Molecular & Biochemical Parasitology 187 (2013) 60–71

Contents lists available at SciVerse ScienceDirect

Molecular & Biochemical Parasitology

Functional profiling of the *Tritrichomonas foetus* transcriptome and proteome

Kuo-Yang Huang[a,b,1], Jyh-Wei Shin[c,1], Po-Jung Huang[d], Fu-Man Ku[a], Wei-Chen Lin[a], Rose Lin[a], Wei-Min Hsu[a], Petrus Tang[a,*]

[a] *Molecular Regulation and Bioinformatics Laboratory, Department of Parasitology, Chang Gung University, Kwei-Shan, Taoyuan 333, Taiwan*
[b] *Graduate Institute of Biomedical Sciences, College of Medicine, Chang Gung University, Kwei-Shan, Taoyuan 333, Taiwan*
[c] *Department of Parasitology, Medical College, National Cheng Kung University, No. 1, Dashiue Road, Tainan 701, Taiwan*
[d] *Bioinformatics Center, Chang Gung University, Kwei-Shan, Taoyuan 333, Taiwan*

ARTICLE INFO

Article history:
Received 30 March 2012
Received in revised form
30 November 2012
Accepted 2 December 2012
Available online 17 December 2012

Keywords:
Tritrichomonas foetus
Expressed sequence tag
Transcriptome
2-DE
Proteome

ABSTRACT

Tritrichomonas foetus is a potent veterinary pathogen, causing bovine and feline trichomoniasis. The principal clinical manifestation of infection in cattle is inflammation of the genital tract and infertility. In feline, the parasite causes large-bowel disease resulting in chronic diarrhea. In contrast to other well-studied protozoan, genetic data regarding the molecular characterization and expression in *T. foetus* is far less understood. In this study, the first large-scale *T. foetus* expressed sequence tag (TfEST) project was conducted on 5064 randomly selected EST clones from a non-normalized unidirectional Tf30924 cDNA library. Assembling of 5064 single-pass sequences from the 5′ end resulted in 713 contigs and 1961 singlets. BLAST search revealed that 53.52% of the unigenes showed significant similarity to known sequences or protein motifs/domains. Functional classifications indicated that most of the unigenes are involved in translation, ribosomal structure and ribosome biogenesis. The average GC content of the *T. foetus* transcriptome is 40.93%. Intriguingly, only 31.29% of the unigenes contain the classical AAUAAA polyadenylation signal sequence at the 3′-UTR region. Furthermore, a panel of potential chemotherapeutic targets was also identified for the first time in *T. foetus*. The protein expression levels were verified by using two-dimensional electrophoresis and matrix-assisted laser desorption ionization time-of-flight mass spectrometry. A total of 68 highly abundant protein spots were successfully identified in the reference 2-DE map based on our *T. foetus*-specific protein database. The EST dataset and the reference 2-DE map established in the present study will provide a foundation for future whole genome sequencing project and comparative transcriptomic/proteomic analyses to provide potential drug targets against *T. foetus* infection.

1. Introduction

Tritrichomonas foetus is the causative agent of bovine and feline trichomoniasis. In bulls, infection is acquired primarily through direct sexual contact. This anaerobic pathogen is confined to the epithelial surface of the penis, prepuce, and urethra. Infection is usually asymptomatic and chronic [1–3], although in some cases early death of the developing foetus is observed. Previous studies showed that the parasite provokes cell death to bovine oviduct epithelial cells [4] and extracellular products secreted by *T. foetus* were cytotoxic to bovine sperm [5]. *T. foetus* has also been used as a model organism to study hydrogenosomes, cellular differentiation, and ultrastructure of Trichomonadidae [6–8]. Study on the epidemiology of *T. foetus* in beef bull populations in Florida revealed an overall 6.0% prevalence [9]. Another report indicated that *T. foetus* infection in Asturiana de la Montaña beef cattle kept in extensive conditions indicated that 32% (33/103) and 41.5% (27/65) of the bulls and herds were infected, respectively [10]. Although the use of artificial insemination in breeding cattle has led to the virtual elimination of this organism from the cattle population in many countries, bovine tritrichomonad infection still causes significant economic losses in the free-ranging cattle industry around the world

Accumulating evidences indicated that in addition to bovine trichomoniasis, *T. foetus* is also the causative agent of feline large

Abbreviations: EST, expressed sequence tag; 2-DE, two-dimensional gel electrophoresis; MALDI-TOF, matrix-assisted laser desorption time-of flight; ORF, open reading frame; ACN, acetonitrile; TC, tentative cluster; GAPDH, glyceraldehyde-3-phosphate dehydrogenate; PEPCK, phosphoenolpyruvate carboxykinase; CP, cysteine proteinase; ADSS, adenylosuccinate synthetase; PDC, pyruvate decarboxylase; ODC, ornithine decarboxylase.

* Corresponding author at: Molecular Regulation and Bioinformatics Laboratory, Department of Parasitology, College of Medicine, Chang Gung University, 259 Wenhwa 1st Road, Kwei-Shan, Taoyuan 333, Taiwan. Tel.: +886 3 2118800x5136; fax: +886 3 2118122.
E-mail address: petang@mail.cgu.edu.tw (P. Tang).
[1] These authors contributed equally.

0166-6851/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.molbiopara.2012.12.001

bowel disease [11–14]. Infections of cats with *T. foetus* have been described in at least 12 countries [15–30]. In cats, *T. foetus* colonizes the ileum, caecum and colon in close proximity to the mucosal surface [31]. Although the presence of *T. foetus* in the feline reproductive tract is highly unlikely [32], there is a single report of a natural *T. foetus* infection in the feline uterus [33]. Epidemiology studies showed that infected domestic cats usually have colitis and chronic, foul-smelling persistent diarrhea [11,15,34]. Feline trichomoniasis is prevalent among cattery cats where transmission *via* the fecal–oral route is suspected. However, the route(s) of transmission of *T. foetus* infection in cats remains unknown.

The host specificity of *T. foetus* remains controversial. The pig commensal *Tritrichomonas suis* have almost identical genomic fingerprints with *T. foetus* [35], suggesting that *T. suis* and *T. foetus* belong to the same species or the variants of the same species. However, clinical signs are not found in swine naturally infected with *T. suis* [36,37]. Moreover, *T. foetus* has also been isolated in dogs [38] and experimental murine model of bovine trichomoniasis was described [39,40]. Recently, in some cases, *T. foetus* infection in humans has been reported [41,42]. The diversity of host specificities (cattle [1], feline [11], swine [36], canine [38], murine [39] and human [41]) and clinical signs induced in different animals indicated that *T. foetus* developed a complex gene regulatory network during evolution to cope with the microenvironments in different hosts. Hence, more information at both genomic and transcriptomic levels is desirable for further studies on the adaptive mechanism(s) of *T. foetus* to different hosts.

Studies on Trichomonadidae have concentrated mainly on the human specific species *Trichomonas vaginalis*. Despite the significance of *T. foetus* as one of the major veterinary diseases in cattles and felines, the basic biology of this parasite is poorly understood at the molecular level. Early study showed that the *T. foetus* genome contains 5 chromosomes [43] and the genome size of the *T. foetus* strain KV-1 (ATCC 30924) was estimated to be 177 MB [44]. This genome size is very close to the recently draft 160 MB *T. vaginalis* G3 genome, which contain 60,000 putative genes [45]. However, less than a hundred nucleotide sequences were deposited in GenBank since the discovery of *T. foetus*. Genome-wide EST analysis is one of the most comprehensive approaches for gene discoveries and gene expression profiling. To gain further insights into the Trichomonadidae transcriptome, we reported here the analysis of more than 5000 ESTs from a non-normalized unidirectional *T. foetus* trophozoite cDNA library. In order to obtain an overview of the *T. foetus* proteome, the direct assessment of biological processes by monitoring expressed proteins, we also performed two-dimensional gel electrophoresis (2-DE) combined with matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry to elucidate the intracellular protein patterns, their identification, and expression profiles of *T. foetus* trophozoite. The proteomic analysis of this study represents the most extensive protein expression profiles to date and will serve as a basis for future comparative proteomic studies to identify the potential virulence factors in *T. foetus*. The EST and protein datasets not only can be exploited for gene expression, protein profiling and genome annotation, but also provided a more throughout understanding on the pathogenesis of *T. foetus*.

2. Materials and methods

2.1. T. foetus culture

T. foetus strain KV-1 (ATCC30924) was maintained in YIS medium, pH 5.8, containing 10% heat-inactivated horse serum at 37 °C [46]. The number of viable cells was determined by Trypan blue exclusion hemocytometer counts. Unsynchronized mid-logarithmic phase parasites with more than 98% viable cells were harvested for the construction of cDNA library and proteomics study.

2.2. Complementary DNA library construction

Total RNA was extracted from harvested cells by using a commercial kit (Pharmacia) and the residual genomic DNA was hydrolyzed with DNase I. PolyA$^+$ RNA was isolated using the PolyA$^+$ tract mRNA isolation kit (Promega). Complementary DNAs primed with oligo-dT were synthesized by using a ZAP-cDNA synthesis kit and directionally cloned into the *Eco*RI and *Xho*I sites of Uni-ZAP XR (Stratagene) vector. The quality of the Tf30924 unidirectional cDNA library was assessed by colony PCR of 96 randomly picked clones to determine the average insert size and efficiency of cloning.

2.3. DNA sequencing and data processing

Plasmids were *in vivo* excised from the Tf30924 cDNA library by using helper phage and transformed into *E. coli* DH10B (Invitrogen) as described by the manufacturer (Stratagene). Individual transformants were picked, grown overnight in LB medium containing IPTG and X-Gal. Plasmid DNA was isolated. Single-pass sequencing of the 5′-end of the cDNA fragment was carried out with T3 primer using the ABI PRISM BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). The sequencing products were resolved and analyzed on MegaBase® 1000 sequencer. (Applied Biosystems). The sequencing trace files obtained were examined by the Phred algorithm for base calling with default parameters [47,48]. Vector sequences were masked by using the Crossmatch program. cDNA sequences longer than 100 bp were used for clustering analysis by using the Phrap algorithm (http://www.phrap.org/) and the Paracel Transcript Assembler (Paracel) with default parameters. Assembled contigs were manually evaluated by the CLC Genomics Suite (CLC Bio) to insure clustering reliability.

2.4. Annotation and functional classification

The annotation was done by BLAST search to identify similarities between TfESTs and sequences deposited in the NCBI non-redundant(nr) protein database and Swiss-Prot database [49]. Sequences with an *e*-value $\leqq 10^{-10}$ were considered as significant sequence similarity to known genes. Functional domains/motifs and signatures of protein families were identified by comparing with the Interpro and Pfam databases [50]. To classify the genes into functional categories, the NCBI identifiers of each best aligned sequence were associated with the Gene Ontology (GO) and the NCBI Cluster of Orthologus Gene (COG) indices [51]. *T. foetus* EST nucleotide sequences reported in this study are available in the GenBank™ dBEST database under the accession numbers CX154307 to CX159216.

2.5. Analysis of mRNA polyadenylation signals and codon usage

Putative coding regions of the ESTs were collected from BLASTx alignment results and compared with the putative open reading frames (ORFs) predicted by Getorf in the EMBOSS package [52]. A Perl script was used to locate polyadenylation signal(s) in ESTs with a poly(A) or poly(T) extremity tail. The pattern match permits for one or two mismatches from the classical polyadenylation signal sequence AAUAAA. The GC content, codon usage and amino acids compositions of each ORF were also determined.