

**Genomic Characterization of Expressed Sequence Tags and Gene Expression Profiling of
the Three Life-Cycle Stages of *Ichthyophthirius multifiliis***

by

Jason Wayne Abernathy

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 14, 2010

Keywords: ciliate, fish, microarray, parasite, protozoa

Copyright 2010 by Jason Wayne Abernathy

Approved by

Zhanjiang Liu, Chair, Professor of Fisheries and Allied Aquacultures
Covadonga Arias, Associate Professor of Fisheries and Allied Aquacultures
Kevin Fielman, Assistant Professor of Biological Sciences
Joanna Wysocka-Diller, Associate Professor of Biological Sciences

Abstract

The ciliate protozoan *Ichthyophthirius multifiliis* (Ich) is an important parasite of freshwater fish that causes 'white spot' disease. Ich is a major contributor of fish mortalities and economic loss to both the ornamental and edible fish stocks around the world. Despite its global importance, very little genetic information is available for Ich. The focus of this study is to create a large-scale genetic resource for Ich and to utilize those resources in a microarray platform to examine global gene expression in the parasite. The first goal was to generate expressed sequence tags (ESTs) for Ich. Toward this goal, a total of 10,368 EST clones were sequenced using a normalized cDNA library made from pooled RNA samples of the trophont, tomont, and theront life-cycle stages, leading to 8,432 high quality sequences. Clustering analysis of these ESTs allowed identification of 4,706 unique sequences containing 976 contigs and 3,730 singletons. This set of ESTs represents a significant proportion of the Ich transcriptome, and provides a material basis for the development of microarrays useful for gene expression studies concerning Ich development, pathogenesis, and virulence.

While analyzing the ESTs, a particular sequence cluster was over-represented in the library. The sequence was chosen for further analysis. Upon first inspection, the sequence appeared to be a putative senescence-related transcript. Upon further inspection, the transcript was of ribosomal origin. However, the genetic library was created by the use of poly (A) tails; conventionally, ribosomal RNAs (rRNA) are not polyadenylated. Reported here is the discovery of polyadenylated rRNAs in Ich. Analysis using multiple sequence alignments revealed four

potential polyadenylation sites including three internal regions and the 3' end of the large subunit rRNA. While the functions of polyadenylation of rRNA in this organism are largely unknown at present, the presence of internal polyadenylation sites, along with the presence of truncated segments of the rRNA, may suggest a role of the polyadenylation in a degradation pathway.

The major goal of this study was to build and utilize a microarray platform for the analysis of gene expression in Ich. An oligo microarray containing all publicly available Ich ESTs representing 9,129 unique genes was developed. In addition to oligo features representing the 9,129 Ich genes, gene coding sequences from two related protozoa, *Tetrahymena thermophila* and *Plasmodium falciparum*, were also included to increase gene content through cross-hybridization. The microarray was used to examine gene expression patterns of the three life stages of Ich: infective theront, parasitic trophont, and reproductive tomont. A total of 173 putative genes were found to be differentially regulated among all three life-stages. Examples of differentially expressed transcripts included immobilization antigens, annexins, and epiplasmin, as well as various other transcripts involved in developmental regulation and host-parasite interactions. The microarray platform was further used to assess differential regulation of genes between early and late serial passages, as the parasite becomes senescent and loses some of its infectivity. A total of 215 transcripts were found to be differentially expressed and common to both tomont and trophont between passage 1 to passage 100, including surface proteins and other likely immunogens potentially involved in host-pathogen interactions. Genes that are differentially expressed during steady-state Ich development, as well as potential candidate genes involved in Ich infectivity will be presented and discussed.

Acknowledgments

I would like to give a special thanks to my father, Jim Abernathy, and my mother, Janet Dean for their continual support. I thank my sister Christi for pushing me along and providing some encouraging words when I needed them most. I thank all of my committee members for their guidance over the years: Dr. Zhanjiang (John) Liu provided support, guidance, and training over the years of my education for which I will be forever grateful; Dr. Covadonga Arias first made me appreciate and enjoy the fields of science and research; Drs. Joanna Wysocka-Diller and Kevin Fielman instructed excellent courses and provided assistance throughout my college career; and Dr. Nannan Liu for kindly providing her expertise as an outside reader. I thank all of the many past and present members of my laboratory group with whom I collaborated, especially Drs. Huseyin Kucuktas and Eric Peatman. I thank my outside collaborators, especially Drs. De-Hai Xu, Phillip Klesius, and Thomas Welker at the United States Department of Agriculture, Aquatic Animal Health Research Laboratory, in Auburn, Alabama for use of their facilities and resources.

Table of Contents

Abstract	ii
Acknowledgments	iv
List of Tables	vii
List of Figures	viii
I. Introduction	1
II. Research Objectives	10
III. Generation and Analysis of Expressed Sequence Tags from the Ciliate Protozoan Parasite <i>Ichthyophthirius multifiliis</i>	11
Abstract	12
Background	13
Results and Discussion	15
Conclusion	25
Materials and Methods	26
Acknowledgments	32
References	33
IV. Transcriptomic Profiling of <i>Ichthyophthirius multifiliis</i> Reveals Polyadenylation of the Large Subunit Ribosomal RNA	40
Abstract	41

Background	42
Results	45
Discussion	53
Conclusion	56
Materials and Methods	57
Acknowledgments	63
References	64
V. Gene Expression Profiling of <i>Ichthyophthirius multifiliis</i> : Insights into Development and Senescence-Associated Avirulence	68
Abstract	69
Background	70
Results and Discussion	73
Conclusion	94
Materials and Methods	95
Acknowledgments	101
References	102
Appendix 1	109
Appendix 2	110
Appendix 3	117

List of Tables

Table 1. A summary of the Expressed Sequence Tag (EST) analysis	15
Table 2. The most abundant ESTs detected from the EST sequencing	17
Table 3. A summary of simple sequence repeats identified from the ESTs.	19
Table 4. The top 10 megablast BLASTN hits to the 28S rRNA transcript	46
Table 5. Polymerase Chain Reaction (PCR) primer sequences for the 28S rRNA transcript	62
Table 6. BLASTX summary of the unique Ich ESTs after clustering and identification of contaminating sequences.	74
Table 7. List of genes selected for quantitative PCR analysis, developmental array.	89
Table 8. List of genes selected for quantitative PCR analysis, pathogenic array	90

List of Figures

Figure 1. Venn diagram summary of sequence comparisons of <i>Ichthyophthirius multifiliis</i> ESTs with <i>Tetrahymena thermophila</i> and <i>Plasmodium falciparum</i> genomes	21
Figure 2. Pie charts of 2nd level Gene Ontology (GO) terms using EST sequences	24
Figure 3. Flowchart of molecular analysis leading to the identification of the polyadenylated 28S rRNA	46
Figure 4. Schematic presentation of the internal polyadenylation sites of the 28S rRNA	48
Figure 5. Graphical results from re-sequencing clones at internal polyadenylated sites of the 28S transcript	50
Figure 6. Results of the polyadenylation test for polyadenylation at the 3' end of the rRNA ..	51
Figure 7. PCR test for the presence of polyadenylation of three internal poly (A) sites of the 28S rRNA	52
Figure 8. Sequence alignment of the 180-bp regions immediately upstream to each of the observed polyadenylation sites	55
Figure 9. Level 2 gene ontology (GO) of the unique Ich sequences, minus contamination.....	76
Figure 10. Quantitative PCR on selected genes, developmental array data	91
Figure 11. Quantitative PCR on selected genes, pathogenic array data	93

I. INTRODUCTION

The ciliate protozoan *Ichthyophthirius multifiliis* is a devastating freshwater teleost parasite. The parasite is widespread; it affects many freshwater fish species around the world. It causes great loss to both the aquaculture and ornamental fish industries. Ich is responsible for the disease ichthyophthiriosis or ‘white spot’ disease with characteristic white spot cysts forming under the fish gill or epidermis. Ich life-cycle biology has been studied since the late 1800’s under Fouquet with much more complete studies in the mid 1990’s by the Matthews [1] and Dickerson [2] labs. Ich displays a cyclical life cycle, with three main developmental stages: the parasitic trophont, the reproductive tomont, and the infective theront. After a period of feeding and growth within the host, the mature trophont bursts from the fish while becoming the tomont, the stage where the cell begins encystment and drops to the sediment or other suitable substrate. The tomont rapidly divides producing several hundred to thousands of daughter cells within the cyst called tomites. After division, the tomont ruptures releasing the tomites into the water where they become the parasitic theronts. The theront is free-swimming and vulnerable; it attempts to rapidly locate a host where it will attach and infect, completing its life cycle and beginning anew. The most current review of the Ich developmental life-cycle stages was completed by R.A. Matthews [3].

Unicellular organisms are classified as ciliates based on their two main characteristics: the possession of cilia in at least one stage and the presence of nuclear dimorphism (also referred to as having nuclear duality or two nuclei) [3]. Typical to members of the phylum Ciliophora, Ich displays cilia in the theront stage. Ich also has nuclear dimorphism; Ich possesses the characteristic macronucleus and micronucleus. Ich has been shown to be highly similar to species of *Tetrahymena* [4-6], such as *T. thermophila*, a free-living non-parasitic ciliate

protozoan. From a comparative view to *Tetrahymena* [5, 7], Ich may also undergo both germ line and somatic cell divisions, generally where the micronucleus contains the germ line DNA that can be replenished through meiotic conjugation, and the macronuclear DNA that is transcriptionally active to support the cell. There has been strong evidence of asexual reproduction in Ich for many years, with possible evidence that some sexual regeneration may occur [3, 8]. Sexual reproduction is essential in other studied ciliate species, for cell rejuvenation and macronucleus regeneration [9, 10]. Further evidence to support sexual regeneration needed for Ich health arises from the major difficulties encountered in maintaining viable isolates of Ich in the laboratory setting. Multiple passages of an Ich isolate on a fish host lead to a significant decline in infectivity, relating to aging or senescence [3, 11-15]. The current knowledge leads to the hypothesis that this type of induced senescence could be due to a lack of recombination of the germ line (reviewed in [3]). Therefore, undertaking a study of senescence-related genes would be an important step in better understanding Ich development, reproduction, and parasitic nature.

Xu et al [15] have demonstrated that Ich theronts lose their ability to infect fish as the number of life-cycle passages / cell divisions increase in the laboratory (purposely related to age or senescence). This is significant to the current study since microarray experiments can be designed using a control (passage 1) and comparing patterns of gene expression as Ich loses its infectivity in subsequent passages.

Most of the studies on Ich to date have been largely based on anatomy, morphology, or biochemistry of the organism, or have concentrated on cell-surface antigens, such as Immobilization antigens (I-antigens) [2, 15-26]. These I-antigens have a role in Ich immunological properties. When these proteins are injected into a host, the host responds by

production of antibodies that immobilize the parasite *in vitro*. Since the majority of Ich studies are comparisons of Ich to properties of *Tetrahymena thermophila* [27, 28], studies of various surface antigens [2, 18, 29, 30], or biochemical measurements [20, 31], very little genomic information is available for Ich. From the March 2007 dbEST release [32], there were only 511 EST sequences deposited in GenBank (<http://www.ncbi.nlm.nih.gov>) and most of those sequences are the previously described I-antigens. There are no large-scale sequencing projects for Ich using genomic data or EST analysis. At the time of this publication, there are relatively few (~125) published works of literature archived at the U.S. National Library of Medicine regarding the parasite. While this list is certainly not comprehensive, there remains a great need for more Ich study, including large-scale genome mining and expression analysis.

One important long-term goal of the discovery of critical developmental and virulence genes is to assist in identifying suitable biomarkers to aid in the development of diagnostic tools and/or control strategies in Ich including downstream DNA-based vaccine development. Prevention of the disease caused by Ich is the best way to avoid fish mortalities. For other aquatic pathogens, DNA-based vaccinations have been developed and used for some time as the main tool in reduction of fish stock losses. However, attainment of sufficient protein-based antigens remains a major problem in commercial development of vaccines against eukaryotic parasites [3]. Acquired protective immunity in fish following injection with whole killed parasites [33] or sonicates [34] of Ich has shown some promise *in vitro*, however these approaches remain difficult for commercial development due to parasite culture difficulties [3]. Also, the use of *T. thermophila*, a similar ciliate parasite based on 18S rDNA sequences [6], as a vaccine against Ich has been proposed in the past [29]; however, cross protection has yet to be established [27, 35]. Interestingly, *T. thermophila* has the potential to be used as a eukaryotic

expression vector for recombinant vaccines but this approach has not been successful in Ich immunizations [28]. Further, in studies utilizing host protective immunity against I-antigens, protection is only directed against homologous serotypes of the protein [3, 26, 33, 34], with limited cross-immunity [25, 36]. To date, five different serotypes of this protein have been identified [3] making vaccination specificity even more of a concern.

To assist in the knowledge of Ich genetics, the goals of this project are to create a repertoire of Expressed Sequence Tags (ESTs) and to analyze the sequences (Chapter III.). Utilizing this EST resource, the most widely expressed transcript will be identified and characterized (Chapter IV). Ultimately, the major goal of this project is to build a microarray platform using the ESTs as a resource in order to examine global gene expression between Ich life-stages and also different passages as the cells lose infectivity (Chapter V).

References

1. Cross ML, Matthews RA: **Localized leucocyte response to *Ichthyophthirius multifiliis* establishment in immune carp *Cyprinus carpio* L.** *Vet Immunol Immunopathol* 1993, **38**(3-4):341-358.
2. Dickerson HW, Clark TG, Leff AA: **Serotypic variation among isolates of *Ichthyophthirius multifiliis* based on immobilization.** *J Eukaryot Microbiol* 1993, **40**(6):816-820.
3. Matthews RA: ***Ichthyophthirius multifiliis* Fouquet and Ichthyophthiriosis in Freshwater Teleosts.** *Adv Parasitol* 2005, **59**:159-241.
4. Abernathy JW, Xu P, Li P, Xu DH, Kucuktas H, Klesius P, Arias C, Liu Z: **Generation and analysis of expressed sequence tags from the ciliate protozoan parasite *Ichthyophthirius multifiliis*.** *BMC Genomics* 2007, **8**(1):176.
5. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM *et al*: **Macronuclear Genome Sequence of the Ciliate *Tetrahymena thermophila*, a Model Eukaryote.** *PLoS Biol* 2006, **4**(9).
6. Wright AD, Lynn DH: **Phylogeny of the fish parasite *Ichthyophthirius* and its relatives *Ophryoglena* and *Tetrahymena* (Ciliophora, Hymenostomatia) inferred from 18S ribosomal RNA sequences.** *Mol Biol Evol* 1995, **12**(2):285-290.
7. Prescott DM: **The DNA of ciliated protozoa.** *Microbiol Rev* 1994, **58**(2):233-267.
8. Matthews R, Matthews, BF., and Ekless LM ***Ichthyophthirius multifiliis*: observations on the life cycle and indications of a possible sexual phase.** *Folia Parasitol* 1996, **43**(3):203-208.

9. Baroin-Tourancheau A, Delgado P, Perasso R, Adoutte A: **A broad molecular phylogeny of ciliates: identification of major evolutionary trends and radiations within the phylum.** *Proc Natl Acad Sci U S A* 1992, **89**(20):9764-9768.
10. Portnoy S, Nanney DL: **Perturbance analysis of nuclear determination in Tetrahymena: analysis of mating type frequency variations with reference to binary-switch models.** *Differentiation* 1980, **16**(1):61-69.
11. Burkart M, Clark T, Dickerson H: **Immunization of channel catfish, *Ictalurus punctatus* Rafinesque, against *Ichthyophthirius multifiliis* (Fouquet): killed versus live vaccines.** *J Fish Biol* 1990, **13**:401-410.
12. Ekless L, Matthews R: ***Ichthyophthirius multifiliis*: axenic isolation and short-term maintenance in selected monophasic media.** *J Fish Dis* 1993, **16**:437-447.
13. Houghton G, Matthews RA: **Immunosuppression of carp (*Cyprinus carpio* L.) to ichthyophthiriasis using the corticosteroid triamcinolone acetonide.** *Vet Immunol Immunopathol* 1986, **12**(1-4):413-419.
14. Noe JG, Dickerson HW: **Sustained growth of *Ichthyophthirius multifiliis* at low temperature in the laboratory.** *J Parasitol* 1995, **81**(6):1022-1024.
15. Xu DH, Klesius PH: **Two year study on the infectivity of *Ichthyophthirius multifiliis* in channel catfish *Ictalurus punctatus*.** *Dis Aquat Organ* 2004, **59**(2):131-134.
16. Clark TG, Dickerson HW: **Antibody-mediated effects on parasite behavior: Evidence of a novel mechanism of immunity against a parasitic protist.** *Parasitol Today* 1997, **13**(12):477-480.
17. Clark TG, Dickerson HW, Findly RC: **Immune response of channel catfish to ciliary antigens of *Ichthyophthirius multifiliis*.** *Dev Comp Immunol* 1988, **12**(3):581-594.

18. Clark TG, Gao Y, Gaertig J, Wang X, Cheng G: **The I-antigens of *Ichthyophthirius multifiliis* are GPI-anchored proteins.** *J Eukaryot Microbiol* 2001, **48**(3):332-337.
19. Clark TG, Lin TL, Dickerson HW: **Surface antigen cross-linking triggers forced exit of a protozoan parasite from its host.** *Proc Natl Acad Sci U S A* 1996, **93**(13):6825-6829.
20. Clark TG, Lin TL, Jackwood DA, Sherrill J, Lin Y, Dickerson HW: **The gene for an abundant parasite coat protein predicts tandemly repetitive metal binding domains.** *Gene* 1999, **229**(1-2):91-100.
21. Clark TG, McGraw RA, Dickerson HW: **Developmental expression of surface antigen genes in the parasitic ciliate *Ichthyophthirius multifiliis*.** *Proc Natl Acad Sci U S A* 1992, **89**(14):6363-6367.
22. Dickerson H, Clark T: ***Ichthyophthirius multifiliis*: a model of cutaneous infection and immunity in fishes.** *Immunol Rev* 1998, **166**:377-384.
23. Dickerson HW, Clark TG, Findly RC: ***Ichthyophthirius multifiliis* has membrane-associated immobilization antigens.** *J Protozool* 1989, **36**(2):159-164.
24. Dickerson HW, Evans DL, Gratzek JB: **Production and preliminary characterization of murine monoclonal antibodies to *Ichthyophthirius multifiliis*, a protozoan parasite of fish.** *Am J Vet Res* 1986, **47**(11):2400-2404.
25. Xu DH, Klesius PH, Panangala VS: **Induced cross-protection in channel catfish, *Ictalurus punctatus* (Rafinesque), against different immobilization serotypes of *Ichthyophthirius multifiliis*.** *J Fish Dis* 2006, **29**(3):131-138.

26. Xu DH, Klesius PH, Shoemaker CA: **Cutaneous antibodies from channel catfish, *Ictalurus punctatus* (Rafinesque), immune to *Ichthyophthirius multifiliis* (Ich) may induce apoptosis of Ich theronts.** *J Fish Dis* 2005, **28**(4):213-220.
27. Everett KD, Dickerson HW: ***Ichthyophthirius multifiliis* and *Tetrahymena thermophila* tolerate glyphosate but not a commercial herbicidal formulation.** *Bull Environ Contam Toxicol* 2003, **70**(4):731-738.
28. Gaertig J, Gao Y, Tishgarten T, Clark TG, Dickerson HW: **Surface display of a parasite antigen in the ciliate *Tetrahymena thermophila*.** *Nat Biotechnol* 1999, **17**(5):462-465.
29. Maki JL, Dickerson HW: **Systemic and cutaneous mucus antibody responses of channel catfish immunized against the protozoan parasite *Ichthyophthirius multifiliis*.** *Clin Diagn Lab Immunol* 2003, **10**(5):876-881.
30. Sigh J, Lindenstrom T, Buchmann K: **The parasitic ciliate *Ichthyophthirius multifiliis* induces expression of immune relevant genes in rainbow trout, *Oncorhynchus mykiss* (Walbaum).** *J Fish Dis* 2004, **27**(7):409-417.
31. Ewing MS, Ewing SA, Zimmer MA: **Sublethal copper stress and susceptibility of channel catfish to experimental infections with *Ichthyophthirius multifiliis*.** *Bull Environ Contam Toxicol* 1982, **28**(6):674-681.
32. Boguski MS, Lowe TM, Tolstoshev CM: **dbEST--database for "expressed sequence tags".** *Nat Genet* 1993, **4**(4):332-333.
33. Wang X, Dickerson HW: **Surface immobilization antigen of the parasitic ciliate *Ichthyophthirius multifiliis* elicits protective immunity in channel catfish (*Ictalurus punctatus*).** *Clin Diagn Lab Immunol* 2002, **9**(1):176-181.

34. Xu DH, Klesius PH, Shelby RA: **Immune responses and host protection of channel catfish, *Ictalurus punctatus* (Rafinesque), against *Ichthyophthirius multifiliis* after immunization with live theronts and sonicated trophonts.** *J Fish Dis* 2004, **27**(3):135-141.
35. Everett KD, Knight JR, Dickerson HW: **Comparing tolerance of *Ichthyophthirius multifiliis* and *Tetrahymena thermophila* for new cryopreservation methods.** *J Parasitol* 2002, **88**(1):41-46.
36. Swennes AG, Findly RC, Dickerson HW: **Cross-immunity and antibody responses to different immobilisation serotypes of *Ichthyophthirius multifiliis*.** *Fish Shellfish Immunol* 2007, **22**(6):589-597.

II. RESEARCH OBJECTIVES

1. To construct a cDNA library using RNA from the freshwater ciliate pathogen *Ichthyophthirius multifiliis* (Ich) in order to sequence, analyze, and characterized large-scale expressed sequence tags (ESTs).
2. To characterize the most abundantly expressed gene found by screening sequences.
3. To build a microarray platform from the ESTs with the intention of utilizing the array for a global assessment of steady-state gene expression in each of the three life-cycle developmental stages of Ich. The array will be further utilized to examine genes that may be involved in infectivity due to the senescence of the organism.

**III. GENERATION AND ANALYSIS OF EXPRESSED SEQUENCE TAGS FROM THE
CILIAE PROTOZOAN PARASITE *ICHTHYOPHTHIRIUS MULTIFILIIS***

Abstract

The ciliate protozoan *Ichthyophthirius multifiliis* (Ich) is an important parasite of freshwater fish that causes 'white spot disease' leading to significant losses. A genomic resource for large-scale studies of this parasite has been lacking. To study gene expression involved in Ich pathogenesis and virulence, the goal was to generate expressed sequence tags (ESTs) for the development of a powerful microarray platform for the analysis of global gene expression in this species. Here, a project was initiated to sequence and analyze over 10,000 ESTs. A total of 10,368 EST clones were sequenced using a normalized cDNA library made from pooled samples of the trophont, tomont, and theront life-cycle stages, and generated 9,769 sequences (94.2% success rate). Post-sequencing processing led to 8,432 high quality sequences. Clustering analysis of these ESTs allowed identification of 4,706 unique sequences containing 976 contigs and 3,730 singletons. These unique sequences represent over two million base pairs (~10% of *Plasmodium falciparum* genome, a phylogenetically related protozoan). BLASTX searches produced 2,518 significant (E-value < 10⁻⁵) hits and further Gene Ontology (GO) analysis annotated 1,008 of these genes. The ESTs were analyzed comparatively against the genomes of the related protozoa *Tetrahymena thermophila* and *P. falciparum*, allowing putative identification of additional genes. All the EST sequences were deposited by dbEST in GenBank (GenBank: **EG957858–EG966289**). Gene discovery and annotations are presented and discussed. This set of ESTs represents a significant proportion of the Ich transcriptome, and provides a material basis for the development of microarrays useful for gene expression studies concerning Ich development, pathogenesis, and virulence.

Background

The ciliate protozoan *Ichthyophthirius multifiliis* (Ich) is one of the most devastating pathogens. It infects fish skin and gills, and causes white spot diseases in many species of freshwater fish worldwide, which leads to significant losses in the aquaculture industry. The ciliate parasite has three main life-cycle stages: the reproductive tomont, the infective theront, and a parasitic trophont [1-3]. The mature trophont drops off the host to become the tomont where it attaches to a substrate, and undergoes multiple divisions to produce hundreds to thousands of tomites within a cyst. Tomites bore their way through the cyst into water, and differentiate into theronts that infect fish. Once they burrow into the fish epithelium, theronts become trophonts that feed and mature in the host.

In spite of great losses caused by Ich to the aquaculture industry, molecular studies of the parasite have been scarce [see a recent review [4]]. Limited studies have concentrated on immune responses of the host and factors affecting them [5-11]. One of the difficulties for the studies of Ich is the problem involved in long-term maintenance of Ich isolates. Ich isolates appear to lose infectivity or become senescent after a certain number of passages [12-15]. Most often a significant decrease in infectivity is observed after about 50 passages [15]. Not only the infectivity decreases with higher numbers of passages, but also the development of the parasite as measured by the period required for trophonts to emerge from fish [15].

The Ich senescence phenomenon is interesting not only as a developmental biology issue, but also as a potential research system to study the virulence factors involved in the parasite pathogenesis. Assuming the life cycles of Ich and its infectivity are controlled by gene products, then it would be of great interest to learn what genes are involved in the loss of infectivity, and in

the slowing down of its development. However, as very limited molecular information is available from Ich, in-depth research is limited by the lack of information and the lack of genomic resources.

EST analysis is one of the most effective means for gene discoveries, gene expression profiling, and functional genome studies [16-23]. It is also one of the most efficient ways for the identification of differentially expressed genes [24-28]. In order to provide genomic resources for the analysis of differentially expressed genes at different developmental stages of the Ich parasite, and for the analysis of genes differentially expressed when infectivity is being lost, the objectives of this study were to create cDNA libraries suitable for the analysis of expressed sequence tags (ESTs) and to generate an EST resource for Ich to allow cDNA-based design of microarrays for the study of gene expression in relation to the passages and development of the parasite. Before this work, there were only 511 Ich sequences in the GenBank dbEST (release 100606) [29]. A brief examination of these existing EST sequences indicated that a large proportion of them were trophont only reads, histones, ribosomal proteins, and immobilization antigen-related sequences. Here is reported the sequencing of 10,368 Ich EST clones, and generation of 8,432 high quality EST sequences. This EST resource should provide the material basis for the development of microarrays for Ich, and serve as a platform for its functional genomic studies including the development and pathogenesis of Ich, and the host-parasite interactions.

Results and Discussion

Generation of the Ich ESTs

As summarized in Table 1, a total of 10,368 clones were sequenced from a normalized Ich library made from pooled cells from all three life cycle stages: theront, tomont, and trophont. Readable sequences were generated with 9,769 clones (94.2% sequencing success rate). After base calling, sequences were processed by using Phred [30, 31] to eliminate low quality sequences below Q20. Sequences passing Q20 were uploaded into Vector NTI Advance 10 (Invitrogen, Carlsbad, CA) for vector trimming and removal of sequences with very short inserts (<100 bp). The post-sequencing processing resulted in 8,432 high quality sequences.

Table 1. A summary of the EST analysis. ¹Percentage of high quality sequences from successful sequences; ²percentage of unique sequences of the high quality sequences; ³percentage of unique sequences

Description	Number	Percentage
Total number of clones sequenced	10,368	
Total number of successful sequences	9,769	94.2%
Number of high quality sequences	8,432	86.3% ¹
Unique sequences	4,706	55.8% ²
Number of contigs	976	
Number of clones included in the contigs	4,702	
Average clones per contig	4.82	
Number of singletons	3,730	
Number of known genes	2,518	53.5% ³
Unique unknown genes	2,188	46.5% ³

The processed sequences were subjected to cluster analysis using Vector NTI to evaluate sequence redundancies. Of the 8,432 sequences, 4,702 sequences fell within 976 contigs while 3,730 sequences were singletons. On average, each contig contained 4.8 sequences. Taken together, the 976 contigs and the 3,730 singletons made up 4,706 unique sequences (Table 1).

The Ich genome expression appeared to be extremely polarized with a few genes expressed at very high levels. In spite of normalization, transcripts from a few genes were sequenced at very high frequencies. The top 20 contigs with the largest number of ESTs are summarized in Table 2. Of the top 20 most abundantly sequenced transcripts, four of them were detected over 0.5% of total sequences. Of these, the most abundantly sequenced EST cluster, cluster 276 with 764 ESTs, accounted for 7.36% of all sequenced clones. BLASTX searches indicated that this transcript was most similar to a hypothetical protein, TTHERM_02141640, from *Tetrahymena thermophila*. The second most abundantly sequenced transcript was cluster 60 with 119 ESTs. It was identified as a transcript most similar to a hypothetical protein, TTHERM_02641280, from *T. thermophila*. The functions of these hypothetical proteins are unknown at present. These are two transcripts sequenced at exceptionally high frequencies. Obviously, the presence of such abundant transcripts suggested a failure in the normalization processes. However, the results are perplexing since it is believed that the overall normalization processes worked based on several other observations: 1) the overall gene discovery rate (unique sequences over all sequences analyzed) was 55.8%, a reasonable rate for the sequencing depth of approximately 10,000 clones; 2) most other anticipated highly expressed genes such as ribosomal protein genes, actin genes, tubulin genes, and dynein genes were not detected at high levels. Nonetheless, this information is relevant and important as these genes should be the subject for additional subtraction for further EST sequencing in this species. In addition, such

information should provide some basic picture about the most abundantly expressed genes in the parasite. As these hypothetical protein genes are transcribed at such high levels, they may be crucially important for the growth and development, or other life-cycle processes of the parasite.

Table 2. The most abundant ESTs detected from the EST sequencing

Cluster	# of Sequences	Putative identities	% of Total
276	764	Hypothetical protein TTHERM_02141640 from <i>Tetrahymena</i>	7.36%
60	119	Hypothetical protein TTHERM_02641280 from <i>Tetrahymena</i>	1.15
636	86	Unknown	0.82
602	78	Unknown	0.75
171	48	Heat shock protein 90	0.46
83	39	Zinc finger ZZ type family protein	0.38
105	38	Unknown	0.37
279	35	Heat shock protein 90	0.34
392	34	Unknown	0.33
354	31	Heat shock protein 70 (dnaK)	0.30
203	31	Conserved hypothetical protein from <i>Paracoccus denitrificans</i>	0.30
219	29	Hypothetical protein PY05925 from <i>Plasmodium yoelii</i>	0.28
932	28	Unknown	0.27
75	27	Unknown	0.26
472	24	ER type HSP70	0.23
351	23	Unknown	0.22
833	23	Unknown protein from <i>Oryza sativa</i>	0.22
131	22	Unknown	0.21
6	21	Dynein heavy chain protein	0.20
45	21	Outer surface protein from <i>Rickettsia typhi</i>	0.20

This work demonstrated that pooling of samples from all three stages of Ich life cycle followed by normalization was an effective way to reduce common messages across all three life stages. As one would expect, many structural genes would be expressed highly abundantly in all stages of the life cycle. In addition to making savings economically, pooling of samples allowed very effective normalization of these common transcripts without going through three rounds of

normalization. This is consistent with previous experience for the generation of a large number of catfish and oyster ESTs [32-36]. It is obvious that the pooling of samples from three developmental stages made it impossible to provide information concerning expression profiling in relation to developmental stages. However, such information would not be highly meaningful in normalized cDNA libraries where the major focus was to develop EST resources, rather than expression profiling. The other limitation caused by construction of a pooled cDNA library is the loss of sequencing flexibility as to the number of clones to be sequenced from each developmental stage library if they had been separately constructed.

The Ich transcribed sequences are highly A/T-rich, similar to the situation in *T. thermophila*. The unique sequences combined contain 2.18 megabases, approximately 10% of the genomic sequence size of the related protozoan *Plasmodium falciparum*, and 2.1% of the *T. thermophila* genomic sequence. As Ich is a ciliate and most closely related to *Tetrahymena*, this EST resource should represent a good sample of the transcribed fraction of the Ich genome for the estimation of its genome contents as compared with *Tetrahymena*. Based on the EST sequences, the average G+C content of Ich transcribed sequences was found to be 33.4%, even more A/T-rich than those of the closely related hymenostome *T. thermophila*, which has an average G+C content of 38% at protein coding regions [37]. The entire genome of *T. thermophila* was much more A/T-rich than the transcribed fraction, with a G+C content of only 22% [38]. It is highly probable that the Ich genome is also highly A/T-rich. To further the analysis, approximately 1% of the unique ESTs sequenced were found to contained simple sequence repeats. The majority of the simple sequence repeats were of di-nucleotide repeats (68.8%) with AC and AG repeats being the majority. Tri-nucleotide and tetra-nucleotide repeats accounted for 23.7% and 7.5% of the identified microsatellites, respectively (Table 3).

Table 3. A summary of simple sequence repeats identified from the Ich ESTs. Percentages indicated in the parentheses are percentage of each type of repeat among all repeats

Total number of sequences analyzed	8,432
Number of dinucleotide repeats	422 (68.8%)
Number of AC repeats	121
Number of AG repeats	108
Number of AT repeats	56
Number of CT repeats	49
Number of GT repeats	88
Number of GC repeats	0
Number of trinucleotide repeats	145 (23.7%)
Number of tetranucleotide repeats	46 (7.5%)
Total number simple sequence repeats	613

The putative identities of the sequenced ESTs were assessed using BLASTX searches against the non-redundant (NR) database in GenBank [39]. All the search results are summarized in supplemental Table 1 (See Appendix 1). Of the 4,706 unique ESTs, 2,518 (53.5%) had significant (E-value < 10⁻⁵) hits. The remaining 2,188 (46.5%) EST sequences were not similar

to any known sequences. Additional searches using the Swiss-Prot database resulted in putative identities for six additional unknown ESTs (Supplemental Table 1).

Identification of putative secretory proteins

Secretory proteins have been shown to be an important component in many biological processes, including pathogenesis of parasites [40-42]. Therefore, transcripts were screened for putative signal peptides (suggestive of peptides of secretory proteins) within the EST set using the program SignalP 3.0 [43]. A total of 314 ESTs with signal peptides were identified, representing 6.7% of the unique sequences. Of these, 180 (3.8%) were from ESTs with no significant (E-value < 10⁻⁵) BLASTX hit to the NR database in GenBank (Supplemental Table 1).

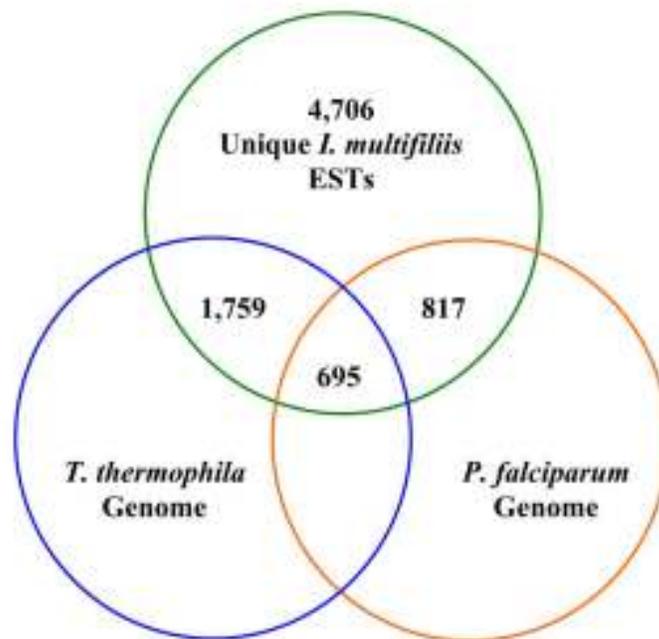
Comparative analysis to related taxa

The parasite Ich is phylogenetically placed between the protozoan's *Plasmodium falciparum* and *Tetrahymena thermophila*. Previous studies using 18S rDNA, histone genes, and I-antigens [44-46] suggested that Ich was more related to *T. thermophila* than to *P. falciparum*. Furthermore, *T. thermophila* and Ich share the ciliate nuclear genetic code, while *P. falciparum* uses the standard genetic code for translation. As the entire genome sequence of *P. falciparum* is available and the macronuclear sequencing project of *T. thermophila* was just recently completed, a comparative BLAST analyses against both genome sequences was made.

The tBLASTx or BLASTX searches of Ich ESTs against the *T. thermophila* and *P. falciparum* genomes are summarized in Supplemental Table 2 (see Appendix 1), and are presented in Figure 1. As expected based on the phylogenetic relationships, more Ich ESTs were similar to the genome sequences of *T. Thermophila* than to that of *P. falciparum*. Of the 4,706

Ich ESTs, 1,759 sequences were similar (E-value < 10⁻⁵) to the *T. thermophila* genome sequences; whereas 817 were similar to the *P. falciparum* genome sequences. In total, 695 ESTs were similar to both *T. thermophila* and *P. falciparum* genomes, and thus are common to all three protists.

Figure 1. Venn diagram summary of sequence comparisons of the Ich ESTs with *Tetrahymena thermophila* and *Plasmodium falciparum* genomes. A total of 4,706 unique Ich ESTs were used as queries yielding 1,759 significant (E-value < 10⁻⁵) hits to the *T. thermophila* genome, and 817 to the *P. falciparum* genome. A total of 695 sequences were ESTs with common hits to both genomes.



Of the 1,759 significant hits against the *T. thermophila* genome, 1,673 had been identified with a putative identity using BLASTX searches against the NR database, while the tBLASTx searches against the *T. thermophila* genome allowed identification of putative identities for additional 86 unique ESTs. Similarly, BLASTX searches against the *P. falciparum* genome allowed identification of 9 additional ESTs. Taken together, the BLAST searches against these two genomes allowed putative identities of 95 additional unique ESTs, bringing the total number of ESTs with significant similarities to known genes to 2,613.

Such genome searches also revealed that of the 2,518 ESTs that had significant hits in BLASTX searches against the NR database, 845 had no significant hits to the *Tetrahymena* genome. Clearly, these ESTs were similar to sequences of organisms other than the ciliate *Tetrahymena*.

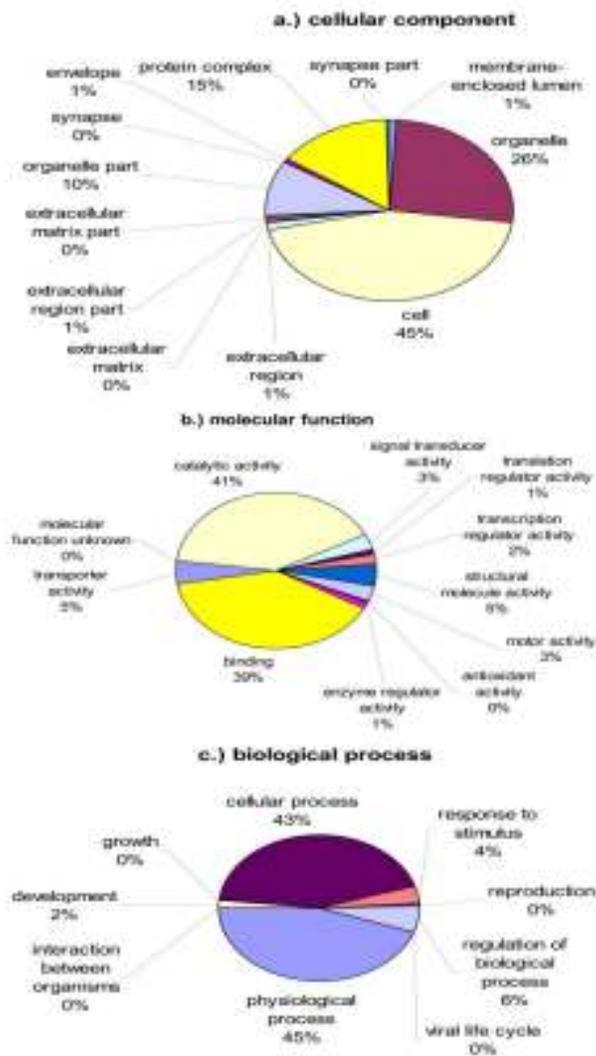
These results clearly suggest conservation of a large fraction of gene sequences among the three protozoa parasites, with a higher level of conservation between Ich and the *T. thermophila* genome than between the Ich genome and the *P. falciparum* genome; although a significant fraction of gene sequences are also shared between the genomes of *T. thermophila* and *P. falciparum*. The results of this comparative analysis are compatible with existing phylogenetic analyses using several molecular markers such as the 18S rDNA, histone genes, and the I-antigens. Obviously, use of a large set of sequences should provide a greater confidence concerning genome evolution. The comparative analysis suggested that the EST resource generated from this study should be useful for phylogenetic analysis and studies concerning genome evolution.

Gene ontology

The unique Ich sequences were compared to annotations through the Gene Ontology Consortium [47] using the automated software Blast2GO [48]. GO terms were obtained for 1,008 unique sequences using this method. Of these, 304 were contigs and 704 were singletons. Sequence descriptions, gene ontology (GO) and enzyme commission (EC) numbers are summarized in Supplemental Table 3 (see Appendix 1). There were 258 sequences with both GO terms and EC numbers.

Gene ontology graphs using percentages of 2nd level GO terms are presented in Figure 2 under the categories of cellular components (Fig. 2A), molecular functions (Fig. 2B), and biological processes (Fig. 2C). Of the cellular component GO terms, 45% and 26% were related directly with cellular and organelle components, respectively. In the category of molecular functions, the vast majority were involved in catalytic activity (41%) and binding activities (39%). Under the category of biological processes, 45% were involved in physiological processes; 43% were involved in cellular processes, 6% in regulation of biological processes, 4% in response to stimuli, and 2% in development (Figure 2).

Figure 2. Pie charts of 2nd level gene ontology (GO) terms. Overall, 1,008 unique sequences were annotated using the Blast2GO software and included in the graphs. Each of the three GO categories is presented including cellular component (a), molecular function (b), and biological process (c).



Conclusion

A total of 8,432 high quality *I. multifiliis* EST sequences were produced. Sequence analysis indicated the presence of 4,706 unique sequences in the EST set. This should represent a significant fraction of the Ich genes, although the exact gene number of the parasite is unknown at present. The majority of the unique EST sequences had similarities to known genes, making them more amenable to functional analysis. The EST sequences should enhance the effectiveness of molecular studies, especially for gene expression profiling and the analysis of genes involved in virulence and infectivity. Microarrays can now be designed using either cDNA microarray or oligo-based platforms using the EST information. Additionally, the cluster and redundancy information should be useful for further subtraction of the most abundant transcripts included in the cDNA library, making further EST analysis in the parasite more effective.

Materials and Methods

Samples

The source of mRNA for this analysis was derived and expanded from a single parasite cloned from the infected fish. The source of the original *Ichthyophthirius multifiliis* was isolated from an infected fish obtained from a local pet shop and the parasite was transmitted to channel catfish held in a 50-l glass aquarium at the USDA-ARS Aquatic Animal Health Research Laboratory, Auburn, AL. The transmission of *I. multifiliis* was achieved through co-habitation of the infected fish with two fingerlings of channel catfish (3 inches in size). When the two catfish were infected, the symptoms of Ich, white spots, started to emerge when trophonts were collected by scraping with a glass slide. Channel catfish infected with maturing trophonts were rinsed in dechlorinated water and the skin was gently scraped to dislodge the parasites. Trophonts were harvested by filtering through a 0.22 μm filter to remove fish skin. The trophonts were placed into a Petri dish to allow them to develop into theronts that were used to infect 8 fish each for the collection of trophonts, tomonts, and theronts, respectively. Trophonts were directly collected from the skin surface of the 8 infected fish. To collect tomonts and theronts, trophonts isolated from fish were placed in Petri dishes and allowed to attach. After replacing the water in the Petri dishes with fresh dechlorinated water to remove contaminating mucus, the trophonts were incubated at 24°C for 8 h to harvest tomonts (32–128 cells/cyst) or 24 h to harvest theronts. Trophonts, tomonts and theronts were washed with PBS (pH 7.4), concentrated with a centrifuge (Beckman Coulter, Inc., Miami, FL) at 228 \times g for 5 min and discarded supernatant. After washing 3 times with PBS, parasite samples from the three life stages were stored in liquid nitrogen and used for the isolation of RNA for the construction of normalized cDNA library.

RNA isolation

Total RNA was isolated from the samples using the TRIzol reagent method from Invitrogen (Carlsbad, CA, USA) according to manufacturer's instructions. Briefly, samples of tomont, theront, and trophont were resuspended after thawing on ice, and 100 μ l each were combined in a sterile tube to provide a total of 300 μ l of Ich samples with equal fractions from each of its three life stages. As the major objective of this study was to generate EST resources with maximal efficiency of gene discovery, a pooled sample followed with normalization would allow inclusion of all transcripts in the library while reducing cost for library construction and increasing gene discovery rate. Three milliliters of TRIzol reagent was added to the sample tube. Cells were lysed by repetitively pipetting up and down. RNA was isolated following the manufacturer's protocol. The RNA pellet was resuspended in 100 μ l of RNase-free double distilled water and divided into 25 μ l aliquots. RNA aliquots were checked for quality using agarose gel electrophoreses containing formaldehyde.

Normalized library construction

The Creator Smart cDNA Library Construction Kit from Clontech (Mountain View, CA) and components from the TRIMMER-DIRECT Kit from Evrogen (Moscow, Russia) were used for the construction of the normalized cDNA library. Total RNA concentration was checked on a spectrophotometer and 1 μ g RNA was combined with 1 μ l of SMART IV oligonucleotide (Clontech) and 1 μ l CDS-3M adapter (Evrogen) for first-strand cDNA synthesis. The reaction was incubated at 72°C for 2 min followed by immediate cooling on ice for 2 min. Next, 2 μ l of 5 \times first strand buffer, 1 μ l of DTT (20 mM), 1 μ l of dNTP mix (10 mM), and 1 μ l of PowerScript reverse transcriptase were added to the tube and incubated at 42°C for 1 h in a thermal cycler

(PTC-100, Bio-Rad, Hercules, CA) then placed on ice. The SMART cDNA cloning system allows the enrichment of full-length cDNA through the use of a 5'-linker with 3'-GGG tails. Reverse transcriptase has terminal transferase activity that preferentially adds three additional Cs at the end of first strand cDNA. As a result, the first strand cDNA is able to base pair with the 5'-linker with 3'-GGG tails. Once base paired, the reverse transcriptase would switch the template and extend into the linker sequences allowing PCR amplification of full-length cDNA using a single primer (the 5'-linker has the same sequences as the linker containing poly T used for the synthesis of the first strand cDNA). Truncated cDNAs are not able to base pair with the 5'-linker and, therefore, get lost in the PCR amplification of the full-length cDNA.

The first strand cDNA was initially amplified by long-distance PCR (LD-PCR) using hot-start amplification. For the reaction, the following were combined in a reaction tube: 1.5 μ l of the first strand cDNA, 60 μ l of sterile deionized water, 7.5 μ l of 10 \times Advantage 2 PCR buffer, 1.5 μ l of 50 \times dNTP mix, 3 μ l of 5' PCR primer and 1.5 μ l of 50 \times Advantage 2 polymerase mix. The tube was mixed and briefly centrifuged and added to a pre-heated (95 $^{\circ}$ C) thermal cycler. Cycle settings were 95 $^{\circ}$ C for 1 min followed by 19 cycles of 95 $^{\circ}$ C for 7 s, 66 $^{\circ}$ C for 20 s, and 72 $^{\circ}$ C for 5.5 min. The product was analyzed on a 1.1% agarose gel to determine the sizes and amount of the cDNA products before proceeding to the next step. The LD-PCR reaction was purified and eluted in 30 μ l of sterile Nanopure water using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA).

For the normalization procedure, the TRIMMER-DIRECT Kit from Evrogen (Moscow, Russia) was used. This system is specially developed to normalize cDNA enriched with full length sequences [49, 50]. The cDNA from the LD-PCR was quantified [\sim 100 ng/ μ l] and 1 μ l was mixed with 1 μ l of the 4 \times hybridization buffer and 2 μ l of sterile water. The mix was

overlaid with mineral oil and incubated for 3 min at 98°C followed by 4 h at 70°C. Then, 5 µl of 2× DSN buffer (preheated to 70°C) and 0.25 Kunitz units of DSN enzyme were added and incubated at 70°C for 20 min. The DSN enzyme specifically degrades double-stranded molecules. The reaction was inactivated by adding 10 µl of DSN stop solution, and sterile water added to a final volume of 40 µl.

Following normalization, two rounds of PCR were performed using 1 µl of the normalization reaction as template. A shorter primer M1 (first 23 bases of the SMART IV oligonucleotide) was used in the first round of PCR with 15 amplification cycles using the same thermal cycling parameters as above; and an even shorter primer M2 (first 20 bases of the SMART IV oligonucleotide) was used in the second round of PCR for 15 amplification cycles of 95°C for 7 s, 64°C for 20 s, and 72°C for 5.5 min. Products were checked on a 1.1% agarose gel. The PCR products were quantified and 3 µg were used for treatment with proteinase K. All the subsequent procedures including proteinase K treatment, restriction digestion with Sfi I, size fractionation, and ligation followed the manufacturer's instructions (Clontech). The cDNA was ligated to the pDNR-LIB vector. Electroporation (MicroPulser, Bio-Rad, Hercules, CA) was performed using DH12S electrocompetent cells following supplier's instructions (Invitrogen). A total of approximately 700,000 primary recombinant clones were obtained, and the library was amplified, titered, and stored in glycerol stocks in a -80°C freezer.

Plasmid isolation and EST sequencing

Independent colonies were picked and grown for 20 h at 37°C in 1.2 ml LB broth containing 30 µg/ml chloramphenicol. Plasmid DNA was isolated using the Perfectprep Plasmid 96 Vacuum Direct Bind Kit from Eppendorf (Westbury, NY). Plasmids were stored at -20°C

until usage. The cDNA inserts were directionally sequenced from the 5'-end of the cDNAs using universal M13(-21) primer and the BigDye terminator sequencing kit version 3.1 from Applied Biosystems (Foster City, CA) on a 3130XL DNA analyzer (Applied Biosystems).

Sequence analysis

Base calling was performed using the Phred program [30, 31] at quality cut-off set at Phred 20. Raw sequences were then imported into the Vector NTI Advance 10 software (Invitrogen) and were subjected to trimming of vector sequences and 5'adapter sequences using default settings. Afterwards, poly (A) tails were trimmed where necessary and sequences less than 100 bases were removed. Contigs were built in Vector NTI ContigExpress using default settings. All unique sequences were compared to the GenBank database using BLASTX in the non-redundant (NR), Swiss-Prot, and *Plasmodium falciparum* 3D7 genome database. For comparison to the *Tetrahymena thermophila* SB210 genome, tBLASTx was used. The cut-off for sequence similarity used was E-value < 10⁻⁵ for all analyses. Ciliate nuclear translation code was used in the BLAST searches. Search results from genome comparisons were summarized using a Venn diagram.

Gene ontology (GO) annotations were assigned using the program Blast2GO [48]. BLASTX results were loaded into the program and the default settings were used to assign GO terms to all unique sequences. From these annotations, pie charts were made using 2nd level GO terms based on biological process, molecular function, and cellular component.

Putative secretory proteins and signal peptides were identified using both neural networks and hidden Markov model methods in SignalP 3.0 [43]. All 4,706 unique ESTs were used as the tester sequences. Open reading frames were predicted using both OrfPredictor [51] and

BLASTX, with ciliate nuclear genetic code for ESTs of known genes and just OrfPredictor with unknown ESTs. The resulting deduced protein sequences from the ORFs were uploaded into SignalP 3.0. Sequences were identified as putatively secretory, predicted with signal peptides if both D-score in the neural network model and prediction probability in the hidden Markov model were significant.

Total lengths of all ESTs, G+C% content and simple repetitive elements were estimated using the Repeatmasker program [52].

Accession numbers

All Ich EST sequences were submitted to the dbEST database of NCBI. Continuous accession numbers are from **EG957858–EG966289**.

Acknowledgements

This project was supported by a Specific Cooperative Agreement with USDA ARS Aquatic Animal Health Laboratory under the Contract Number 58-6420-5-030, and in part by CSREES from a grant of USDA NRI Animal Genome Tools and Resources Program (award # 2006-35616-16685). An equipment grant from the National Research Initiative Competitive Grant no. 2005-35206-15274 from the USDA Cooperative State Research, Education, and Extension Service also contributed. Special thanks to the co-authors of this study: Peng Xu, Ping Li, De-Hai Xu, Huseyin Kucuktas, Phillip Klesius, Covadonga Arias and Zhanjiang Liu.

References

1. Hines RS, Spira DT: **Ichthyophthiriasis in the mirror carp *Cyprinus carpio* (L.) V. Acquired Immunity.** *J Fish Biology* 1974, **6**:373-378.
2. MacLennan RF: **Observations on the life-cycle of *Ichthyophthirius multifiliis*, a ciliate parasitic on fish.** *Northwestern Scientist* 1935, **9**:12-14.
3. Nigrelli RF, Pokorny KS, Ruggieri GD: **Notes on *Ichthyophthirius multifiliis*, a ciliate parasitic on fresh-water fishes, with some remarks on possible physiological races and species.** *Trans Am Microsc Soc* 1976, **95**(4):607-613.
4. Matthews RA: ***Ichthyophthirius multifiliis* Fouquet and ichthyophthiriosis in freshwater teleosts.** *Adv Parasitol* 2005, **59**:159-241.
5. Dickerson HW, Clark TG, Findly RC: ***Ichthyophthirius multifiliis* has membrane-associated immobilization antigens.** *J Protozool* 1989, **36**(2):159-164.
6. Dickerson HW, Clark TG, Leff AA: **Serotypic variation among isolates of *Ichthyophthirius multifiliis* based on immobilization.** *J Eukaryot Microbiol* 1993, **40**(6):816-820.
7. Xu DH, Klesius PH: **Protective effect of cutaneous antibody produced by channel catfish, *Ictalurus punctatus* (Rafinesque), immune to *Ichthyophthirius multifiliis* Fouquet on cohabited non-immune catfish.** *J Fish Dis* 2003, **26**(5):287-291.
8. Xu DH, Klesius PH, Panangala VS: **Induced cross-protection in channel catfish, *Ictalurus punctatus* (Rafinesque), against different immobilization serotypes of *Ichthyophthirius multifiliis*.** *J Fish Dis* 2006, **29**(3):131-138.
9. Xu DH, Klesius PH, Shelby RA: **Immune responses and host protection of channel catfish, *Ictalurus punctatus* (Rafinesque), against *Ichthyophthirius multifiliis* after**

- immunization with live theronts and sonicated trophonts. *J Fish Dis* 2004, **27**(3):135-141.**
10. Xu DH, Klesius PH, Shoemaker CA: **Effect of lectins on the invasion of *Ichthyophthirius* theront to channel catfish tissue.** *Dis Aquat Organ* 2001, **45**(2):115-120.
 11. Xu DH, Klesius PH, Shoemaker CA: **Cutaneous antibodies from channel catfish, *Ictalurus punctatus* (Rafinesque), immune to *Ichthyophthirius multifiliis* (Ich) may induce apoptosis of Ich theronts.** *J Fish Dis* 2005, **28**(4):213-220.
 12. Burkart M, Clark T, Dickerson H: **Immunization of channel catfish, *Ictalurus punctatus* Rafinesque, against *Ichthyophthirius multifiliis* (Fouquet): killed versus live vaccines.** *J Fish Biol* 1990, **13**:401-410.
 13. Ekless L, Matthews R: ***Ichthyophthirius multifiliis*: axenic isolation and short-term maintenance in selected monophasic media.** *J Fish Dis* 1993, **16**:437-447.
 14. Houghton G, Healey LJ, Matthews RA: **The cellular proliferative response, humoral antibody response, and cross reactivity studies of *Tetrahymena pyriformis* with *Ichthyophthirius multifiliis* in juvenile carp (*Cyprinus carpio* L.).** *Dev Comp Immunol* 1992, **16**(4):301-312.
 15. Xu DH, Klesius PH: **Two year study on the infectivity of *Ichthyophthirius multifiliis* in channel catfish *Ictalurus punctatus*.** *Dis Aquat Organ* 2004, **59**(2):131-134.
 16. Deng Y, Dong Y, Thodima V, Clem RJ, Passarelli AL: **Analysis and functional annotation of expressed sequence tags from the fall armyworm *Spodoptera frugiperda*.** *BMC Genomics* 2006, **7**:264.

17. Gonzalez SF, Chatziandreou N, Nielsen ME, Li W, Rogers J, Taylor R, Santos Y, Cossins A: **Cutaneous immune responses in the common carp detected using transcript analysis.** *Mol Immunol* 2007, **44**(7):1664-1679.
18. Govoroun M, Le Gac F, Guiguen Y: **Generation of a large scale repertoire of Expressed Sequence Tags (ESTs) from normalised rainbow trout cDNA libraries.** *BMC Genomics* 2006, **7**:196.
19. Hagen-Larsen H, Laerdahl JK, Panitz F, Adzhubei A, Hoyheim B: **An EST-based approach for identifying genes expressed in the intestine and gills of pre-smolt Atlantic salmon (*Salmo salar*).** *BMC Genomics* 2005, **6**:171.
20. Kim TH, Kim NS, Lim D, Lee KT, Oh JH, Park HS, Jang GW, Kim HY, Jeon M, Choi BH *et al*: **Generation and analysis of large-scale expressed sequence tags (ESTs) from a full-length enriched cDNA library of porcine backfat tissue.** *BMC Genomics* 2006, **7**:36.
21. La Claire JW, 2nd: **Analysis of expressed sequence tags from the harmful alga, *Prymnesium parvum* (Prymnesiophyceae, Haptophyta).** *Mar Biotechnol (NY)* 2006, **8**(5):534-546.
22. Perez F, Ortiz J, Zhinaula M, Gonzabay C, Calderon J, Volckaert FA: **Development of EST-SSR markers by data mining in three species of shrimp: *Litopenaeus vannamei*, *Litopenaeus stylirostris*, and *Trachypenaeus biridy*.** *Mar Biotechnol (NY)* 2005, **7**(5):554-569.
23. Vizcaino JA, Gonzalez FJ, Suarez MB, Redondo J, Heinrich J, Delgado-Jarana J, Hermosa R, Gutierrez S, Monte E, Llobell A *et al*: **Generation, annotation and analysis of ESTs from *Trichoderma harzianum* CECT 2413.** *BMC Genomics* 2006, **7**:193.

24. Brenner ED, Katari MS, Stevenson DW, Rudd SA, Douglas AW, Moss WN, Twigg RW, Runko SJ, Stellari GM, McCombie WR *et al*: **EST analysis in *Ginkgo biloba*: an assessment of conserved developmental regulators and gymnosperm specific genes.** *BMC Genomics* 2005, **6**:143.
25. Gandhe AS, Arunkumar KP, John SH, Nagaraju J: **Analysis of bacteria-challenged wild silkmoth, *Antheraea mylitta* (lepidoptera) transcriptome reveals potential immune genes.** *BMC Genomics* 2006, **7**:184.
26. Hecht J, Kuhl H, Haas SA, Bauer S, Poustka AJ, Lienau J, Schell H, Stiege AC, Seitz V, Reinhardt R *et al*: **Gene identification and analysis of transcripts differentially regulated in fracture healing by EST sequencing in the domestic sheep.** *BMC Genomics* 2006, **7**:172.
27. Nelson RT, Shoemaker R: **Identification and analysis of gene families from the duplicated genome of soybean using EST sequences.** *BMC Genomics* 2006, **7**:204.
28. Ribichich KF, Georg RC, Gomes SL: **Comparative EST analysis provides insights into the basal aquatic fungus *Blastocladiella emersonii*.** *BMC Genomics* 2006, **7**:177.
29. Boguski MS, Lowe TM, Tolstoshev CM: **dbEST--database for "expressed sequence tags".** *Nat Genet* 1993, **4**(4):332-333.
30. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res* 1998, **8**(3):186-194.
31. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment.** *Genome Res* 1998, **8**(3):175-185.

32. Cao D, Kocabas A, Ju Z, Karsi A, Li P, Patterson A, Liu Z: **Transcriptome of channel catfish (*Ictalurus punctatus*): initial analysis of genes and expression profiles of the head kidney.** *Anim Genet* 2001, **32**(4):169-188.
33. Ju Z, Karsi A, Kocabas A, Patterson A, Li P, Cao D, Dunham R, Liu Z: **Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain.** *Gene* 2000, **261**(2):373-382.
34. Karsi A, Cao D, Li P, Patterson A, Kocabas A, Feng J, Ju Z, Mickett KD, Liu Z: **Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin.** *Gene* 2002, **285**(1-2):157-168.
35. Kocabas AM, Li P, Cao D, Karsi A, He C, Patterson A, Ju Z, Dunham RA, Liu Z: **Expression profile of the channel catfish spleen: analysis of genes involved in immune functions.** *Mar Biotechnol (NY)* 2002, **4**(6):526-536.
36. Quilang J, Wang S, Li P, Abernathy J, Peatman E, Wang Y, Wang L, Shi Y, Wallace R, Guo X *et al*: **Generation and analysis of ESTs from the eastern oyster, *Crassostrea virginica* Gmelin and identification of microsatellite and SNP markers.** *BMC Genomics* 2007, **8**:157.
37. Wuitschick JD, Karrer KM: **Analysis of genomic G + C content, codon usage, initiator codon context and translation termination sites in *Tetrahymena thermophila*.** *J Eukaryot Microbiol* 1999, **46**(3):239-247.
38. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM *et al*: **Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote.** *PLoS Biol* 2006, **4**(9):e286.

39. **The National Center for Biotechnology Information** [<http://www.ncbi.nlm.nih.gov/>].
40. O'Donnell RA, Blackman MJ: **The role of malaria merozoite proteases in red blood cell invasion.** *Curr Opin Microbiol* 2005, **8**(4):422-427.
41. Prato M, Giribaldi G, Polimeni M, Gallo V, Arese P: **Phagocytosis of hemozoin enhances matrix metalloproteinase-9 activity and TNF-alpha production in human monocytes: role of matrix metalloproteinases in the pathogenesis of *falciparum* malaria.** *J Immunol* 2005, **175**(10):6436-6442.
42. Tosini F, Trasarti E, Pozio E: **Apicomplexa genes involved in the host cell invasion: the Cpa135 protein family.** *Parassitologia* 2006, **48**(1-2):105-107.
43. Bendtsen JD, Nielsen H, von Heijne G, Brunak S: **Improved prediction of signal peptides: SignalP 3.0.** *J Mol Biol* 2004, **340**(4):783-795.
44. Clark TG, McGraw RA, Dickerson HW: **Developmental expression of surface antigen genes in the parasitic ciliate *Ichthyophthirius multifiliis*.** *Proc Natl Acad Sci U S A* 1992, **89**(14):6363-6367.
45. Van Den Bussche RA, Hooper SR, Drew CP, Ewing MS: **Characterization of histone H3/H4 gene region and phylogenetic affinity of *Ichthyophthirius multifiliis* based on H4 DNA sequence variation.** *Mol Phylogenet Evol* 2000, **14**(3):461-468.
46. Wright AD, Lynn DH: **Phylogeny of the fish parasite *Ichthyophthirius* and its relatives *Ophryoglena* and *Tetrahymena* (Ciliophora, Hymenostomatia) inferred from 18S ribosomal RNA sequences.** *Mol Biol Evol* 1995, **12**(2):285-290.
47. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: **Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.** *Nat Genet* 2000, **25**(1):25-29.

48. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674-3676.
49. Liu Z: **Transcriptome characterization through the generation and analysis of expressed sequence tags: Factors to consider for a successful EST project.** *Israel J Aquaculture - Bamidgeh* 2006, **58**(4):328-340.
50. Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz MV, Meleshkevitch E, Moroz LL, Lukyanov SA *et al*: **Simple cDNA normalization using kamchatka crab duplex-specific nuclease.** *Nucleic Acids Res* 2004, **32**(3):e37.
51. Min XJ, Butler G, Storms R, Tsang A: **OrfPredictor: predicting protein-coding regions in EST-derived sequences.** *Nucleic Acids Res* 2005, **33**(Web Server issue):W677-680.
52. Smit A, Hubley R, Green P: **RepeatMasker Open - 3.0.** [\[http://www.repeatmasker.org\]](http://www.repeatmasker.org). 1996.

**IV. TRANSCRIPTOMIC PROFILING OF *ICHTHYOPHTHIRIUS MULTIFILIIS*
REVEALS POLYADENYLATION OF THE LARGE SUBUNIT RIBOSOMAL RNA**

Abstract

Polyadenylation of eukaryotic transcripts is usually restricted to mRNA, providing transcripts with stability from degradation by nucleases. Conversely, an RNA degradation pathway can be signaled through poly (A) tailing in prokaryotic, archeal, and organellar biology. Recently polyadenylated transcripts have also been discovered in rRNA in some eukaryotes including humans and yeast. Here is reported the discovery of polyadenylated rRNAs in the ciliate teleost parasite *Ichthyophthirius multifiliis*, an important fish pathogen. Through large-scale analysis of ESTs, a large contig composed of the 28S rRNA with poly (A) tails was identified. Analysis using multiple sequence alignments revealed four potential polyadenylation sites including three internal regions and the 3' end of the rRNA. Further analysis using a polyadenylation test, re-sequencing, and gene-specific PCR using primers flanking the presumed poly (A) sites confirmed the presence of polyadenylated rRNA in this parasite. The functions of polyadenylation of rRNA in this organism are largely unknown at present, but the presence of internal polyadenylation sites, along with the presence of truncated segments of the rRNA, may suggest a role of the polyadenylation in the degradation pathway, a function typical of prokaryotes, archaea, and organelles. These results are in congruence with reports of a similar phenomenon in humans and yeast.

Background

Three main classes of RNA exist in eukaryotes consisting of ribosomal RNA (rRNA), messenger RNA (mRNA), and transfer RNA (tRNA). In protein synthesis, rRNA and ribosomal proteins form the ribosomes where protein synthesis occurs; mRNAs encode the primary sequence of proteins for which the proper amino acids are transferred to the growing peptide chain by tRNA. In addition to these three major classes of RNAs, several other classes of RNAs have been reported including hybrid transfer-messenger RNA (tmRNA), small nuclear and small nucleolar RNA (snRNA and snoRNA), and micro-RNA (miRNA).

Important features of RNA are the modifications that occur during or immediately after transcription; in particular 5' methylated reverse capping and 3' polyadenylation. Polyadenylation is evident not only in the mRNA of all domains of eukaryotes but also in some organellar RNA: mitochondrial RNA (mtRNA) or chloroplast RNA (cpRNA). Typically, eukaryotic mRNA is polyadenylated on the mature transcript with implications for stability, signaling, and nuclear transport in compartmentalized cells, and can also serve as a binding domain [1-5]. In contrast, polyadenylation of mRNA in prokaryotes and archaea, as well as polyadenylation with certain mtRNA or cpRNA, signals RNA degradation [6-9]. In recent RNA examinations, exceptions to the stability versus decay rule are rising and coexisting mechanisms seem to occur [9-14].

Polyadenylation is typically thought to be exclusive to translated RNA. Conversely, non-translated rRNAs have recently been shown to exhibit polyadenylation in some eukaryotes. Discoveries have been forthcoming of rRNA polyadenylation in humans (Slomovic et al., 2006), yeasts *Candida albicans* and *Saccharomyces cerevisiae* [15-17], and the protozoan Kinetoplastid

Leishmania [18]. Here is reported the polyadenylation of the large subunit rRNA in the ciliate protozoan *Ichthyophthirius multifiliis* (Ich).

The ciliate protozoan *I. multifiliis* is a devastating freshwater teleost parasite. The parasite is widespread; it affects many freshwater fish species around the world. It causes great loss to both the aquaculture and ornamental fish industries. Ich is responsible for the disease ichthyophthiriosis or ‘white spot’ disease with characteristic white spot cysts forming under the fish gill or epidermis. Also, typical of unicellular members of the phylum *Ciliophora*, Ich exhibits nuclear dimorphism. It possesses the characteristic macronucleus and micronucleus. Ich has been shown to be closely related phylogenetically to species of *Tetrahymena* [19-21] such as *T. thermophila*, a free-living non-parasitic ciliate protozoan. From a comparative view to *Tetrahymena* [20, 22], Ich may also undergo both germ line and somatic cell divisions, generally where the micronucleus contains the germ line DNA that can be replenished through meiotic conjugation, and the macronuclear DNA that is transcriptionally active to support the cell. There has been evidence for asexual reproduction in Ich for many years, with possible evidence that some sexual regeneration may occur [for a review see [23]]. One such piece of evidence arises from the major difficulties encountered in maintaining viable isolates of Ich in the laboratory setting. Multiple passages of an Ich isolate on a fish host lead to a significant decline in infectivity, relating to senescence [23, 24]. It is speculated that this type of induced senescence could be due to a lack of recombination of the germ line. Therefore, undertaking a study of senescence-related genes would be an important step in better understanding Ich development, reproduction, and parasitic nature.

A set of highly abundant transcripts containing poly (A) tails at the 3' end was initially identified during EST analysis [19]. BLASTX similarity comparisons of these transcripts yielded

hypothetical proteins related to senescence proteins in antisense orientation, with low to moderate similarity ($e^{-6} \geq E\text{-value} \geq e^{-78}$ for various segmented alignments). These transcripts were studied further to test for senescence-related expression, and it was found that the antisense transcripts did not exist using Northern blot analysis, but its related sense strand transcripts were highly expressed in all three stages and all ages of Ich tested. Such results led to more extensive BLAST searches using both BLASTX and BLASTN. BLASTN searches revealed that these transcripts were highly similar (E-value = 0) to ribosomal RNAs, in particular to portions of the large subunit ribosomal RNA in many organisms, including closely related *Tetrahymena* species, or to precursor-rRNA in *T. thermophila* with alignment to the large subunit. However, the putative identity as rRNA is in contradiction of the fact that the transcripts were found to be polyadenylated. Therefore, the transcripts were further characterized to obtain more of their sequences and test for polyadenylation. Here is presented the 28S rRNA sequence from the ciliate protozoan *I. multifiliis*, and evidence is provided for polyadenylation of the large subunit rRNA.

Results

Identification of the largest contig in the EST assembly as the 28S rRNA

The overall strategy for the demonstration of the presence of polyadenylated 28S rRNA in *I. multifiliis* is shown in Fig. 3. As introduced above, initial BLASTX analysis generated several significant hits to existing proteins named senescence-related proteins, but in an antisense orientation. However, Northern blot analysis revealed that there was no antisense RNA present. Further BLAST analysis was conducted using both BLASTX and BLASTN. The BLASTN searches revealed that the identity of the contig under study was the large subunit of rRNA (Table 4). All E-values for the top 100 megablast hits were zero and identities were $\geq 81\%$. As expected, the 26S rRNA of the closely related ciliate *Tetrahymena pyriformis* and *T. thermophila* had the highest similarity, with $\geq 91\%$ identity covering the entire rRNA transcripts. Sequence alignments further confirm the identity of the involved transcripts. Obviously, in the initial BLASTX searches, rRNA was not included in the protein database. However, this result also indicated the polyadenylation of the 28S rRNA transcripts in Ich, an unusual but reported phenomenon in several organisms such as yeast and humans [15-17, 25]. A consensus sequence of the complete Ich 28S rRNA was generated by multiple sequence alignment using all ESTs within this contig, and the consensus sequence has been submitted to GenBank (accession number **EU185635**).

Figure 3. Flowchart of molecular analysis leading to the identification of the polyadenylated 28S rRNA.

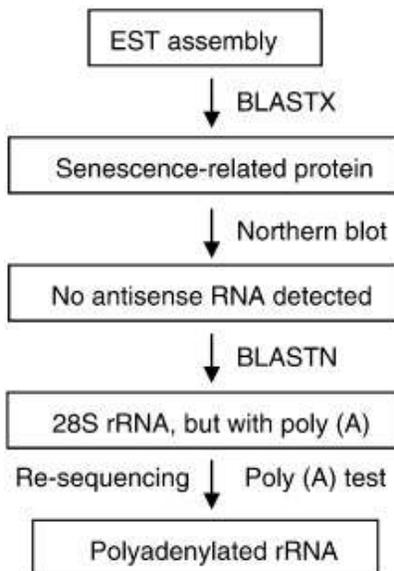


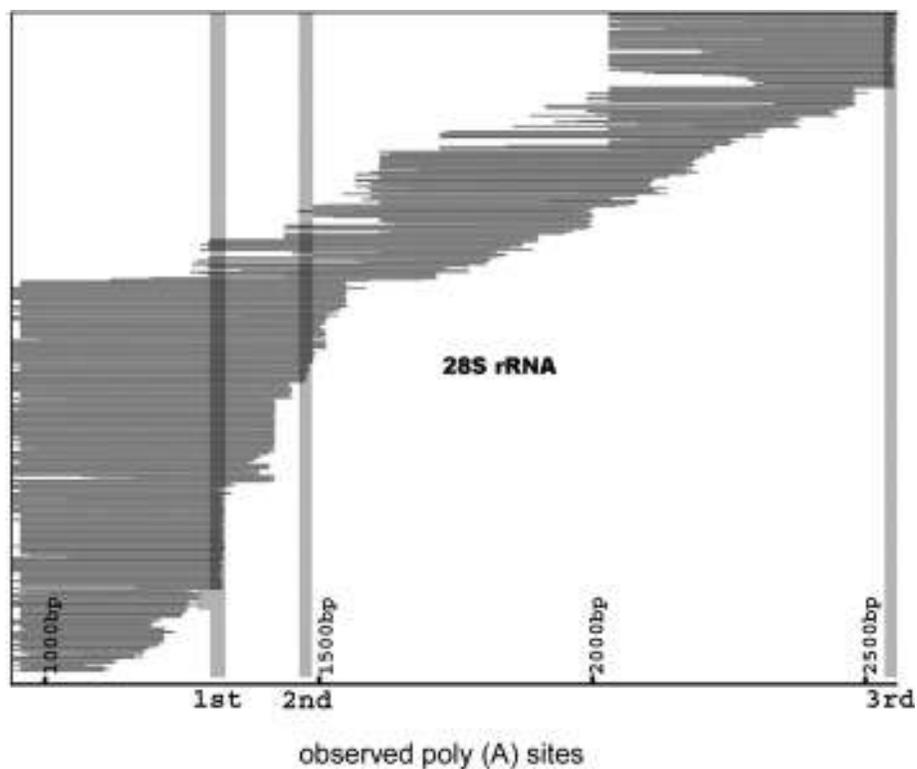
Table 4. The top 10 megablast BLASTN hits to the Ich transcript.

Accession number	Description	Identity	E-value
X54004	<i>Tetrahymena pyriformis</i> gene for 26S large subunit ribosomal RNA	92%	0
AY210458	<i>Crisia</i> sp. YJP-2003 28S ribosomal RNA gene, partial sequence	84%	0
X54512	<i>T. thermophila</i> rdn A+ gene for pre-RNA, 17S rRNA, 5.8S rRNA, 26S	91%	0
DQ273766	<i>Rozella</i> sp. JEL347 isolate AFTOL-ID 16 28S ribosomal RNA gene,	82%	0
DQ273770	<i>Rhizophyidium brooksianum</i> isolate AFTOL-ID 22 28S ribosomal RNA	81%	0
DQ273825	<i>Rhizophyidium</i> sp. PL 42 isolate AFTOL-ID 691 28S ribosomal RNA	81%	0
DQ273823	<i>R. macroporosum</i> isolate AFTOL-ID 689 28S ribosomal RNA gene,	81%	0
AF149979	<i>Paramecium tetraurelia</i> macronuclear X gene, complete sequence	87%	0
XM_001471443	<i>T. thermophila</i> SB210 hypothetical protein (TTHERM_02141639)	91%	0
DQ273835	<i>Rhizophyidium</i> sp. JEL316 isolate AFTOL-ID 1535 28S ribosomal RNA	86%	0

Polyadenylation of the Ich large subunit rRNA

The initial contig involving the transcript under study was assembled using ESTs previously generated [19] and is depicted in Fig. 4. A total of 207 (27%) polyadenylated ESTs were found in this contig of 764 ESTs (note that all these 764 ESTs could contain poly (A) tails, but because the ESTs were sequenced from the 5' end, many clones were not sequenced through the whole inserts). The regions containing the poly (A) tails fall into three regions, 1220–1279 bp, 1427–1450 bp, and 2468–2501 bp along the consensus sequence (Fig. 4). Further searches of GenBank (now containing over 33,000 EST sequences) allowed the identification of additional polyadenylated sequences that fall under these three regions, as well as polyadenylated sequences containing poly (A) tails at the real 3' end of the rRNA transcripts. These results suggested the presence of polyadenylated 28S rRNA with poly (A) tails both at the 3' end of the transcript and at three internal sites of the transcript.

Figure 4. Schematic presentation of the internal polyadenylation sites. Initial contig was assembled using Ich ESTs generated from GenBank dbEST sequences from EG957858-EG966289. Vertical shaded regions labeled as 1st, 2nd, and 3rd on the X-axis represent regions of observed internal polyadenylation of ESTs. Nucleotide positions are given based on the complete rRNA consensus sequence assembled with ContigExpress (Vector NTI 10.3.0 software, Invitrogen) with poly (A) tails included. Poly (A) sites were not detected within the first 1000 bp of the rRNA sequences, and therefore the schematic presentation is truncated for better exhibition of the detected poly (A) sites.



Of the 207 observed clones with poly (A) tails, 113 clones had poly (A) at the first site; 8 clones had poly (A) at the second site; and 86 clones had poly (A) at the third site (Fig. 4). The majority of poly (A) tails had lengths of 9–23 bp after subtraction of the oligo (dT) primer of 12 bases. Sequence alignments of all the 764 clones clearly suggest that the poly (A) was not part of the gene internal sequences because many clones were sequenced through the poly (A) junction, but there were no poly (A/T) being observed. Furthermore, the consensus sequence [without poly (A)] aligned nicely with the large subunit ribosomal RNA from closely related species (not shown). All these direct sequence data provide very strong evidence that polyadenylation was real rather than being derived from artifacts due to internal priming of the oligo (dT) primers during cDNA synthesis.

A subset of the original clones was re-sequenced to test if the insert contained a poly (A) tail at the internal sites. Out of the 10 clones chosen for re-sequencing, 7 were found to contain poly (A) tails after stringent post-sequencing processing. All 3 internal sites were represented by these 7 clones. BLASTN was performed on each sequence, and all were determined to be a portion of the 28S rRNA gene. These sequences were then re-clustered against the consensus 28S sequence, and it was clear that these transcripts align to the 28S. As with the original 764 EST cluster described above, these sequences produced an alignment file with a consensus sequence containing transcripts with poly (A) overhangs (Fig. 5).

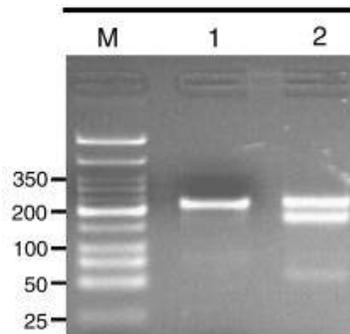
Figure 5. Graphical results from re-sequencing clones at internal polyadenylated sites. The top sequence is the 28S rDNA consensus sequence (GenBank accession no. EU185635). Numbers along the top represent the base-pair position along the consensus 28S sequence. A total of 7 out of 10 clones that were re-sequenced contained a poly (A) tail after base-calling at > 99% accuracy using phred > Q20. The 1st, 2nd, and 3rd polyadenylation regions are each represented. Double bars indicate areas where the sequence was shortened for brevity.



In addition to direct sequence alignments, additional tests were conducted to demonstrate the presence of polyadenylation in 28S rRNA transcripts. First, a polyadenylation test [26] was used to determine if the rRNA was truly 3' polyadenylated, or the sequenced ESTs were just an artifact. PCR amplification was conducted using one rRNA gene-specific primer and one poly (T) primer, with the gene-specific primer designed proximal to the suspected poly (A) sites. The 3' end of the 28S, determined after clustering and alignments with other similar species, was tested for potential polyadenylation with the polyadenylation test at the consensus sequence position 2,677 bp (the full 3' end). As shown in Fig. 6, the PCR product was amplified and correctly digested. The PCR product had the correct sizes as predicted from the potential poly (A) site, confirming the correctness of the consensus sequence and the presence of polyadenylated transcripts within the pool of the 28S rRNA transcripts.

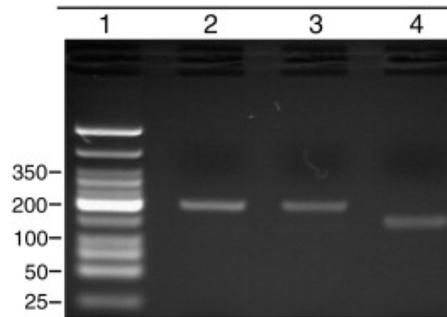
Figure 6. Results of the polyadenylation test for polyadenylation at the 3' end of the rRNA.

Lane 1, undigested amplicon of the polyadenylation test; lane 2, the amplicon digested with restriction endonuclease *Acu* I, note that not all the amplified products were digested completely; and M, low molecular weight marker from New England Biolabs (NEB #N3233).



The poly (A) test is widely used in the literature for the validation of the presence of polyadenylation at the 3' end of the transcripts [15-17, 25, 26]. However, for internal polyadenylation sites, the argument can be made that the PCR is supported by internal priming of the oligo (dT) primers. Therefore, another set of PCR tests was conducted to demonstrate that the observed poly (A) tails were not derived from internal priming of the oligo (dT) primers during cDNA synthesis. In this experiment, PCR primers were designed flanking the three presumed poly (A) sites such that the amplified segments would contain the poly (A) region. If the 28S rRNA gene contained internal poly (A/T) sequences, the PCR products would be longer than the consensus sequence. As shown in Figure 7 and Table 5, the PCR products were exactly the same size as predicted from the consensus sequence without internal poly (A/T) sequences, providing strong supporting evidence that the poly (A)-containing clones from the EST analysis were not derived from internal priming.

Figure 7. PCR test for the presence of polyadenylation using gene-specific primers flanking each of the three internal poly (A) sites. Lane 1, low molecular weight (in bp) marker from NEB; lane 2–4, PCR products from the first, second, and third internal poly (A) sites, respectively.



Discussion

Polyadenylation typically occurs with mRNA of eukaryotic transcripts. Recent studies, however, have discovered polyadenylation in non-translated ribosomal RNA in a few organisms [15-18, 25]. In humans and yeast, polyadenylated rRNA was reported for both 18S and 28S rRNAs [15-17, 25]. In the protozoan parasite *Leishmania*, polyadenylation was found at several positions within the large subunit rRNA [18]. Here is reported the polyadenylation of the large subunit rRNA in the ciliate protozoan parasite *I. multifiliis*, at both the 3' end and three internal positions within the 28S rRNA. A current review of the literature suggests this is the first report of this phenomenon in ciliate parasites.

The strongest supporting evidence for the presence of polyadenylated 28S rRNA comes from multiple sequence alignment where a consensus sequence can be assembled based on several hundreds of sequences, and among them, over 270 sequences were found to harbor poly (A) tails. The presence of EST sequences of clones with poly (A) at the 3' end, and also clones spanning through the three internal poly (A) sites but without any poly (A) tract clearly demonstrates a polyadenylation in the internal sites. In order to confirm the polyadenylation of the 28S rRNA at the 3' end and internally, a polyadenylation test was conducted using PCR as well as re-sequencing. The positive result of the polyadenylation test PCR provided further evidence for polyadenylation of the 3' end of the 28S rRNA in *I. ch.* Re-sequencing of internal clones helped determine the presence of polyadenylated transcripts at internal sites as well. The polyadenylation test and re-sequencing analysis showed that this phenomenon was not simply an artifact of the original data.

The poly (A) test, however, still uses oligo (dT) primers. Arguments can be made that the PCR in the poly (A) test is supported by internal priming of the oligo (dT) primers. However, PCR reactions using gene-specific primers designed flanking the poly (A) sites generated products with sizes exactly the same as predicted from the consensus sequence without internal poly (A/T) sequences, providing strong support evidence that the poly (A)-containing clones from the EST analysis were not derived from internal priming.

This study demonstrated that polyadenylation within the 28S rRNA of *Ich* occurs at three internal positions of the complete transcript as well as at the 3' end of the transcripts. It is not clear why polyadenylation occurs at both internal and end of the transcripts. One possibility is that polyadenylation could serve dual functions as both a stability signal and degradation signal. For mRNA, it is widely believed that polyadenylation protects the mRNA from degradation in eukaryotes [1-3, 5]. However, with rRNA, polyadenylation could signal the degradation of the molecules, especially when they are present as truncated transcripts [9, 11, 12, 14, 16, 17, 25].

One obvious question is why polyadenylation occurs at the specific sites as observed. Are the processes sequence-dependent? Alignment of the sequence intervals between the observed polyadenylation positions failed to reveal any conserved sequences (Fig. 8) nor any conserved secondary structures, suggesting that it is sequence independent, as similarly reported in humans [25].

Figure 8. Sequence alignment of the 180-bp regions immediately upstream to each of the observed polyadenylation sites using ClustalW. The 1st, 2nd, and 3rd regions correspond to 1st, 2nd, and 3rd regions of polyadenylation sites in Fig. 4, whereas 3' end indicates sequences immediately upstream from the 3' end of the rRNA. Black highlights are conserved nucleotides and grey highlights are similar nucleotides. No shading indicates non-similarity.

```

1st_region 1 ASAGCAGCTCACCCCAAGAGCTOTCAGTTCCCATCTGATGACCCCTSAARATCTG--GGGAGGCTAAGTATACACCCAGGT-CGTACCCCTAACCCCATC-----G
2nd_region 1 -----CATATTGGATCCCTAAGCTTCGGAAAAGGATTCGCTCTGAGGATGGGGTGAATGGGGCTTACATGAACTCTAAGCTTGGCATGGGCTGGAAAC---A
3rd_region 1 -----TGAAGAAATGATATTCCACCTATTTAACTTATCTG-----GAGGAGCCGAGAACAGTAAATGGTAATACGGTTCCTGAAAGACAA
3'end      1 -CTTAAGTTCGAATTGGTCCGAGTGGCTCAAGACGAAAGAGTCCAGGCTCATC----GCATGAACTCGAATTCGAATTATCA-TCAACATGAATCTTTTGT-----A

1st_region 104 TCTCAAGGTTAGCAGCC--ICTGGCA--ATAGAACAA--GTAGATAAGGGAAGTCGGC--ATTTGGATCCTAACTTCGGG----- 188
2nd_region 93 TGCCTGCCATAGCTGCCGAGGAATT--GTAGAGTGGCTCCGGCTTCCTCCATCAATGANCATCAACTGAGAAGTGAAGC----- 188
3rd_region 92 TGCCTGAAAGCTACCATCTAAGAAATATGACTGAAAGGCCCTAAGTCAGAAATCTATGCTGGAAACCAATGTCTGGTGTATGATAAACG 188
3'end      101 CGAATTAGTTCAAGGAGGAGGATATGAAA--GTATGAGG--AGAAATTTTCTACGATCTGCTGAGTTTCATCCCTTTTCCTAAG----- 188

```

Conclusion

While the roles and mechanisms of polyadenylation in rRNA are unknown at present, polyadenylation of eukaryotic messages has a role in transcript stability, among other functions. Interestingly, there is evidence that the contrast is true of polyadenylation in some organellar RNA as well as prokaryotic and archeal messages; polyadenylation in these systems stimulates a decay or degradation cascade [6-9]. Additionally, polyadenylation could paradoxically serve as a signal for decay and to increase stability of a mature transcript in the same organism, specifically in humans and possibly in yeasts [9-14]. In *Ich*, there is perhaps evidence of a similar dual function for polyadenylation. The evidence supporting polyadenylation as a signal for degradation was the detection of truncated polyadenylated transcripts along various regions of the 28S rRNA, suggesting that many of these transcripts could be captured as intermediary degradation fragments. Since the contigs were assembled using normalized sequences, the scope of polyadenylated rRNAs cannot be assessed *in vivo*. Further studies are needed to reveal the mechanisms and functions of polyadenylation in rRNAs of *Ich*.

Materials and Methods

Cells, RNA isolation, and analysis

A single Ich isolate was obtained from a local pet shop with an outbreak of ichthyophthiriosis. Ich was cultured on fish as previously described [19]. Samples were washed in phosphate-buffered saline (PBS; pH 7.2) and flash-frozen in liquid nitrogen. Cells were stored in a – 80 °C freezer until usage. Total RNA was isolated from each of the three Ich life-stages using the RNeasy Plus Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol. The kit uses a spin column to eliminate DNA contamination. RNA quality was assessed by using denaturing agarose gel electrophoresis containing formaldehyde.

Sequence analysis

The Ich ESTs used for assembly of the initial contiguous sequences (contigs) are from GenBank dbEST accession numbers **EG957859-EG966289**.

The ESTs were generated by sequencing a unidirectional cDNA library enriched for full-length cDNAs [19, 27, 28]. All ESTs were sequenced from the 5' end. The sequences were assembled into contigs using ContigExpress in the Vector NTI version 10.3.0 software (Invitrogen, Carlsbad, CA, USA) with an overlap length cutoff set at 40 bp and sequence identity cutoff of the overlap set at 90%. All other settings were at the default values. The contig used for this study was the largest contig in the assembly containing 764 ESTs. The contig was visually inspected for polyadenylated regions. During the course of this study, EST resources grew significantly in the dbEST database of the GenBank. Using the latest release, (containing 25,084

ESTs), the contigs were assembled to include the mature 3' end of the rRNA molecule as compared with the top 4 megablast hits.

Sequence identity was determined using BLASTN at the National Center for Biotechnology Information (NCBI). Multiple sequence alignments were conducted using ClustalW with the 2,677 bp Ich consensus sequence from the contig (herein 28S rRNA), along with the top homologous sequences generated from BLAST searches.

Northern blot

Two separate Northern blot analyses were performed on the three life-stages of Ich. Briefly, total RNA was isolated using the RNeasy Plus kit according to the manufacturer's provided protocol. Total RNA (3 µg each of the three Ich life-stages) was separated by electrophoreses on a 1% agarose gel containing formaldehyde. The gel was UV visualized, then rinsed twice for 10 min in DEPC-treated water to destain and remove excess formaldehyde. The gel was dipped in 20× SSC buffer. A downward capillary transfer was performed for 4 h using positively charged nylon membranes (Millipore, Bedford, MA) and 20× SSC buffer. After transfer, the membranes were dipped in 20× SSC buffer, and RNA was fixed to the membrane by UV-crosslink using the UV Stratalinker 2400 (Stratagene, La Jolla, CA, USA) using the auto-crosslink setting. The membranes were pre-hybridized in 5 mL ULTRAhyb-Oligo buffer (Ambion, Austin, TX, USA) for 1.5 h at 42 °C. Labeled probes were added and hybridized overnight at 42 °C in a hybridization oven. The membranes were washed twice for 30 min each in a wash buffer (2× SSC, 0.5% SDS) and exposed to X-ray film for 4 h at room temperature. As the initial BLASTX analysis indicated the contig under consideration was similar to senescence proteins, but in an antisense orientation, two probes were used to determine the presence of

sense-strand transcript (using an antisense probe: 5'-GACCAGAGGCTGCTAACCTTGGAGACCTGATGCGGTTATG-3') and antisense transcript (using a sense probe: 5'-CATAACCGCATCAGGTCTCCAAGGTTAGCAGCCTCTGGTC-3'). The probes were end-labeled using ATP [γ -³²P] (MP Biomedicals, Solon, OH, USA).

Testing for polyadenylation at the full 3' end

One-hundred nanograms of total RNA from each of the three Ich life-stages was pooled and reverse transcribed using an oligo (dT) adapter primer 5'-GGTGAGCCCGCGTCACGG(T)₁₂-3' as designed elsewhere [18] using the SuperScript First-Strand Synthesis System for RT-PCR (Invitrogen) according to the manufacturer's protocol. A polyadenylation test was conducted as previously described [26]. Briefly, RNA is reverse transcribed using an oligo (dT) adapter primer. This reaction was used as a template for the polyadenylation test. To test for polyadenylation, a PCR reaction was performed where the oligo (dT) adapter primer and a gene-specific primer designed close to the polyadenylated site, were used to amplify the gene. The oligo (dT) adapter primer had the ability to hybridize at any position along the poly (A) tail. In this regard, the polyadenylation test allowed the determination of the presence/absence as well as size differentiation of the poly (A) tail [26]. The amplicon is subsequently digested by a restriction enzyme with a unique restriction site close to the 5' end and analyzed on an agarose gel. For the analyses, a 6-bp cutter (*Acu* I, New England Biolabs, Ipswich, MA) was chosen as a restriction enzyme for its uniqueness and position in the amplicon. Restriction enzyme digestion reactions were conducted under standard conditions according to the manufacturer's instructions.

For Ich, the polyadenylation test was used to test polyadenylation at the full 3' end of the 28S rRNA. Primers for the polyadenylation test were designed to be located at nucleotide position 2,482–2,502 of the consensus sequence. At this position, the test would help to determine if the added 3' end was truly part of the gene at the 3' end after clustering, and test for polyadenylated transcripts. One-hundred nanograms of total RNA from each of the three Ich life-stages was pooled and reverse transcribed using the oligo (dT) adapter primer. This first-strand reaction was used as a template for the polyadenylation test. A PCR reaction was performed using reaction guidelines from the polyadenylation test [26] with a gene-specific forward primer (GSP1, 5'-TAAGCGCAAGCTTAAGTTCGA-3') and the oligo (dT) adapter primer as a reverse primer. The thermo profiles for the PCR reaction were: an initial heat treatment at 94 °C for 3 min, 94 °C for 30 s, 55 °C for 30 s, 72 °C for 3 min, for 35 cycles, followed by a final 7 min at 72 °C. Additional PCR cycles were added from the polyadenylation test in increments of 5 where needed to achieve an amplicon. The expected amplicon size was of a minimum size of 208 bp. The resulting product was digested with *Acu* I to produce a single band representing the 5' end of the amplicon (expected size of 58 bp) and a 3' polyadenylated end of at least 150 bp. The digested and undigested products were electrophoresed on a 2% agarose gel for visualization.

Testing for polyadenylation at internal sites

To test the internal sites of the 28S for polyadenylation, re-sequencing and re-analyzing some of the original clones that contained putative poly (A) tails and that clustered internally of the 28S was the strategy chosen. A total of 10 clones from a cDNA library [19] were randomly selected for re-sequencing to test whether these clones truly contained poly (A) tails, and to test for sequencing artifacts from the original trace files. All clones selected were isolated from

single bacterial colonies from $-80\text{ }^{\circ}\text{C}$ stocks, plasmid DNA obtained using the Qiaprep Spin Miniprep kit (Qiagen), and sequenced in triplicate using M13 (- 21) universal primer and BigDye v3.1 on an ABI 3130XL DNA analyzer according to the manufacturer's recommendation (Applied Biosystems, Foster City, CA). To give the highest confidence that the sequences contain poly (A) tails, post-sequencing processing included base-calling using Phred [29, 30] statistics at $> Q20$ ($> 99\%$ accuracy). Vector sequences were eliminated by VecScreen BLAST. The consensus of the triplicate sequence was used for each clone. These sequences were re-analyzed by clustering against the consensus 28S and using BLAST tools.

In addition to the direct sequencing analysis and informatic analysis, PCR analysis was also conducted on the region containing the presumed poly (A) sites using flanking PCR primers, with one PCR primer upstream and the other downstream of the presumed poly (A) sites. The rationale is that if poly (A/T) sequences do exist internally, then the size of the PCR products would be larger than the predicted size without internal poly (A/T) sequences based on the consensus sequences. PCR primer pairs were designed for the three presumed poly (A) sites, and PCR reactions were conducted with thermo cycling profiles of $94\text{ }^{\circ}\text{C}$ for 30 s, $55\text{ }^{\circ}\text{C}$ for 30 s, $72\text{ }^{\circ}\text{C}$ for 1 min with a total of 35 cycles. An initial 3 min of heat treatment at $94\text{ }^{\circ}\text{C}$ was used. The PCR products were analyzed on both 1.8% agarose gels and on 8% polyacrylamide gels. PCR primers used for the test are shown in Table 5.

Table 5. PCR primer sequences and product sizes assuming there are no internal poly (A/T) tracts as predicted from the consensus sequences.

	Upper primer	Lower primer	Predicted PCR product sizes	Observed PCR product sizes
First poly (A) site	GGTCTCCAAGGTTAGCAGC	AGGCCGAAGCCACTCTAC	200 bp	200 bp
Second poly (A) site	ACATGCCTGCGCATAAG	GTTGAATTGCGTCACTTTGA	200 bp	200 bp
Third poly (A) site	TGCCGTGAAGCTACCATC	GACTCTTTCGTCTTCAGCC	150 bp	150 bp

Acknowledgements

This project was supported in part by a Specific Cooperative Agreement with USDA ARS Aquatic Animal Health Laboratory under Contract Number 58-6420-5-030, and in part by CSREES from a grant of USDA NRI Animal Genome Tools and Resources Program (award # 2006-35616-16685). An equipment grant from the National Research Initiative Competitive Grant no. 2005-35206-15274 from the USDA Cooperative State Research, Education, and Extension Service also contributed. Special thanks to the co-authors of this study: De-Hai Xu, Ping Li, Phillip Klesius, Huseyin Kucuktas and Zhanjiang Liu.

References

1. Beelman CA, Parker R: **Degradation of mRNA in eukaryotes.** *Cell* 1995, **81**(2):179-183.
2. de Moor CH, Richter JD: **Translational control in vertebrate development.** *Inl rev cytol* 2001, **203**:567-608.
3. Edmonds M: **A history of poly A sequences: from formation to factors to function.** *Prog Nucleic Acid Res Mol Biol* 2002, **71**:285-389.
4. Mangus DA, Evans MC, Jacobson A: **Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression.** *Genome biol* 2003, **4**(7):223.
5. Shatkin AJ, Manley JL: **The ends of the affair: capping and polyadenylation.** *Nature Struct Biol* 2000, **7**(10):838-842.
6. Dreyfus M, Regnier P: **The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria.** *Cell* 2002, **111**(5):611-613.
7. Kushner SR: **mRNA decay in prokaryotes and eukaryotes: different approaches to a similar problem.** *IUBMB life* 2004, **56**(10):585-594.
8. O'Hara EB, Chekanova JA, Ingle CA, Kushner ZR, Peters E, Kushner SR: **Polyadenylation helps regulate mRNA decay in *Escherichia coli*.** *Proc Natl Acad Sci U S A* 1995, **92**(6):1807-1811.
9. Slomovic S, Laufer D, Geiger D, Schuster G: **Polyadenylation and degradation of human mitochondrial RNA: the prokaryotic past leaves its mark.** *Mol Cell Biol* 2005, **25**(15):6427-6435.

10. Kao CY, Read LK: **Opposing effects of polyadenylation on the stability of edited and unedited mitochondrial RNAs in *Trypanosoma brucei*.** *Mol Cell Biol* 2005, **25**(5):1634-1644.
11. LaCava J, Houseley J, Saveanu C, Petfalski E, Thompson E, Jacquier A, Tollervey D: **RNA degradation by the exosome is promoted by a nuclear polyadenylation complex.** *Cell* 2005, **121**(5):713-724.
12. Vanacova S, Wolf J, Martin G, Blank D, Dettwiler S, Friedlein A, Langen H, Keith G, Keller W: **A new yeast poly(A) polymerase complex involved in RNA quality control.** *PLoS Biol* 2005, **3**(6):e189.
13. West S, Gromak N, Norbury CJ, Proudfoot NJ: **Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells.** *Mol Cell* 2006, **21**(3):437-443.
14. Wyers F, Rougemaille M, Badis G, Rousselle JC, Dufour ME, Boulay J, Regnault B, Devaux F, Namane A, Seraphin B *et al*: **Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase.** *Cell* 2005, **121**(5):725-737.
15. Fleischmann J, Liu H: **Polyadenylation of ribosomal RNA by *Candida albicans*.** *Gene* 2001, **265**(1-2):71-76.
16. Fleischmann J, Liu H, Wu CP: **Polyadenylation of ribosomal RNA by *Candida albicans* also involves the small subunit.** *BMC Mol Biol* 2004, **5**:17.
17. Kuai L, Fang F, Butler JS, Sherman F: **Polyadenylation of rRNA in *Saccharomyces cerevisiae*.** *Proc Natl Acad Sci U S A* 2004, **101**(23):8581-8586.

18. Decuypere S, Vandesompele J, Yardley V, De Donckeri S, Laurent T, Rijal S, Llanos-Cuentas A, Chappuis F, Arevalo J, Dujardin JC: **Differential polyadenylation of ribosomal RNA during post-transcriptional processing in *Leishmania***. *Parasitol* 2005, **131**(Pt 3):321-329.
19. Abernathy JW, Xu P, Li P, Xu DH, Kucuktas H, Klesius P, Arias C, Liu Z: **Generation and analysis of expressed sequence tags from the ciliate protozoan parasite *Ichthyophthirius multifiliis***. *BMC Genomics* 2007, **8**(1):176.
20. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM *et al*: **Macronuclear Genome Sequence of the Ciliate *Tetrahymena thermophila*, a Model Eukaryote**. *PLoS Biol* 2006, **4**(9).
21. Wright AD, Lynn DH: **Phylogeny of the fish parasite *Ichthyophthirius* and its relatives *Ophryoglena* and *Tetrahymena* (Ciliophora, Hymenostomatia) inferred from 18S ribosomal RNA sequences**. *Mol Biol Evol* 1995, **12**(2):285-290.
22. Prescott DM: **The DNA of ciliated protozoa**. *Microbiol Rev* 1994, **58**(2):233-267.
23. Matthews RA: ***Ichthyophthirius multifiliis* Fouquet and Ichthyophthiriosis in Freshwater Teleosts**. *Adv Parasitol* 2005, **59**:159-241.
24. Xu DH, Klesius PH: **Two year study on the infectivity of *Ichthyophthirius multifiliis* in channel catfish *Ictalurus punctatus***. *Dis Aquat Org* 2004, **59**(2):131-134.
25. Slomovic S, Laufer D, Geiger D, Schuster G: **Polyadenylation of ribosomal RNA in human cells**. *Nucleic Acid Res* 2006, **34**(10):2966-2975.
26. Salles FJ, Richards WG, Strickland S: **Assaying the polyadenylation state of mRNAs**. *Methods* 1999, **17**(1):38-45.

27. Liu ZJ: **Transcriptome characterization through the analysis of expressed sequence tags**. In: Liu Z.J. (Ed.), 1st Ed. Aquaculture Genome Technologies, vol. 20. Blackwell Publishing, Ames, IA 2007, 339-354.
28. Liu ZJ: **DNA sequencing technologies**. In: Liu Z.J. (Ed.), 1st Ed. Aquaculture Genome Technologies, vol. 20. Blackwell Publishing, Ames, IA 2007, 463-474.
29. Ewing B, Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities**. *Genome Res* 1998, **8**(3):186-194.
30. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment**. *Genome Res* 1998, **8**(3):175-185.

**V. GENE EXPRESSION PROFILING OF *ICHTHYOPHTHIRIUS MULTIFILIIS*:
INSIGHTS INTO DEVELOPMENTAL AND SENESCENCE-ASSOCIATED
AVIRULENCE**

Abstract

The ciliate parasite *Ichthyophthirius multifiliis* (Ich) infects many freshwater fishes, causing white spot disease that leads to heavy economic losses to aquaculture and ornamental industries. Despite its economic importance, molecular studies examining fundamental processes such as life stage regulation and infectivity have been scarce. In this study, analysis of Ich expressed sequence tags (EST) was conducted; an oligo microarray using the EST information was developed, and gene expression profiling was carried out. All available ESTs of Ich were analyzed to obtain a unigene set. An oligo microarray was developed, and the microarray was used to examine gene expression patterns of the three life stages of Ich: infective theront, parasitic trophont, and reproductive tomont. The microarray was further used to explore expression of senescence-related genes. A total of 173 putative genes were found to be differentially expressed among all three life-stages. Examples of differentially expressed transcripts included immobilization antigens and epiplasmin, as well as various other transcripts involved in developmental regulation and host-parasite interactions. Also, a total of 215 transcripts were found to be differentially expressed and common to both tomont and trophont between the lower and higher passages as Ich loses its infectivity. The oligo microarray constructed from sequences representing EST-derived unigenes was highly informative in the analysis of transcript expression, and provided reproducible expression data as validated by real time RT-PCR. Many of the identified differentially expressed are examined and discussed in relation to functions in development, senescence, and senescence-related avirulence.

Background

The protozoan *Ichthyophthirius multifiliis* (Ich) is one of the most widespread ciliate parasites of freshwater fish worldwide [1], causing ichthyophthiriosis, or white spot disease, characterized by white spot cysts covering the host skin and gills. The parasite is responsible for high mortalities and severe economic losses to both aquaculture and ornamental fish industries. Ich has a life-cycle with three main developmental stages: a reproductive tomont, an infective theront, and a parasitic trophont [1]. The trophont lives within the host, where it feeds and grows; once matured, the trophont emerges from the host, drops to a substrate, and forms a cyst wall. The tomont rapidly divides producing daughter cells within the cyst; these cells develop into theronts. When fully developed, the cyst bursts, releasing the free-swimming theronts that actively locate a host, where the parasite invades the fish epithelium and the life-cycle continues [1, 2].

Ichthyophthiriosis begins when a free-swimming theront locates a fish, penetrates the host epithelium and develops into trophonts. The trophont grows within the host for approximately 5 to 6 days when the mature trophont emerges from the host after completion of normal development [1]. The trophont stage is an obligate parasitic stage and must be propagated on a susceptible host fish. A difficulty in long-term maintenance of any Ich isolate is the reduction in infectivity after a number of life-cycles, presumably related to senescence as the isolate undergoes serial passages from fish to fish in the laboratory [3-5]. Currently, *in vitro* methods used to culture Ich remain elusive as no media or cryopreservation technique has been developed that is needed for long-term maintenance. Xu et.al. [5] described the effect of senescence on an isolate of Ich maintained *in vivo* on channel catfish over a two-year period involving a total of 105 serial passages. The authors noted that infectivity of Ich declined as the

number of serial passages increased, causing lower and lower fish mortalities. Furthermore, the time period required for trophont emergence from the infected fish was found to be significantly increased at the later passages (91-105 cycles) as compared to that at earlier passages [5].

The role of senescence as occurs in the serial passages of laboratory Ich isolates remains unclear. Genomic characteristics of Ich may play a role in the senescence effect observed. Ich, typical of the phylum Ciliophora, has nuclear dimorphism; the genome consists of a macronucleus and a micronucleus [1]. The characteristics of ciliate nuclear dimorphism include both germ line and somatic cell divisions, where the micronucleus contains the germ line that can undergo recombination through meiotic conjugation, and the macronucleus that is transcriptionally active for cellular functioning. As such, ciliates can reproduce both sexually and asexually. Asexual cell division through binary fission has been studied for quite some time. While the mechanisms of Ich sexual reproduction are unresolved, some evidence exists to support sexual conjugation [1]. It has been suggested that senescence of laboratory isolates may be related to sexual reproduction [1, 6], or lack thereof, due to a lack of recombination of the germ line. Several biological processes are known to be capable of inducing senescence, including changes in gene expression, activation of oncogenes, epigenetic effects, and DNA damage and replicative damage in the form of telomeric shortening [7-10]. Therefore, undertaking a study of senescence-associated gene expression would be an important step in the identification of factors of senescence in the Ich macronuclear genome.

Despite these intriguing observations, follow up studies have been difficult due to the lack of necessary genomic tools. The objectives of this study were to develop a unique set of sequences using all existing ESTs, to construct an oligo microarray using the unigenes, and to

utilize the microarray for capture of expression profiles of Ich related to its developmental stages and passages.

Results and Discussion

EST Analysis

A total of approximately 33,000 Ich ESTs were available for analysis. Cluster analysis of these ESTs indicated the presence of 9,129 unique sequences. These ESTs were further subjected to BLASTX analysis to obtain their putative identities. As summarized in Table 6, of the 9,129 unique sequences, 4,730 had significant BLASTX hits. A close examination of putative identities suggested the presence of three main types of transcripts: those with highest similarities to protozoa sequences; those with highest similarities to teleost sequences; and those with highest similarities to bacterial sequences (Table 6), suggesting that the Ich EST database may actually contain ESTs with a mixture of their origin.

The vast majority, i.e., 3,717 (78.6%) of the Ich ESTs with significant BLASTX hits had the highest similarities to protozoan sequences, as expected (Table 6). However, a fraction of 387 (8.2%) ESTs with significant BLASTX hits had the highest similarities to teleost sequences. Similarly, a fraction of 469 (9.9%) ESTs with significant BLASTX hits had the highest similarities to bacterial sequences. While it is difficult to draw conclusions based on informatic analysis alone, it is likely that the Ich EST database contained sequences of various origins: those of Ich as intended, those from fish hosts, and those from bacteria. It is reasonable to assume that those ESTs with the highest similarities to protozoan sequences were likely of Ich origin, representing truly Ich ESTs, while those with the highest similarities to teleost sequences may have been derived from collection of host fish tissues along with parasitic trophont samples for cDNA library creation. Those ESTs with the highest similarities to bacterial sequences may have been derived from bacterial origin. Indeed, symbiotic bacteria have been recently reported for this parasite [11]. Since Ich is an obligate parasite, complete removal of contaminating fish

RNA remains difficult with the best of efforts. Removal of endosymbiotic signals is improbable, particularly since it is currently unclear if Ich can survive without the symbionts, and the trophont and tomont life-stages both contain symbiotic bacteria within their cytoplasm [11]. Nonetheless, caution must be exercised in dealing with ESTs from parasites such as Ich.

Table 6 - BLASTX summary of the unique Ich ESTs.

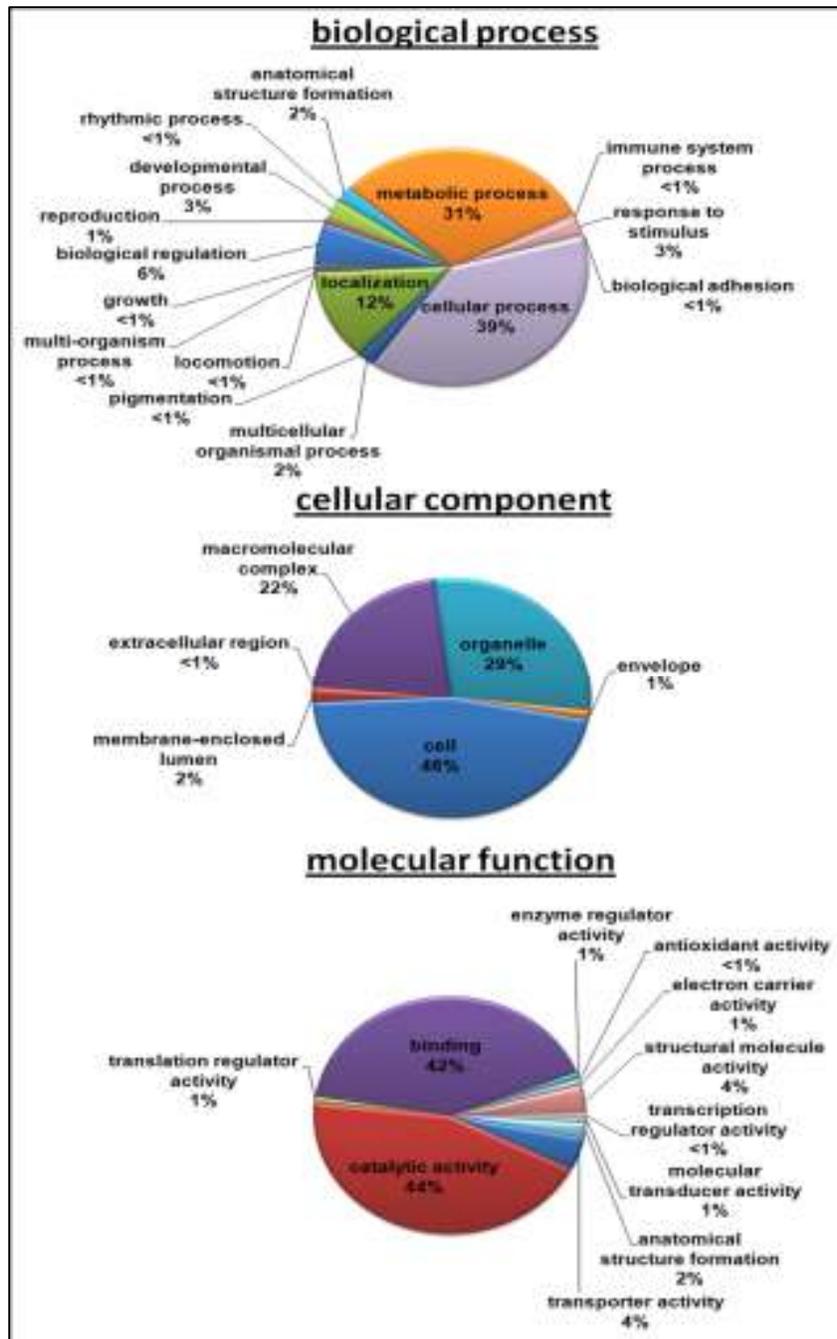
<u>Description</u>	<u>Value</u>	<u>Taxonomy</u>	<u>% of significant hits</u>
Total unique Ich ESTs	9,129		
BLASTX significant (<1e-5) hits	4,730		
I. Number of top hits to eukaryotes	4,259		90%
Hit distribution	3,519	<i>Tetrahymena</i> sp.	74%
	152	<i>Paramecium tetraurelia</i>	3%
	25	<i>Ichthyophthirius multifiliis</i>	<1%
	2	<i>Plasmodium</i> sp.	<1%
	174	others	4%
i. Number of top hits to teleosts	387		8%
Hit distribution	236	<i>Danio rerio</i>	5%
	36	<i>Tetraodon nigroviridis</i>	<1%
	33	<i>Salmo salar</i>	<1%
	25	<i>Ictalurus punctatus</i>	<1%
	1	<i>Ictalurus furcatus</i>	<1%
	56	others	1%
II. Number of top hits to prokaryotes	469		10%
Hit distribution	191	<i>Rickettsia</i> sp.	4%
	95	<i>Chryseobacterium gleum</i>	2%
	13	<i>Flavobacterium</i> sp.	<1%
	6	<i>Wolbachia</i> endosymbiont	<1%
	4	Candidatus bacteria	<1%
	160	others	3%
III. Other			
Hit distribution	2	Synthetic constructs	<1%

Subsequent to the above analysis, the Ich ESTs with a significant BLASTX hit were filtered to remove potential teleost and bacterial homologues, and the remaining sequences were analyzed using gene ontology (GO) searches (Fig. 9). A total of 1,816 sequences were assigned

GO terms within the biological processes, with 31% and 39% being involved in metabolic and cellular processes, respectively; 1,284 sequences were assigned functions under the cellular component category; and 1,948 sequences were assigned under the molecular functions category, with 42% for binding functions, and 44% for catalytic activities. The overall functional diversity of the sequences remains high based on analysis using 2nd level GO terms. All sequences assigned to GO categories did not exceed 44% assigned to any one category (Fig. 9).

Figure 9 - Gene ontology. Level 2 gene ontology (GO) of the unique Ich sequences.

BLASTX analysis was performed on all unique Ich ESTs, and then results filtered to remove teleost and bacterial homologues. Level 2 GO categories include biological process (top graph), cellular component (middle graph), and/or molecular function (lower graph).



Microarray design and construction

A microarray was constructed using the unique 9,129 Ich EST sequences. Because of excess capacity on the microarray for additional features, gene coding sequences from *Tetrahymena thermophila* and *Plasmodium falciparum* were also included on the microarray to potentially provide more information via cross-hybridization to the related protozoa: one highly-related and non-pathogenic ciliate (*T. thermophila*) and one highly pathogenic apicomplexan (*P. falciparum*). The resultant microarray included a total of 38,728 features, with 7,271 features designed from the Ich sequences, and 31,457 designed from the related organisms. Some of the 9,129 unique Ich EST sequences (1,858) were not able to be utilized for probe design after quality control.

Differentially expressed genes among different life stages

The cDNA samples derived from theront, trophont, and tomont stages were used as probes to explore gene expression profiles at the three different stages of Ich. The genes whose expression was at least two-fold different among all three stages were defined as differentially expressed genes among different stages. Comparison of hybridization data indicated that a total of 173 transcripts were differentially expressed among all three life-cycle stages (Appendix 2). The 173 transcripts represented 115 singletons and 57 contigs of the Ich ESTs, and one feature representing a hypothetical protein sequence of *T. thermophila*. Of the 173 differentially expressed transcripts, 95 were assigned putative identities by BLASTX at E-value of $E < 1e-5$.

Examination of the 173 transcripts reveals some prominent genes involved in protozoan cellular regulation including immobilization antigens (I-antigens), epiplasmin, cysteine protease, and annexins (Appendix 2). Ich I-antigens are surface proteins associated with cilia and cell cortex that strongly invoke the host immune response. The presence of host antibodies to these

antigens renders the parasite immobile [12] and provides the host teleost some protective immunity against subsequent exposure [13, 14]. As such, I-antigens have long been studied for their potential as an immunogen source (reviewed in [1]). However, differential expression of I-antigens across all three life stages has not been previously examined. The results of the microarray experiments, as also confirmed with qRT-PCR (see below), suggested that I-antigen was expressed at low level in the parasitic trophont, at a slightly increased level in the reproductive tomont, and at a highly elevated level in the infective theront life-stage (Appendix 2 and Fig. 10). These results confirmed a previous report indicating differential regulation of this protein [15], and further indicate relative amounts. A similar expression pattern was observed with epiplasmin 1, a gene representing the major component of the ciliate membrane skeleton and member of a multigenic family of high-interest in evolutionary and genomic studies [16, 17]. Epiplasmins were demonstrated to be of critical importance for normal cell development and morphogenesis in the ciliate *Paramecium tetraurelia* [16], and may play a similar role in Ich.

Cathepsin L cysteine protease was also found to be differentially expressed among all life stages (Appendix 2). This gene was previously found to be differentially regulated between Ich life-stages using quantitative RT-PCR [18]. Cathepsin L appears to play important roles in host-pathogen interactions. Proteases from parasitic protozoa have been demonstrated to function in host cell invasion and emergence, encystment and excystment, cytoadherence, stimulation and evasion of host immune responses, and catabolism of host proteins for a nutrient source [19, 20]. In ciliate fish parasites, proteases have been implicated in host invasion strategies and degradation of host cells from both Ich [18] and *Philasterides dicentrarchi* [21].

An interesting set of genes known as annexins, particularly annexin A1 and A5, are another group that were found to be differentially expressed among the three life stages of Ich

(Appendix 2). Annexins are calcium ion dependent phospholipid-binding proteins common to all eukaryotic cells. Annexin A1 is correlated with apoptosis, and has been implicated in endocytosis and exocytosis [22]. Annexin A5 expression has been linked to an apoptosis pathway in the protozoan *Leishmania major* [23]. The presence of apoptotic *L. major* increased the survival of non-apoptotic *L. major*, resulting in disease development. The presence of Annexin A5⁺ *Leishmania* likely delayed the host immune response and onset of inflammation, leading to a survival advantage for the wild-type parasite [23]. In Ich, it has been shown that annexin A5 levels increase in apoptotic theronts [24]. It must be pointed out that the origin of annexin A1 and A5 is uncertain at present because the annexin A1 and A5 ESTs had the highest similarities to fish annexins. Future studies are warranted to determine the expression of annexins in both the parasite and in the fish host, and the interplay between the two.

A total of 30 of the 173 genes identified as differentially expressed between all three life-cycle stages had top BLASTX hit and highest sequence identity to *Tetrahymena* species (Appendix 2). There is highest confidence in the homology of these transcripts to Ich, as these sequences are identified from a closely related ciliate. Many of these genes identified are involved in cell structure and cell regulation. They include genes involved in protein assembly, folding, and translocation such as with HSP70, HSP90, DnaJ, DnaK, prefoldin, Ras family and TCP1 gamma family proteins. A few core structural proteins were identified including ribosomal protein S25, TPR domains, and histones. Several other transcripts involved in cell-cycle functions were identified including proteins involved in fatty acid oxidation and the citric acid cycle, amino acid biosynthesis, ion binding and transport, phosphorylation, nucleic acid modification and translation termination. Three transcripts also had homology to hypothetical proteins. One final transcript was identified as a proteasome A-type and B-type family protein

(Appendix 2). Proteasomes represent a ubiquitous central component in eukaryotic cells involved in protein turnover. These proteins have been extensively characterized in protozoan parasites including *Giardia*, *Entamoeba*, *Leishmania*, *Trypanosoma*, *Plasmodium* and *Toxoplasma* species [25]. The proteasome has been demonstrated to be critical for cell differentiation and replication in protozoa; as such, proteasomes have been studied for their use as a therapeutic target to help the control of such parasites [25]. The current study shows that the transcript for the Ich proteasome is differentially regulated in all three life-cycle stages and consequently critical to cell regulation; therefore, perhaps this gene could also be a potential target for a therapeutic agent in the biocontrol of the parasite.

Differentially expressed genes between life stages

Microarray hybridization results were also examined between each life-stage: tomont versus theront, theront versus trophont, and trophont versus tomont. These data represent a significant portion of the steady-state transcriptome between the life-cycle stages of Ich and further provide a material basis for future developmental candidate gene expression studies. Comparison between tomont and theront yielded 576 putative differently expressed genes, between theront and trophont 787 putative differently expressed genes, and between trophont and tomont 162 putative differently expressed genes. A total of 32, 335, and 310 transcripts identified as differentially expressed and with highest BLASTX similarities to *Tetrahymena* species were identified between trophont versus tomont, theront versus trophont, and tomont versus theront, respectively. In the trophont -tomont dataset, the two most abundantly identified transcripts were hypothetical proteins and dynein heavy chain proteins; the remaining transcripts included cell structural and regulatory proteins such as granule lattice, tubulin, and histone, and dehydrogenase, lyase, and intraflagellar transport proteins. In the theront-trophont dataset, the

three most abundantly identified transcripts were hypothetical proteins, protein kinases, and leishmanolysins. Various other family proteins were identified including zinc-finger domains family, EF hand family, and ATPase family proteins. In the tomont-theront dataset, the three most abundantly identified transcripts were identical to the theront to trophont dataset: hypothetical proteins, protein kinases, and leishmanolysins.

The identity of the transcripts with BLASTX similarity to *Tetrahymena* hypothetical proteins remains unclear; interestingly however, protein kinases and leishmanolysins are the most abundant differentially expressed genes in two of the three datasets. Protein kinases are proteins that modify other proteins through kinase activity, or phosphorylation. They are one of the largest families of proteins in the cells of most eukaryotes; they comprise at least 2% of human genes [26, 27] and upwards of 3.9% of *T. thermophila* genes [27, 28]. The importance of overwhelming identification of protein kinases could be useful from Ich, since these proteins have been studied as potential therapeutic targets for control of other protozoan parasites [27, 29]. Leishmanolysins are proteases originally identified from *L. major* and are believed to be involved in processing surface proteins [28]. These proteins have since been identified from *T. thermophila* [28] and *Cryptocaryon irritans* [30], both ciliates highly related to Ich. As previously shown in Ich, cysteine proteases can be utilized for detection of the pathogen [18]. Parasitic proteases are highly immunogenic, and may be useful as biomarkers, vaccine candidates, and/or therapeutic targets [20]. Based on these findings, further data is warranted on the characterization of leishmanolysins in species other than *L. major* to determine their potential roles as critical virulence factors in pathogenic ciliates.

Differentially expressed genes identified through cross-hybridization

Given additional capacity on the array, features designed from genes of *T. thermophila* and *P. falciparum* were included. The microarray hybridization data suggested that a total of 35 transcripts were identified as differentially expressed genes through cross-hybridization with *T. thermophila* or *P. falciparum* sequences. This included 16 transcripts by tomont-theront comparison, 16 transcripts by theront-trophont comparison, and 3 transcripts by trophont-tomont comparison. Transcript identities include a translation initiation factor and an erythrocyte membrane protein from *P. falciparum*, and an immobilization antigen and IBR domain containing protein from *T. thermophila*. There were also multiple identities to hypothetical proteins identified from *T. thermophila*.

As expected due to the evolutionary relationships of the three protozoa [31, 32], the largest degree of cross-hybridization was observed with *T. thermophila* sequences as compared to *P. falciparum* sequences with the *I. multifiliis* probes. Additional comparative genomic data was obtained when the stringency was reduced from a 5-fold cut-off to a 2-fold cut-off; a total of 118 additional differentially expressed transcripts were identified, including 23 from *P. falciparum* and 95 from *T. thermophila*. Results from comparative analyses should facilitate gene discovery and annotation in Ich, particularly since relatively little expression data is known and no whole genome sequence is currently available from Ich whereas the whole genome sequences are available for both *T. Thermophila* and *P. falciparum*.

Differentially expressed genes between low and high Ich passages

The ciliate parasite Ich was found to lose its infectivity and thereby its virulence upon a high number of passages [5]. Ich infectivity decreased significantly after 26-105 passages in the lab on fish hosts [5]. In this study, the microarray was used to assess gene expression patterns in low and high passages. Probes from Ich tomont and trophont at passage 1 (P1) and passage 100

(P100) were used to study the differentially expressed genes. A total of 215 transcripts were found to be differentially expressed and common to both tomont and trophont between passage 1 to passage 100 (Appendix 3). Of the 215 transcripts differentially expressed, 176 were assigned putative identities by BLASTX at E-value of $E < 1e-5$. Of the 176 transcripts assigned putative identities, 130 of these were identified as having highest similarity to *Tetrahymena* species by BLASTX analysis. One transcript was also identified through cross-hybridization and found to be a hypothetical protein in *T. thermophila*. Interestingly, all 176 shared transcripts were differentially expressed in concert; all trophont genes down-regulated in P100 were also down-regulated in P100 of tomont, and likewise all trophont genes up-regulated in P100 were also up-regulated in P100 of tomont. The four most commonly found transcripts identified were hypothetical proteins, leishmanolysins, protein kinases, and major facilitator superfamily proteins. All protein kinases were found to be up-regulated in the later passage (P100) of both trophont and tomont life-cycle stages. Similarly, all leishmanolysin family proteins were identified as up-regulated in P100 when compared to P1. All major facilitator superfamily proteins likewise were up-regulated in the later passages. Major facilitator superfamily proteins comprise one of the largest families of membrane transport proteins known. These proteins transport a wide array of substances including sugars, nucleosides, ions, metabolites, amino acids, and drugs across the cell membranes [33]. This family has long been studied as a means of multidrug resistance in many pathogens from bacteria to eukaryotes, as these proteins are involved in transmembrane transport and are actively removing cytotoxic compounds out of the cell [34]. As with other protozoa, these proteins may be found to play a critical role in the biocontrol of Ich, especially since Ich harbors endosymbiotic bacteria [11] that could possibly be used as a target for a therapeutic agent.

Microarray hybridization results were also examined between each passage separately: tomont P1 versus tomont P100, and also trophont P1 versus trophont P100. Comparison between tomont P1 to tomont P100 yielded 579 putative differentially expressed genes and 1,537 putative differentially expressed genes between trophont P1 to trophont P100 were identified. This data provides a large repertoire of genes putatively involved in Ich infectivity or the loss-of-virulence in tomont and trophont life-cycle stages. In the tomont passage analysis (P1 to P100), 308 transcripts were down-regulated while 271 transcripts were up-regulated. In the trophont passage analysis (P1 to P100), 139 transcripts were down-regulated while 1,398 transcripts were up-regulated. A total of 420 transcripts were assigned putative identities by BLASTX when comparing tomont P1 to P100, and 1,285 transcripts were assigned putative identities by BLASTX when comparing trophont P1 to P100. Unfortunately, a sufficient quantity of RNA was not achieved in the theront life-cycle stage to obtain microarray results for theront passages P1 and P100.

Eleven transcripts in this analysis were identified as differentially regulated by cross-hybridization of Ich cDNA to *T. thermophila* sequences on the microarray. This included 5 transcripts by tomont P1 to P100 analysis and 6 by trophont P1 to P100 analysis. All were identified as hypothetical proteins except one in the trophont analysis. That sequence was identified as a NAC domain containing protein. No *P. falciparum* sequences were identified through cross-hybridization of probes by Ich passage analysis.

Some frequently observed transcripts involved in protozoan cellular regulation identified as differentially expressed between the passages include I-antigens and other surface proteins. I-antigens and other surface proteins are again found to be differentially expressed and surface proteins could have a role in loss of infectivity if these proteins affect Ich attachment and/or

emergence from the host. Other frequently observed transcripts include leishmanolysin, zinc-finger domains, insect antifreeze proteins and neurohypophysial hormones. Interestingly, these types of similar proteins are characterized as meckelin transmembrane family proteins in mammals. These proteins are believed to be associated with ciliatory diseases [35]. It would be highly interesting to understand the function of these similar sequences in Ich and if, in fact, they are related to cilia damage or dysfunction in the later passages and if this correlates with a loss of infectivity.

The complete molecular mechanisms of senescence remain poorly understood. Several biological processes have been suggested as capable of inducing senescence, including changes in gene expression, activation of oncogenes, epigenetic effects, and DNA damage and replicative damage in the form of telomeric shortening [10, 36]. Telomerase has long been identified from *T. thermophila* [37], and later from other protozoa such as *Leishmania* [38]. As such, telomerase activity is likely in Ich, particularly due to the ciliate genomic organization. A clear homolog in Ich was not identified through the analysis of EST data and therefore telomerase gene activity could not be assessed by cDNA microarray analysis. However, known senescence proteins were found to be differentially regulated in Ich. The silent information regulator 2 (*sir2*) gene was found to be highly (>33-fold) up-regulated in the later passage, P100, of Ich trophonts. Sir2 is a histone deacetylase that functions in yeast in silencing of transcription at repetitive DNA including mating-type loci, telomeres, ribosomal DNA (rDNA), and also represses recombination in rDNA [39-41]. Overexpression of Sir2 in yeast has been shown to extend the replicative lifespan of this organism [42], and similar observations were later found in fruit fly and worms [39, 41]. The current findings in Ich suggest that selection of trophont cells that survive to later ages may have resulted, at least in part, through an overexpression of Sir2.

Future cytogenetic studies would be useful to determine if telomeric shortening and/or alterations of cilia are occurring in the laboratory and if these processes lead to a loss of infectivity in Ich.

Recently in *P. falciparum*, the protozoan responsible for malaria in humans, two Sir2 paralogs were characterized, with one Sir2 gene shown to affect telomere length [43]. Further, the two Sir2 paralogs also play a role in antigenic gene variation by regulation of *var*, the gene family that encodes the major *P. falciparum* antigen: erythrocyte membrane protein 1. Approximately 60 subtelomeric *var* genes encode different variants of the protein; with one type being expressed at a time. Sir2 homologs were found to be major effectors on silencing the *var* genes [43]. Gene switches in *var* expression cause variation of the protein enabling the parasite to evade the host immune system [43, 44]. Other pathogenic protozoa, including *Cryptosporidium* and *Trypanosoma brucei*, have surface protein gene families associated with subtelomeric regions and possess some gene switching activity [45]. As such, there remains great interest in elucidating the complete mechanism of Sir2 as related to Ich senescence and/or loss of infectivity. Similar to *P. falciparum*, a single serotype of Ich surface antigen (I-antigen) appears to be expressed at any given time [46]; however, there is no current evidence of antigen switching in clonal isolates of Ich [13]. Also, it has yet to be demonstrated whether the I-antigen gene family is located adjacent to telomeres. The question therefore arises as to which of the mechanism(s) are responsible for Ich senescence and loss of infectivity in the lab passages. Two mechanisms are described here relating to changes in gene expression: a replicative senescence effect as the cells age presumably due to an evolutionarily similar mechanism of telomere shortening, or partial loss of function in the mechanisms that regulate antigenic variation of the immunogens. Current knowledge of Ich biology indicates that antigenic switching may not occur; rather, that senescence could be an artifact of cloning [1, 6]. A particular Ich serotype is

passed clonally in the laboratory setting and there is no opportunity for the isolate to regenerate its cell or recombine its DNA through sexual conjugation, as occurs in related free-living ciliates such as *Tetrahymena* [1]. Since Ich bridges the gap between free-living ciliates and parasitic protozoa, the mechanisms that affect gene regulation as related to senescence and virulence may have characteristics of both, somewhat revealed through analysis of Ich genes ([30-32, 47], this study). Overall, the complete mechanisms related to Ich senescence and loss of infectivity evaluated by gene expression warrant further study; the current findings based on gene expression analysis provide the material basis for such future studies.

Endosymbiotic bacteria may be another factor to consider in relation to Ich senescence. Endosymbiotic bacteria, particularly *Rickettsia* species and sphingobacteria, have been recently discovered in Ich [11]. Could changes in gene expression of the endosymbionts be a contributing factor in the decline of infectivity? To address this question, the microarray data was further examined for bacterial sequences. Since a portion of the ESTs on the microarray likely represent the endosymbiotic bacteria based on sequence homology, their gene expression can be assessed. All putative genes from *Rickettsia* species were dramatically ($-8.4 \geq \text{fold-change} \leq -68.9$) down-regulated in Ich P100, including surface antigens, surface proteins, outer membrane proteins, ribosomal protein S15, and GRoEL (Appendix 3). A DnaK transcript with sequence similarity (E-value = $4e^{-35}$) to *Acinetobacter* species was also sharply down-regulated (-19.9 fold-change) in the tomont P100. Obviously, if surface antigens from bacterial symbionts are strong immunogens during infection of fish, a decrease in the amount of antigen could modulate the host immune response, potentially explaining a previous report of longer periods of trophont emergence in high passage Ich [5]. The interplay of the Ich and symbiotic genomes may also affect fish immunity. Also, both DnaK and GRoEL are chaperones that have been

indicated in bacterial senescence [48, 49]. While interesting, further study is needed to elucidate the contribution of the host-symbiont relationship to senescence.

Validation using quantitative RT-PCR

In order to validate the microarray and the results from the microarray study, 21 differentially expressed genes were analyzed using quantitative real time reverse transcription PCR (qRT-PCR). For the developmental microarray study, based on the microarray data and BLAST analysis, a subset of sequences with putative identities and BLASTX hit to other protozoa was selected for qRT-PCR. This included two genes (I-antigen and epiplasmin 1) that were differentially regulated among all three life-cycle stages. Validation by qRT-PCR was also performed for various other developmental genes represented in the microarray when comparing one-to-one life-stage expression changes (Table 7). At least two genes differentially expressed between each of the three life-cycle stages were also chosen for validation, to represent at least one unique up-regulated and one unique down-regulated transcript.

Table 7 - Ich developmental array quantitative RT-PCR

List of genes selected for qRT-PCR analysis. Information includes putative gene identities based on BLAST analysis, GenBank accession numbers, life-stages examined, and primer sequences. The fold-change data are \log_2 transformed values from the microarray and from qRT-PCR, using one life-stage as baseline data (B) and another life-stage as experimental data (E) for comparisons. Probe names (ID) are also listed when the probe was constructed from contigs or using in-house sequences since submitted to GenBank and provided in the microarray (GEO accession no. **GSE18556**).

Gene name	Accession / ID	Comparison	Fold-change (μ array)	Fold-change (qRT-PCR)	Primer sequence (5'-3'), F: forward; R: reverse
18S rRNA gene	U17354	all			F: GTGACAAGAAATAGCAAGCC R: CCCAGCTAAATAGGCAGAAG
epiplasmin I	EL906016 / Ich_Contig_670	Theront (B) → Trophont (E) Trophont (B) → Tomont (E) Tomont (B) → Theront (E)	-3.84 1.60 2.24	-7.38 2.67 7.03	F: TGTTCACTAACCTGTACAAGAAGC R: AAACAAGTGGTGCTCTACAGTC
I-antigen	EL905772	Theront (B) → Trophont (E) Trophont (B) → Tomont (E) Tomont (B) → Theront (E)	-2.81 1.56 1.24	-7.64 1.67 6.33	F: TTTGCGAAAGTGGAAGCTGG R: AACATCTGGTTTTGCAGCAG
p25-alpha family protein	EL922674 / Ich_contig_479	Theront (B) → Trophont (E)	-5.61	-7.64	F: CTCTGGAAAAGCTGAAATGGA R: TTAAGGACCTCCAGCTGAAGTAA
pyruvate kinase family protein	EL925641	Theront (B) → Trophont (E)	-5.48	-9.97	F: TTGCAGACGGTACTTTAGTATGTATTG R: TCTTCTTTTCGGTTACTGTTGG
MCM2/3/5 family protein	EL912787	Theront (B) → Trophont (E)	4.25	3.17	F: TCCTTCGCCTTCTGATGATACAGGAAT R: TGGCAGGCAGAAGAAACAATCGTG
Sec7 domain containing protein	EL907900	Tomont (B) → Theront (E)	-4.27	-0.89	F: GGAGAATTATTTGGGTCTGATAATG R: AAGCACCAGTAGCCGATTG
transketolase	EL916987 / Ich_contig_1786	Tomont (B) → Theront (E)	4.88	3.96	F: TGAAGGGTCAGTTTGGGAAG R: GCCATCCAAAAGCTTCAAATC
similar to Importin-3/5	EL909221	Trophont (B) → Tomont (E)	-2.11	-2.97	F: TCAGAAGCTCTTGAAAGTTTCCTT R: TTCATTGCATCTGGTTTGC
SGS domain containing protein	EG962593 / JAIch_018B_H09	Trophont (B) → Tomont (E)	3.72	2.03	F: GAATATGGTGAAGATCCTATGAATG R: GGAGGTTTCAGGTCTATCTTTTCC

For the passage microarray study, at least one common down-regulated and one common up-regulated transcript in both the tomont and trophont passages were selected for validation (Table 8). An additional subset of genes was also chosen for validation by qRT-PCR, to represent at least two unique up-regulated transcripts and two unique down-regulated transcripts between the tomont and trophont passages (Table 8).

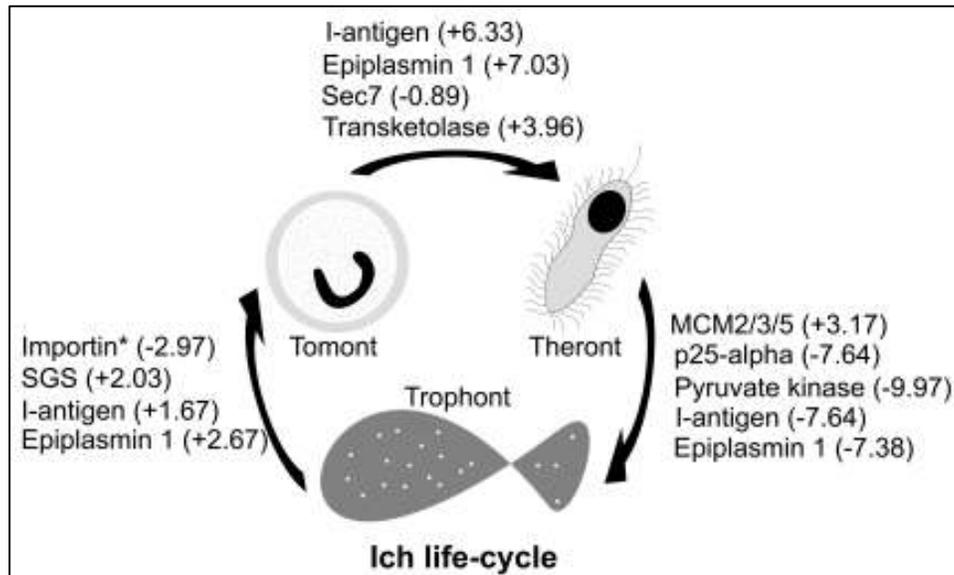
Table 8 - Ich passages array quantitative PCR

List of genes selected for quantitative PCR analysis using RNA of Ich passage 1 (P1) and passage 100 (P100). Detailed description of data is given in Table 7.

Gene name	Accession / ID	Comparison	Fold-change (μ array)	Fold-change (qRT-PCR)	Primer sequence (5'-3'), F: forward; R: reverse
18S rRNA gene	U17354	all			F: GTGACAAGAAATAGCAAGCC R: CCCAGCTAAATAGGCAGAAG
high mobility group protein	EG963277 / Ich_Contig_20	Tomont P1(B) →Tomont P100 (E)	-4.74	-9.97	F: GGCTAAACCCTCTGCATCAA R: GCAACTTTCATTAATCAGCAAT
casein kinase II	EL923519	Tomont P1(B) →Tomont P100 (E)	-4.39	-5.16	F: TGTCGTGATATTGATGGATGCT R: CATCTTCAGCAAATCTTGATCCT
TPR domain containing protein	EL912984	Tomont P1(B) →Tomont P100 (E)	3.49	2.35	F: GCAGTATTTGACGGCTTGGT R: TACACCCGACCTTGGTTTA
prenyltransferase and squalene oxidase repeat family protein	EL907482	Tomont P1(B) →Tomont P100 (E)	3.31	5.89	F: TGGGGAGAAGTGGATACTCG R: ATACGCCCATGACTTCTG
Sec61beta family protein	EL914941 / Ich_Contig_1766	Trophont P1(B) →Trophont P100 (E)	-2.87	-6.27	F: TGGTGGATCAAAAAGTTCTGG R: TGCAGCAGATGAAGTTCAC
UDP-sugar pyrophosphorylase	EL908359	Trophont P1(B) →Trophont P100 (E)	-2.65	-3.61	F: TGCCATTCACAATTACGAACA R: AAATGTGCTGAATGCAATATTGT
heat shock protein 90	EG961165 / JAIch_014C_D08	Trophont P1(B) →Trophont P100 (E)	-2.48	-5.21	F: TCTTGGGTGTGAACCTGGAT R: CGAAAGGAGATTAAGCTACAGAGG
Acyl-CoA oxidase family protein	EL910216	Trophont P1(B) →Trophont P100 (E)	6.39	5.72	F: TTGGAACCCCTCCCACTGTA R: AACATTGGAACCATGGCCTA
similar to WD repeat domain	EG958609 / JAIch_006D_C03	Trophont P1(B) →Trophont P100 (E)	6.35	4.97	F: AGACGCGGTTTCATAGGTTG R: TTTAGAATCTCCCCCTGCTG
cathepsin z	EL923418 / Ich_Contig_487	Trophont P1(B) →Trophont P100 (E)	6.10	4.32	F: CAATGCAGGAGGATCATGTG R: TTACAGCCCAGCATCTACCA
ankyrin 2,3/unc44	EG965438 / JAIch_026A_E02	Tomont P1(B) →Tomont P100 (E)	-2.97	0.96	F: TATATACGGCGTGCCAGGA R: CACCCGCTGTGAGTAACGTA
		Trophont P1(B) →Trophont P100 (E)	-3.55	-8.97	
phosphoglucomutase / parafusin	EL912504 / Ich_Contig_910	Tomont P1(B) →Tomont P100 (E)	2.58	4.00	F: GGGATTGCAAGAAGGCAATA R: AGTCGCTCCTTCAGATCCAG
		Trophont P1(B) →Trophont P100 (E)	4.55	5.05	

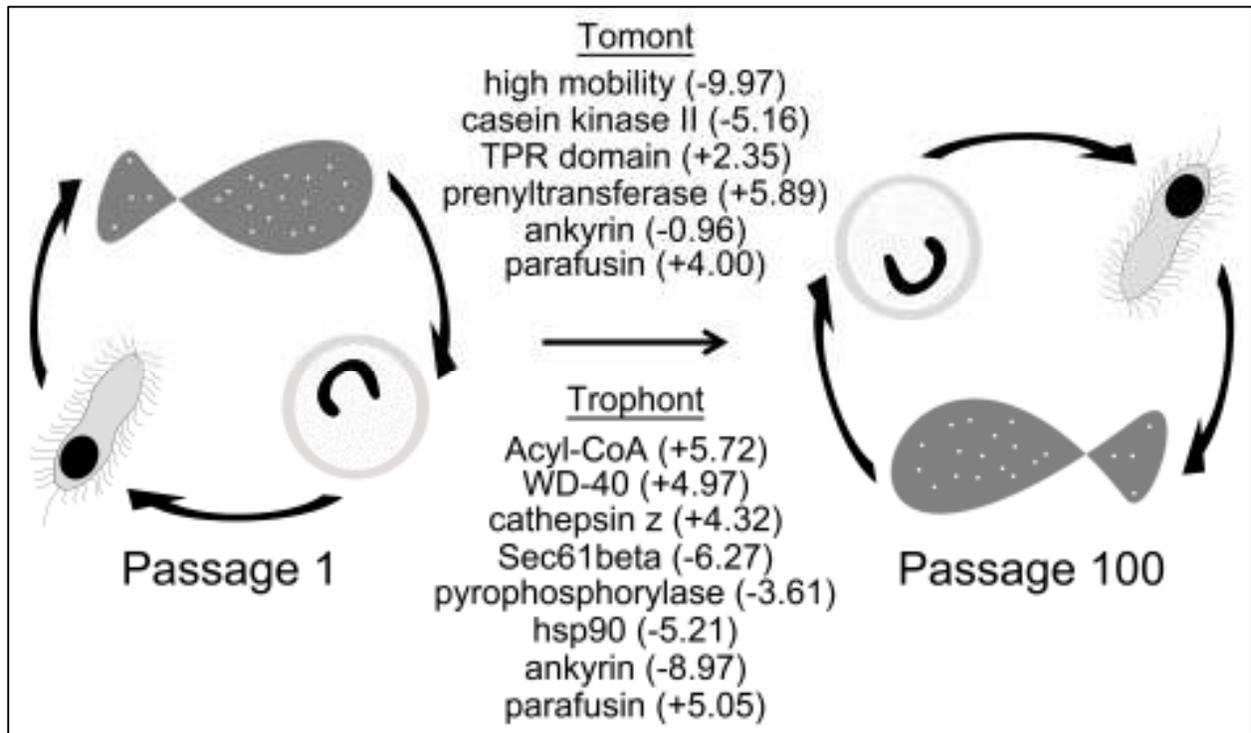
Transcripts tested by qRT-PCR for the developmental study concurred with the microarray hybridization results (Table 7 and Fig. 10). For comparisons between theront and trophont life-stages, p25-alpha and pyruvate kinase were determined to be down-regulated and MCM2/3/5 was determined to be up-regulated in trophont. For comparisons between tomont and theront life-stages, Sec7 was found to be down-regulated and transketolase was found to be up-regulated in theront. For comparisons between trophont and tomont life-stages, importin was determined to be down-regulated and SGS was found to be up-regulated in tomont (Fig. 10). The results from quantitative RT-PCR (Fig. 10) confirmed those from the microarray hybridization experiments, whether comparing expression levels from all three Ich life-stages simultaneously (Appendix 2) or with comparisons between each life-stage, validating the design and reproducibility of the microarray between developmental life stages.

Figure 10 - Ich developmental stages quantitative PCR. Quantitative real-time RT-PCR of selected transcripts from *I. multifiliis*. Values indicated are log₂ transformed data generated from the REST 2009 software. Asterisk indicates a hypothetical protein with putative similarity. Full gene information is listed in Table 7.



Transcripts tested by qRT-PCR for the passages study also concurred with the microarray hybridization results (Fig. 11). Two transcripts were found to be up-regulated in the later (P100) passages of Ich tomonts including a high mobility group protein and casein kinase II. Two transcripts were found to be down-regulated in the later (P100) passages of Ich tomonts including a TPR domain containing protein and a prenyltransferase / squalene oxidase repeat family protein. In the later passage of trophont, three transcripts were validated to be down-regulated including a Sec61beta family protein, UDP-sugar pyrophosphorylase, and heat shock protein 90. In the later passage of trophont, three transcripts were validated to be up-regulated including an Acyl-CoA oxidase family protein, a WD repeat domain, and cathepsin z. A common down-regulated gene (ankyrin 2,3/unc44) and a common up-regulated gene (phosphoglucomutase / parafusin) were also validated (Fig. 11). To test the utility of the microarray for the theront life-stage, real time RT-PCR was also performed on the two common differentially expressed genes. The transcript for parafusin was also up-regulated in the theront stage (data not shown).

Figure 11 - Ich passages quantitative PCR. Quantitative real-time RT-PCR of selected transcripts from *I. multifiliis*. Values indicated are log₂ transformed data generated from the REST 2009 software. All data were significant at $p < 0.05$. Full gene information is listed in Table 8.



In all cases, transcripts tested by qRT-PCR concurred with the microarray hybridization results, thus demonstrating the feasibility, reproducibility, and usefulness of the microarrays in the study of *Ich* gene regulation.

Conclusion

Analysis of Ich ESTs was conducted and an oligo cDNA microarray was constructed using the EST information. The microarrays were used to assess the gene expression profiles on each of the three life-stages. Microarray hybridization results were confirmed by results from real time quantitative RT-PCR using a selected set of genes, providing assessment for the reproducibility of the microarray design. This work therefore provides a technological platform for future gene expression studies of Ich at the genome scale. Currently, genome resources are limited in Ich, and no whole-genome sequence is available; however, the use of cDNA microarrays based on ESTs in species with limited genomic information has been highly effective [50, 51]. The application of microarrays to address parasitological questions involving developmental and life-stage biology, host-pathogen interactions, virulence factors, and comparative genomics has been well-established [52]. The current work provides a framework and platform which can facilitate further transcriptional studies in this important fish parasite.

Materials and Methods

Biological samples

Ich was isolated from an outbreak of ichthyophthiriosis at a local pet shop. Ich was cultured by serial infections of channel catfish as previously described [2, 24]. The tomont, trophont, and theront samples were collected into separate tubes. Because of small biological samples that can be collected from each fish, tomont, trophont, and theront samples were pooled from multiple fish, and three distinct pools were collected to serve as biological replicates for the microarray analysis. Samples of each pool were collected from approximately 8 Ich infected fish: 2 for trophonts, 3 for theronts and 3 for tomonts. Trophonts were placed in Petri dishes and allowed to attach and develop into tomonts and theronts.

Each life stage was also collected over approximately a two-year period of serial passages, based on the observation that Ich loses infectivity over the course of serial passages [5]. The first passage after wild isolation (passage 1) and passage 100 samples were collected as described above. All samples were suspended in phosphate buffered saline (PBS, pH 7.4) and flash-frozen in liquid nitrogen. Samples were then stored at -80 °C until RNA extraction.

Ich genetic information

The Ich microarray was constructed using EST sequence information. ESTs available in the dbEST at the GenBank were analyzed to determine a set of unique genes. The Ich EST sequences were first downloaded to a local database, and then contiguous sequences (contigs) were assembled. A total of 33,515 Ich ESTs was used for sequence assembly including 8,432 ESTs from a previous sequencing effort [31] and an additional 25,083 ESTs in the public domain. Before sequence assembly, all Ich ESTs were trimmed of poly A/T sequences at both

the 5' and 3' ends using the default settings of the Vector NTI Advance version 10 software (Invitrogen, Carlsbad, CA). Next, the trimEST feature at EMBOSS (<http://emboss.sourceforge.net/>) was employed using the settings of minimum 4 poly A/T and mismatch setting at 2 bases. Ich ESTs were then clustered to create contigs using the ContigExpress program of the Vector NTI software (Invitrogen) with the settings of 90% identity and 40 bp overlap. All other cluster settings used were the default. This produced a set of 9,129 total unique Ich expressed sequences, comprising 2,824 contigs and 6,305 singletons. Ich contig sequences can be provided upon request. Unique Ich sequences were used in BLAST analyses and gene ontology comparisons, and for the construction of the microarray.

Microarray construction

High-density *in situ* 385 K oligonucleotide microarrays were constructed by Nimblegen (Madison, WI). All unique Ich EST sequences were submitted for probe design. Since the unique Ich ESTs comprised 9,129 sequences, extra space was available on the 385 K microarrays to place additional features. To facilitate comparative genomic analysis and to potentially increase gene content through cross-hybridization, gene coding sequences of *Tetrahymena thermophila* and *Plasmodium falciparum* were also added to the microarrays. *T. thermophila* and *P. falciparum* sequences were selected since both organisms are phylogenetically close to Ich [15, 31, 32, 53], and both have a whole-genome sequence available. A total of 26,273 *T. thermophila* and 5,184 *P. falciparum* gene coding sequences were included in the array for probe design. The probe design strategy was to create 60 base pair (60mer) oligonucleotide probes, including 12 60mer oligos per unique Ich sequence, and 10 60mer oligos for both *T. thermophila* and *P. falciparum* sequences. The complete design of the array can be accessed at the NCBI Gene Expression Omnibus (GEO) database under the accession number of **GSE18556**.

RNA extraction and synthesis of cDNA

Total RNA was extracted using the TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. Replicates from each life-stage were processed separately to avoid cross-contamination. Quantity and quality of total RNA was assessed on a spectrophotometer and by denaturing agarose gel electrophoresis containing formaldehyde.

Total RNA was used for the synthesis of first-strand cDNA, using an oligo(dT)₁₂₋₁₈ primer (Invitrogen) and components from the SuperScript™ Double-Stranded cDNA Synthesis kit (Invitrogen) according to protocols from Nimblegen Systems, Inc. (Roche Nimblegen, Madison, WI). For first-strand cDNA synthesis, 10 µg of each Ich total RNA was combined with 100 pmol of oligo(dT) primer in total reaction volumes of 11 µl. Samples were heated to 70 °C for 10 min to denature and immediately placed on ice. Then, 4 µl of 5x First Strand Buffer, 2 µl of 0.1 M DTT, and 1 µl of 10 mM dNTP mix were added to each reaction. Reactions were mixed, incubated for 2 min at 42 °C, and then 2 µl of RT was added (200 U/µl SuperScript II, Invitrogen). Incubations were continued at 42 °C for 1 h and placed on ice.

Next, second-stranded cDNA was synthesized, also with kit components from the SuperScript™ Double-Stranded cDNA Synthesis kit (Invitrogen). A master mix was made for each first-strand reaction sample in a total volume of 150 µl as follows: 20 µl of first-strand reaction was combined with 91 µl of nanopure water, 30 µl of 5x Second Strand Buffer, 3 µl of 10 mM dNTP mix, 1 µl of 10 U/µl DNA ligase, 4 µl of 10 U/µl DNA polymerase I, and 1 µl of 2 U/µl RNase H. Samples were gently mixed and incubated at 16 °C for 2 h. Then, 2 µl of 5 U/µl T4 DNA polymerase was added and the incubation was continued for an additional 5 min. Samples were then placed on ice and 10 µl of 0.5 M EDTA added to each reaction to complete

the synthesis of double-stranded cDNA. One-microliter of 10 mg/ml RNase A Solution was added to each reaction and incubated at 37 °C for 10 min. The samples were extracted with TE-saturated 25:24:1 phenol:chloroform:isoamyl alcohol (Sigma-Aldrich, St. Louis, MO) and the aqueous phase saved for cDNA precipitation. Second strand cDNA was precipitated by adding 16 µl of 7.5 M ammonium acetate to each sample. Samples were mixed by repeated inversion and 7 µl of 5 mg/ml glycogen was added and mixed. Cold (-20 °C) absolute ethanol (326 µl) was added and mixed by inversion. Samples were centrifuged at 14,000 x g for 20 min and the supernatants removed and discarded. To wash the samples, 500 µl of 80% ethanol (-20 °C) was added to each tube and mixed by inversion. Samples were centrifuged at 14,000 x g for 5 min and the supernatants removed. Washing with ice-cold 80% ethanol was repeated. Pellets were air-dried under a laminar flow hood for 10 min and reconstituted with 20 µl of nanopure water. Concentration of each cDNA sample was determined by spectrophotometry, and ≥ 2 µg cDNA from each sample was provided to Nimblegen (Madison, WI) for labeling, hybridization, and image acquisition.

Microarray hybridization and analysis

The microarray was used to assess gene expression profiles on each of the three life-stages of Ich, and to assess gene expression in early to late life-cycles based on serial passages. For a comparison between life-cycle stages, a total of nine hybridization experiments were performed utilizing the microarray design, including three biological replicates from each of the three Ich life-stages (tomont, trophont, and theront life-stages). For the infectivity study, a total of 18 hybridization experiments were performed to include three biological replicates at passage 1 and passage 100 of each of the three Ich life-stages. Second-stranded cDNA created as described above from each was hybridized to microarrays and used to generate gene calls. Gene

call data for each array was provided by Nimblegen (Madison, WI). NimbleScan software was used to normalize the raw images by quantile normalization [54]. Gene calls of the normalized data were generated using the Robust Multichip Average (RMA) algorithm [55]. Normalized RMA call expression values and standard errors were loaded into the DNA-chip analyzer (dChip) software [56]. This software was used to perform data comparisons and generate fold-change of expression values. All default comparison criterion statistics were selected in dChip, along with a cut-off value of 2-fold expression difference minimum among all three-life stages. A 2-fold cut-off difference was chosen to increase the likelihood of gene discovery. Additional dChip analyses were performed comparing one-to-one life-stage expression changes with a 5-fold difference cut-off. This included an analysis comparing trophont to tomont life-stage, tomont to theront life-stage, and theront to trophont life-stage, mimicking the continual nature of the parasite development. A 5-fold difference cut-off was chosen to increase the confidence in hybridization results between the life-stages and also between the passage data. A global false discovery rate of 10% was assessed using 500 permutations for all analyses.

Quantitative RT-PCR

A selected subset of genes was chosen for microarray validation using quantitative PCR. For quantitative real-time RT-PCR (qRT-PCR), total RNA of Ich was extracted using the RNeasy Plus Mini kit (Qiagen, Valencia, CA). Briefly, aliquots of Ich cells from each life-stage were homogenized in lysis buffer (6 μ l of 14.3 M β -mercaptoethanol; 600 μ l Buffer RLT Plus) using a 20-gauge needle attached to a sterile syringe according to the protocol from the RNeasy Plus Mini kit (Qiagen). Each homogenized lysate was centrifuged to remove cellular debris, and the supernatants were filtered through a gDNA eliminator spin column to eliminate the genomic DNA. The total RNA was further extracted using the protocol supplied from the manufacturer.

Total RNA from each life-stage replicate was isolated for qRT-PCR. Each of the total RNA samples was adjusted to 100 ng/μl to use as template. One-step qRT-PCR was performed on a LightCycler 1.0 real-time instrument (Roche Applied Science, Indianapolis, IN). The qRT-PCR reactions were made using the LightCycler RNA Master SYBR Green I kit (Roche Applied Science), with modifications. A 10 μl total reaction volume was used (9 μl reaction mix + 1 μl RNA template) for each sample. Each reaction mix contained 2.65 μl of PCR grade water, 0.65 μl of 50 mM Mn[OAc]₂, 1 μl of 5 μM forward primer, 1 μl of 5 μM reverse primer, and 3.7 μl of SYBR Green I. The thermal cycling profile used was as follows: (i) Reverse transcription, 61 °C for 20 min; (ii) Denaturation, 95 °C for 2 min; (iii) Amplification, 50 cycles of 95 °C for 5 s, 56 °C for 10 s, and 72 °C for 20 s. Melting curve analysis was performed on each run and confirmed single product amplification. Cycle threshold (Ct) values were generated for each triplicate sample, both for the genes of interest and 18S rRNA gene. The 18S rRNA primers were modelled from [57].

Relative expression levels were determined using Ct values and the Relative Expression Software Tool (REST) version 2009 [58, 59], assuming 100% PCR efficiencies and statistics based on 1000 randomizations. The 18S rRNA gene was used as the normalizer. Negative control PCR reactions (PCR using the total RNA templates, minus RT) were also performed to test for amplification of genomic DNA.

Acknowledgements

This project was supported by a grant from the USDA NRI Animal Genome Basic Genome Reagents and Tools Program (USDA/NRICGP award# 2009-35205-05101), and partially by a Specific Cooperative Agreement with the USDA ARS Aquatic Animal Health Laboratory under the contract number 58-6420-5-030. Special thanks to the co-authors of this study: De-Hai Xu, Eric Peatman, Huseyin Kucuktas, Phillip Klesius, and Zhanjiang Liu.

References

1. Matthews RA: ***Ichthyophthirius multifiliis* Fouquet and ichthyophthiriosis in freshwater teleosts.** *Adv Parasitol* 2005, **59**:159-241.
2. Xu D, Klesius PH, Shoemaker CA, Evans JJ: **The early development of *Ichthyophthirius multifiliis* in channel catfish *in vitro*.** *J Aquat Anim Health* 2000, **12**(4):290-296.
3. Burkart M, Clark T, Dickerson H: **Immunization of channel catfish, *Ictalurus punctatus* Rafinesque, against *Ichthyophthirius multifiliis* (Fouquet): killed versus live vaccines.** *J Fish Biol* 1990, **13**:401-410.
4. Houghton G, Matthews RA: **Immunosuppression of carp (*Cyprinus carpio* L.) to ichthyophthiriasis using the corticosteroid triamcinolone acetonide.** *Vet Immunol Immunopathol* 1986, **12**(1-4):413-419.
5. Xu DH, Klesius PH: **Two year study on the infectivity of *Ichthyophthirius multifiliis* in channel catfish *Ictalurus punctatus*.** *Dis Aquat Organ* 2004, **59**(2):131-134.
6. Scholz T: **Parasites in cultured and feral fish.** *Vet Parasitol* 1999, **84**(3-4):317-335.
7. Ben-Porath I, Weinberg RA: **When cells get stressed: an integrative view of cellular senescence.** *J Clin Invest* 2004, **113**(1):8-13.
8. Campisi J: **Senescent cells, tumor suppression, and organismal aging: good citizens, bad neighbors.** *Cell* 2005, **120**(4):513-522.
9. Herbig U, Sedivy JM: **Regulation of growth arrest in senescence: telomere damage is not the end of the story.** *Mech Ageing Dev* 2006, **127**(1):16-24.
10. Peters G: **An INKlination for epigenetic control of senescence.** *Nat Struct Mol Biol* 2008, **15**(11):1133-1134.

11. Sun HY, Noe J, Barber J, Coyne RS, Cassidy-Hanley D, Clark TG, Findly RC, Dickerson HW: **Endosymbiotic bacteria in the parasitic ciliate *Ichthyophthirius multifiliis***. *Appl Environ Microbiol* 2009, **75**(23):7445-7452.
12. Dickerson HW, Clark TG, Findly RC: ***Ichthyophthirius multifiliis* has membrane-associated immobilization antigens**. *J Protozool* 1989, **36**(2):159-164.
13. Swennes AG, Findly RC, Dickerson HW: **Cross-immunity and antibody responses to different immobilisation serotypes of *Ichthyophthirius multifiliis***. *Fish Shellfish Immunol* 2007, **22**(6):589-597.
14. Xu DH, Klesius PH, Shoemaker CA: **Protective immunity of Nile tilapia against *Ichthyophthirius multifiliis* post-immunization with live theronts and sonicated trophonts**. *Fish Shellfish Immunol* 2008, **25**(1-2):124-127.
15. Clark TG, McGraw RA, Dickerson HW: **Developmental expression of surface antigen genes in the parasitic ciliate *Ichthyophthirius multifiliis***. *Proc Natl Acad Sci U S A* 1992, **89**(14):6363-6367.
16. Damaj R, Pomel S, Bricheux G, Coffe G, Vignes B, Ravet V, Bouchard P: **Cross-study analysis of genomic data defines the ciliate multigenic epiplasmin family: strategies for functional analysis in *Paramecium tetraurelia***. *BMC Evol Biol* 2009, **9**:125.
17. Pomel S, Diogon M, Bouchard P, Pradel L, Ravet V, Coffe G, Vignes B: **The membrane skeleton in *Paramecium*: Molecular characterization of a novel epiplasmin family and preliminary GFP expression results**. *Protist* 2006, **157**(1):61-75.
18. Jousson O, Di Bello D, Donadio E, Felicioli A, Pretti C: **Differential expression of cysteine proteases in developmental stages of the parasitic ciliate *Ichthyophthirius multifiliis***. *FEMS Microbiol Lett* 2007, **269**(1):77-84.

19. Klemba M, Goldberg DE: **Biological roles of proteases in parasitic protozoa.** *Annu Rev Biochem* 2002, **71**:275-305.
20. Sajid M, McKerrow JH: **Cysteine proteases of parasitic organisms.** *Mol Biochem Parasitol* 2002, **120**(1):1-21.
21. Parama A, Iglesias R, Alvarez MF, Leiro J, Ubeira FM, Sanmartin ML: **Cysteine proteinase activities in the fish pathogen *Philasterides dicentrarchi* (Ciliophora: Scuticociliatida).** *Parasitology* 2004, **128**(Pt 5):541-548.
22. Monastyrskaya K, Babiychuk EB, Draeger A: **The annexins: spatial and temporal coordination of signaling events during cellular stress.** *Cell Mol Life Sci* 2009, **66**(16):2623-2642.
23. van Zandbergen G, Bollinger A, Wenzel A, Kamhawi S, Voll R, Klinger M, Muller A, Holscher C, Herrmann M, Sacks D *et al*: ***Leishmania* disease development depends on the presence of apoptotic promastigotes in the virulent inoculum.** *Proc Natl Acad Sci U S A* 2006, **103**(37):13837-13842.
24. Xu DH, Klesius PH, Shoemaker CA: **Cutaneous antibodies from channel catfish, *Ictalurus punctatus* (Rafinesque), immune to *Ichthyophthirius multifiliis* (Ich) may induce apoptosis of Ich theronts.** *J Fish Dis* 2005, **28**(4):213-220.
25. Paugam A, Bulteau AL, Dupouy-Camet J, Creuzet C, Friguet B: **Characterization and role of protozoan parasite proteasomes.** *Trends Parasitol* 2003, **19**(2):55-59.
26. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S: **The protein kinase complement of the human genome.** *Science* 2002, **298**(5600):1912-1934.
27. Doerig C, Meijer L, Mottram JC: **Protein kinases as drug targets in parasitic protozoa.** *Trends Parasitol* 2002, **18**(8):366-371.

28. Eisen JA, Coyne RS, Wu M, Wu D, Thiagarajan M, Wortman JR, Badger JH, Ren Q, Amedeo P, Jones KM *et al*: **Macronuclear genome sequence of the ciliate *Tetrahymena thermophila*, a model eukaryote.** *PLoS Biol* 2006, **4**(9):e286.
29. Doerig C: **Protein kinases as targets for anti-parasitic chemotherapy.** *Biochim Biophys Acta* 2004, **1697**(1-2):155-168.
30. Lokanathan Y, Mohd-Adnan A, Wan KL, Nathan S: **Transcriptome analysis of the *Cryptocaryon irritans* tomont stage identifies potential genes for the detection and control of cryptocaryonosis.** *BMC Genomics*, **11**:76.
31. Abernathy JW, Xu P, Li P, Xu DH, Kucuktas H, Klesius P, Arias C, Liu Z: **Generation and analysis of expressed sequence tags from the ciliate protozoan parasite *Ichthyophthirius multifiliis*.** *BMC Genomics* 2007, **8**:176.
32. Wright AD, Lynn DH: **Phylogeny of the fish parasite *Ichthyophthirius* and its relatives *Ophryoglena* and *Tetrahymena* (Ciliophora, Hymenostomatia) inferred from 18S ribosomal RNA sequences.** *Mol Biol Evol* 1995, **12**(2):285-290.
33. Pao SS, Paulsen IT, Saier MH, Jr.: **Major facilitator superfamily.** *Microbiol Mol Biol Rev* 1998, **62**(1):1-34.
34. Sa-Correia I, Tenreiro S: **The multidrug resistance transporters of the major facilitator superfamily, 6 years after disclosure of *Saccharomyces cerevisiae* genome sequence.** *J Biotechnol* 2002, **98**(2-3):215-226.
35. Smith UM, Consugar M, Tee LJ, McKee BM, Maina EN, Whelan S, Morgan NV, Goranson E, Gissen P, Lilliquist S *et al*: **The transmembrane protein meckelin (MKS3) is mutated in Meckel-Gruber syndrome and the wpk rat.** *Nat Genet* 2006, **38**(2):191-196.

36. Takahashi A, Ohtani N, Hara E: **Irreversibility of cellular senescence: dual roles of p16INK4a/Rb-pathway in cell cycle control.** *Cell Div* 2007, **2**:10.
37. Collins K, Gorovsky MA: **Tetrahymena thermophila.** *Curr Biol* 2005, **15**(9):R317-318.
38. Giardini MA, Lira CB, Conte FF, Camillo LR, de Siqueira Neto JL, Ramos CH, Cano MI: **The putative telomerase reverse transcriptase component of Leishmania amazonensis: gene cloning and characterization.** *Parasitol Res* 2006, **98**(5):447-454.
39. Blander G, Guarente L: **The Sir2 family of protein deacetylases.** *Annu Rev Biochem* 2004, **73**:417-435.
40. Imai S, Armstrong CM, Kaeberlein M, Guarente L: **Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase.** *Nature* 2000, **403**(6771):795-800.
41. Kennedy BK, Smith ED, Kaeberlein M: **The enigmatic role of Sir2 in aging.** *Cell* 2005, **123**(4):548-550.
42. Kaeberlein M, McVey M, Guarente L: **The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms.** *Genes Dev* 1999, **13**(19):2570-2580.
43. Tonkin CJ, Carret CK, Duraisingh MT, Voss TS, Ralph SA, Hommel M, Duffy MF, Silva LM, Scherf A, Ivens A *et al*: **Sir2 paralogues cooperate to regulate virulence genes and antigenic variation in Plasmodium falciparum.** *PLoS Biol* 2009, **7**(4):e84.
44. Borst P, Genest PA: **Parasitology: switching like for like.** *Nature* 2006, **439**(7079):926-927.
45. Templeton TJ: **Whole-genome natural histories of apicomplexan surface proteins.** *Trends Parasitol* 2007, **23**(5):205-212.

46. Clark TG, Gao Y, Gaertig J, Wang X, Cheng G: **The I-antigens of *Ichthyophthirius multifiliis* are GPI-anchored proteins.** *J Eukaryot Microbiol* 2001, **48**(3):332-337.
47. Abernathy JW, Xu D-H, Li P, Klesius P, Kucuktas H, Liu Z: **Transcriptomic profiling of *Ichthyophthirius multifiliis* reveals polyadenylation of the large subunit ribosomal RNA.** *Comp Biochem Physiol Part D Genomics Proteomics* 2009, **4**(3):179-186.
48. Ballesteros M, Fredriksson A, Henriksson J, Nystrom T: **Bacterial senescence: protein oxidation in non-proliferating cells is dictated by the accuracy of the ribosomes.** *EMBO J* 2001, **20**(18):5280-5289.
49. Nystrom T: **Translational fidelity, protein oxidation, and senescence: lessons from bacteria.** *Ageing Res Rev* 2002, **1**(4):693-703.
50. Chen YA, McKillen DJ, Wu S, Jenny MJ, Chapman R, Gross PS, Warr GW, Almeida JS: **Optimal cDNA microarray design using expressed sequence tags for organisms with limited genomic information.** *BMC Bioinformatics* 2004, **5**:191.
51. Jenny MJ, Chapman RW, Mancia A, Chen YA, McKillen DJ, Trent H, Lang P, Escoubas JM, Bachere E, Boulo V *et al*: **A cDNA microarray for *Crassostrea virginica* and *C. gigas*.** *Mar Biotechnol (NY)* 2007, **9**(5):577-591.
52. Boothroyd JC, Blader I, Cleary M, Singh U: **DNA microarrays in parasitology: strengths and limitations.** *Trends Parasitol* 2003, **19**(10):470-476.
53. Van Den Bussche RA, Hooper SR, Drew CP, Ewing MS: **Characterization of histone H3/H4 gene region and phylogenetic affinity of *Ichthyophthirius multifiliis* based on H4 DNA sequence variation.** *Mol Phylogenet Evol* 2000, **14**(3):461-468.

54. Bolstad BM, Irizarry RA, Astrand M, Speed TP: **A comparison of normalization methods for high density oligonucleotide array data based on variance and bias.** *Bioinformatics* 2003, **19**(2):185-193.
55. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP: **Exploration, normalization, and summaries of high density oligonucleotide array probe level data.** *Biostatistics* 2003, **4**(2):249-264.
56. Li C, Wong WH: **Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection.** *Proc Natl Acad Sci U S A* 2001, **98**(1):31-36.
57. Jousson O, Pretti C, Di Bello D, Cognetti-Varriale AM: **Non-invasive detection and quantification of the parasitic ciliate *Ichthyophthirius multifiliis* by real-time PCR.** *Dis Aquat Organ* 2005, **65**(3):251-255.
58. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR.** *Nucleic Acids Res* 2001, **29**(9):e45.
59. Pfaffl MW, Horgan GW, Dempfle L: **Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR.** *Nucleic Acids Res* 2002, **30**(9):e36.

Appendix 1

Section III. Generation and Analysis of Expressed Sequence Tags from the Ciliate Protozoan Parasite *Ichthyophthirius multifiliis*.

The electronic version of this article is the complete one and can be found online at: <http://www.biomedcentral.com/1471-2164/8/176>. All Supplemental Files are available in the public domain using the Digital Object Identifier (DOI) number of doi:10.1186/1471-2164-8-176.

Appendix 2

Section V. A table of the microarray results when comparing expression among all three life-stages of Ich at 2-fold differential expression cut-off. Data includes GenBank accession numbers or contig ID, putative sequence identity, and fold-change in expression.

<u>Accession number/Contig ID</u>	<u>Putative identity</u>	<u>Fold change, Tomont (baseline) and Theront (experiment)</u>	<u>Fold change, Theront (baseline) and Trophont (experiment)</u>	<u>Fold change, Trophont (baseline) and Tomont (experiment)</u>
EL905232	histone h2b	-9.21	3.64	2.53
EL905249	protein kinase domain containing protein	-22.07	7.91	2.79
EL905772	immobilization antigen variant b	2.36	-6.99	2.95
EL906059	Acyl-coenzyme a dehydrogenase	3.97	-12.18	3.07
EL907192	proteasome a-type and b-type family protein	-16.65	6.17	2.7
EL909156	---NA---	-7.62	30.71	-4.03
EL912238	prefoldin subunit family protein	-8.29	3.06	2.71
EL916934	dnak protein	4.03	-11.83	2.93
EL922591	cathepsin L cysteine protease	4.65	-28.17	6.06
EL926730	ribosomal protein l36	-9.87	71.24	-7.22
EL926770	ribosomal protein l36	-11.53	91.35	-7.92
EL927085	---NA---	-5.02	35.59	-7.09
EL927754	---NA---	-7.05	56.5	-8.01
EL928059	diamine acetyltransferase 1	-4.03	15.82	-3.92
EL928170	---NA---	-5.76	48.23	-8.38
EL928193	---NA---	-5.16	31.53	-6.11
EL928216	ribosomal protein s14	-10.77	65.32	-6.06
EL928230	ribosomal protein s14	-18.27	89.68	-4.91
EL928291	receptor expression-enhancing protein 5	-5.87	26.71	-4.55
Ich_Contig_1056	40s ribosomal protein s23	3.38	-9.27	2.74
Ich_Contig_113	dnaj protein	3.61	-12.76	3.53

Ich_Contig_12	---NA---	-11.21	37.25	-3.32
Ich_Contig_1288	hsp90	3.12	-7.69	2.46
Ich_Contig_1298	---NA---	-21.29	7.78	2.74
Ich_Contig_137	peroxiredoxin 5	-7.82	44.61	-5.71
Ich_Contig_1434	t-complex protein 1	-9.48	3.43	2.76
Ich_Contig_1444	annexin a5	-3.48	56.9	-16.33
Ich_Contig_1476	son DNA-binding protein	-8.62	57.05	-6.62
Ich_Contig_15	---NA---	-5.48	38.32	-7
Ich_Contig_1532	cytochrome p450 like_tbp	3.51	-14.19	4.05
Ich_Contig_1536	calcium-translocating p-type pmca-type family protein	2.63	-23.07	8.76
Ich_Contig_1599	egf-like domain containing protein	-2.81	9.01	-3.21
Ich_Contig_170	nucleoside diphosphate kinase b	-7.39	33.77	-4.57
Ich_Contig_1753	hypothetical protein THERM_00136120 [Tetrahymena thermophila]	-6.18	15.32	-2.48
Ich_Contig_1767	---NA---	-13.96	2.93	4.77
Ich_Contig_1769	histone h3	-44.99	11.44	3.93
Ich_Contig_1771	ras family protein	2.77	-8.31	3
Ich_Contig_180	---NA---	-12.62	54.45	-4.32
Ich_Contig_1816	cAMP-dependent protein kinase	3.51	-9.53	2.72
Ich_Contig_1961	---NA---	-4.18	39.09	-9.34
Ich_Contig_1987	similar to germinal histone H4	-15.81	47.75	-3.02
Ich_Contig_2024	---NA---	-6.71	37.89	-5.65
Ich_Contig_2055	sjchg03012 protein	-5.42	21.29	-3.93
Ich_Contig_2096	---NA---	7.35	-24.42	3.32
Ich_Contig_2099	s100 calcium binding protein a16	-5.69	45.69	-8.03
Ich_Contig_2128	ribosomal protein s9	-8.33	63.18	-7.58
Ich_Contig_2262	loc569167 protein	-6.97	27.43	-3.94
Ich_Contig_2325	phosphatidylinositol-4-phosphate 5-kinase	2.32	-11.9	5.12

Ich_Contig_2326	transposable element tc1 transposase	-19.4	71.75	-3.7
Ich_Contig_2338	thioredoxin-like 1	-3	12.39	-4.14
Ich_Contig_2419	---NA---	-9.58	39.75	-4.15
Ich_Contig_2482	---NA---	-8.6	52.93	-6.15
Ich_Contig_2514	immobilization antigen isoform	24.96	-109.83	4.4
Ich_Contig_256	SOUL heme-binding protein	4.78	-28.53	5.97
Ich_Contig_2584	transposase	-36.5	105.93	-2.9
Ich_Contig_260	integrin beta 1 binding protein 2	4.26	-24.39	5.73
Ich_Contig_2644	keratin	-17.46	51.99	-2.98
Ich_Contig_2693	---NA---	-42.65	175.27	-4.11
Ich_Contig_2708	---NA---	-20.87	5.57	3.75
Ich_Contig_2723	---NA---	-5.58	18.36	-3.29
Ich_Contig_2733	---NA---	-4.37	25.04	-5.72
Ich_Contig_2752	---NA---	-7.08	35.6	-5.03
Ich_Contig_2788	cystathionine gamma-lyase	3.06	-21.7	7.08
Ich_Contig_2794	aconitate hydratase 1 family protein	4.59	-18.18	3.96
Ich_Contig_329	---NA---	-9.27	42.86	-4.62
Ich_Contig_40	protein kinase domain containing protein	18.8	-6.59	-2.85
Ich_Contig_481	fumarylacetoacetase	5.6	-19.82	3.54
Ich_Contig_538	stress-induced-phosphoprotein 1 (hsp70 hsp90-organizing protein)	2.51	-9.19	3.66
Ich_Contig_665	dihydropyridyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	2.96	-8.25	2.78
Ich_Contig_670	epiplasmin 1	4.74	-14.33	3.03
Ich_Contig_701	protein kinase domain containing protein	2.41	-7	2.91
Ich_Contig_743	erf1 domain 3 family protein	3.17	-8.85	2.79
Ich_Contig_813	cdc2 protein kinase	3.64	-9.3	2.56
Ich_Contig_818	dnak protein	3.42	-11.89	3.48

Ich_Contig_895	---NA---	-10.2	2.7	3.78
Ich_Contig_898	DEAD/DEAH box helicase family protein	-7.85	3.08	2.55
EG961300	orf2-encoded protein	-47.95	133.14	-2.78
EG958491	---NA---	-12.42	59.71	-4.81
EG958591	transposable element tcb1 transposase	-35.15	107.38	-3.05
EG958713	---NA---	-19.12	63.95	-3.34
EG958924	keratin K10	-7.35	22.86	-3.11
EG958935	reverse transcriptase-like protein	-25.37	73.71	-2.9
EG960190	---NA---	-11.86	50.93	-4.29
EG961600	conserved hypothetical protein	-15.72	-13.12	206.36
EG961911	---NA---	-12.03	61.67	-5.13
EG962656	transposase	-14.28	61.33	-4.29
EG964333	---NA---	-3.78	20.34	-5.37
EG964556	transposase	-18.96	46.28	-2.44
EG964622	keratin 19	-16.68	43.66	-2.62
EG965645	atp-dependent protease la	3.44	-11.08	3.22
EG966176	---NA---	-3.42	48.95	-14.3
EG966259	---NA---	-18.26	67.23	-3.68
EG957901	ribosomal protein l6	-10.78	55.04	-5.1
EG957953	---NA---	-4.63	13.65	-2.95
EG957977	40s ribosomal protein sa	-5.78	83.21	-14.4
EG958024	dhhc zinc finger domain containing protein	7.62	-29.31	3.85
EG958026	---NA---	-8.14	25.47	-3.13
EG958028	---NA---	-40.11	143.57	-3.58
EG958089	orf2-encoded protein	-70.46	207.84	-2.95
EG958094	---NA---	-13.1	51.35	-3.92
EG958134	Hydrolethalus syndrome protein 1	-3.47	20.02	-5.77
EG958373	---NA---	-13.46	61.53	-4.57

EG958434	---NA---	-3.9	19.43	-4.99
EG958449	keratin 15	-11.46	30.38	-2.65
EG958618	---NA---	-9.25	31.35	-3.39
EG958901	s100 calcium binding protein a14	-11	55.42	-5.04
EG959018	ribosomal protein s26	-6.02	54.81	-9.11
EG959050	---NA---	-10.74	45.45	-4.23
EG959228	---NA---	-9.22	26.55	-2.88
EG959530	---NA---	-7.51	23.44	-3.12
EG959547	---NA---	-3.59	43.11	-12.02
EG959763	epithelial membrane protein 2	-18.84	81.22	-4.31
EG959786	40s ribosomal protein s2	-8.7	58.4	-6.71
EG959915	predicted protein	-2.42	7.9	-3.27
EG959993	---NA---	-50.84	139.42	-2.74
EG960061	casein kinase i	3.03	-10.41	3.44
EG960100	transposable element tcb1 transposase	-8.31	22.5	-2.71
EG960143	---NA---	-3.05	18.87	-6.19
EG960172	pol-like protein	-4.46	28.03	-6.28
EG960199	transposable element tcb1 transposase	-39.5	114.69	-2.9
EG960463	---NA---	-6.58	58.59	-8.91
EG960593	---NA---	-6.56	17.37	-2.65
EG960627	---NA---	-4.25	12.09	-2.84
EG960708	---NA---	-16.67	101.47	-6.09
EG960906	---NA---	-3.82	20.69	-5.41
EG961151	---NA---	-6.11	71.57	-11.71
EG961194	---NA---	-4.34	27.92	-6.43
EG961276	---NA---	-7.97	32.34	-4.06
EG961279	---NA---	-6.82	31.86	-4.67
EG961308	unnamed protein product	-13.76	37.37	-2.72
EG962032	annexin a1	-10.96	51.98	-4.74

EG962060	---NA---	-13.99	54.65	-3.91
EG962112	---NA---	-5.74	18.89	-3.29
EG962451	---NA---	-9.35	36.79	-3.93
EG962490	---NA---	-10.27	34.41	-3.35
EG962494	---NA---	-34.69	101.97	-2.94
EG962577	---NA---	-4.82	25.7	-5.33
EG962684	---NA---	-7.1	29.89	-4.21
EG962753	transposase	-14.36	47.08	-3.28
EG962929	ribosomal protein s16	-7.11	38.87	-5.46
EG962932	---NA---	-7.93	36.52	-4.6
EG963061	phospholipid hydroperoxide glutathione peroxidase	-11.26	42.86	-3.81
EG963088	---NA---	-25.07	91.54	-3.65
EG963124	transposable element tcb1 transposase	-27.51	99.76	-3.63
EG963215	---NA---	-27.79	86.77	-3.12
EG963227	---NA---	-5.18	50.82	-9.81
EG963268	pol-like protein	-9.69	34.67	-3.58
EG963317	ribosomal protein l4	-4.64	45.17	-9.73
EG963404	---NA---	-4	19.56	-4.89
EG963442	---NA---	-17.23	73.52	-4.27
EG963512	---NA---	-75.38	205.48	-2.73
EG963704	---NA---	-19.97	106.4	-5.33
EG963902	---NA---	-5.05	18.32	-3.63
EG963970	---NA---	-5.61	20.26	-3.61
EG964088	---NA---	-5.08	12.01	-2.36
EG964153	elongation factor 2	-7.74	85.36	-11.03
EG964167	---NA---	-18.59	58.6	-3.15
EG964314	transposable element tcb1 transposase	-22.5	81.93	-3.64
EG964383	elongation factor 1-alpha	-4.99	30.09	-6.03
EG964450	elongation factor-1 alpha	-4.77	31.7	-6.64

EG964454	transgelin	-19.62	48.8	-2.49
EG964472	unknown [Schistosoma japonicum]	-58.71	158.32	-2.7
EG964705	---NA---	-63.21	182.33	-2.88
EG964813	---NA---	-12.75	44.42	-3.49
EG964841	---NA---	-14.92	56.8	-3.81
EG965149	---NA---	-10.38	34.27	-3.3
EG965423	---NA---	-6.79	25.08	-3.7
EG965584	NADH dehydrogenase subunit 5	-32.72	102.44	-3.13
EG965597	lambda-recombinase-like protein	-40.27	111.57	-2.77
EG965699	actin	-7.7	30.3	-3.93
EG965977	---NA---	-11.04	38.71	-3.51
EG966049	---NA---	-6.87	44.65	-6.49
Tetrahymena XM_001013532.1	---NA---	-14.8	49.52	-3.35

Appendix 3

Section V. A table of the microarray results of common transcripts when comparing expression among tomont and trophont passages P1 to P100 at 5-fold differential expression cut-off. Data includes GenBank accession numbers or contig ID, putative sequence identity, and fold-change in expression.

<u>Ich Accession number / Contig ID</u>	<u>BLASTX Description</u>	<u>Trophont P1 to P100 fold change</u>	<u>Tomont P1 to P100 fold change</u>
BQ134884	Transmembrane amino acid transporter protein	42.05	7.28
BQ134924	hypothetical protein TTHERM_00647070	54.6	11.31
BQ134942	DEAD/DEAH box helicase family protein	32.88	6.78
EL904069	--NA--	-12.28	-12.01
EL905174	hypothetical protein TTHERM_00294750	-11.09	-6.48
EL906204	hypothetical protein TTHERM_00052140	21.8	8.43
EL906498	Na,H/K antiporter P-type ATPase, alpha subunit family	8.97	16.81
EL907428	AAA family ATPase, CDC48 subfamily protein	-11.98	-7.68
EL907545	Phosphatidylinositol 3- and 4-kinase family protein	7.78	7.65
EL908060	major facilitator superfamily protein	9.77	6.68
EL908097	hypothetical protein TTHERM_00455630	11.42	6.58
EL908572	--NA--	14.76	15.89
EL908691	Leishmanolysin family protein	63.55	8.76
EL908702	ABC transporter N-terminus family protein	14.44	8.18
EL908880	i-antigen [<i>Ichthyophthirius multifiliis</i>]	-8.12	-24.24
EL908881	leucyl-tRNA synthetase family protein	10.07	6.41
EL908998	Glycosyl hydrolase family 20, catalytic domain	6.95	9.22
EL909161	SET domain containing protein	35.71	6.28
EL909187	--NA--	46.37	7.67
EL909320	Protein kinase domain containing protein	20.7	11.89
EL909379	CPSF A subunit region family protein	28.61	12.28
EL909413	Leishmanolysin family protein	9.03	6.88
EL909512	oxidoreductase, aldo/keto reductase family protein	10.69	25.39
EL910508	hypothetical protein TTHERM_00334410	61.57	7.27
EL910640	hypothetical protein TTHERM_00443100	18.19	7.1
EL911297	--NA--	18.18	10.26
EL911381	nucleolar phosphoprotein, putative	20.19	9.3
EL911438	Sec23/Sec24 trunk domain containing protein	22.65	13.36
EL911517	Sec23/Sec24 trunk domain containing protein	42.95	12.61
EL911526	hypothetical protein TTHERM_00354870	27.3	7.43

EL911889	fimbrin-like 71 K protein [Tetrahymena thermophila]	17.61	9.32
EL912026	--NA--	31.83	9.49
EL912077	rRNA pseudouridine synthase, putative family protein	37.22	11.74
EL912566	Cleft lip and palate transmembrane protein 1	10.43	8.54
EL912952	Oxalate/Formate Antiporter protein	13.89	7.48
EL912967	--NA--	31.59	7.7
EL912984	TPR Domain containing protein	21.4	11.24
EL913058	ATP-dependent protease La	12.23	7.8
EL913068	--NA--	8.56	-7.98
EL913180	Class-II DAHP synthetase family protein	13.29	10.48
EL913363	--NA--	-55.5	-23.62
EL913469	--NA--	20.89	7.49
EL913548	hypothetical protein SORBIDRAFT_0285s002020	-11.6	-9.69
EL913586	--NA--	43.93	7.17
EL913691	putative leishmanolysin-like protein	48.48	9.4
EL914051	Leishmanolysin family protein	10.4	11.06
EL914071	--NA--	30.36	6.82
EL914147	--NA--	102.41	6.53
EL914561	--NA--	-73.72	-46.7
EL914591	--NA--	37.82	8.01
EL914783	--NA--	11.49	6.34
EL915354	enoyl-CoA hydratase/isomerase family protein	10.04	8.09
EL915459	Eukaryotic porin family protein	26.69	10.24
EL915477	--NA--	-37.92	-25.13
EL915642	FtsJ-like methyltransferase family protein	15.89	9.35
EL915963	Copper/zinc superoxide dismutase family protein	18.43	7.88
EL915965	MBOAT family protein [Tetrahymena thermophila]	17.11	7.32
EL916191	Leishmanolysin family protein	8.05	7.47
EL916455	E1-E2 ATPase family protein [Tetrahymena thermophila]	9.6	6.34
EL916969	--NA--	-34.69	-28.35
EL917381	hypothetical protein [Paramecium tetraurelia strain d4-2]	66.09	7.96
EL917534	hypothetical protein TTHERM_00629910	304.41	491.09
EL917960	DNA-dependent RNA polymerase family protein	32.97	11.29
EL918340	Leishmanolysin family protein	38.34	9.98
EL918380	hypothetical protein TTHERM_00289410	10.95	6.8
EL918446	pyruvate kinase family protein	12.21	10.38
EL919908	hypothetical protein	10.29	6.51
EL920140	Leishmanolysin family protein	8.69	6.58
EL920339	major facilitator superfamily protein	19.95	13.95
EL921722	hypothetical protein TTHERM_00242270	18.24	15.27
EL921847	Leishmanolysin family protein	17.05	7.28

EL922190	Adenylate and Guanylate cyclase catalytic domain protein	12.65	7.01
EL922199	--NA--	-25.11	-16.75
EL923575	E1-E2 ATPase family protein	27.32	10.36
EL923854	Sodium/hydrogen exchanger family protein	19.93	8.47
EL924484	Pyridoxal-dependent decarboxylase conserved domain	11.83	12.13
EL924619	--NA--	9.95	9.86
EL925371	High cysteine membrane protein VSP-like	12.74	10.01
EL926080	--NA--	-38.73	-20.09
EL926086	conserved hypothetical protein	22.71	13.2
EL926200	Ammonium Transporter Family protei	9.21	10.25
EL926363	Protein kinase domain containing protein	8.09	7.14
EL926567	Protein kinase domain containing protein	20.09	5.62
EL927061	--NA--	15.52	9.39
EL927950	tryptophanyl-tRNA synthetase family protein	19.94	10.09
EL928101	major facilitator superfamily protein	33.87	9.63
Ich_Contig_103	viral A-type inclusion protein [Trichomonas vaginalis G3]	13.05	7.07
Ich_Contig_1085	--NA--	-39.37	-25.22
Ich_Contig_1099	hypothetical protein TTHERM_00446190	-6.81	-7.41
Ich_Contig_1101	Sodium/hydrogen exchanger family protein	74.92	9.11
Ich_Contig_1134	hypothetical protein TTHERM_02141640	-13.34	-17.97
Ich_Contig_1221	--NA--	30.91	6.94
Ich_Contig_1238	hypothetical protein TTHERM_00320150	15.56	7.47
Ich_Contig_1243	--NA--	10.94	9
Ich_Contig_1251	hypothetical protein TTHERM_00276060	11.29	7.41
Ich_Contig_1316	Protein kinase domain containing protein	10.97	8.41
Ich_Contig_1338	hypothetical protein TTHERM_00550700	13.88	8.37
Ich_Contig_1365	ribose-phosphate pyrophosphokinase family protein	8.5	9.49
Ich_Contig_1387	isoleucyl-tRNA synthetase family protein	14.22	7.37
Ich_Contig_1438	Pyridoxal-dependent decarboxylase conserved domain	20.85	7.59
Ich_Contig_1460	hypothetical protein CLOHIR_02084	-12.69	-29.23
Ich_Contig_1565	conserved hypothetical protein	26.41	8.51
Ich_Contig_1575	Region in Clathrin and VPS family protein	17.85	7.81
Ich_Contig_1589	Leishmanolysin family protein	13.93	10.67
Ich_Contig_1615	glutaminyl-tRNA synthetase family protein	7.51	6.43
Ich_Contig_1620	Ribosomal protein L7Ae containing protein	-7.32	-16.2
Ich_Contig_1629	AMP-binding enzyme family protein	42.14	6.79
Ich_Contig_1636	Nucleoside transporter family protein	8.09	10.73
Ich_Contig_1644	hypothetical protein TTHERM_00695660	-12.39	-26.98
Ich_Contig_1655	hypothetical protein SORBIDRAFT_1368s002010	-16.22	-21.03
Ich_Contig_1664	GTP-binding protein YchF containing protein	17.44	5.77
Ich_Contig_1677	hypothetical protein TTHERM_00052140	10.19	14.52

Ich_Contig_1711	hypothetical protein [Paramecium tetraurelia strain d4]	-12.8	-30.83
Ich_Contig_1716	hypothetical protein THERM_00382350	-7.46	-8.98
Ich_Contig_1717	valyl-tRNA synthetase family protein	13.65	8.13
Ich_Contig_1725	U-box domain containing protein	9.98	7.37
Ich_Contig_1769	histone H3 [Tetrahymena thermophila]	-10.23	-8.84
Ich_Contig_1795	--NA--	-34.06	-26.59
Ich_Contig_1797	Cytochrome P450 family protein	12.98	6.91
Ich_Contig_1804	Pyruvate kinase, barrel domain containing protein	14.61	7.69
Ich_Contig_1827	Leishmanolysin family protein	27.79	7.89
Ich_Contig_1844	--NA--	11.68	6.48
Ich_Contig_1852	histone H4	-6.35	-7.34
Ich_Contig_19	hypothetical protein RT0201	-21.42	-9.85
Ich_Contig_1941	Ribosomal protein L13e containing protein	-6.59	-8.99
Ich_Contig_1997	outer surface protein [Rickettsia typhi str. Wilmington]	-68.86	-48.52
Ich_Contig_20	RecName: Full=High mobility group protein	-37.09	-26.76
Ich_Contig_2030	outer membrane protein B (cell surface antigen sca5)	-23.82	-14.25
Ich_Contig_2074	outer membrane protein [Chryseobacterium gleum]	-32.56	-36.24
Ich_Contig_2082	phosphatidylinositol 4-kinase [Tetrahymena thermophila]	11.76	9.42
Ich_Contig_2089	Glutamate racemase [Bacillus thuringiensis]	-49.54	-68.67
Ich_Contig_2090	--NA--	41.67	7.41
Ich_Contig_2098	hypothetical protein BACCAP_03832	-9.99	-8.37
Ich_Contig_2110	--NA--	-14.62	-8.53
Ich_Contig_2130	ACV synthetase PcbAB [Aspergillus flavus]	-32.49	-66.42
Ich_Contig_2133	transposase IS4 [Legionella longbeachae D-4968]	-357.98	-401.22
Ich_Contig_2136	conserved hypothetical protein [Escherichia coli]	-14.45	-12.53
Ich_Contig_2166	--NA--	-11.31	-11.89
Ich_Contig_2256	CHK1 checkpoint-like protein [Helicoverpa armigera]	-8.69	-7.72
Ich_Contig_2278	--NA--	-365.91	-422.98
Ich_Contig_2295	Histone deacetylase family protein	42.54	7.81
Ich_Contig_2301	alcohol dehydrogenase, putative [Perkinsus marinus]	13.56	8.05
Ich_Contig_2329	outer surface protein [Rickettsia Canadensis]	-14	-17.1
Ich_Contig_2356	histone H4	-9.02	-8.4
Ich_Contig_2360	hypothetical protein THERM_00446190	-11.32	-8.01
Ich_Contig_2373	hypothetical protein [Curvibacter putative symbiont]	-6.4	-8.82
Ich_Contig_2382	--NA--	-290.74	-298.69
Ich_Contig_2390	EGF-like domain containing protein	39.05	10.6
Ich_Contig_2398	putative surface antigen [Rickettsia akari str. Hartford]	-10.45	-8.65
Ich_Contig_2473	--NA--	7.21	6.92
Ich_Contig_2519	major facilitator superfamily protein	23.44	9.28
Ich_Contig_2537	RecName: Full=Cold shock-like protein cspA	-18.07	-10.99
Ich_Contig_2541	hypothetical protein [Paramecium tetraurelia strain d4-2]	-8.57	-6.85

Ich_Contig_2572	GroEL [Rickettsia endosymbiont of Bemisia tabaci]	-8.4	-6.26
Ich_Contig_2576	--NA--	-291.69	-113.92
Ich_Contig_2579	--NA--	-362.48	-396.65
Ich_Contig_2592	C2 domain containing protein [Tetrahymena thermophila]	20.91	6.58
Ich_Contig_2610	--NA--	-28.26	-9.09
Ich_Contig_2637	hypothetical protein FAEPRAM212_01166	-16.87	-9.48
Ich_Contig_2652	--NA--	-11.22	-10.18
Ich_Contig_2708	--NA--	-9.66	-22.74
Ich_Contig_2713	conserved hypothetical protein [Magnetospirillum gryphiswaldense]	-206.35	-194.39
Ich_Contig_2714	hypothetical protein [Paramecium tetraurelia strain d4-2]	18.58	16.9
Ich_Contig_2756	GroEL [Rickettsia endosymbiont of Bemisia tabaci]	-13.66	-9.18
Ich_Contig_2781	--NA--	-252.59	-241.78
Ich_Contig_2792	Ubiquitin carboxyl-terminal hydrolase family protein	36.52	6.61
Ich_Contig_2793	Leishmanolysin family protein	19.13	8.04
Ich_Contig_2807	RNA polymerase Rpb1, domain 2 family protein	16.4	8.89
Ich_Contig_2811	30S ribosomal protein S15	-8.52	-5.7
Ich_Contig_2813	glycosyl transferase, group 1 family protein	-16.83	-8.1
Ich_Contig_2821	--NA--	14.07	6.51
Ich_Contig_316	Zinc finger, ZZ type family protein	-10.83	-6.42
Ich_Contig_325	hypothetical protein TTHERM_00912230	21.36	-9.85
Ich_Contig_355	Adenylate and Guanylate cyclase catalytic domain protein	48.54	8.04
Ich_Contig_366	hypothetical protein TTHERM_00474830	49.97	11.26
Ich_Contig_376	Dynein heavy chain family protein	11.81	9.15
Ich_Contig_381	--NA--	-28.16	-19.16
Ich_Contig_422	hypothetical protein TTHERM_00467910	13.33	12.91
Ich_Contig_439	hypothetical protein TTHERM_00961840	21.07	10.22
Ich_Contig_463	RecName: Full=High mobility group protein	-13.83	-11
Ich_Contig_479	p25-alpha family protein [Tetrahymena thermophila]	11.26	9.01
Ich_Contig_482	hypothetical protein TTHERM_00629910	12.04	23.27
Ich_Contig_500	--NA--	9.47	9.6
Ich_Contig_53	--NA--	-16.87	-7.69
Ich_Contig_557	hypothetical protein TTHERM_01246720	9.97	7.95
Ich_Contig_561	hypothetical protein TTHERM_00629910	16.47	28.66
Ich_Contig_606	hypothetical protein TTHERM_00884630	10.77	10.36
Ich_Contig_608	Leishmanolysin family protein	17.01	17.64
Ich_Contig_610	conserved hypothetical protein	25.89	10.99
Ich_Contig_645	histone H4	-8.71	-9.18
Ich_Contig_680	--NA--	74.33	9.89
Ich_Contig_696	Ribosomal protein L24e containing protein	15.91	14.39
Ich_Contig_698	predicted protein [Naegleria gruberi]	-9.64	-7.69
Ich_Contig_716	hypothetical protein TTHERM_00951920	143.8	9.36

Ich_Contig_720	--NA--	17.41	7.26
Ich_Contig_732	hypothetical protein	21.69	8.06
Ich_Contig_759	hypothetical protein TTHERM_00591660	39.51	9.39
Ich_Contig_766	hypothetical protein TTHERM_00277570	18.4	7.36
Ich_Contig_769	hypothetical protein TTHERM_00390120	9.56	8.94
Ich_Contig_804	hypothetical protein TTHERM_00433480	7.01	7.15
Ich_Contig_815	Nucleoside diphosphate kinase family protein	7.63	5.57
Ich_Contig_828	hypothetical protein TTHERM_00473260	7.87	6.8
Ich_Contig_840	Nucleoside transporter family protein	24.73	7.89
Ich_Contig_850	hypothetical protein TTHERM_00543590	11.54	7.29
Ich_Contig_868	Dynein heavy chain family protein	8.7	9.56
Ich_Contig_874	60S ribosomal protein L31,	-6.55	-5.73
Ich_Contig_878	histone H3 [Tetrahymena thermophila]	-8.07	-7.31
Ich_Contig_883	hypothetical protein TTHERM_00550700	9.39	7.31
Ich_Contig_895	--NA--	-10.38	-14.83
Ich_Contig_910	Phosphoglucomutase/phosphomannomutase,	23.36	5.98
Ich_Contig_911	Leishmanolysin family protein	40.3	9.02
Ich_Contig_980	Glutathione S-transferase, N-terminal domain containing protein	17.33	6.98
Ich_Contig_985	Thioredoxin-dependent peroxide reductase	10.21	7.52
Ich_Contig_995	hypothetical protein TTHERM_00550700	14.39	12.34