy J, et al. Comparative analysis of *Tritrichomonas foetus* (Riedmuller, 1928) cat genotype, *T. foetus* (Riedmuller, 1928) cattle genotype and *Tritrichomonas suis* (Davaine, 1875) at 10 DNA loci. International journal for parasitology. 2012;42(13–14):1143–9.
12. Stroud LJ, Slapeta J, Padula MP, Druery D, Tsiotsioras G, Coorssen JR, et al. Comparative proteomic analysis of two pathogenic *Tritrichomonas foetus* genotypes: there is more to the proteome than meets the eye. International journal for parasitology. 2017;47(4):203–13.

13. Morin-Adeline V, Lomas R, O'Meally D, Stack C, Conesa A, Slapeta J. Comparative transcriptomics reveals striking similarities between the bovine and feline isolates of *Tritrichomonas foetus*: consequences for in silico drug-target identification. *BMC Genomics*. 2014;15:955.
14. Morin-Adeline V, Mueller K, Conesa A, Slapeta J. Comparative RNA-seq analysis of the *Tritrichomonas foetus* PIG30/1 isolate from pigs reveals close association with *Tritrichomonas foetus* BP-4 isolate 'bovine genotype'. *Veterinary parasitology*. 2015;212(3–4):111–7.
15. Benchimol M, de Almeida LGP, Vasconcelos AT, de Andrade Rosa I, Reis Bogo M, Kist LW, et al. Draft Genome Sequence of *Tritrichomonas foetus* Strain K. *Genome announcements*. 2017;5(16).
16. Lucas JJ, Hayes GR, Kalsi HK, Gilbert RO, Choe Y, Craik CS, et al. Characterization of a cysteine protease from *Tritrichomonas foetus* that induces host-cell apoptosis. *Arch Biochem Biophys*. 2008;477(2):239–43.
17. Gould EN, Giannone R, Kania SA, Tolbert MK. Cysteine protease 30 (CP30) contributes to adhesion and cytopathogenicity in feline *Tritrichomonas foetus*. *Veterinary parasitology*. 2017;244:114–22.
18. Tolbert MK, Stauffer SH, Brand MD, Gookin JL. Cysteine protease activity of feline *Tritrichomonas foetus* promotes adhesion-dependent cytotoxicity to intestinal epithelial cells. *Infect Immun*. 2014;82(7):2851–9.
19. Arroyo R, Cardenas-Guerra RE, Figueroa-Angulo EE, Puente-Rivera J, Zamudio-Prieto O, Ortega-Lopez J. *Trichomonas vaginalis* Cysteine Proteinases: Iron Response in Gene Expression and Proteolytic Activity. *BioMed research international*. 2015;2015:946787.
20. Cardenas-Guerra RE, Arroyo R, Rosa de Andrade I, Benchimol M, Ortega-Lopez J. The iron-induced cysteine proteinase TvCP4 plays a key role in *Trichomonas vaginalis* haemolysis. *Microbes infection*. 2013;15(13):958–68.
21. Mendoza-Lopez MR, Becerril-Garcia C, Fattel-Facenda LV, Avila-Gonzalez L, Ruiz-Tachiquin ME, Ortega-Lopez J, et al. CP30, a cysteine proteinase involved in *Trichomonas vaginalis* cytoadherence. *Infect Immun*. 2000;68(9):4907–12.
22. Weston K. Myb proteins in life, death and differentiation. *Curr Opin Genet Dev*. 1998;8(1):76–81.
23. Leinonen R, Sugawara H, Shumway M. International Nucleotide Sequence Database C. The sequence read archive. *Nucleic acids research*. 2011;39(Database issue):D19–21.
24. Andrews S. FASTQC. A quality control tool for high throughput sequence data 2010 [cited 2020 November]. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
26. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature protocols*. 2016;11(9):1650–67.
27. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*. 2012;7(3):562–78.

28. Davies DL, Bouldin DW. A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell.* 1979;1(2):224–7.
29. Rawlings ND, Waller M, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic acids research.* 2014;42(Database issue):D503-9.
30. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic acids research.* 2014;42(Database issue):D222-30.
31. D'Haeseleer P. How does gene expression clustering work? *Nature biotechnology.* 2005;23(12):1499–501.
32. Altman N, Krzywinski M. Clustering. *Nat Methods.* 2017;14(6):545–6.
33. Mallinson DJ, Livingstone J, Appleton KM, Lees SJ, Coombs GH, North MJ. Multiple cysteine proteinases of the pathogenic protozoon *Tritrichomonas foetus*: identification of seven diverse and differentially expressed genes. *Microbiology.* 1995;141(Pt 12):3077–85.
34. Rada P, Kelleraova P, Verner Z, Tachezy J. Investigation of the Secretory Pathway in *Trichomonas vaginalis* Argues against a Moonlighting Function of Hydrogenosomal Enzymes. *J Eukaryot Microbiol.* 2019;66(6):899–910.
35. Leitsch D, Williams CF, Hrdy I. Redox Pathways as Drug Targets in Microaerophilic Parasites. *Trends Parasitol.* 2018;34(7):576–89.
36. Hernandez HM, Marcet R, Sarracent J. Biological roles of cysteine proteinases in the pathogenesis of *Trichomonas vaginalis*. *Parasite.* 2014;21:54.
37. Nagamune K, Sibley LD. Comparative genomic and phylogenetic analyses of calcium ATPases and calcium-regulated proteins in the apicomplexa. *Molecular biology evolution.* 2006;23(8):1613–27.
38. Kobe B, Kajava AV. The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol.* 2001;11(6):725–32.
39. Yan N. Structural Biology of the Major Facilitator Superfamily Transporters. *Annual review of biophysics.* 2015;44:257–83.
40. Thomford JW, Talbot JA, Ikeda JS, Corbeil LB. Characterization of extracellular proteinases of *Tritrichomonas foetus*. *J Parasitol.* 1996;82(1):112–7.
41. Mazumdar R, Endler L, Monoyios A, Hess M, Bilic I. Establishment of a de novo Reference Transcriptome of *Histomonas meleagridis* Reveals Basic Insights About Biological Functions and Potential Pathogenic Mechanisms of the Parasite. *Protist.* 2017;168(6):663–85.
42. Padilla-Vaca F, Anaya-Velazquez F. Insights into *Entamoeba histolytica* virulence modulation. *Infect Disord Drug Targ.* 2010;10(4):242–50.
43. Hirt RP. *Trichomonas vaginalis* virulence factors: an integrative overview. *Sex Transm Infect.* 2013;89(6):439–43.
44. Weber C, Koutero M, Dillies MA, Varet H, Lopez-Camarillo C, Coppee JY, et al. Extensive transcriptome analysis correlates the plasticity of *Entamoeba histolytica* pathogenesis to rapid phenotype changes depending on the environment. *Scientific reports.* 2016;6:35852.

45. Sajid M, McKerrow JH. Cysteine proteases of parasitic organisms. *Mol Biochem Parasitol.* 2002;120(1):1–21.
46. Jacobs T, Bruchhaus I, Dandekar T, Tannich E, Leippe M. Isolation and molecular characterization of a surface-bound proteinase of *Entamoeba histolytica*. *Mol Microbiol.* 1998;27(2):269–76.
47. De Jesus JB, Cuervo P, Britto C, Saboia-Vahia L, Costa ES-FF, Borges-Veloso A, et al. Cysteine peptidase expression in *Trichomonas vaginalis* isolates displaying high- and low-virulence phenotypes. *J Proteome Res.* 2009;8(3):1555–64.
48. Kummer S, Hayes GR, Gilbert RO, Beach DH, Lucas JJ, Singh BN. Induction of human host cell apoptosis by *Trichomonas vaginalis* cysteine proteases is modulated by parasite exposure to iron. *Microb Pathog.* 2008;44(3):197–203.
49. Alvarez-Sanchez ME, Avila-Gonzalez L, Becerril-Garcia C, Fattel-Facenda LV, Ortega-Lopez J, Arroyo R. A novel cysteine proteinase (CP65) of *Trichomonas vaginalis* involved in cytotoxicity. *Microb Pathog.* 2000;28(4):193–202.
50. Ramon-Luing Lde L, Rendon-Gandarilla FJ, Puente-Rivera J, Avila-Gonzalez L, Arroyo R. Identification and characterization of the immunogenic cytotoxic TvCP39 proteinase gene of *Trichomonas vaginalis*. *Int J Biochem Cell Biol.* 2011;43(10):1500–11.
51. Atkinson HJ, Babbitt PC, Sajid M. The global cysteine peptidase landscape in parasites. *Trends Parasitol.* 2009;25(12):573–81.
52. Rendon-Gandarilla FJ, Ramon-Luing Lde L, Ortega-Lopez J, Rosa de Andrade I, Benchimol M, Arroyo R. The TvLEGU-1, a legumain-like cysteine proteinase, plays a key role in *Trichomonas vaginalis* cytoadherence. *BioMed research international.* 2013;2013:561979.
53. Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, et al. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science.* 2007;315(5809):207–12.
54. Ubersax JA, Ferrell JE Jr. Mechanisms of specificity in protein phosphorylation. *Nature reviews Molecular cell biology.* 2007;8(7):530–41.
55. Anamika K, Bhattacharya A, Srinivasan N. Analysis of the protein kinome of *Entamoeba histolytica*. *Proteins.* 2008;71(2):995–1006.
56. De Cadiz AE, Jeelani G, Nakada-Tsukui K, Caler E, Nozaki T. Transcriptome analysis of encystation in *Entamoeba invadens*. *PloS one.* 2013;8(9):e74840.
57. Huang KY, Chien KY, Lin YC, Hsu WM, Fong IK, Huang PJ, et al. A proteome reference map of *Trichomonas vaginalis*. *Parasitol Res.* 2009;104(4):927–33.
58. Ehrenkaufer GM, Hackney JA, Singh U. A developmentally regulated Myb domain protein regulates expression of a subset of stage-specific genes in *Entamoeba histolytica*. *Cellular microbiology.* 2009;11(6):898–910.
59. Yang H, Chung HJ, Yong T, Lee BH, Park S. Identification of an encystation-specific transcription factor, Myb protein in *Giardia lamblia*. *Mol Biochem Parasitol.* 2003;128(2):167–74.

60. Waldman BS, Schwarz D, Wadsworth MH 2nd, Saeij JP, Shalek AK, Lourido S. Identification of a Master Regulator of Differentiation in *Toxoplasma*. *Cell*. 2020;180(2):359–72. e16.
61. Hsu HM, Ong SJ, Lee MC, Tai JH. Transcriptional regulation of an iron-inducible gene by differential and alternate promoter entries of multiple Myb proteins in the protozoan parasite *Trichomonas vaginalis*. *Eukaryot Cell*. 2009;8(3):362–72.
62. Ramsay RG, Gonda TJ. MYB function in normal and cancer cells. *Nature reviews Cancer*. 2008;8(7):523–34.
63. Hsu HM, Lee Y, Indra D, Wei SY, Liu HW, Chang LC, et al. Iron-inducible nuclear translocation of a Myb3 transcription factor in the protozoan parasite *Trichomonas vaginalis*. *Eukaryot Cell*. 2012;11(12):1441–50.

Figures

Figure 1

A principal component analysis (PCA) for *Tritrichomonas foetus* strains expression data. Colored points represent a dimensional reduction of gene expression data for each strain. As can be seen, the feline strain (G10/1) shows numerous differences with the rest of the strains (porcine strain PIG30/1 and bovine strain BP-4).

Figure 2

Heatmap representation for gene clustering analysis of *Tritrichomonas foetus* strains. Clusters of genes are visualized in heatmap representation and grouped into five groups named A, B, C, D, E by dendrogram inspection. Zoom views for group B and C are shown. Clusters that contain cysteine proteases 7 (592) and 8 (498) are highlighted in red colour. The cluster are represented as log2 (FPKM)

Figure 3

Cysteine proteases expression patterns. A) A heatmap representation of clusters that contain almost one cysteine protease gene. Clusters highlighted in red are those that exhibit similar expression patterns to cluster 592 (CP7). Clusters in black are those with similar expression patterns to cluster 498 (CP8); B), C) Line plots for each cluster that contains almost one cysteine protease gene. Clusters that exhibit similar expression patterns with cluster 592(CP7) and 498 (CP8) are highlighted.

Figure 4

Gene ontology analysis of differentially expressed genes between strains. Plots that show the five most represented biological processes and molecular functions in each comparison between strains.

Figure 5

Comparison between up and downregulated kinases and DNA binding proteins. A) Protein phosphorylation process is the most represented, but different genes related with that process are specifically expressed; B) same particularity is shown for genes related with c-Myb DNA binding proteins.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile2.pdf](#)
- [GA.tiff](#)
- [TableS1.xls](#)
- [TableS2.xls](#)
- [TableS3.xls](#)
- [TableS4.xls](#)
- [TableS5.xls](#)