University of Tennessee, Knoxville
TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations | Graduate School

5-2018

Next Generation Sequencing, Assembly, and Analysis of Bovine and Feline *Tritrichomonas foetus* Genomes Toward Taxonomic Clarification And Improved Therapeutic and Preventive Targets

Ellen Ann Fleetwood
University of Tennessee, emesseng@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Recommended Citation

Fleetwood, Ellen Ann, "Next Generation Sequencing, Assembly, and Analysis of Bovine and Feline *Tritrichomonas foetus* Genomes Toward Taxonomic Clarification And Improved Therapeutic and Preventive Targets. " PhD diss., University of Tennessee, 2018.
https://trace.tennessee.edu/utk_graddiss/4925

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Ellen Ann Fleetwood entitled "Next Generation Sequencing, Assembly, and Analysis of Bovine and Feline *Tritrichomonas foetus* Genomes Toward Taxonomic Clarification And Improved Therapeutic and Preventive Targets." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Comparative and Experimental Medicine.

Rebecca P. Wilkes, Major Professor

We have read this dissertation and recommend its acceptance:

Richard W. Gerhold Jr., Stephen A. Kania, Scott C. Lenaghan

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Next Generation Sequencing, Assembly, and Analysis of Bovine and Feline *Tritrichomonas foetus* Genomes Toward Taxonomic Clarification And Improved Therapeutic and Preventive Targets

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Ellen Ann Fleetwood

May 2018

Copyright © by Ellen Ann Fleetwood

All rights reserved.

Dedication

To Cougar and Nash, you are in my heart forever.

Acknowledgments

This dissertation work would not have been possible without constant support, oversight, and encouragement from my dissertation advisor, Dr. Rebecca Wilkes. I would like to thank both Dr. Wilkes and Dr. Lenaghan for giving me the opportunity to work on a project with such significance for veterinary health. Additionally, Dr. Stephen A. Kania has been a steady advocate for my research career and provided the resources and project for me to begin my sequencing work. Dr. Rick Gerhold was another valued source of knowledge and insight into working in the laboratory with veterinary parasites. Finally, Dr. David Bemis provided very helpful discussions and experimental support with bacteriology issues.

I would like to extend a special thanks to the UT Knoxville Molecular Genomics Sequencing Core – Joe May and Veronica Brown. Joe and Veronica were instrumental in assisting and troubleshooting many aspects of sequencing work and provided facilities that became my laboratory home away from home. Also, Sujata Agarwal at the UT Agriculture Laboratory Core and Dr. Daniel Johnson at the UTHSC Genomics Core were very helpful with computational resources. Also, Kim Rutherford and the Comparative and Experimental Medicine Program at UTCVM have helped me keep on track and sort out obstacles great and small alike. Additionally, James Stanton provided MinION data that will be used in future work.

Additionally, I would like to thank Dr. Patricia Coan for providing clinical training opportunities, support, and advice throughout my time at UTCVM. Thank you for being a wonderful mentor and friend.

Last, but certainly not least, funding for this work was graciously supplied by the Morris Animal Foundation.

Abstract

Tritrichomonas foetus is a bovine and feline parasite and a porcine commensal. This organism is the causative agent of bovine and feline trichomonosis. In cattle, the parasite colonizes the urogenital tract and causes similar symptoms to those caused by *Trichomonas vaginalis* in humans. In cats, the parasite colonizes the gastrointestinal tract and produces a protracted watery diarrhea. In cattle, this parasite can lead to abortions and substantial herd loss due to culling of infected animals, whereas in cats prolonged courses of diarrhea can lead to abandonment or euthanasia.

At the inception of this dissertation work, no genomic data was available for *T. foetus*. The parasitology community has debated the taxonomic relationship between bovine and feline-associated strains of *T. foetus* and other Trichomonad parasites. Some have hypothesized that different pathotypes of *T. foetus* constitute wholly separate species based on a limited number of cross-infectivity studies, a scant amount of genomic DNA and protein sequences, and non-targeted nuclease-based strategies. Still, the community has been slow to adopt this idea.

We aimed to use next-generation sequencing (NGS) technology to: sequence the genomes of bovine and feline isolates, utilize the genomes to determine the taxonomic relationship between bovine and feline-associated *T. foetus*, and determine whether there were detectable genomic differences that might lead to host-specific targets. We hypothesized that significant genomic changes would be detectable and would lead to host-specific targets for future therapeutics.

We successfully extracted genomic DNA and produced *de novo* draft genome assemblies for two *Tritrichomonas foetus* isolates: strain Beltsville and strain Auburn. Our resulting genomic analyses reveal that these are two members of the same species at the molecular level. These results ran contrary to our initial hypothesis, showing that the difference between these two pathotypes may be subtler than previously believed. We used numerous house-keeping, gold standard phylogenetic markers in addition to bioinformatic and phylogenetic analyses to highlight the profound similarity between these two samples. This work should lay the foundation for a multitude of future investigations into *Tritrichomonas foetus* in the hopes of producing better therapeutic strategies and clinical outcomes for bovine and feline populations alike.

Preface

This dissertation utilized Next Generation Sequencing (NGS) technologies to explore veterinary pathogens and parasites with a specific focus on bovine and feline isolates of the Protozoan parasite *Tritrichomonas foetus.*

The dissertation contains two chapters. The first comprises a review on *Tritrichomonas foetus* with a special emphasis on taxonomic, phylogenetic, and molecular studies done to date. The second chapter contains wet lab and computational work and results that answer the major questions posed by the first chapter: 1) Are bovine and feline *T. foetus* the same species? and 2) Are there genomic differences in bovine and feline *T. foetus* that can be identified as potential therapeutic targets? The individual components of this chapter consist of: cell culture and genomic DNA extraction, genome sequencing via multiple NGS platforms (Ion Torrent Personal Genome Machine and Illumina MiSeq), sequencing quality control and trouble-shooting, computational genome assembly and quality assessment, and computational bioinformatics and comparative genomics analysis of our two draft genome sequences. Finally, we discuss the implications of our results and project where future work may lead.

While our analyses have not yet yielded a molecular "smoking gun" that clearly explains or hints at the feline vs. bovine host specificity of *T. foetus* as we had initially hoped, we have generated several promising leads, hypotheses and valuable information nonetheless that will be of substantial to the parasitology community. We also provide substantial evidence for continuing to maintain bovine and feline *T. foetus* as two genotypes within the same species.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

1. NGS	Next Generation Sequencing
2. DNA	Deoxyribonucleic acid
3. RNA	Ribonucleic acid
4. *T. foetus*	*Tritrichomonas foetus*
5. AF	Anterior Flagella
6. PF	Posterior Flagella
7. UM	Undulating Membrane
8. Mb	megabases
9. nM	Nanomolar
10. dsDNA	Double Stranded DNA
11. EDTA	Ethylenediamine Tetra-acetic Acid
12. Taq polymerase	Polymerase from *Thermus aquaticus*
13. PCR	Polymerase Chain Reaction
14. gDNA	Genomic DNA
15. QC	Quality Control
16. GC	Guanine-Cytosine
17. AT	Adenine-Thymine
18. -seq	-sequencing
19. ChIP	Chromatin Immuno-Precipitation
20. bp	Base Pairs
21. BLAST	Basic Local Alignment Search Tool
22. PFAM	Protein Families Database
23. rRNA	Ribosomal RNA
24. dNTPs	Deoxynucleotide Triphosphates
25. ng	Nanogram
26. FU	Fluorescence Intensity Units
27. HMW	High Molecular Weight (DNA)
28. RAPD	Random Amplified Polymorphic DNA
29. ITS	Internal Transcribed Spacer
30. TR	Tandem Repeat
31. CP	Cysteine Protease
32. ug	Microgram
33. ml	Milliliter
34. ul	Microliter
35. MiST	Microbial Signal Transduction (Database)
36. NR	Non-Redundant
37. FISH	Fluorescence *in situ* Hybridization

List of Attachments

1. Tritrichomonas_foetus_tRNAscanSE_Results.xlsx
2. AUTF OrthoVenn Unique Clusters.xlsx
3. Beltsville OrthoVenn Unique Clusters.xlsx
4. Tritrichomonas foetus OrthoVenn SingleCopyOrthologs.xlsx
5. Tritrichomonas Foetus Cysteine Protease 30clearOrthologs PhylogeneticTree.pdf

Chapter 1: Targeted Review of Bovine and Feline *Tritrichomonas foetus*

Introduction:

Tritrichomonas foetus is a protozoan parasite that has significant health and economic impact on cattle and cats. The parasite is difficult to treat with anti-protozoal therapies, and infection can lead to serious consequences including abortion in cattle and euthanasia due to intractable diarrhea in cats. Infected cattle are also culled leading to a magnification of this parasite's effects on herd health. Additionally, the only effective treatment in cats has a narrow therapeutic index and a high incidence of serious neurologic side effects. Thus, knowledge upon how this parasite infects, persists, and parasitizes cattle and cats is desirable to produce new, efficacious therapeutic and preventive strategies. Unfortunately, this mission has been complicated by the wholly different modes of transmission and pathology in cattle and cats affected by *T. foetus*. In cattle, *T. foetus* is a sexually transmitted infection that affects the urogenital tract. In cats, *T. foetus* is transmitted via a fecal-oral route and results in a gastrointestinal infection instead. To further complicate matters, *T. foetus* has also been recently determined to be synonymous with *Tritrichomonas suis*, which is a commensal of the nares, stomach and gastrointestinal tract of pigs. Therefore, *T. foetus* has narrow and peculiar host-specific behavior that prevents generalization between study of each case. Moreover, differing host-specificity has prompted a protracted debate over whether bovine *T. foetus* and feline *T. foetus* are different genotypes (pathotypes) of the same species or are, in fact, two separate species. These challenges and more have been severely hampered by a lack of sequenced genomes and other molecular resources.

Epidemiology:

T. foetus infects and persists in the bovine urogenital tract, which can lead to bovine trichomonosis. Signs associated with this sexually transmitted infection can include vaginitis, abortion and infertility (1). In bovine trichomoniasis, treatment can be attempted with common anti-protozoal drugs such as fenbendazole and metronidazole, but this is not standard practice as infected bulls are typically culled. In cats, neither fenbendazole or metronidazole are effective (2), and while ronidazole has proven to provide limited effectiveness, it also carries serious neurological side effects and is prescribed as an off-label treatment in the US (3). Due to the significant health and economic impact of this parasite, any further characterization of this organism will be of tremendous value to production and veterinary medicine especially with regards to vaccination candidates.

Associations between Trichomonads and felines have been reported for almost a century, though it was not until 2001 that *T. foetus* was shown to be an experimental cause of feline trichomonosis.(4, 5) *Tritrichomonas foetus* has recently been shown to infect the gastrointestinal tract of cats, where it causes chronic large-bowel diarrhea that is refractory to treatment (6). *T. foetus* symptoms are variable in cats, but the disease can manifest as prolonged and intractable diarrhea with foul-smelling feces containing fresh blood and mucus (7). The parasite is transmitted via a fecal-oral route through shared litter box use and mutual grooming behaviors. The parasite can persist for several days in wet stools, and infections are more common in situations with many animals in close quarters such as multi-cat households, catteries, or shelters. Younger cats less than 12 months of age are most frequently affected, and older cats can serve as asymptomatic carriers. The long term prognosis for cats is largely positive, yet the resolution of symptoms can range from as short as 2 months to as long as 3 years (8). In

addition, low levels of parasites are continually shed after resolution of symptoms, serving as a potential source of infection for other cats in a household (6, 7). Unfortunately, due to the cost and commitment required to care for these cats, many end up being relinquished or euthanized.

Morphology:

Tritrichomonas foetus is a challenging organism to work with from an experimental standpoint. As there are few molecular tools currently available for *T. foetus*, most of what we know about this organism has been determined through microscopic observation and cell culture studies (which sharply contrasts "post-genomic" organisms (9)). *T. foetus* is categorized as a parabasalid (a member of *Parabasalia*) within the order *Tritrichomonadida* through both phenotypic observation and also through 18S rRNA taxonomy. Parabasalids are characterized by the presence of a unique Golgi body attached to striated fibers (the namesake parabasal body), unique hydrogenosome organelles, and a mastigont architecture containing a specific arrangement of two kinetosomes (10). *Tritrichomonadida* possess four flagella (three anterior flagella (AF) and a single posterior flagellum (PF)) that are anchored by an A-type costa (as opposed to a B-type costa in *Trichomonadida*). *T. foetus* is further differentiated by a comb-like structure and a rail-type undulating membrane (UM) that spans approximately three-quarters of the length of the cell (10) (see **Figure 1**).

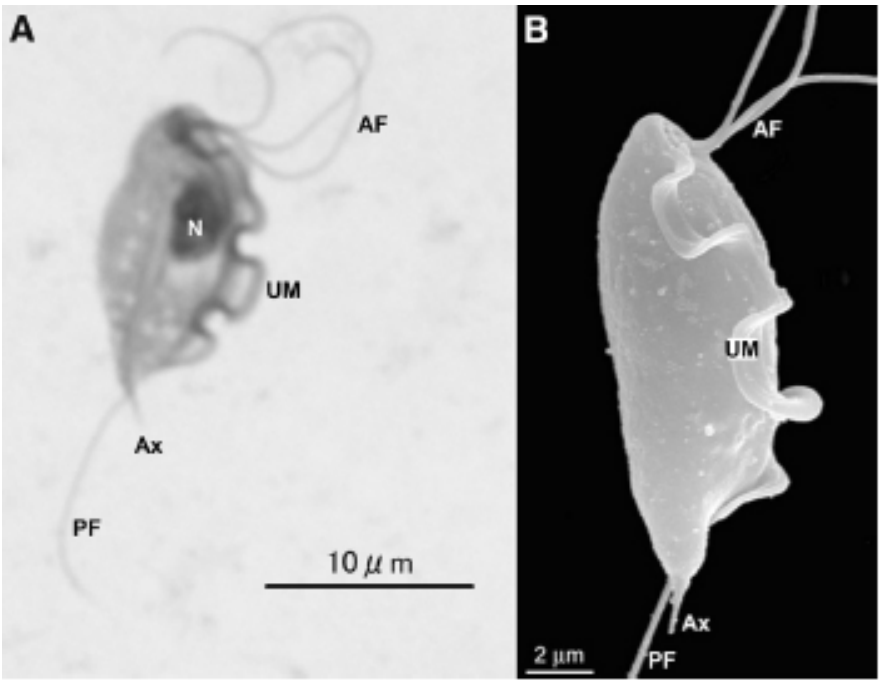

Figure 1. Microscopic Images of *Tritrichomonas foetus* Trophozoites from Yao and Koster.(11) A) Giemsa-stain and B) scanning electron microscopy images of *T. foetus*.

Pathophysiology:

While the exact mechanism of pathology is not entirely clear, a growing body of research has developed in recent years owing to experimental models, including *in vivo* studies and *in vitro* studies in porcine intestinal epithelial cells (of which *T. foetus* is a natural commensal). Specific gains have been made in elucidating the ever-increasing prominent role of cysteine proteases in *T. foetus* infections of all types. Cysteine proteases in *T. foetus* are numerous for both bovine and feline isolates and are involved in mediating adhesion and proteolyzing immunoglobulins to promote survival in the host.(12, 13) Recently, cysteine proteases have been demonstrated to promote adhesion-dependent cytotoxicity in intestinal epithelium.(14) To strengthen the critical role that this protein family plays in *T. foetus* pathophysiology, inhibitors specific for cysteine proteases have also been shown to offer protective effects *in vitro*.(15) In the past year, specific cysteine proteases, such as CP30, have become the subjects of further scrutiny and have been shown to play a role in adhesion and cytotoxicity as well.(16)

Other potential virulence factors have been briefly studied or inferred from other parasites, especially *Trichomonas vaginalis*. Examples of these virulence factors include leishmanolysins, sialidases, glycolytic enzymes such as alpha-mannosidases, phospholipases, and tetraspanins.(17, 18) However, very little has been inferred from research outside of cysteine proteases in *T. foetus* excluding isolated studies. Clearly, much work remains and experimental targets from molecular and genomic sources are in demand.

Molecular Characterization:

The genome of *T. foetus* has not yet been sequenced for either bovine or feline isolates. However, it has been reported that the haploid genome of *T. foetus* is expected to be 177 Mb +/- 11 (strains KV-1 and LUB), consisting of 5 chromosomes (19). This genome size is consistent with the 160 Mb draft genome of another related parabasalid, *Trichomonas vaginalis,* which was sequenced in 2007 and is the causative agent of human trichomoniasis (20). These organisms are often compared and are collectively referred to as Trichomonads (21), but it is important to note that molecular taxonomic methods show that several other characterized organisms may be more closely related to *T. foetus* than *T. vaginalis* (e.g. *Histomonas meleagridis* and *Dientamoeba fragilis*) (see **Figure 2**). Of these closest relatives, only *Histomonas* has been partially sequenced as a limited cDNA library (22).

In the late 1990s, debate over the taxonomic relationship between *T. foetus* and *T. suis* spurred the initial targeted molecular phylogenetics studies for *T. foetus*. These studies began by investigating the 5.8S rRNA internal transcribed spacers (ITS1 and ITS2), which are loci that serve as the gold standard for molecular taxonomic classification and differentiation of closely related parasite species.(23, 24) The 5.8S rRNA ITS regions were first shown to be either identical or nearly identical amongst *T. foetus* and *T. suis* isolates.(25) Based on this work, improved PCR protocols for detecting *Tritrichomonas foetus* 5.8S rRNA ITS regions in both bovine preputial and feline fecal samples were developed.(26-28) Additionally, random amplified polymorphic DNA (RAPD) profiling demonstrated that *T. foetus* and *T. suis* belonged to the same distinct complex apart from a third close relative isolated from squirrel monkeys, *Tritrichomonas mobilensis*.(29) Ultimately, Tachezy *et al.* demonstrated in 2002 that *T. suis* and *T. foetus* were the same species utilizing combined molecular and morphologic techniques.(30) This point was then re-emphasized in a 2005 article by Lun *et al.(31)*

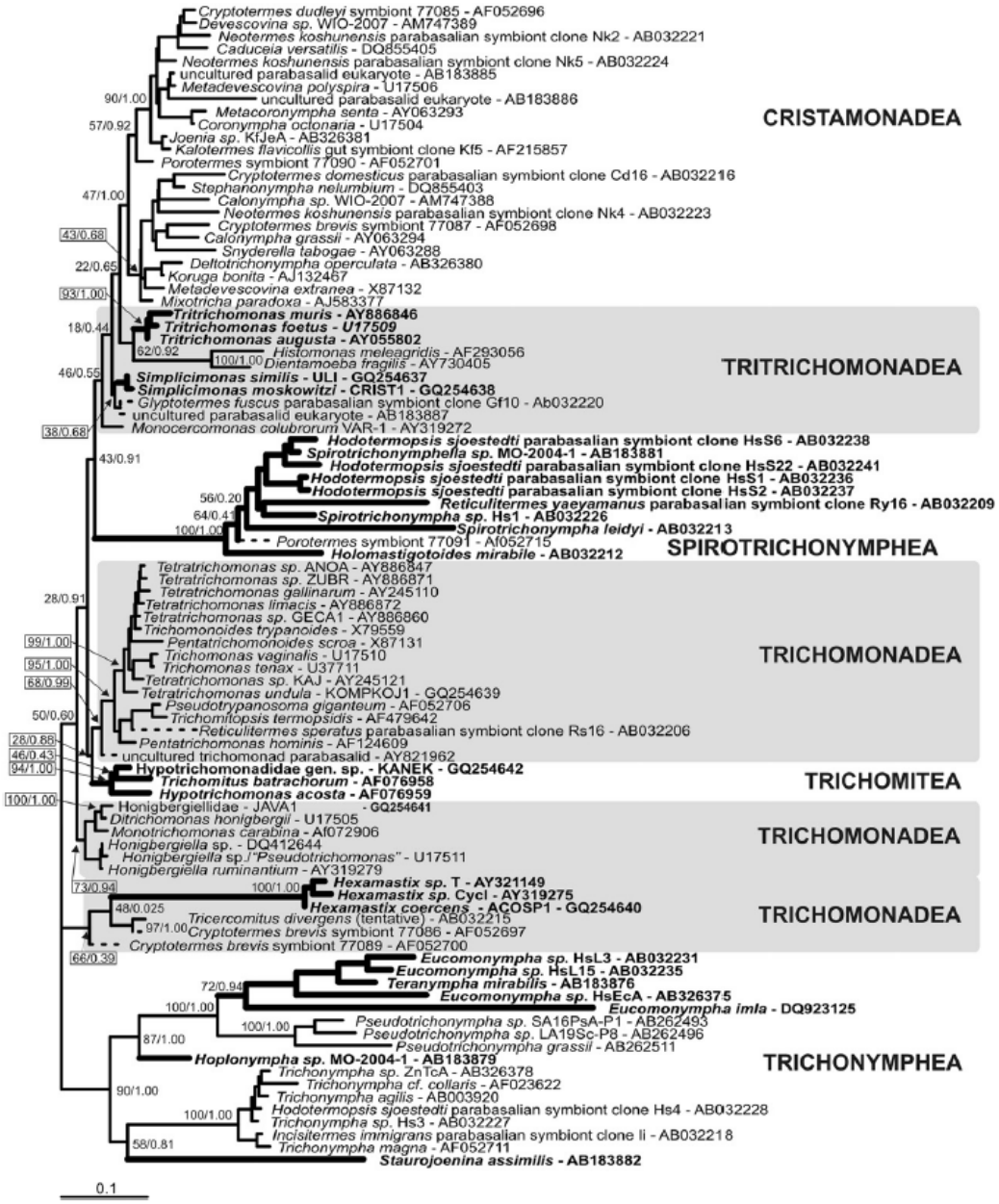

Figure 2. SSU rRNA Phylogenetic Tree of Parabasalid Organisms from Cepicka *et al.* (10). *Tritrichomonas foetus* branches with *T. muris* and *T. augusta* in a clade with *Histomonas meleagridis* and *Dientamoeba fragilis*. *Trichomonas vaginalis* belongs to a separate taxonomic order despite morphologic similarities.

Two Species or Two Pathotype/Genotype Debate:

In 2004, Kleina *et al.* conducted the first large scale molecular phylogenetic analysis of Trichomonadidae family using ITS-1, 5.8S rRNA, and ITS-2 sequences and placed bovine *T. foetus*, feline *T. foetus*, *T. suis*, and *T. mobilensis* in a single clade within the larger *Trichomonadidae* family.(32) In 2007, Stockdale *et al.* launched the first of two cross-infectivity studies where two sets of eight heifers were infected with either the D-1 bovine *T. foetus* strain or the AUTf-1 feline *T. foetus* strain. For this study, both sets of heifers were infected with what the authors determined to be "comparable, but not identical" infections. A direct quote from this manuscript summarizes the results as follows: "this suggests that the Trichomonads induce the same pathology, regardless of origin."(33) The primary reported difference between the two infections was the presence or absence of endometrial surface epithelium, though what effect this created clinically remains unclear.(33) For the second study, Stockdale *et al.* infected six cats with the D-1 bovine *T. foetus* strain and a single cat with the AUTf-1 feline *T. foetus* strain. The results indicated that it was more difficult for a bovine *T. foetus* strain to cause GI pathology in a cat, as only one cat presented with diarrhea and two cats had parasites recoverable from the cecum versus the control cat that both had diarrhea and culture positive small and large bowel results.(34)

Following the cross-infectivity studies, Slapeta *et al.* reported in 2010 that *T. foetus* from domestic cats and cattle were genetically distinct and should be referred to as different "genotypes" of the same species based on 5.8s rRNA ITS1 and TR7/TR8 variable length repeat regions.(35) The same group expanded their analysis in 2012 and published a molecular analysis of bovine and feline *T. foetus*, porcine *T. suis*, and *T. mobilensis* using 10 DNA loci (sequences from eight cysteine proteases, cytosolic malate dehydrogenase 1, and ITS1 + ITS2). They reported that bovine and feline sequences from their study were identical within group and were dissimilar to one another by **1%**. Thus, they ultimately concluded that bovine, feline, and porcine *T. foetus* should be incorporated under the umbrella of *Tritrichomonas foetus* as one species based on the 10 loci that they assessed.(36) In a contemporaneous review, Frey and Muller separately assessed the molecular designations of *Tritrichomonas* spp. and determined that "morphological, physiological and transmission studies" were "not yet adequate to identify and delineate triTrichomonad species." They further posited that whole genome sequencing would be "highly desirable".(1)

The two genotypes within one species consensus was upended in 2013 by a single report from Walden (Stockdale) *et al.* who also performed the previously discussed pair of cross-infectivity studies. In their work, they conducted RAPD profiling of five bovine isolates (D-1 and four Florida strains) and six feline isolates (all AUTf strains from Auburn) (see **Figure 3**). They also produced two phylogenetic analyses: one quantifying the relationships based on the RAPD results and the other depicting relationships inferred from the ITS1 region of 5.8S rRNA. Based primarily upon their previous cross-infectivity work and the results of this RAPD profile, the authors determined that feline *T. foetus* constitutes a distinct organism that justified a new species designation as *T. blagburni.(37)*

The designation of feline *T. foetus* as *T. blagburni* has not been readily or widely accepted in the current literature and conflicting experimental results have also recently been published. Two recent review articles on *T. foetus* have weighed in upon this designation. The first from Yao and Koster in 2015 states: "Based on the molecular sequence differences along with disparity of experimental cross-infections and divergence in pathogenicity Walden (Stockdale) *et al.* proposed a new name for the feline isolate. They named it *T. blagburni*, a new Trichomonad species. However, these authors did not show a clear separation of feline from bovine isolate, a requirement for naming a new species. The new nomenclature, if holding up, still waits to be accepted, which may take a while, especially among veterinarians.

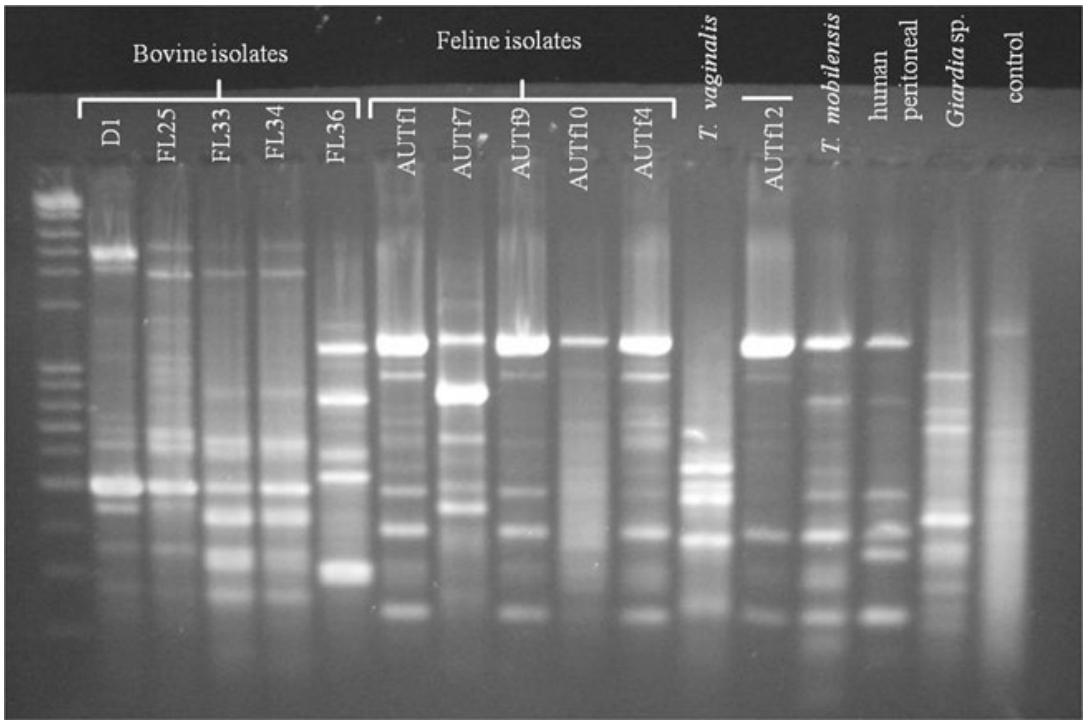

Figure 3. RAPD DNA Profiles of Bovine and Feline *T. foetus* from Walden *et al*.(37) Key analysis from 2013 report leading to the suggestion that feline *T. foetus* should be designated as a new species, *T. blagburni*.

Nevertheless, comparative transcriptomics revealed near identical functional category distribution of expressed genes with no indication of molecular level divergence, which strongly suggested feline and bovine isolates were taxonomically two isolates of one species."(11) Yao and Koster refer to the 2014 BMC Genomics paper on comparative transcriptomics that found few to no differences across the entire transcriptome except for CP8 and CP7, which were found to be the most highly differentially transcribed genes between bovine and feline *T. foetus* isolates.(38) The second review that commented on this issue in 2017 from Gookin *et al.* carefully considers the body of literature in *T. foetus* and states: "it is the authors' opinion that renaming the feline infection at this time may be premature."(4) Finally, a proteomic comparison using two-dimensional gel electrophoresis and tandem mass spectrometry corroborated transcriptomics results - bovine and feline *T. foetus* show tremendous similarity except for in cysteine protease expression, agreeing with the single species theory.(39) All authors from both sides of the issue seem to agree that sequenced genomes will help to definitively decide this issue.

Summary:

The lack of universal agreement on the taxonomic and phylogenetic designation of bovine and feline *T. foetus* stems from its differing host specificities and a paucity of genomic resources. Those arguing for the different species scenario place more weight on host specificity and cross-infectivity experiments, while those arguing for the same species scenario place more weight on identical morphology, phylogenetic markers, and recent "omics" studies results. Without genomes, progress toward resolving this matter will be slow, resource-intensive, and inconclusive. If genomes were to be produced, comparative genomics and whole genome analyses would be able to create thousands of points of comparison between the two organisms rather than relying on the handful that are currently available in *T. foetus.* Additionally, these comparisons will not be limited to this binary dispute, but can inform and be informed by relationships to numerous other sequenced and better characterized organisms.

While it might appear that genomic sequencing would provide a quick fix to this dispute, next-generation sequencing is not as robust a tool for parasitologists as it is for many other microbiologists. Next-generation sequencing has been available for many years and yet no genomes of *T. foetus* or other closely related parasites have been produced aside from the important human pathogen *Trichomonas vaginalis.* With the advent of mature Next-Generation Sequencing (NGS) technologies, some biologists often take for granted the ease of preparing a biological sample for successful sequencing. Indeed, the production of a high-quality draft genome assembly from a previously un-sequenced organism is still an enormous undertaking in molecularly neglected organisms. The first key bottleneck to sequencing, regardless of the type of organism, lies in DNA extraction and library preparation. Without stable, high quality nucleic acids that are free from contaminants, downstream sequencing can yield little data, produce erroneous results, or simply fail entirely. This is especially true for organisms that are wholly dissimilar to widely used model systems like *Escherichia coli* or heavily sequenced data types like *Homo sapiens. Tritrichomonas foetus* falls squarely in the neglected pool of organisms, and protozoan genomes are still an underrepresented group even with the current advances in today's sequencing technology (see **Table 1**).

Table 1. Few Protozoan Genomes Sequenced as of 2014. Table adapted from Ellegren (40). *T. foetus'* taxonomic lineage is highlighted in bold to show few sequenced genomes relative to other eukaryotes.

Kingdom	Phylum	Class	Number of Genomes
Animalia	Annelida	Clitellata	1
		Polychaeta	1
	Arthropoda	Arachnida	5
		Branchiopoda	1
		Chilopoda	1
		Insecta	69
		Maxillopoda	1
	Chordata	Actinopterygii	1
		Amphibia	1
		Aves	11
		Mammalia	73
		Reptilia	6
		Leptocardii	1
	Tunicata	Appendicularia	2
		Ascidiacea	1
	Cnidaria	Anthozoa	2
		Cubozoa	1
		Hydrozoa	1
	Echinodermata	Asteroidea	1
		Echinoidea	2
	Hemichordata		1
	Mollusca	Bivalvia	1
		Gastropoda	2
	Placozoa		1
	Porifera	Demospongiae	1
	Platyhelminthes	Trematoda	3
		Turbellaria	1
	Nematoda	Secernentea	21
		Chromadorea	2
Fungi	Ascomycota		178
	Basidiomycota		48
	Other fungi		22
Rhizaria	Cercozoa	Chlorarachnea	1
Archaeplastida	Rhodophyta	Florideophyceae	1
		Cyanidiophyceae	2
Chromalveolata	Cryptophyta	Cryptophyceae	1
	Heterokontophyta	Bacillariophyceae	1
		Coscinodiscophyceae	2
		Eustigmatophyceae	2
		Oomycetes	12
Alveolata	Apicomplexa		20
	Ciliophora	Ciliatea	1
		Spirotrichea	6
		Oligohymenophorea	1
	Perkinsozoa	Perkinsea	1
Excavata	Euglenozoa	Kinetoplastea	13
	Percolozoa	Heterolobosea	1
Choanoflagellatea			2
Unikonta	Amoebozoa	Mycetozoa	2
	Metamonada	**Parabasalia**	**1**
Viridiplantae	Chlorophyta	Chlorophyceae	2
		Trebouxiophyceae	1
		Trebouxiophyceae	1
		Prasinophyceae	4

Chapter 2: Sequencing and Comparative Genomics of *Tritrichomonas foetus*

The taxonomic relationship between bovine *T. foetus* and feline *T. foetus* is unclear and this point of contention is the major focus of this dissertation work. *T. foetus* from both host species have long been considered the same organism, but recent findings have cast doubt on this assumption. However, while it is has become well-accepted that feline and bovine *T. foetus* isolates are genetically distinct in a handful of marker loci, designation as a new species may require more evidence to support. Clearly, no consensus has been achieved on this subject, and thus we have endeavored to sequence the genomes of both types of *T. foetus* isolates to definitively answer this question. At the inception and for the majority of this dissertation work, there were no genome sequences of any type available for any strain of *T. foetus*.

Our primary motivation for studying *Tritrichomonas foetus* in this project is to determine whether bovine and feline *T. foetus* isolates, which constitute separate genotypes, contain significant differences in genomic content. Furthermore, upon delineating those differences at the whole genome level, we hoped to identify differential genetic potential between the isolates that might provide a hypothesis for the differences in host specificity, which in turn might eventually lead to improved therapeutic targets and interventions for feline and bovine cases alike.

To accomplish this, each isolate must produce a sequenced genome. To sequence a genome, there are multiple steps: first, high molecular weight DNA must be extracted; second, a DNA library must be prepared; third, the actual sequencing chemistry must make base calls to generate raw sequence reads; and finally, software and bioinformatics tools must assemble reads into overlapping contigs in to begin building the genome sequence before the data can be analyzed. Even still, the analysis of whole genomes can be less than straight forward when there is little molecular information available for a newly sequenced organism that is not related to many other sequenced and well characterized species.

This dissertation work culminated in the production of two draft genomes for *Tritrichomonas foetus*. The genotypes sequenced were a bovine specific parasite, *T. foetus* from Beltsville, Maryland (henceforward referred to as Beltsville strain/genotype), and a feline specific parasite from Auburn, Alabama (henceforward AUTF for Auburn *T. foetus*). After our genomes were sequenced and assembled, a separate draft genome from Benchimol *et al.* for the bovine *Tritrichomonas foetus* Strain K was deposited into the NCBI Genome Database and a draft genome announcement was published in February 2017.(41) Comparison with this genome was conducted in order to leverage the status quo of genomic information available for *T. foetus*.

This chapter and dissertation will conclude with the formulation and discussion of working hypotheses for potential differences between bovine and feline *T. foetus* and their differing host specificities, general predictions for the direction of future work in this area of study, and specific suggestions for future investigations of these significant parasites beyond the debate over genotype and species relationships.

Introduction:

Tritrichomonas foetus is a protozoan parasite with two very different hosts: cattle and cats. It also has two very different modes of infection: the bovine type is a sexually transmitted infection and the feline type is a gastrointestinal infection. These differences have led some within the *T. foetus* research community to contest that the bovine and feline types belong to separate species. However, while the amount of data is small, the current molecular evidence (including gold standard phylogenetic markers) supports designation of bovine and feline *T. foetus* as different genotypes of the same species. While the taxonomic relationship is in dispute, there is clear agreement that bovine and feline *T. foetus* cause significant morbidity and mortality for their respective hosts. This significant problem is further worsened by a lack of safe and effective treatment options for most *T. foetus* infections. Furthermore, at the inception of this work, *T. foetus* also lacked a critical molecular resource: a sequenced genome.

There are several pieces of evidence that support the single species concept. First, Random Amplified Polymorphic DNA (RAPD) profiling from Walden *et al.* showed that the banding patterns of random amplified fragments from a set of bovine and a set of feline isolates have different profiles, indicating that they have diverged significantly enough to constitute new species (see **Figure 3** in **Chapter 1**).(37) For specific differences, *T. foetus* from bovine and feline isolates have been shown to be genetically distinct in ITS (internal transcribed spacer) and TR7/TR8 (tandem repeat) regions.(35) Additionally, *T. foetus* bovine and feline isolates demonstrate minor genetic differences in both ITS-2 and EF-1α.(42) Furthermore, distinctions have also been noted during comparisons of the highly conserved ITS regions of 5.8S rRNA that are comparable to those that distinguish other closely related Trichomonad genera.(25) Additionally, minor differences were noted across 10 separate loci (8 of which were cysteine proteases.(35)

However, despite many reported differences, the level of these differences (between 0-1% in most instances) and conflicting molecular and systems biology data argue against a two-species model. At the systems biology level, *T. foetus* transcriptomes from bovine and feline isolates have been shown to be exceedingly similar, with differences primarily noted in only a small handful of genes (e.g. differentially expressed cysteine proteases (CP) CP7 and CP8).(38) Likewise, a 2017 proteomics study comparing feline and bovine *T. foetus* protein expression yielded much the same results: remarkable conservation and few notable differences except in cysteine protease expression.(39) Perhaps most importantly, two cross-infectivity studies with small numbers of isolates and animals appear to come down on both sides of this issue. Feline *T. foetus* isolates have been shown to readily infect cattle and recapitulate a comparable urogenital infection, whereas bovine isolates seem to have more difficulty creating gastrointestinal disease in cats but still manage to do so in some instances.(33, 34)

To resolve this argument and move the *T. foetus* research community closer to a better understanding of *T. foetus* pathophysiology and potential therapeutic strategies, it has become evident that genomes for both types are now necessary to bring the study of *T. foetus* into the post-genomic era. Genomic sequencing and molecular biology of *Tritrichomonas foetus* and most veterinary parasites is a relatively unexplored frontier. However, using genomics has allowed us to study veterinary parasitology and pathogens on a more quantitative, integrative, and larger scale.(43) While the information gained in parasite biology continues to increase, there have been many challenges to the widespread adoption of genomics in veterinary parasitology research.(44, 45). The fundamental obstacle to conducting genomics research is producing genomes, and these have proved technically difficult to produce and slower to

accumulate (relative to bacterial pathogens, for instance) in eukaryotic parasites. Protozoan parasites present numerous challenges for genomic projects such as through various life cycles, difficulty cultivating and obtaining isolates, and handling and amplifying parasite genetic material.(40, 46). Furthermore, many veterinary parasites have inherently complex genomes with high AT or GC content, heterochromatin blocks, and repetitive regions of low complexity that highlight shortcomings of current sequencing technology.(40) If a low complexity repetitive region is longer than the read length it cannot often be assembled or leads to gaps, wrong copy numbers, or collapse of regions into one single locus.(46) Because of a combination of all of the aforementioned factors, many veterinary parasitic genomes often have incomplete assemblies with poor annotation.(46)

Difficulty with genome sequencing in parasitology may seem counter-intuitive to some because it is easy to rationalize parasites have stream-lined genomes due to reliance on host genes for survival. The extension of this is that many parasites have a smaller number of genes compared to their free living relatives.(44) Unfortunately, while this can be true, some parasites are the exception to this rule. This includes the approximately 160 Mb *Trichomonas vaginalis*, one of the largest protozoan parasites ever sequenced. While it is unclear why *T. vaginalis* does not have a reduced gene size, some of the large genome size can be explained. One element contributing to the expanded genome size in *T. vaginalis* is the presence of a large number of repeats and transposable elements. Additionally, it has a large number of gene families that most likely came from prokaryotic lateral gene transfer.(44) As *T. vaginalis* is the closest related genome to *T. foetus,* we expected that all of the challenges that *T. vaginalis* presented would extend to *T. foetus*. Throughout this work, that expectation was met and exceeded, as even *T. vaginalis* is not so closely related to *T. foetus* from a molecular evolutionary standpoint.

Genome sequencing is a multi-disciplinary endeavor that requires the utilization of tools and methods from molecular biology, biochemistry, bioinformatics, and genomics (at a bare minimum). DNA must be extracted from the organism, the DNA must be prepared into a chemically stable library, the library must be sequenced, and the sequences must be computational assembled, assessed, and analyzed. Once stitched into a single file, a genome stands as an authoritative blueprint for the biological potential of that given organism, which makes it the invaluable resource missing from the *T. foetus* debate. However, to produce the *T. foetus* genome (and in this case two genomes), the standard genomic approaches and experimental protocols did not apply. Furthermore, since *T. vaginalis* was not closely related enough, the *T. foetus* genome required *de novo* assembly. *De novo* assembly of a eukaryote like *Tritrichomonas foetus* is a process fraught with computational pitfalls and difficulties that are wholly confined, separate, and/or even deeply intertwined with the novel challenges in the wet lab, too.

Fortunately, this work reports the successful sequencing of two *T. foetus* genomes. The bovine genome was obtained from an isolate from Beltsville, Maryland, and the feline genome was obtained from a clinical isolate from Auburn, Alabama. The feline *T. foetus* genome is the first and only feline *T. foetus* genome that has been produced to our knowledge, and the bovine *T. foetus* genome joins another genome from the bovine *strain K,* which was submitted to the NCBI database by another research group while this work was being completed. The two genomes produced in this work provide compelling evidence for bovine and feline *T. foetus* to continue as different genotypes of the same species. Additionally, comparative analyses of these genomes have highlighted several of the minor genotypic differences that may contribute to host specificity and virulence, and these will serve as the first of many novel and promising therapeutic candidates that will require further investigation. These genomes stand to serve as important resources for the *T. foetus* community and for veterinary parasitologists studying Trichomonads, Parabaslids, and other Protozoa as well.

Materials and Methods:

Cell Culture and Genomic DNA Extraction:

Tritrichomonas foetus isolates initially utilized in this work were *T. foetus* strain D1 (bovine) and *T. foetus* strain AUTF (feline). Later in the project, we obtained another bovine strain (*T. foetus* strain Beltsville) that was ultimately sequenced. Stock cultures were handled using the following protocol. Stock cultures for isolates were stored in -80° C freezer. Cells were "awoken" in a warm water bath until thawed. Samples were then transferred into autoclaved Diamond's Media. Samples were incubated at 37° C upright in 15ml Falcon screw-cap tubes. Isolates were assessed daily for growth under light microscopy. Cell lines were passaged in fresh media until cells reached carrying capacity. Cells were counted with a hemocytometer.

PCR was employed for initial species identification and confirmation. The primers that we used have been shown to be specific for the ITS1 and ITS2 regions (internal transcribed spacers) of the 5.8S rRNA gene of *Tritrichomonas foetus*.(26) Our materials and primers for the PCR reaction included Taq PCR Master Mix kit (Qiagen), water, forward primer (THIIHf) and reverse primer (THIIHr).

To produce sequence data, high molecular weight (HMW) DNA must be successfully extracted from *Tritrichomonas foetus.* The metrics that we used to assess the quality of our HMW DNA were the 260/280 ratio, the 260/230 ratio, and then the concentration in ng/ul. The 260/280 ratio should be ~1.8 for pure DNA and ~2.0 for pure RNA. This ratio can be disrupted if the concentration of nucleic acids is low or if there is contamination by proteins or chemicals involved in extraction such as phenols or guanidine. The 260/230 ratio is used to determine the purity of the nucleic acids, and a value of 2.0-2.2 generally indicates a "clean" sample. There are numerous protocols and kits available for extracting HMW DNA, and these methods can be further subdivided into the types of organisms that they target and the types of sample background from which the organism is derived. However, there are no protocols or kits that have been developed for *T. foetus*. Most work on nucleic acid extraction for *T. foetus* has been targeted towards PCR and diagnostic applications only,(28, 47, 48) which is quite different from HMW DNA extraction because the latter must be large, stable, and a comprehensive representation of the genome as opposed to small, targeted fragments. Thus, we began our search for HMW DNA extraction protocols in the closest related organisms that have been successfully sequenced using next-generation sequencing technology (*Trichomonas vaginalis, Giardia lamblia,* and *Plasmodium falciparum*).(49-51)

The custom-designed protocol that was utilized to produce high quality genomic DNA extraction began with a series of rapid freeze-thaw cycles, which were thought to more effectively disrupt protozoan cells to increase the yield of DNA during extractions. While this protocol included both Proteinase K and RNase A to knock down enzymatic activity and remove unwanted RNA, this protocol also used a phenol-chloroform based lysis and separation method to protect the DNA, which proved to be the most effective technique for our samples. Additionally, this protocol used multiple phenol-chloroform extractions with isoamyl alcohol, which were managed more effectively by specialized MaXtract collection tubes that cut down on cross-contamination and sample loss. The aqueous phase of this protocol contained all nucleic acids rather than separating the DNA into the interphase, as in TRIzol. Finally, additional rounds of RNA removal were performed at the end to ensure the purity of the final DNA. The details of the protocol were as follows with a starting cell count of 5×10^6. Materials included a wash buffer (50 mM Tris-HCl, pH 7.6, 50 mM EDTA, 50 mM D-glucose, 0.5mg/ml Proteinase K (Qiagen)), 10% SDS (Thermo Fisher Scientific),

DNAse-free RNase A (Thermo Fisher Scientific), Phenol, Chloroform, Isoamyl Alcohol, 3 M Sodium Acetate pH 5.0, 100% and 70% EtOH, TE Buffer, pH 8.0 (Illumina), and RiboShredder RNase Blend (Epicentre).

Next Generation Sequencing:

For our *T. foetus* samples, we produced results with mechanical (Bioruptor) and enzymatic digestion as well as transposase-based Tagmentation (Illumina). Ion Torrent libraries for the Ion Torrent Personal Genome Machine (PGM) were constructed for 300bp read lengths using Ion Torrent 318 Chips. Illumina MiSeq libraries were constructed for 250bp read lengths with paired-ends. Size and Quality of samples were periodically assessed at protocol checkpoints using a Bioanalyzer (Agilent). Ion Torrent raw data was quality-assessed to optimize well-loading and minimize poly-clonality. Illumina MiSeq runs were optimized to prevent over-clustering.

Pseudomonas aeruginosa *contamination*:

Samples from stock cultures stored at -80° C were brought up in aliquots of autoclaved Diamond's Media and submitted for blood agar sterility checks at the University of Tennessee College of Veterinary Medicine (UTCVM) Bacteriology Lab. Samples (approximately 0.5-1.0 mL) were inoculated on a blood agar plate which was then incubated at 37°C for 3 days. For contaminated samples, additional samples were provided for gram stain and pyocanin-plate sterility checks specific for *Pseudomonas aeruginosa*. Additionally, at each time we turned in a sample for gram stain and sterility check, we also took a separate aliquot of autoclaved Diamond's media to rule out media contamination.

Once we recognized the presence of *P. aeruginosa* in several of our *T. foetus* isolates, we attempted to purify the *T. foetus* by employing antibiotics specific to gram negative bacteria to which *P. aeruginosa* is susceptible (e.g. Gentamicin). Ultimately, we were unable to remove *P. aeruginosa* without substantially damaging *T. foetus* as well in these samples, which led us to consider alternative methods of separating the organisms. We employed a differential centrifugation protocol to mechanically separate the organisms based on the differences in their size that can be exploited by generating a density gradient.

We utilized *T. foetus* culture at exponential trophozoite growth from D1 (which was confirmed to contain *P. aeruginosa* via blood agar, pyocyanin, and sequencing results). We introduced different concentrations of sucrose/glycerol into the 15ml tubes (10%, 20%, 30%, 40%, and 50%). After mixing to equilibrate the culture, we aliquoted our sample into 15ml tubes. We then spun each tube at 1,000 x g for 5 minutes. We then removed the supernatant (containing *P. aeruginosa*), careful to avoid disturbing the pellet with our *T. foetus*. We then re-suspended the pellet in fresh Diamond's Media. After re-suspending, we then assessed residual presence of bacteria via light microscopy.

After five to seven technical replicates at the maximum density gradient, we thinned the numbers of *P. aeruginosa*. However, we also experienced a significant loss of the numbers of our *T. foetus* isolates to achieve this result. This effect was more substantial than even the loss from our Gentamicin protocol and rendered the samples unusable for downstream sequencing leading us to employ a digital method of contaminant removal.

Sequencing QC:

We utilized both FastQC (52) and CLC Genomics Workbench to assess the GC content of our samples after ever sequencing run. We used BLAST-based methods to determine the closest sequenced organism for our reads. Additionally, we employed read mapping software within the CLC Genomics

Workbench to "purify" contaminating sequences when several samples proved intractable to antibiotic isolation. Specifically, when we detected *Pseudomonas aeruginosa* within several *T. foetus* sequencing runs, we utilized the previously sequenced genome of *Pseudomonas aeruginosa* PAO1 as a "digital magnet" to extract those sequences from our data. The software mapped any reads in our data to the known bacterial genome, and we retained any non-mapping reads as presumptively belonging to *T. foetus*. In this way, we combined the concept of templated assembly to improve our *de novo* assembly efforts. Through combining a combination of all three quality control methods when necessary: GC content, BLAST, and mapping to a known "contaminant", we could then confidently proceed to assembly even with sub-optimal raw inputs (when necessary). *De novo* assembly was conducted in the CLC Genomics Workbench using a *de Bruijn* graph algorithm.

Annotation:

We employed GeneMarkES (53) and Companion (54) as our gene-finder and annotation software of choice for our *de novo* assemblies. While GeneMarkES is a stand-alone *de novo* annotation program for eukaryotic genomes, Companion is a web-based, parasite specific annotation pipeline that utilizes AUGUSTUS (55) as the annotation program and targets the results toward a closely sequenced relative. However, annotation programs routinely use BLAST and generic threshold values to determine when the annotation of the best hit will be transferred to each query sequence, which makes this a very coarse resolution process. GeneMarkES was used with default parameters, whereas Companion required the selection of a sequenced parasite genome template (*Plasmodium falciparum* produced the highest quality annotation).

Functional and Comparative Genomics:

To gain a better view into the functional genomics of our genomes, we utilized the ghostKOALA (KEGG) (56) platform with default parameters as our primary method for determining the gene ontology of individual gene products as well as the predicted biological capabilities of our sequenced organisms. KEGG provides the added benefit of fitting annotated hits into modules and functional pathways that have been curated and possess numerous different sequences from a diverse range of organisms, which allows for a higher confidence in the accuracy of the functional genomic assignments. Still, this process requires very closely related organisms be sequenced for the highest accuracy, and in the case of *T. foetus,* hits were still inherently based on less than ideal BLAST results.

We used OrthoVenn to compare our predicted protein sequences as clusters of orthologous groups (57). Clusters of orthologous groups allow binning based upon sequence similarity; however, careful analysis of the sequences within each group are necessary downstream to determine evolutionary history (e.g. true orthologs, paralogs, or homologs).

Protein Sequence Analysis:

Sequences of virulence factors were extracted using annotation-based searches and custom PERL scripts, aligned with MAFFT (58), and then a maximum likelihood phylogenetic tree was constructed using the JTT matrix-based model (59) in MEGA 7.0.26.(60) A representative of each protein family of virulence factors that we assessed was manually verified to correspond to its functional annotation via BLASTp (61) and PFAM (62) searches. In all cases, we included the recent bovine *T. foetus* strain K in our analyses to increase our ability to determine whether a gene was different between feline and bovine genomes. When either a sequence from AUTF and Beltsville, AUTF and strain K, or all three sequences grouped together as a clade with negligible evolutionary distance separating them, they were deemed to be true orthologs. When sequences branched alone, they were identified as differentiating candidates. These

candidate sequences were then used as queries in local BLASTp searches against the AUTF or Beltsville genome that they were initially missing from to verify that they are indeed absent from our genomes and were not missed by less specific initial methods.

Results:

Extraction of High Quality Genomic DNA for Tritrichomonas foetus*:*

We encountered two major technical challenges during the process of obtaining genomic DNA from our *T. foetus* isolates. First, it was exceedingly difficult to extract stable, high quality DNA for next-generation sequencing that would survive downstream applications including NGS. Second, and possibly related to the extraction difficulties were challenges encountered in constructing an optical map using HMW DNA. The optical map would have been beneficial to guide higher level scaffolding of *de novo* assembly. We overcame the first obstacle with the labor and time-intensive implementation of numerous protozoan DNA extraction protocols and extraction techniques both alone and in combination. However, despite numerous attempts, our samples would not yield an optical map, which prompted us to utilize another sequencing technology, the Illumina NGS platform, to complement (and exceed) our Ion Torrent sequencing data.

We used PCR to confirm the species identity of our isolates as *Tritrichomonas* foetus (see **Figure 4**). To confirm that our initial stock cultures contained an adequate amount of nucleic acid starting material, we employed pulse-field gels after HMW DNA extraction protocols and fragmentation (e.g. bioruption) (see **Figure 5**).

Figure 4. PCR Verification of *T. foetus*. Lanes 1 and 2 contained *T. foetus* and lane 3 contained growth medium as a negative control.

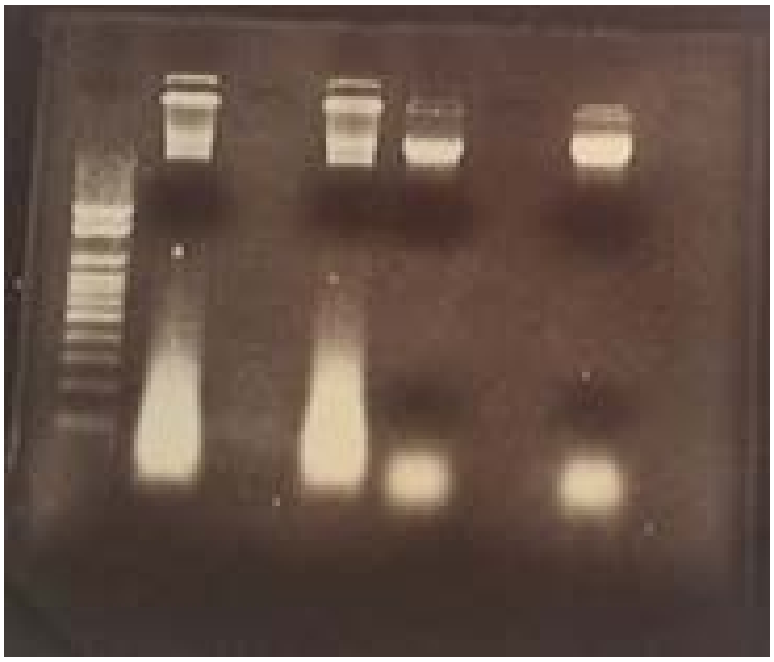

Figure 5. Pulse-Field Gel Confirming HMW DNA is Present in Both D1 and AUTF. Lane 1 (D1) and Lane 3 (AUTF) show HMW DNA from stock cultures. This was compared to the samples that had recently undergone HMW DNA extraction protocols – Lane 4 (D1) and Lane 6 (AUTF). The extraction process yielded HMW DNA in addition to small sized nucleic acids (100-200bp). These smaller bands indicated possible degradation of nucleic acids due to nuclease activity.

After numerous DNA extraction protocol attempts, we report the development of a robust and reproducible protocol for extracting stable high-quality genomic DNA from *T. foetus*. This protocol was based upon a customized protocol for *Giardia lamblia* genomic DNA extraction (63) and involved numerous aspects of previously employed protocols (see **Materials and Methods** for final protocol details). Our initial results for AUTF were a concentration of 250 ng/ul with a 260/280 ratio of 1.84 indicating nearly pure DNA and a 260/230 of 2.03 indicating a very clean sample. We sequenced this sample on both the Ion Torrent and Illumina MiSeq platforms.

Identification and Treatment of Pseudomonas aeruginosa *Contamination*:

After the first round of successful next-generation sequencing, we conducted a preliminary quality check of our sequencing data. This quality control check revealed a bimodal peak for the GC content for several of our sequenced *T. foetus* samples. The expected GC content for *T. foetus* was between 30%-40%, which was one of these peaks (see **Figure 6**). The second peak was determined to belong to *Pseudomonas aeruginosa* based on GC content (66.6%), BLAST-based searches of raw sequences, and bacterial culture on blood agar and pyocyanin plates.

Since we were constantly visually inspecting our samples with light microscopy, we believed that the second organism would most likely be bacterial (as we would have noticed a larger organism). Interestingly, we did not observe the typical signs of bacterial contamination (i.e. cloudiness, altered growth). To identify this second organism, we first consulted the MiST (Microbial Signal Transduction Database) v2.2 (64), which has a database of microbial genomes that can be sorted by GC content. While there are numerous bacteria with a GC content around 67%, one of the most likely candidate microbes was *Pseudomonas aeruginosa* (66.6% GC content). We created a custom Perl script to access our raw sequence reads, and then we conducted BLASTn searches with randomly selected sequences against the NCBI non-redundant (NR) database. The best BLAST hits were to *Pseudomonas aeruginosa* with identity scores of 99%. We also identified sequences that matched previously sequenced *T. foetus* reads, confirming its presence as well. We confirmed our computational findings of *Pseudomonas aeruginosa* contamination by submitting samples from our stock cultures for bacterial culture sterility checks, which yielded the presence of gram negative rods (not shown) which were positive for growth on both blood agar and *P. aeruginosa* specific pyocyanin plates as well (see **Figure 7**).

We obtained two additional feline *T. foetus* samples from the Tolbert Laboratory at UTCVM as well as two bovine strains (one from the Tolbert Lab and one from the ATCC) to see if other *T. foetus* samples might be harboring this co-contaminant. These additional isolates were sourced from Auburn University, NC State University, and UTCVM. Almost all samples we obtained from outside of our laboratory tested positive for *Pseudomonas aeruginosa* (see **Table 2**). All media cultures assessed were negative for all tests including blood agar plates.

To address this contamination, we employed a variety of antibiotics and centrifugation techniques to create an axenic *T. foetus* culture (see **Materials and Methods**). Unfortunately, our attempts were unsuccessful as killing or removing the *P. aeruginosa* uniformly resulted in a loss of viable *T. foetus* as well. Thus, we endeavored to remove the *P. aeruginosa* from our sequences via computational means (see **Materials and Methods** and further **Results** and **Discussion**).

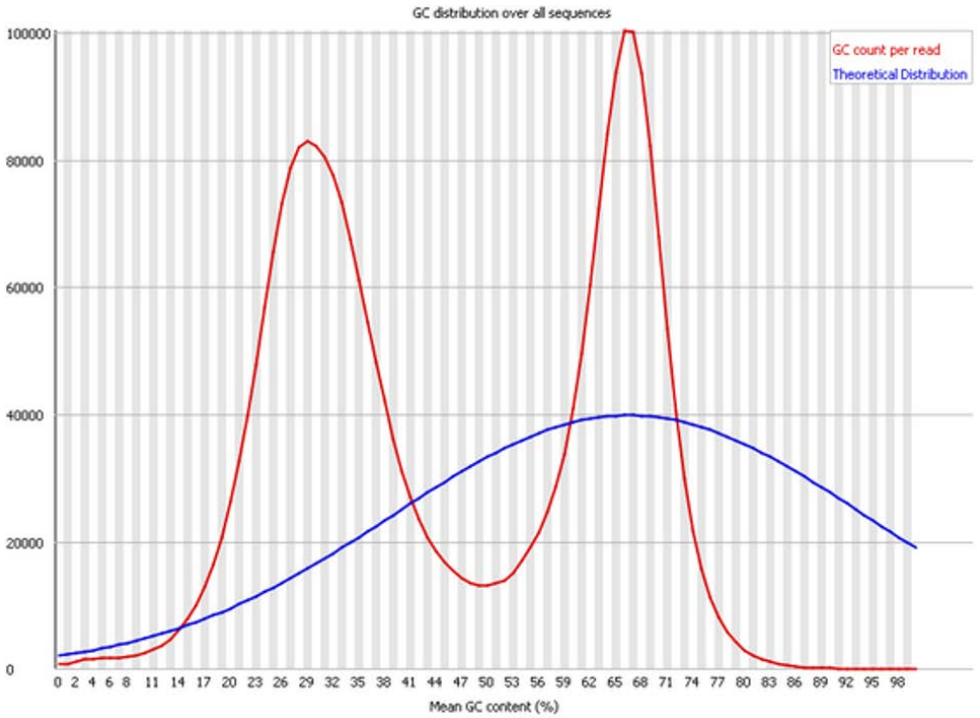

Figure 6. FastQC Output for GC Content of AUTF Isolate Ion Torrent Sequence Data. Two distinct, Gaussian peaks indicate two clonal populations rather than a mixed contamination scenario. The GC content peak on the left centered at approximately 30% corresponds to *T. foetus*, while the peak on the right centered at approximately 67% was confirmed to belong to *Pseudomonas aeruginosa*.

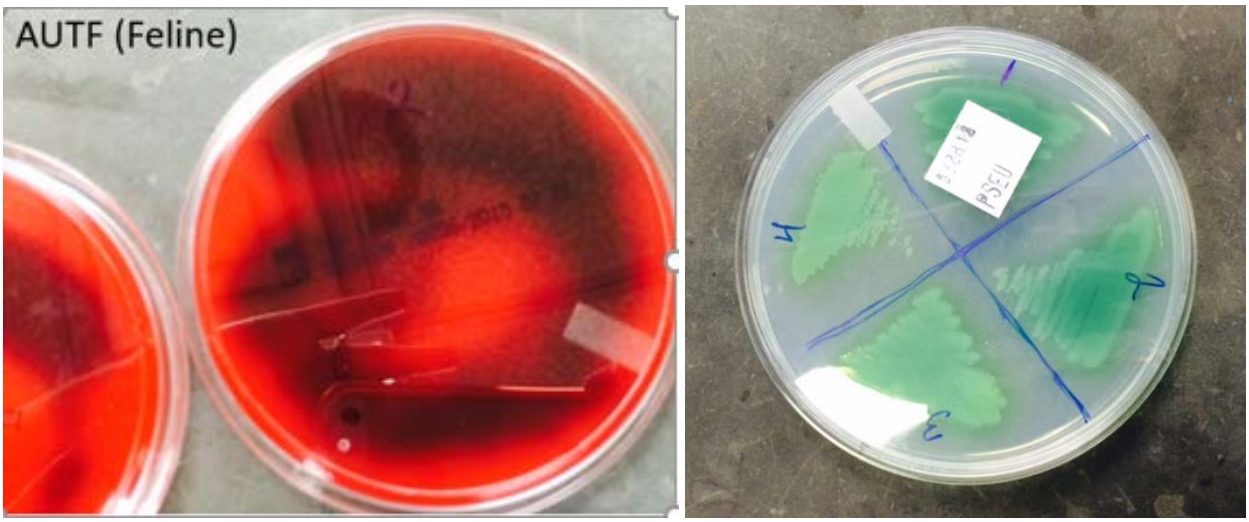

Figure 7. Feline *T. foetus* Strain AUTF Stock Cultures are Contaminated with *Pseudomonas* aeruginosa. Blood agar (left) and pyocyanin plate (right) results for *T. foetus* strain AUTF isolates (June 2015). Cultures were conducted in the UTCVM Bacteriology Lab after detection of second GC content peak at 67%.

Table 2. Sterility Check Results for Bovine and Feline *Tritrichomonas foetus* Isolates. Sterility Check Results for Bovine and Feline *Tritrichomonas foetus* Isolates**.** Several strains were a kind gift from the Tolbert Lab at the University of Tennessee College of Veterinary Medicine, but none of these were available for sequencing or further analyses.

Sample	Species	Source	Stock Culture	Media Culture
D1	-	ATCC	*P. aeruginosa* (6/2015)	Negative (7/13/2015)
AUTF	Feline	Auburn	*P. aeruginosa* (6/2015)	Negative (7/13/2015)
Abby (Tolbert)	Feline	NC State	*P. aeruginosa* (7/15/2015)	Negative (7/13/2015)
Moxie (Tolbert)	Feline	UTK	*P. aeruginosa* (10/2015)	Negative (10/2015)
Bovine (Tolbert)	Bovine	Auburn	*P. aeruginosa* (7/15/2015)	Negative (7/13/2015)
Beltsville1 30003	Bovine	ATCC	Negative (12/9/2015)	Negative (12/9/2015)

Genome Assembly, Annotation and Quality Assessment:

Unfortunately, in the case of *Tritrichomonas foetus*, this genome has never been sequenced and the closest relatives are from different genera, so reference-guided assembly was not an option. Thus, *de novo* assembly, or building the genome *in silico* is the only method that is available for this project. *De novo* assembly is not straightforward and is extremely computationally intensive in terms of both physical resources (e.g. RAM) and time. For *de novo* assembly projects, producing large low resolution "maps" by sequencing barcoded landmarks can greatly expedite this process.(65-67)

These genomes were sequenced and assembled before any other *Tritrichomonas foetus* was publicly available, though toward the end of this work and after assembly was completed, a bovine specific parasite, *T. foetus* strain K, had a genome deposited in the NCBI database by Benchimol and colleagues (41). The only other similar organism sequenced beforehand was *Trichomonas vaginalis*, the causative agent of human trichomoniasis (20). A reference-guided assembly approach using *T. vaginalis* as the reference was attempted, but this produced significantly unsatisfactory results compared to *de novo* assembly. Thus, genomes were *de novo* assembled using CLC Genomics Workbench (Qiagen).

DNA extraction and NGS work produced four Ion Torrent and four Illumina MiSeq sequenced libraries. Unfortunately, the quality and coverage of Ion Torrent libraries was too low to stand alone or significantly improve MiSeq data when combined. Furthermore, Ion Torrent preferentially sequenced *Pseudomonas aeruginosa* over *T. foetus*. Additionally, combining MiSeq runs also did not improve, but rather diminished the quality of *de novo* assembly due to significant over-coverage. Thus, the best Beltsville and AUTF Paired-End MiSeq runs were selected to produce the final draft genomes.

A *Pseudomonas aeruginosa* genome was also assembled using a reference-guided assembly from both Ion Torrent and Illumina data. An additional step for the AUTF library was the computational removal of *Pseudomonas aeruginosa* sequences by initial mapping to the complete genome *Pseudomonas aeruginosa* PAO1. All reads that mapped to *P. aeruginosa* were discarded, while reads that did not map were retained. To perform a quality control analysis of this step, the GC content pre-mapping and post-mapping were analyzed. Satisfactorily, the *P. aeruginosa* GC content peak around ~67% was no longer present post-mapping (see **Figure 8**).

These libraries contained 43,727,120 and 41,231,358 sequenced reads respectively, which amounted to 8,877,470,719 and 9,524,495,094 sequenced nucleotide bases (~9 gbp for each). Using the experimentally estimated *T. foetus* genome size of ~160 mbp (19), this equated to roughly **56X** coverage for each genome.

To prepare the raw sequences for assembly, paired-end reads were merged to produce longer reads when possible (see **Figure 9**), Illumina MiSeq adapter sequences and low-quality ends were trimmed (see **Figure 10**), and sequences below 50 bp were discarded. The *de bruijn* graph assembly algorithm adaptively selected k-mer sizes between 24-40 bp, so sequences greater than 50 bp would allow for a full k-mer plus overlapping whereas sequences lower than this would not have productively added to the assembly.

Even with the presence of *P. aeruginosa* in culture with the AUTF *T. foetus*, the two draft genomes were assembled from very similar raw sequence libraries (both ~9 gbp and ~56X coverage). After identical sequence processing and assembly in CLC Genomics Workbench, we report the genome statistics for

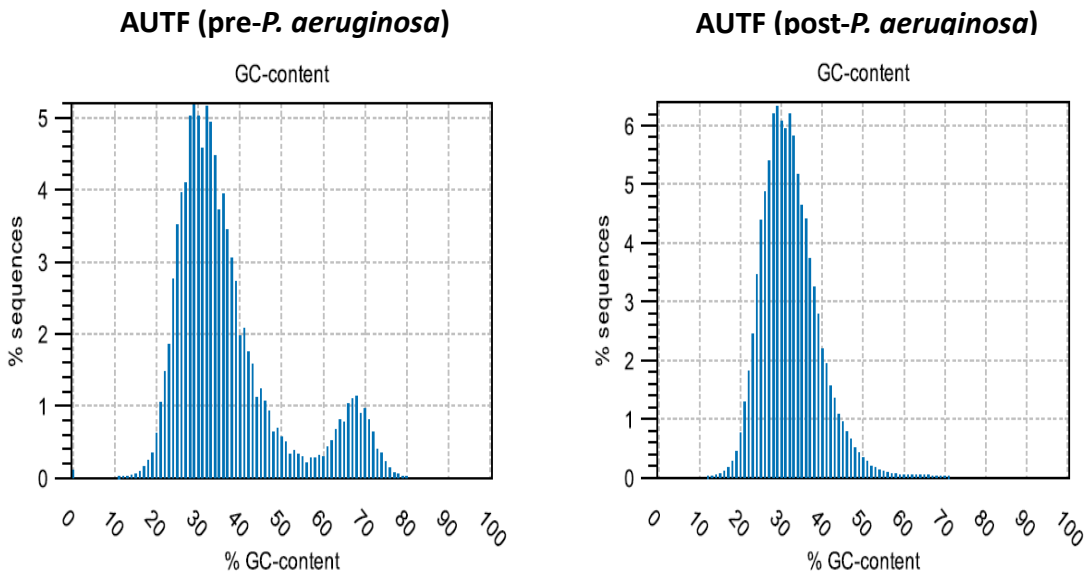

Figure 8. **GC Content Pre- and Post-*Pseudomonas aeruginosa* Mapping to AUTF *T. foetus*.** Left shows bimodal peaks at 33% and 67% GC Content before mapping to and removing *P. aeruginosa* PAO1 genome content from sequencing data. Right shows the elimination of the peak at 67%, providing preliminary evidence that *P. aeruginosa* sequences have been successfully computationally removed.

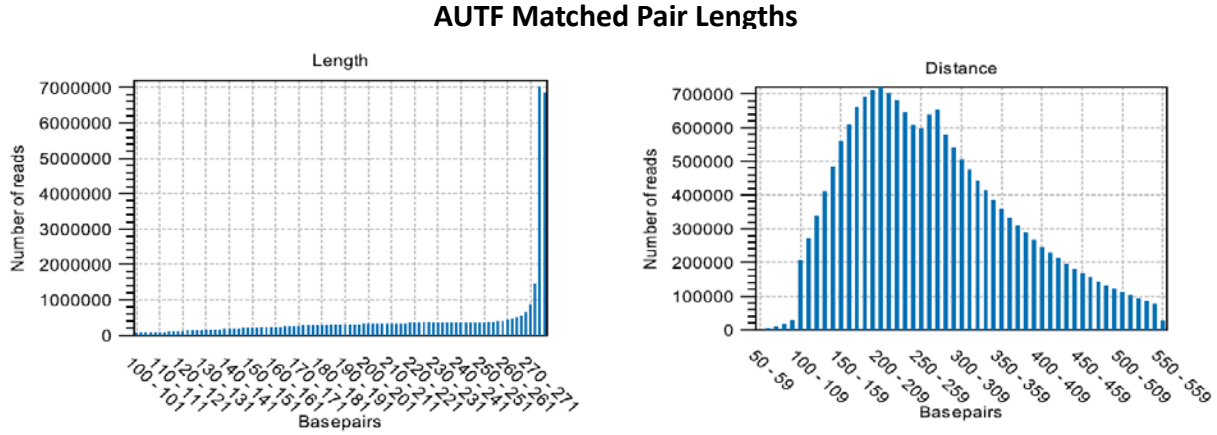

Figure 9. Paired-End Length Distributions from AUTF *T. foetus* MiSeq Library. Pre-merge sequence lengths shown on left and post-merge sequence lengths shown on right.

AUTF *T. foetus* (Feline) **Beltsville *T. foetus* (Bovine)**

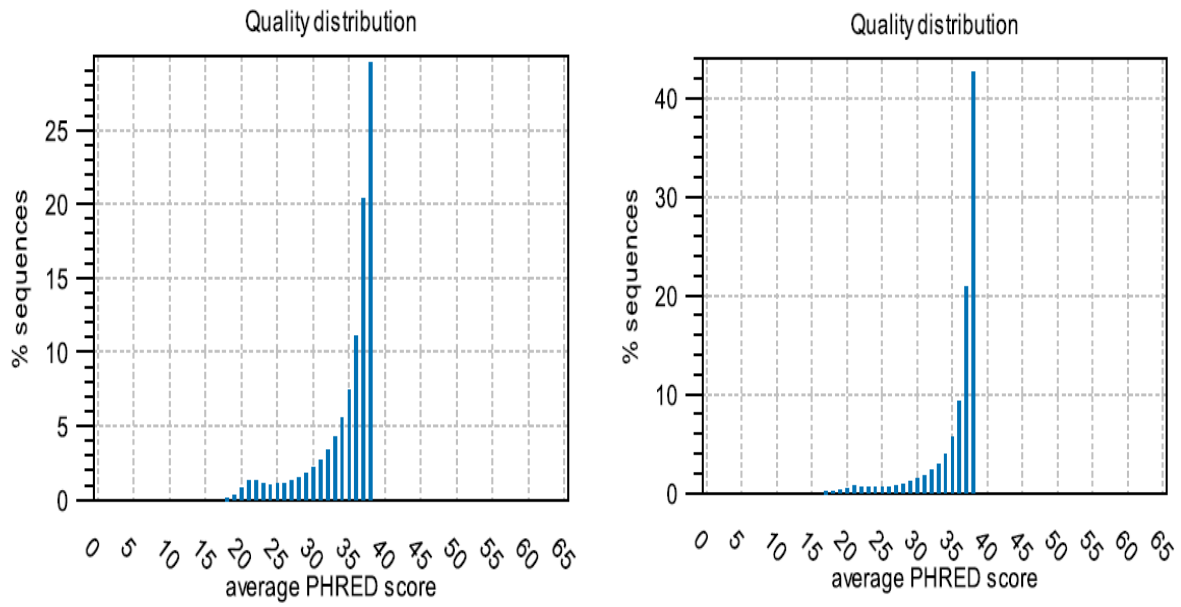

Figure 10. Quality Distributions for *T. foetus* AUTF and Beltsville MiSeq Libraries. >Q20 bases shown.

AUTF and Beltsville genotypes of *Tritrichomonas foetus* in **Table 3**. These genomes were very similar in terms of overall size (~68 mbp), though the AUTF genotype scaffolded better. This is apparent in AUTF with fewer contigs, higher average contig size, better N25, N50, and N75 metrics, and higher maximum scaffold size. For comparison, the recent *T. foetus* strain K genome is also ~68 mbp (41).

Phylogenetics and Taxonomy:

Once the genomes were assembled and consensus assemblies were produced, we were then able to begin data mining at the nucleotide level in parallel with the annotation process. The first goal of the former work was to compare the standard highly conserved phylogenetic markers: ribosomal RNA (rRNA) sequences. We used BLASTn to retrieve the rRNA 5S, 5.8S, 18S and 23S from each of our genomes to compare these to each other and to place our genomes within the context of other Trichomonads and protozoa. These genotypes have been distinguished based upon subtle differences in their ITS (internal transcribed spacer) regions (42), but we sought out further markers to determine the degree of their relatedness and confirm where these organisms fit within the context of other Trichomonads and Protozoa (42). **The 5.8S, 18S, and 23S rRNA sequences were all identical.** We found one 5S rRNA sequence in our AUTF assembly that differed from Beltsville and other AUTF 5S rRNA sequences at 94.74% identity, yet there were multiple copies of the AUTF 5S rRNA sequence and three of the four were also **100% identical**. See **Figure 11** for pairwise comparison of differences in 5S rRNA sequences.

To continue our analysis of conserved nucleotide markers, we identified transfer RNA (tRNA) sequences using tRNA-scan SE 2.0 (68). From this data, we generated tRNA profiles of our genomes to assess whether there were any differences in this critical area of codon determination that might explain subtle differences in genotype. While we did not expect substantial differences as these genomes belong to the same species, it is prudent to assess and tRNA codon potential might inform downstream investigations and molecular techniques in future work. Our results show both genomes are predicted to encode the same 47 tRNA anti-codon types, and furthermore, similar numbers were predicted for each anti-codon type (see **Attachment 1,** "Tritrichomonas foetus tRNAscanSE Results.xlsx"**)**.

Table 3. Genome Statistics for AUTF and BELT Draft Genomes.

Assembly	Beltsville (BELT)	Auburn *T. foetus* (AUTF)
N75	6905	21,452
N50	15525	47,239
N25	28291	88,234
Minimum	37	66
Maximum	156754	335,594
Average	2947	4,634
Count	23163	14,651
Total	68,260,219	67,893,133

		1	2	3	4	5	6	7	8	9	10	11	12	13
AUTF 5S contig 4346	1		100.00	100.00	94.74	100.00	97.44	97.44	90.60	90.60	91.45	93.16	54.70	55.56
AUTF 5S contig 223	2	100.00		100.00	94.74	100.00	97.44	97.44	90.60	90.60	91.45	93.16	54.70	55.56
AUTF 5S contig 222	3	100.00	100.00		94.74	100.00	97.44	97.44	90.60	90.60	91.45	93.16	54.70	55.56
AUTF 5S contig 8093	4	94.74	94.74	94.74		94.74	92.31	92.31	85.47	85.47	86.32	88.03	52.14	52.99
Beltsville 5S contig 7611	5	100.00	100.00	100.00	94.74		97.44	97.44	90.60	90.60	91.45	93.16	54.70	55.56
Tritrichomonas foetus FJ492750.1	6	97.44	97.44	97.44	92.31	97.44		100.00	90.68	90.68	91.53	93.22	56.41	57.26
Tritrichomonas foetus FJ492749.1	7	97.44	97.44	97.44	92.31	97.44	100.00		90.68	90.68	91.53	93.22	56.41	57.26
Trichomonas vaginalis DQ029070.1	8	90.60	90.60	90.60	85.47	90.60	90.68	90.68		100.00	97.44	95.73	53.39	54.24
Trichomonas vaginalis FJ492751.1	9	90.60	90.60	90.60	85.47	90.60	90.68	90.68	100.00		97.44	95.73	53.39	54.24
Trichomonas tenax FJ492748.1	10	91.45	91.45	91.45	86.32	91.45	91.53	91.53	97.44	97.44		96.58	52.54	53.39
Trichomonas tenax FJ492747.1	11	93.16	93.16	93.16	88.03	93.16	93.22	93.22	95.73	95.73	96.58		51.69	52.54
Tetrahymena thermophila emb|X00475.1	12	54.70	54.70	54.70	52.14	54.70	56.41	56.41	53.39	53.39	52.54	51.69		99.14
Tetrahymena thermophila gb|J01893.1	13	55.56	55.56	55.56	52.99	55.56	57.26	57.26	54.24	54.24	53.39	52.54	99.14	

Figure 11. Pairwise Comparison of 5S rRNA sequences from AUTF, Beltsville and Trichomonads. Other Trichomonad sequences were obtained from the NCBI Nucleotide database. Unrelated Protozoan sequences from *Tetrahymena thermophila* were included as internal controls for the pairwise alignment.

As a final nucleotide level analysis, we targeted our comparisons at the level of individual gene sequences for known virulence factors of Trichomonads (18). mRNA and genomic DNA sequences for numerous cysteine proteases, for instance, have been some of the only publicly available data for *T. foetus* for many years. Recently, transcriptomics work has shown that cysteine protease virulence factor CP7 is upregulated in bovine *T. foetus* and CP8 is upregulated in feline *T. foetus* (38). We identified several of the major cysteine proteases and adhesins from *T. foetus* and *T. vaginalis* respectively using BLASTn. We then conducted pairwise alignments with reference nucleotide sequences from a feline-specific genotype, *T. foetus* IPA-Berne, that was previously used for comparative analysis at 10 DNA loci between feline and cattle *T. foetus* (36). **Table 4** contains the results of this analysis, as well as results for two adhesin genes prominently implicated in *Trichomonas vaginalis* virulence (18). Our results for *T. foetus* cysteine proteases show similar results to previous analyses (~1% difference) except for cysteine protease 2 (CP2), which only shares 96.71% nucleotide identity (a significant difference at the within-species level). As for the adhesin genes AP33 and AP65, identities of ~79% and ~75% respectively indicate potential homologs or even functional orthologs based on the degree of evolutionary distance between *T. foetus* and *T. vaginalis*.

We then proceeded to annotate our genome utilizing two independent methods, GeneMark-ES (53) and the parasite specific annotation pipeline Companion (54) that uses AUGUSTUS gene-calling software (55). GeneMark-ES provides *de novo* annotation, whereas Companion utilizes another reference parasite genome selected by the user to add homology-based annotation. We report our genome as annotated protein sequences from our Companion results; we utilized GeneMark-ES *de novo* gene calling annotations to resolve conflicts when necessary. This proved to be a good compromise solution, as our results are less likely to be biased by annotating with respect to *T. vaginalis* (which was unavailable for use with Companion). **Table 5** summarizes the results of annotation of our feline AUTF and bovine Beltsville *T. foetus* genomes, as well as the difference in values. Our results were consistent for each genome, with the only major differences being in number of sequences annotated and number of pseudogenes. Both are likely due to better scaffolding of AUTF relative to Beltsville.

Comparative Genomics (Whole Genome Analyses):

Our first comparison was between our feline AUTF genome and bovine Beltsville genome, and our results are summarized in **Figure 12**.

As expected, most orthologous clusters (16,199) are shared between two genomes from the same species. 15,653 of these were single copy orthologs and are not likely to manifest significant differences except at the level of SNPs that produce gain or loss of function mutations. Only, 265 clusters of orthologous groups were determined to be unique to the AUTF genome, whereas 222 clusters were determined to be unique to the Beltsville genome. We report these differences as supplementary data in **Attachments 2 and 3** ("AUTF OrthoVenn Unique Clusters.xlsx", "Beltsville OrthoVenn Unique Clusters.xlsx"). Clusters that matched annotations of well described Trichomonad virulence factors are reported individually later in this section.

Next, our comparison was expanded to include the results from the bovine parasite *T. foetus* strain K. We expected that both bovine parasite genotypes would cluster heavily together and further delineate feline specific clusters. Unfortunately, our expectations were only partially met, as all three of these genomes shared almost identical numbers of clusters in pairwise comparisons when set against one another (~2300 clusters). See **Figure 13** for results.

Table 4. BLASTn Results for Key Virulence Factors in *Tritrichomonas foetus.* Percentage identity reported as compared to virulence factors sequenced and reported in NCBI nucleotide database.

Virulence Factor Name and Source	AUTF	BELT
CP1 (Cysteine Protease, IPA-Berne feline isolate)	100.00	99.40
CP2 (Cysteine Protease, IPA-Berne feline isolate)	100.00	96.71
CP3 (Cysteine Protease, IPA-Berne feline isolate)	98.93	100.00
CP4 (Cysteine Protease, IPA-Berne feline isolate)	100.00	99.27
CP5 (Cysteine Protease, IPA-Berne feline isolate)	100.00	99.17
CP6 (Cysteine Protease, IPA-Berne feline isolate)	100.00	98.11
CP7 (Cysteine Protease, IPA-Berne feline isolate)	100.00	99.73
CP8 (Cysteine Protease, IPA-Berne feline isolate)	100.00	99.34
CP9 (Cysteine Protease, IPA-Berne feline isolate)	100.00	99.31
AP33 Adhesin (*Trichomonas vaginalis*)	78.76	78.65
AP65 Adhesin (*Trichomonas vaginalis*)	75.35	75.41

Table 5. Comparison of Annotation of AUTF and BELT *T. foetus* Genomes via Companion.

Genomic Features Assessed	AUTF Annotation	BELT Annotation	Difference
Annotated regions/sequences	5701	8797	3096
Number of genes	23317	22911	406
Gene density (genes/Mb)	566.105	512.76	53.345
Number of coding genes	22997	22619	378
Number of pseudogenes	15	42	27
Number of genes with function	10909	10730	179
Number of pseudogenes with function	0	26	26
Number of non-coding genes	320	292	28
Number of genes with multiple CDSs	2947	2805	142
Overall GC%	31.59	31.16	0.43
Coding GC%	33.545	33.65	0.105

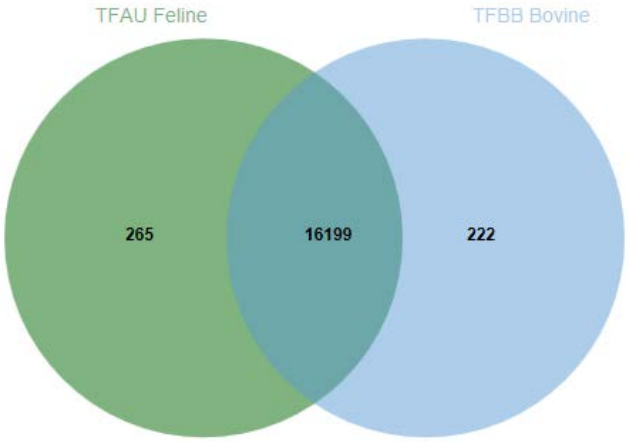

Figure 12. Clusters of Orthologous Groups between two *T. foetus* Genomes. There were 23,000 AUTF proteins and 22,651 Beltsville proteins that were initially assessed. 16,199proteins (at least 70% of each genome) belonged to clusters of orthologous groups that were shared between the two genomes. Of these, 15,653 are single-copy gene clusters which serve as true orthologous genes in most cases (there can be exceptions). 265 clusters were unique for the AUTF genome and 222 clusters were unique to the Beltsville strain. Each genome also contained over 5,000 singleton genes that did not fit any clusters. These results were produced by OrthoVenn (57).

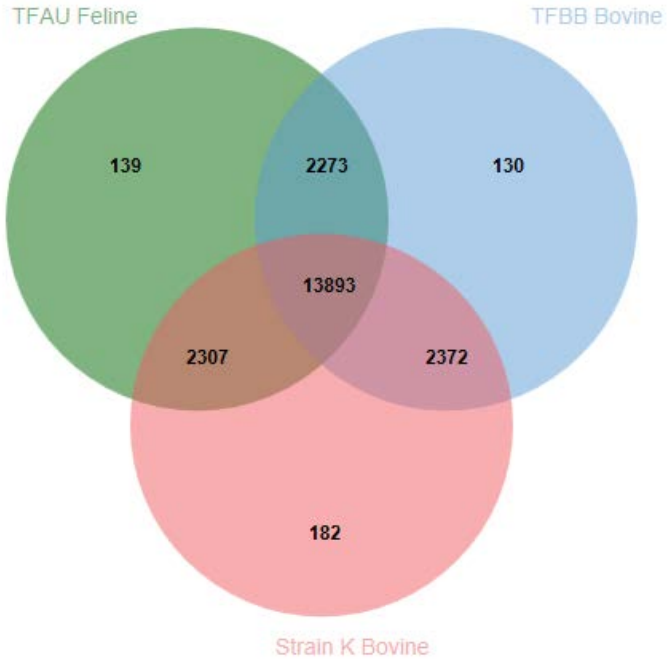

Figure 13. OrthoVenn Results for Three *T. foetus* Genomes. TFAU (AUTF) feline genotype, TFBB (Beltsville) bovine genotype, and strain K bovine genotype were included in this analysis. Of the 13,893 clusters of orthologous groups that were shared, 13,135 were shared as single copy orthologs.

139 clusters were found to be specific to the feline genotype, 130 clusters were found to be specific to our Beltsville bovine genotype, and 182 clusters were found to be specific to the strain K bovine genotype. Additionally, 13,135 clusters were found to contain single copy orthologs (1:1:1). We present this dataset as the most comprehensive list of markers for **identical speciation comparison for feline and bovine *T. foetus* isolates**. This list also serves as the most comprehensive list of candidate genes that can be **ruled out** for large scale changes (e.g. presence vs. absence) that might contribute to host specificity (see **Attachment 4,** "Tritrichomonas foetus OrthoVenn SingleCopyOrthologs.xlsx").

Seeking further specific differences in highly conserved house-keeping genes and phylogenetic markers to distinguish these two genotypes, we extended our analysis to the protein level and examined annotated ribosomal proteins from our two genomes (see **Table 6**).

These results show that there are numerous potential markers to differentiate these genotypes at the protein level in conserved ribosomal proteins. Additionally, 60S ribosomal protein L15 only shares 82.42% amino acid identity between our feline and bovine genotypes, standing as a substantially diverged marker between these two genotypes. However, 25 of 32 ribosomal proteins are greater than 98% amino acid identity over the full length of the ribosomal protein and 17 of these phylogenetic markers are 100% identical at the amino acid level.

We next analyzed our whole genome protein sequences using the KEGG Database to assess for molecular pathway presence and gene ontology. The KEGG Database compiles enzyme pathways and uses BLAST analyses to identify the components of whole gene systems (69). Our expectation was that this type of analysis might highlight different pathways that would allow us to compare metabolic capabilities and systems between the two organisms at a very broad level. Furthermore, the differential encoding of virulence factors or pathways involved in virulence might also help to explain host specificity in *T. foetus*. Still, since these genomes are the same species, we further did not anticipate tremendous differences between them. We report the functional gene ontology distribution for our two genomes in **Figure 14**.

These results show comparable annotation via the KEGG Database: ~5,000 entries or ~22% of the predicted genes were detected based on this comparative analysis. Unfortunately, but not unexpectedly, the distributions shown in **Figure 14** indicate highly similar gene ontology between the two genomes. It is interesting to note that the section for "Glycan biosynthesis and metabolism" comprises a slightly larger percentage in Beltsville *T. foetus* than in AUTF *T. foetus*. Glycans, both the synthesis of proteoglycans and the enzymatic cleavage of extracellular matrix and other glycosylated proteins, stand as a leading physiological arena of *T. foetus* virulence.

We then assessed the taxonomic relationships of our genomes through ghostKOALA which used BLAST to report the best BLAST hit taxa for each sequence in **Figure 15**. Clearly, the majority of sequences in each genome were most related to Parabasalids like *Trichomonas vaginalis* (green). Critically, the KEGG database does not include the recent *T. foetus* genome. If it did this section would be much higher and the undefined region (white) would be much smaller. Additionally, and satisfactorily, *Gammaproteobacteria* (e.g. *Pseudomonas aeruginosa*) was not noted to be a significant source of taxonomic hits for the AUTF genotype, further ensuring that the *P. aeruginosa* genomic content was effectively removed pre-assembly. Finally, while small in total terms, the significant enrichment for *Bacteroidetes* is part of an interesting trend in *T. foetus* and *T. vaginalis* that will be expanded upon in the next section.

Table 6. Ribosomal Protein Pairwise Comparison Based on OrthoVenn Cluster Results.

Ortholog Clusters	Annotation	Pairwise (%ID)	Length (AAs)
846	40S ribosomal protein S24	100	140
985	40S ribosomal protein S30	100	79
1365	40S ribosomal protein S2	100	275
1371	40S ribosomal protein S18	99.35	154
1447	60S ribosomal protein L14	100	130
2081	60S ribosomal protein L13	95.15	217
2136	60S ribosomal protein L15	82.42	252
2143	60S ribosomal protein L5	99.29	281
2262	40S ribosomal protein S26	100	117
2301	60S ribosomal protein L11	100	175
2601	60S ribosomal protein L17	100	162
2674	60S ribosomal protein S24	100	140
3101	60S ribosomal protein L7	99.58	239
3126	Ubiquitin-60S ribosomal protein L40	100	153
3250	60S ribosomal protein L21	100	159
3426	60S acidic ribosomal protein P0	97.46	315
3507	60S ribosomal protein L14	99.23	130
3669	60S ribosomal protein L11	99.52	209
3771	40S ribosomal protein S6	99.18	243
3995	40S ribosomal protein S21	100	97
4742	60S ribosomal protein L38	94.59	74
4761	60S acidic ribosomal protein P2-A	100	106
4859	60S ribosomal protein L35	94.31	123
5050	60S ribosomal protein L38	96.3	81
5121	60S ribosomal protein L36	100	123
5997	60S ribosomal protein L6	100	154
6190	40S ribosomal protein S20	99.17	120
6322	40S ribosomal protein S7	100	186
6442	Ubiquitin-40S ribosomal protein S27a	97.35	151
6450	60S acidic ribosomal protein P0	99.68	315
6694	Putative 40S ribosomal protein S4-like	100	182
6866	40S ribosomal protein S2	100	306

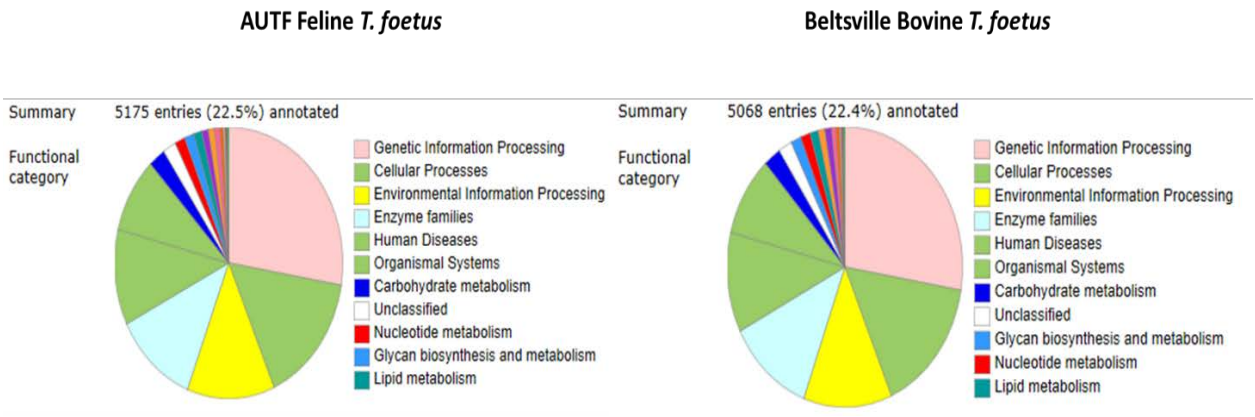

Figure 14. Functional Genomics and Gene Ontology from KEGG Database. These results were obtained from the KEGG Database using the ghostKOALA bioinformatics webserver.

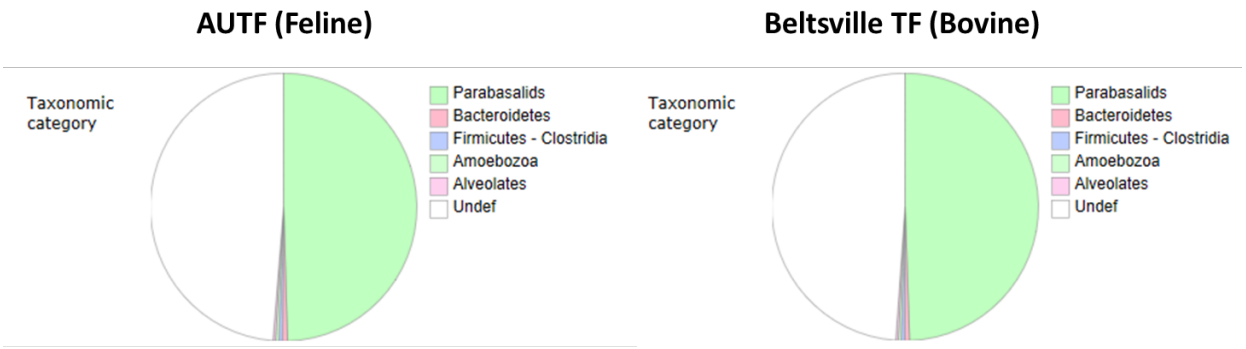

Figure 15. Taxonomic Information for Best BLAST hits for Two *T. foetus* Genomes. The top 5 sources of best BLASTp hits from taxa were chosen for this figure. After Parabasalids, Bacteroidetes (a phylum rich in pathogenic bacteria) was the next highest source of hits with close to 100 sequences represented.

Comparative Genomics (Targeted Analysis of Parasite Virulence Factors):

As our final and most significant comparative effort, we mined our genomes for sequences annotated as or similar to specific virulence factors of *T. foetus* and *T. vaginalis*. For the time and manpower available, these results are extensive if not comprehensive. The virulence factors assessed were: **cysteine proteases**, **leishmanolysins**, **glycan metabolism (alpha-mannosidases)**, **sialidases**, **phospholipases**, **tetraspanins**, and **viral elements** (17, 18). These results were obtained through three analyses: 1) differential cluster presence via OrthoVenn, 2) differential detection via KEGG, and 3) annotation-based mining directly from protein sequence files.

1. Cysteine Proteases:

First, we examined the numerous cysteine proteases that are hallmark virulence factors of Trichomonads. Cysteine proteases are critical for *T. foetus* virulence as they mediate adhesion and digest key factors of the immune response including immunoglobulins (12-14, 17). To mine these protein sequences, we utilize numerous annotation keywords (cysteine protease, cysteine proteinase, cysteine peptidase, cathepsin, papain, falcipain, actinidain, and crustapain) as there are numerous variations for annotating this large protein family.

We extracted over two hundred sequences from all three genomes. From these, we identified 30 sets of 1:1:1 orthologous sequences that have greater than 90% amino acid sequence identity over the length of full length (or partial) pairwise alignments (see **Attachment 5**, "Tritrichomonas Foetus Cysteine Protease 30clearOrthologs PhylogeneticTree.pdf"). Of the 30 sets, the majority had >98% amino acid identity conservation between bovine and feline sequences, indicating ostensibly identical function as well. Significant differences at the N-terminus, which is a well-known deficiency of gene calling algorithms, were attributed to inaccurate start site calling when the remainder of the alignment was >98% amino acid identity as well. The only group that had a significant difference in this set of orthologs was the group comprised of TFAU_001245800.1, TFBB_110250400.1, and OHT17152.1, which were variably annotated as Papain cysteine proteinases and Cathepsin-B proteases. These were 100% identical over 246aa for both bovine sequences but shared only 91% amino acid identity with the feline cognate sequence. The major difference in the feline sequence was the addition of a 6 amino acid sequence motif that was repeated 8 times at the C-terminus in place of the last 25aa of the bovine sequences (see **Figure 16**). It is possible that this change could affect the expression or function of this particular feline cysteine protease, and this serves as a potential target for further investigation.

Figure 16. Multiple Sequence Alignment of *T. foetus* Cysteine Proteases with C-Terminal Change. Feline *T. foetus* strain AUTF displays the c-terminal change of unknown significance.

As for the remainder of cysteine protease sequences, the lack of clear 1:1:1 orthology in our dataset will require more extensive manual curation of the multiple sequence alignment and individual local BLAST searches beyond the scope of this project.

2. Leishmanolysins:

Next, we mined all protein sequences annotated as leishmanolysin surface-expressed proteases from all three genomes. Leishmanolysins are key virulence factors originally studied in *Leishmania spp.* but that are present in numerous protozoan parasites. The prototypical leishmanolysin is GP63, and an ortholog of this leishmanolysin has been annotated in *Trichomonas vaginalis* and served as an outgroup for our comparative analysis. We aligned leishmanolysin sequences using MAFFT (58) multiple sequence alignment software, and then (59) in MEGA 7.0.26 (60). Through this comparative analysis, we have

detected two putative leishmanolysin proteins that are potentially present in only bovine *T. foetus* genomes, not AUTF (see **Figure 17).**

3. Sialidases:

Sialic acid lectin binding and cleavage has been suggested to be potential mediators for virulence in *T. foetus* as well (17, 70-72). We identified 48 clear sialidase, neuraminidase, or sialidase-like protein sequences between all three *T. foetus* genomes. 15 of these have clear 1:1:1 orthologs amongst all three *T. foetus* genomes. However, from *T. foetus* strain K, **OHS95222.1** (annotated as a secreted sialidase), **OHT00383.1** (annotated as a secreted sialidase), and **OHS99826.1** (annotated as a neuramidase) branch alone and also do not contain detectable orthologs in either of our feline or bovine genomes. All three of these sequences are most closely related to the sialidase group corresponding to TFAU_000054500.1, TFBB_000908900.1, and OHS9762.1 and may represent a lineage specific expansion or recent duplications within strain K alone.

4. Glycan Metabolism and Bacterial Lateral Gene Transfer:

Closely related to sialidases are many other glycan metabolism enzymes. Many of these have been noted in *T. vaginalis* and found to be most likely products of lateral gene transfer from *Bacteroidetes* bacteria (18). This is part of the connection between *Bacteroidetes* and Trichomonads. Both AUTF and Beltsville contain an alpha-mannosidase D orthologous group according to our OrthoVenn results (98.29% pairwise identity). However, in differently shared orthologous groups, only the Beltsville genotype contains two putative mannosyl-oligosaccharide 1,2-alpha-mannosidases that we are unable to detect in the AUTF genome: **TFBB_000515100.1** and **TFBB_000137600.1**.

5. Phospholipases:

Another virulence factor that has been proposed for *T. foetus* is the expression of secreted phospholipases (17). OrthoVenn detected two orthologous clusters of phospholipases (3 Phospholipase D C, 3 Phospholipase B-like protein G) in AUTF *T. foetus* that are not found in the Beltsville genotype. These phospholipases are: **TFAU_000921700.1** (Phospholipase D C), **TFAU001549200.1** (Phospholipase B-like protein G), **TFAU_001385400.1** (Phospholipase B-like protein G), and **TFAU_002114800.1** (Phospholipase B-like protein G).

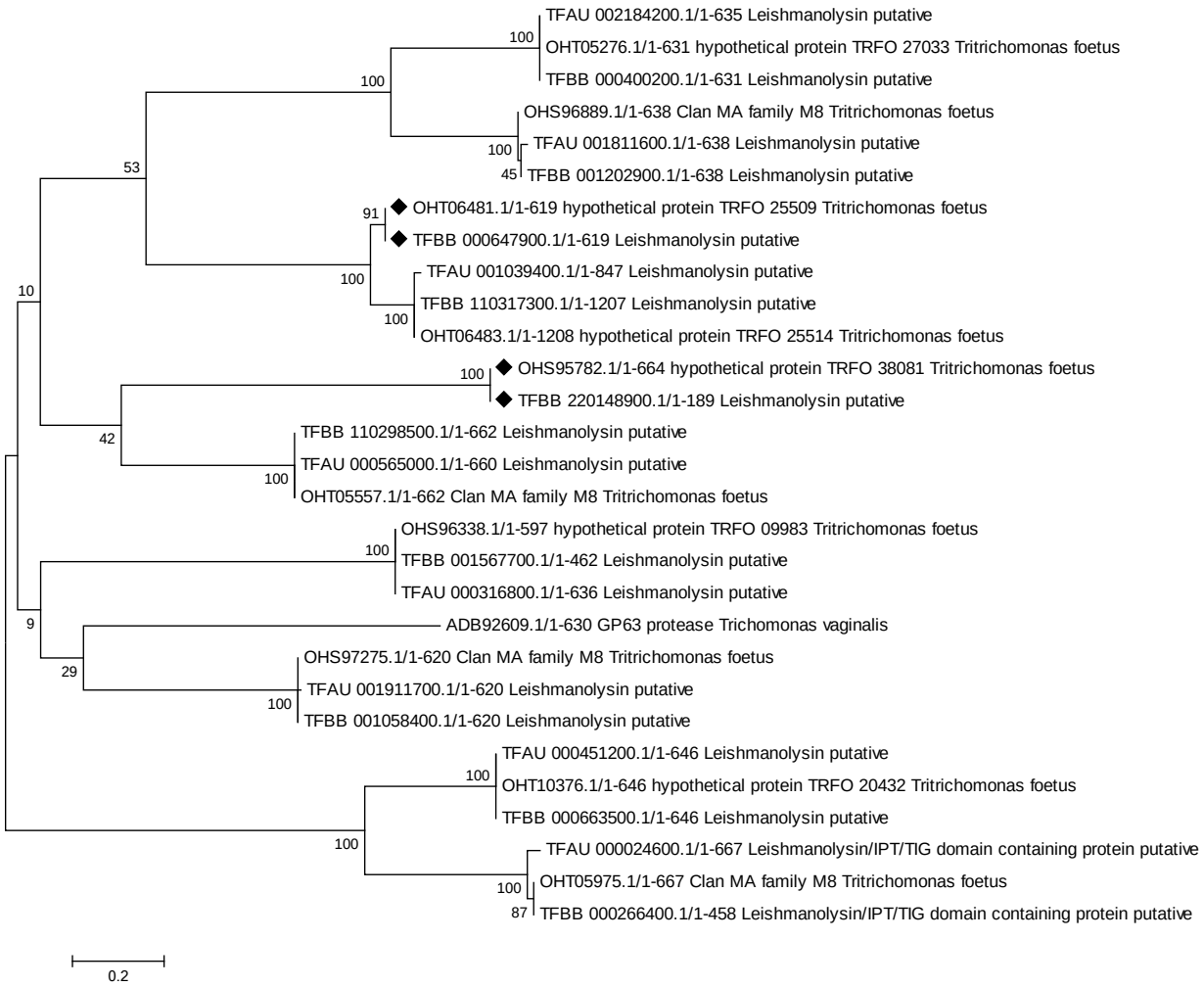

Figure 17. Max. Likelihood Phylogenetic Tree of Leishmanolysin Homologs from *T. foetus* genomes. Clades marked with black diamonds highlight putative leishmanolysin proteins that we have detected in bovine *T. foetus* genomes only. TFBB 220148900.1 (189aa) is a partial protein sequence but is the clear ortholog of OHS95782.1 (664aa) and we assume that the full protein coding sequence is present and a missed start call is responsible for the n-terminal truncation. Phylogeny was tested using bootstrap method (n=500).

6. Tetraspanins:

Finally, while it has not yet been suggested in *T. foetus*, a virulence factor present in *T. vaginalis* that we considered are tetraspanin proteins (18). According to Hirt, tetraspanins are "membrane proteins involved in signaling modulating adhesion, motility and tissue invasion." We detected 31 putative tetraspanin sequences within *T. foetus* genomes and assessed whether any were absent or different. There were 8 sets of 1:1:1 or 1 bovine:1 feline sets of orthologs between these three genomes. Additionally, we have uncovered 5 putative tetraspanin protein sequences (2 in Beltsville, 2 in AUTF, and one in strain K) that are unique and not detectable in the other genomes. These potential differentiating virulence factors are: **TFBB_001381300.1** (Tetraspanin family, putative), **TFBB_000362200.1** (Tetraspanin family, putative), **TFAU_000391800.1** (Tetraspanin family, putative), **TFAU_001697800.1** (Tetraspanin family, putative), and **OHT03547.1** (Tetraspanin family protein) (see **Figure 18).**

7. Viral Presence:

Hirt also reviewed findings on the presence of Trichomonas associated viruses (TVV), which can alter metronidazole sensitivity and the immunogenicity of *T. vaginalis* (18). Interestingly, we detected viral elements (10 mariner mos1 transposases, 3 retrovirus-related Pol polyprotein from type-1 retrotransposable element R2s) in the Beltsville genotype that we did not detect in the AUTF genotype.

Discussion:

While we can successfully report the sequencing of two *T. foetus* genomes, this organism proved to be a tremendous challenge for traditional molecular and sequencing methods that are geared toward model organisms. *T. foetus* proved refractory to an extensive list of genomic DNA extraction protocols that were shown in the literature to be effective for other Trichomonads or protozoa. We believe that the many nucleases that feature heavily in general Trichomonad virulence and pathogenicity might account for the challenges that we observed (73, 74). Each step in our final genomic DNA extraction protocol were necessary to produce samples of high enough quality that were stable enough to survive several days of Ion Torrent sequencing. This was not a hurdle that those conducting PCR or Illumina sequencing would have faced, so we believe that this may be an under-appreciated facet of *T. foetus* biology that we have encountered during this dissertation work.

As for our findings, our genomic and phylogenetic results clearly support that bovine and feline *T. foetus* belong to the same species at nearly every level of conserved molecular markers assessed. 5S, 5.8S, 18S, and 23S rRNA regions were all identical except for one of four copies of the AUTF 5S rRNA gene (**Figure 11**). While we observed the same small level of strain level variation in the 5.8S rRNA ITS regions that others have reported, our genomes encode the same tRNA repertoire of 47 codons (**Attachment 1**), exhibit minimal amino acid differences across 32 ribosomal protein sequences (**Table 6**), and display nearly identical nucleotide sequences for cysteine proteases (**Table 4**) consistent with what has been reported in the literature. At the whole genome level, the assemblies of both bovine and feline *T. foetus* genomes had remarkably similar sizes and properties (**Table 3**), remarkably similar annotation results (**Table 5**), few unique clusters of orthologous groups (**Figures 12-13 and Attachments 2-4**), strikingly similar functional gene ontology distributions across ~5000 assessed genes (**Figure 14**), and near

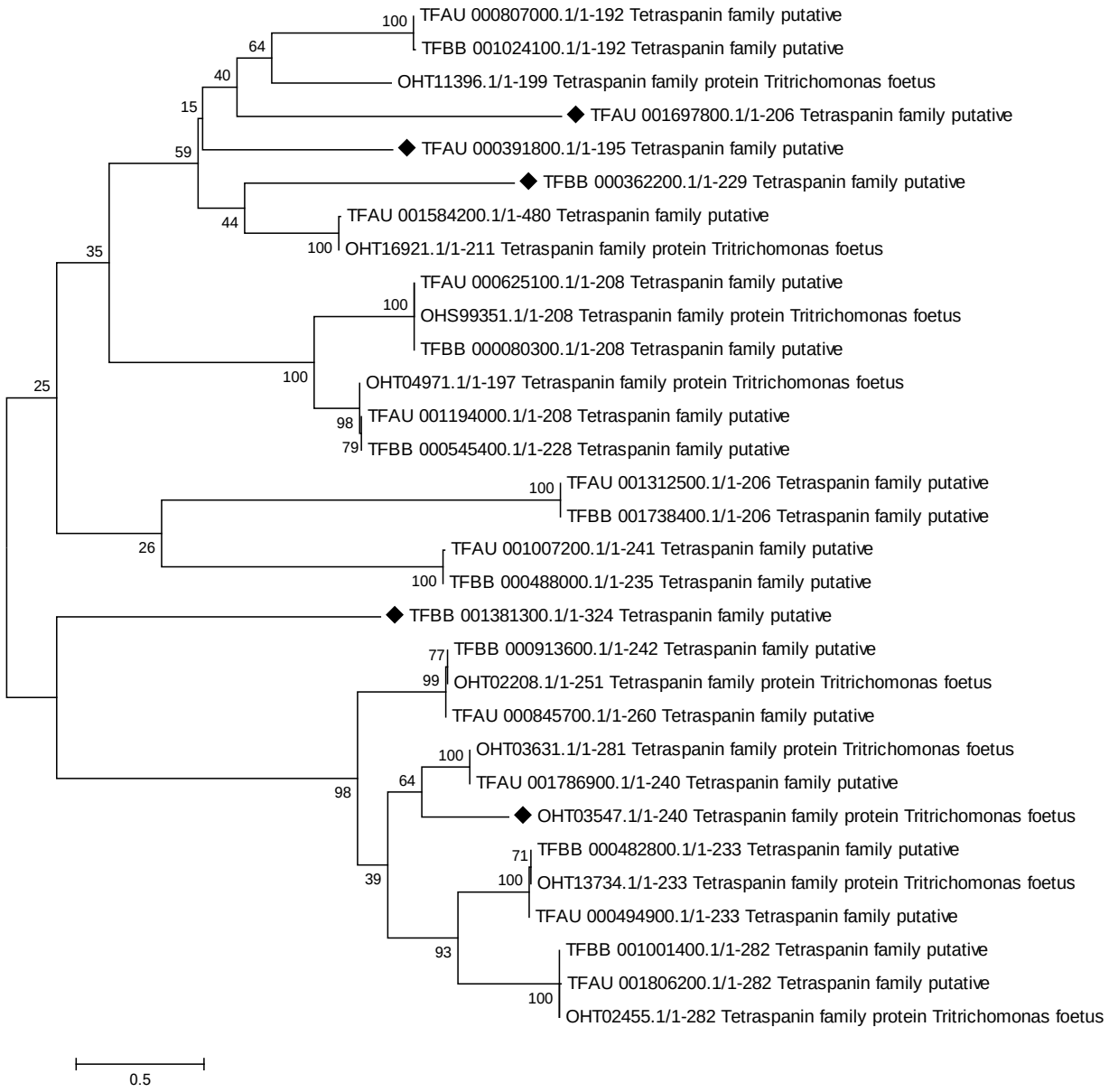

Figure 18. *T. foetus* Putative Tetraspanin Family Proteins. Phylogenetic tree that places potential investigational targets within the context of other *T. foetus* tetraspanin clades. Sequences were aligned using MAFFT (58) multiple sequence alignment software. Maximum likelihood phylogenetic tree was constructed using the JTT matrix-based model (59) with bootstrap test of phylogeny (n=500) in MEGA 7.0.26 (60)

identical taxonomic relationships across the whole genome (**Figure 15**). At the level of protein sequence analysis, the two genomes exhibited few notable differences across six major families of virulence factors while clear 1:1 orthologs were easy to identify and, most importantly, often identical in sequence (**Figures 16-18 and Attachment 5**).

Still, there are genomic differences, and the different host specificities and experimental findings for these genotypes cannot be ignored. However, we conclude that these differences are consistent with genotypic variation amongst strains of the same species. These minute differences are exceedingly important and may help unlock the keys to improved therapeutics for *T. foetus* and other Trichomonads and parasites in the future. Thus, we propose four separate but by no means mutually exclusive hypotheses for why bovine and feline genotypes of *T. foetus* differ in their host specificity and pathophysiology based on our knowledge of the literature, expertise in the field, and results of this dissertation.

Future Work: Hypotheses for Tritrichomonas foetus Host Specificity

Virulence factors for *T. foetus*, like in *T. vaginalis*, are not singular genes, but appear to belong to large homologous groups of proteins that have undergone significant functional expansion in these organisms. With over 200 cysteine proteases, 28 leishmanolysins, 48 sialidases, 31 tetraspanins, several phospholipases, alpha-mannosidases, etc., it may likely be that subtle differences in the repertoire may contribute to host specificity. However, with the close amino acid identity of many of these sequences, care must be taken to separate true orthologs from within “orthologous groups”. From this dissertation work, in addition to producing several novel genomes that will benefit veterinary parasite and pathogen research for years to come, we have identified many novel candidate genes for investigation into differing host specificity for *Tritrichomonas foetus*. It is our earnest hope that, through our future efforts or through the works of others that revisit this work and utilize these genomes, new therapeutic interventions can be devised to benefit animals that might suffer from this terrible parasite.

1. Virulence factor paralogs are effectors that subtly modulate host specificity.

Tritrichomonas foetus life-cycle, transmission, infection, and immune-system evasion are all products of an extremely complex chain of simultaneous events and systems working in concert. While it is tempting (and still possibly the case) to envision a small handful of genes that switch from bovine to feline host specificity, we hypothesize that the repertoire of genes that constitute such a switch is slightly larger than a handful and consists of virulence factors that are extremely closely related to key effectors of shared mechanisms of virulence.

We have detected numerous instances of virulence factor paralogs that are, as far as we can determine at this point, specific for a given pathotype (bovine vs. feline). Paralogs are homologs that arise from duplication events and acquire new functions that remain closely related to that of their original sequence. It is our view that a likely means for producing host-specific virulence factors is to take an existing virulence factor and subtly alter the function so that it now acts in a more efficient and advantageous manner within the current host. For example, a cysteine protease that is highly effective at cleaving general IgG molecules is duplicated, and the second new copy finely tunes itself to cleave bovine specific sequence patterns when under the selective pressure of constant bovine IgG bombardment. When faced with the analogous scenario in a feline host, a similar duplication and neofunctionalization occurs, leaving us to sort through four cysteine proteases that all look exceedingly

similar in terms of sequence identity and biochemical properties. We propose from the analysis of our genomes that virulence-specialization and host-specificity may have occurred in several protein families and the summative effect of all of these changes is a predilection for bovine urogenital epithelium versus feline gastrointestinal epithelium.

Unfortunately, draft genomes are not definitive statements on the presence or absence of a protein encoding gene. Rather, they are jumping off points for hypothesis generation and further investigation. As such, we would look forward to the opportunity to use PCR, RT-PCR, and Western Blot or SDS-PAGE methods to confirm the presence or absence of genomic DNA, transcribed mRNA, and expressed protein, respectively, for any of the targets that we have identified in our analyses. That said, in every group of protein families that we have extracted and analyzed, our ability to detect the presence of high quality 1:1:1 orthologs between our Beltsville bovine *T. foetus* strain, our AUTF feline *T. foetus* strain, and the bovine strain K give us a measure of confidence in our predictions that there are indeed some unique features.

As the field of *Tritrichomonas foetus* virulence research moves forward, the question of closely related homologs (orthologs vs. paralogs) may continue to thwart rapid movement towards clear, definitive candidates for experimentation. One primary impact that we hope for this dissertation work is the caution that these families of proteins must be taken within their evolutionary context and very carefully sequenced and verified. Without highly specific probes for immunohistochemistry, protein extraction, and PCR-identification, many of these sequences may cross-react and obscure their true effects. Additionally, others still may have alternative splice variant isoforms that further confound direct analysis.

2. Subtle SNPs may play a role in adjusting host specificity.

Unfortunately, we did not have a reference genome that could be used to generate variant calls to determine the rate and location of SNPs in our genomes when we were sequencing and assembling them. Additionally, SNPs, unless clearly causing a loss of function in the case of a premature stop codon, for instance, are difficult to assign phenotypes without experimental work. We postulate that subtle SNPs might alter host specificity akin to the minute changes that alter the serotype and ultimate infectivity of viruses.

Therefore, future work can focus on assessing for SNPs between these two genomes (*T. foetus* Beltsville, *T. foetus* AUTF) and *T. foetus* strain K. When paired with molecular tools (which are slim to none for *T. foetus)*, this may produce extremely productive results and therapeutic targets. However, this work had to conclude before thorough SNP analysis could be conducted. Thorough SNP analysis would take a lifetime's work to track down every lead, which makes these new draft genomes produced in this work of such high value to the *T. foetus* experimental community.

3. Host specificity may involve genes/proteins of unknown function.

A third, and highly unsatisfying but also likely hypothesis for host-specificity is that the determinants have been detected in our comparative analysis and simply correspond to no known protein domains, annotated functions, or putative gene ontology groups. Further exposition on this point does not seem necessary except to state that work outside of *T. foetus,* even in the most distantly related veterinary

parasite or pathogen from *T. foetus,* might unlock the key to *T. foetus* host-specificity through future comparative techniques.

4. Parasite-microbe (*Pseudomonas aeruginosa)* interactions may feature prominently in host specificity.

Finally, we have reported in this dissertation a close and seemingly mutualistic association between *T. foetus* and *Pseudomonas aeruginosa* in many of our isolates *in vitro*. *P. aeruginosa* appeared to be beneficial if not essential for the growth of the feline AUTF *T. foetus* strain. We attempted numerous methods of removing *P. aeruginosa* from our cultures to produce axenic *Tritrichomonas foetus* cultures for the D1 and AUTF strains. Serial passaging, treatment with antibiotics (PenStrep, Gentamicin) and/or differential centrifugation protocols have not yielded long-term negative sterility check results.

Tritrichomonas foetus is resistant to metronidazole treatment, which is unlike human pathogenic Trichomonads such as *Trichomonas vaginalis.(75-78)* Interestingly, *Pseudomonas aeruginosa* is well-known for its inherent multi-drug resistance properties, which include resistance to metronidazole. Furthermore, *Pseudomonas aeruginosa* is also reported to "confer" resistance to other microbial community members (e.g. methicillin-resistant *Staphylococcus aureus* within biofilms). If a strain of *Pseudomonas aeruginosa* has become preferentially selected as an endosymbiont of *Tritrichomonas foetus*, then this would be a possible explanation for *T. foetus'* metronidazole resistance. From a therapeutic standpoint, a strategy for eliminating *T. foetus* under this scenario would involve first targeting the endosymbiotic *Pseudomonas* with antibiotics specific for drug resistant gram-negative bacteria (e.g. aminoglycosides like gentamicin, tobramycin, etc.), followed by metronidazole to clear the weakened/vulnerable *T. foetus*.

Our results indicate that it is possible that *T. foetus* is involved in a symbiotic relationship with *P. aeruginosa* in our *in vitro* cultures. It is possible that the larger protozoa has phagocytized the bacteria, where it is able to harness them for increased nutrition (e.g. faster growth). *In vivo,* it is also possible that this relationship would confer drug resistance that would be of extreme benefit against first-line anti-Trichomonad treatments such as Flagyl (metronidazole). *Tritrichomonas foetus* is a Parabasalid, as it belongs to the Class *Parabasalia*. Parabasalids are named for the Parabasal body, which is known to house endosymbiotic bacteria. This body is a modified Golgi complex organelle. We expect that this may be a potential location for localization of *P. aeruginosa* within *T. foetus.* The ramification of this relationship for treatment of affected animals is that treating with gentamicin to damage the *P. aeruginosa* endosymbiont may weaken the *T. foetus* and increase the penetrance and potentiate previously ineffective treatments like Flagyl. Trichomonad-bacteria interactions are not unheard of - the ability of *T. vaginalis* to consume consuming vaginal microbiota to alter the pH of the urogenital tract and augment colonization has heavily studied since 1942 (79-82). Furthermore, bacterial endosymbionts have been visualized using a variety of techniques within another Parabasalid (*Trichonympha agilis*), which is itself an endosymbiont of the termite gut (see **Figure 19**). We propose that future work might start with an epidemiological survey to assess the co-localization and true mutualistic existence of *T. foetus* with *P. aeruginosa* (or other microbiota).

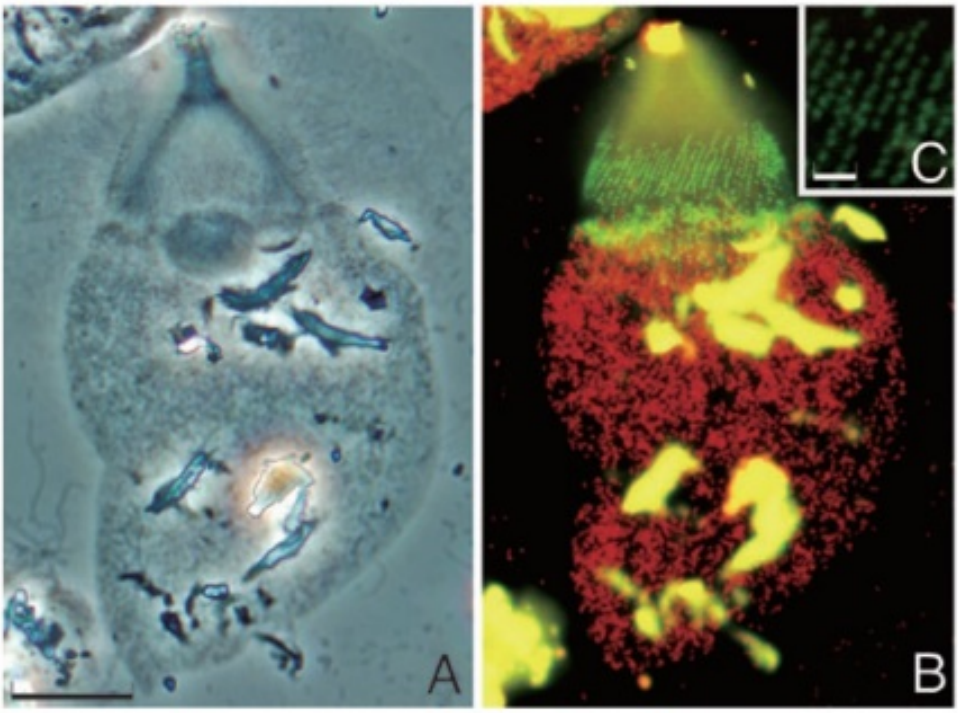

Figure 19. Example of FISH Results for a Bacterial Symbiont of a Protozoa. *In situ* detection of symbionts of the flagellate *Trichonympha agilis.* A. Phase-contrast image. B. FISH of bacteria (red and green) in *T. agilis*. C. Endosymbiont magnified and shown to reside in ordered structures within its protozoan host. Figure from Sato *et al.* 2009 (83).

References

1. Frey CF & Muller N (2012) Tritrichomonas--systematics of an enigmatic genus. *Molecular and cellular probes* 26(3):132-136.
2. Kather EJ, Marks SL, & Kass PH (2007) Determination of the in vitro susceptibility of feline tritrichomonas foetus to 5 antimicrobial agents. *J Vet Intern Med* 21(5):966-970.
3. Gookin JL, Stauffer SH, Dybas D, & Cannon DH (2010) Documentation of in vivo and in vitro aerobic resistance of feline Tritrichomonas foetus isolates to ronidazole. *J Vet Intern Med* 24(4):1003-1007.
4. Gookin JL, Hanrahan K, & Levy MG (2017) The conundrum of feline Trichomonosis. *J Feline Med Surg* 19(3):261-274.
5. Gookin JL*, et al.* (2001) Experimental infection of cats with Tritrichomonas foetus. *Am J Vet Res* 62(11):1690-1697.
6. Gookin JL*, et al.* (2004) Prevalence of and risk factors for feline Tritrichomonas foetus and giardia infection. *J Clin Microbiol* 42(6):2707-2710.
7. Foster DM, Gookin JL, Poore MF, Stebbins ME, & Levy MG (2004) Outcome of cats with diarrhea and Tritrichomonas foetus infection. *J Am Vet Med Assoc* 225(6):888-892.
8. Gunn-Moore DA, McCann TM, Reed N, Simpson KE, & Tennant B (2007) Prevalence of Tritrichomonas foetus infection in cats with diarrhoea in the UK. *J Feline Med Surg* 9(3):214-218.
9. Conrad MD, Bradic M, Warring SD, Gorman AW, & Carlton JM (2013) Getting trichy: tools and approaches to interrogating Trichomonas vaginalis in a post-genome world. *Trends Parasitol* 29(1):17-25.
10. Cepicka I, Hampl V, & Kulda J (2010) Critical taxonomic revision of Parabasalids with description of one new genus and three new species. *Protist* 161(3):400-433.
11. Yao C & Koster LS (2015) Tritrichomonas foetus infection, a cause of chronic diarrhea in the domestic cat. *Vet Res* 46:35.
12. Lucas JJ*, et al.* (2008) Characterization of a cysteine protease from Tritrichomonas foetus that induces host-cell apoptosis. *Arch Biochem Biophys* 477(2):239-243.
13. Singh BN, Hayes GR, Lucas JJ, Beach DH, & Gilbert RO (2005) In vitro cytopathic effects of a cysteine protease of Tritrichomonas foetus on cultured bovine uterine epithelial cells. *Am J Vet Res* 66(7):1181-1186.
14. Tolbert MK, Stauffer SH, Brand MD, & Gookin JL (2014) Cysteine protease activity of feline Tritrichomonas foetus promotes adhesion-dependent cytotoxicity to intestinal epithelial cells. *Infect Immun* 82(7):2851-2859.
15. Tolbert MK, Brand MD, & Gould EN (2016) In vitro effects of cysteine protease inhibitors on Trichomonas foetus-induced cytopathic changes in porcine intestinal epithelial cells. *Am J Vet Res* 77(8):890-897.
16. Gould EN, Giannone R, Kania SA, & Tolbert MK (2017) Cysteine protease 30 (CP30) contributes to adhesion and cytopathogenicity in feline Tritrichomonas foetus. *Veterinary parasitology* 244:114-122.
17. Toblert MK & Gookin JL (2016) Mechanisms of Tritrichomonas foetus Pathogenicity in Cats with Insights from Venereal Trichomonosis. *J Vet Intern Med* 30(2):516-526.
18. Hirt RP (2013) Trichomonas vaginalis virulence factors: an integrative overview. *Sexually transmitted infections* 89(6):439-443.
19. Zubacova Z, Cimburek Z, & Tachezy J (2008) Comparative analysis of trichomonad genome sizes and karyotypes. *Mol Biochem Parasitol* 161(1):49-54.
20. Carlton JM*, et al.* (2007) Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. *Science* 315(5809):207-212.

21. Oyhenart J & Breccia JD (2014) Evidence for repeated gene duplications in Tritrichomonas foetus supported by EST analysis and comparison with the Trichomonas vaginalis genome. *Veterinary parasitology* 206(3-4):267-276.
22. Klodnicki ME, McDougald LR, & Beckstead RB (2013) A genomic analysis of Histomonas meleagridis through sequencing of a cDNA library. *J Parasitol* 99(2):264-269.
23. Hillis DM & Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. *Q Rev Biol* 66(4):411-453.
24. Perkins SL, Martinsen ES, & Falk BG (2011) Do molecules matter more than morphology? Promises and pitfalls in parasites. *Parasitology* 138(13):1664-1674.
25. Felleisen RS (1997) Comparative sequence analysis of 5.8S rRNA genes and internal transcribed spacer (ITS) regions of trichomonadid protozoa. *Parasitology* 115 (Pt 2):111-119.
26. Felleisen RS*, et al.* (1998) Detection of Tritrichomonas foetus by PCR and DNA enzyme immunoassay based on rRNA gene unit sequences. *J Clin Microbiol* 36(2):513-519.
27. Gookin JL, Birkenheuer AJ, Breitschwerdt EB, & Levy MG (2002) Single-tube nested PCR for detection of tritrichomonas foetus in feline feces. *J Clin Microbiol* 40(11):4126-4130.
28. Mukhufhi N, Irons PC, Michel A, & Peta F (2003) Evaluation of a PCR test for the diagnosis of Tritrichomonas foetus infection in bulls: effects of sample collection method, storage and transport medium on the test. *Theriogenology* 60(7):1269-1278.
29. Felleisen RS (1998) Comparative genetic analysis of tritrichomonadid protozoa by the random amplified polymorphic DNA technique. *Parasitol Res* 84(2):153-156.
30. Tachezy J*, et al.* (2002) Cattle pathogen tritrichomonas foetus (Riedmuller, 1928) and pig commensal Tritrichomonas suis (Gruby & Delafond, 1843) belong to the same species. *J Eukaryot Microbiol* 49(2):154-163.
31. Lun ZR, Chen XG, Zhu XQ, Li XR, & Xie MQ (2005) Are Tritrichomonas foetus and Tritrichomonas suis synonyms? *Trends Parasitol* 21(3):122-125.
32. Kleina P, Bettim-Bandinelli J, Bonatto SL, Benchimol M, & Bogo MR (2004) Molecular phylogeny of Trichomonadidae family inferred from ITS-1, 5.8S rRNA and ITS-2 sequences. *Int J Parasitol* 34(8):963-970.
33. Stockdale H*, et al.* (2007) Experimental infection of cattle with a feline isolate of Tritrichomonas foetus. *J Parasitol* 93(6):1429-1434.
34. Stockdale HD*, et al.* (2008) Experimental infection of cats (Felis catus) with Tritrichomonas foetus isolated from cattle. *Veterinary parasitology* 154(1-2):156-161.
35. Slapeta J, Craig S, McDonell D, & Emery D (2010) Tritrichomonas foetus from domestic cats and cattle are genetically distinct. *Experimental parasitology* 126(2):209-213.
36. Slapeta J*, et al.* (2012) Comparative analysis of Tritrichomonas foetus (Riedmuller, 1928) cat genotype, T. foetus (Riedmuller, 1928) cattle genotype and Tritrichomonas suis (Davaine, 1875) at 10 DNA loci. *Int J Parasitol* 42(13-14):1143-1149.
37. Walden HS*, et al.* (2013) A new species of Tritrichomonas (Sarcomastigophora: Trichomonida) from the domestic cat (Felis catus). *Parasitol Res* 112(6):2227-2235.
38. Morin-Adeline V*, et al.* (2014) Comparative transcriptomics reveals striking similarities between the bovine and feline isolates of Tritrichomonas foetus: consequences for in silico drug-target identification. *BMC Genomics* 15:955.
39. Stroud LJ*, et al.* (2017) Comparative proteomic analysis of two pathogenic Tritrichomonas foetus genotypes: there is more to the proteome than meets the eye. *Int J Parasitol* 47(4):203-213.
40. Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends Ecol Evol* 29(1):51-63.
41. Benchimol M*, et al.* (2017) Draft Genome Sequence of Tritrichomonas foetus Strain K. *Genome Announc* 5(16).

42. Reinmann K*, et al.* (2012) Tritrichomonas foetus isolates from cats and cattle show minor genetic differences in unrelated loci ITS-2 and EF-1alpha. *Veterinary parasitology* 185(2-4):138-144.
43. Coppel RL & Black CG (2005) Parasite genomes. *Int J Parasitol* 35(5):465-479.
44. Forrester SJ & Hall N (2014) The revolution of whole genome sequencing to study parasites. *Mol Biochem Parasitol* 195(2):77-81.
45. Talavera-Lopez C & Andersson B (2017) Parasite genomics-Time to think bigger. *PLoS Negl Trop Dis* 11(4):e0005463.
46. Wit J & Gilleard JS (2017) Resequencing Helminth Genomes for Population and Genetic Studies. *Trends Parasitol* 33(5):388-399.
47. Stauffer SH, Birkenheuer AJ, Levy MG, Marr H, & Gookin JL (2008) Evaluation of four DNA extraction methods for the detection of Tritrichomonas foetus in feline stool specimens by polymerase chain reaction. *Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc* 20(5):639-641.
48. Effinger L*, et al.* (2014) Pooling of cultured samples and comparison of multistate laboratory workflows with the MagMAX sample preparation system and VetMAX quantitative polymerase chain reaction reagents for detection of Tritrichomonas foetus-colonized bulls. *Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc* 26(1):72-87.
49. Rotureau B, Gego A, & Carme B (2005) Trypanosomatid protozoa: a simplified DNA isolation procedure. *Experimental parasitology* 111(3):207-209.
50. Adam RD (2000) The Giardia lamblia genome. *Int J Parasitol* 30(4):475-484.
51. Riley DE & Krieger JN (1992) Rapid and practical DNA isolation from Trichomonas vaginalis and other nuclease-rich protozoa. *Mol Biochem Parasitol* 51(1):161-163.
52. Brown J, Pirrung M, & McCue LA (2017) FQC Dashboard: integrates FastQC results into a web-based, interactive, and extensible FASTQ quality control tool. *Bioinformatics*.
53. Borodovsky M & Lomsadze A (2011) Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr Protoc Bioinformatics* Chapter 4:Unit 4 6 1-10.
54. Steinbiss S*, et al.* (2016) Companion: a web server for annotation and analysis of parasite genomes. *Nucleic Acids Res* 44(W1):W29-34.
55. Hoff KJ & Stanke M (2013) WebAUGUSTUS--a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Res* 41(Web Server issue):W123-128.
56. Kanehisa M, Sato Y, & Morishima K (2016) BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol* 428(4):726-731.
57. Wang Y, Coleman-Derr D, Chen G, & Gu YQ (2015) OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research* 43(Web Server issue):W78-W84.
58. Katoh K, Rozewicki J, & Yamada KD (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*.
59. Jones DT, Taylor WR, & Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 8(3):275-282.
60. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7):1870-1874.
61. Altschul SF*, et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17):3389-3402.
62. Finn RD*, et al.* (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44(D1):D279-285.

63. Adam RD*, et al.* (2013) Genome sequencing of Giardia lamblia genotypes A2 and B isolates (DH and GS) and comparative analysis with the genomes of genotypes A1 and E (WB and Pig). *Genome Biol Evol* 5(12):2498-2511.
64. Ulrich LE & Zhulin IB (2007) MiST: a microbial signal transduction database. *Nucleic Acids Res* 35(Database issue):D386-390.
65. Ramos RT*, et al.* (2013) Tips and tricks for the assembly of a Corynebacterium pseudotuberculosis genome using a semiconductor sequencer. *Microb Biotechnol* 6(2):150-156.
66. Nagarajan N, Read TD, & Pop M (2008) Scaffolding and validation of bacterial genome assemblies using optical restriction maps. *Bioinformatics* 24(10):1229-1235.
67. Heesch S*, et al.* (2010) A sequence-tagged genetic map for the brown alga Ectocarpus siliculosus provides large-scale assembly of the genome sequence. *New Phytol* 188(1):42-51.
68. Lowe TM & Chan PP (2016) tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44(W1):W54-57.
69. Zhu XQ*, et al.* (2015) Genetic blueprint of the zoonotic pathogen Toxocara canis. *Nat Commun* 6:6145.
70. Pakandl M & Grubhoffer L (1994) Some properties of sialic-acid binding systems in Tritrichomonas suis and Tritrichomonas foetus. *Comp Biochem Physiol Biochem Mol Biol* 108(4):529-536.
71. Babal P, Pindak FF, Russell LC, & Gardner WA, Jr. (1999) Sialic acid-specific lectin from Tritrichomonas foetus. *Biochim Biophys Acta* 1428(1):106-116.
72. Babal P & Russell LC (1999) Sialic acid-specific lectin-mediated adhesion of Tritrichomonas foetus and Tritrichomonas mobilensis. *J Parasitol* 85(1):33-40.
73. Greenwell P, Younes M, & Rughooputh S (2008) Purification and analysis of DNases of Tritrichomonas foetus: evidence that these enzymes are glycoproteins. *Int J Parasitol* 38(7):749-756.
74. Chou CF & Tai JH (1996) Simultaneous extraction of DNA and RNA from nuclease-rich pathogenic protozoan Trichomonas vaginalis. *Biotechniques* 20(5):790-791.
75. Maritz JM, Land KM, Carlton JM, & Hirt RP (2014) What is the importance of zoonotic trichomonads for human health? *Trends Parasitol* 30(7):333-341.
76. Poole DN & McClelland RS (2013) Global epidemiology of Trichomonas vaginalis. *Sexually transmitted infections* 89(6):418-422.
77. Edwards T, Burke P, Smalley H, & Hobbs G (2014) Trichomonas vaginalis: Clinical relevance, pathogenicity and diagnosis. *Crit Rev Microbiol*:1-12.
78. Nanda N, Michel RG, Kurdgelashvili G, & Wendel KA (2006) Trichomoniasis and its treatment. *Expert review of anti-infective therapy* 4(1):125-135.
79. Brady L & Reid RD (1942) The Treatment of Trichomonas Vaginalis Vaginitis with the Lactobacillus. *Ann Surg* 115(5):840-848.
80. McGrory T & Garber GE (1992) Mouse intravaginal infection with Trichomonas vaginalis and role of Lactobacillus acidophilus in sustaining infection. *Infect Immun* 60(6):2375-2379.
81. McGrory T, Meysick K, Lemchuk-Favel LT, & Garber GE (1994) The interaction of Lactobacillus acidophilus and Trichomonas vaginalis in vitro. *J Parasitol* 80(1):50-54.
82. Lushbaugh WB*, et al.* (2000) Use of intravaginal microbicides to prevent acquisition of Trichomonas vaginalis infection in Lactobacillus-pretreated, estrogenized young mice. *Am J Trop Med Hyg* 63(5-6):284-289.
83. Sato T*, et al.* (2009) Candidatus Desulfovibrio trichonymphae, a novel intracellular symbiont of the flagellate Trichonympha agilis in termite gut. *Environ Microbiol* 11(4):1007-1015.

Vita

Ellen Ann Fleetwood earned a bachelor of engineering degree at Vanderbilt University and a doctor of veterinary medicine degree from the University of Tennessee College of Veterinary Medicine. She has pursued a Ph.D. in Comparative and Experimental Medicine to solve translational medicine questions, especially those pertaining to feline health. She hopes to continue her career as a principal investigator in comparative medical research and teach the next generation of scientists and veterinarians as well.