DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Isogenic sets of T. vaginalis harboring different combinations of trichomonasviruses reveal insights into host–virus interactions

Citation

Hetzel, Carrie Ann. 2025.Isogenic Sets of T. vaginalis Harboring Different Combinations of Trichomonasviruses Reveal Insights Into Host–Virus Interactions. Doctoral Dissertation, Harvard University Graduate School of Arts and Sciences.

Link

https://dash.harvard.edu/handle/1/42719595

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at
https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

HARVARD
Kenneth C. Griffin

GRADUATE SCHOOL
OF ARTS AND SCIENCES

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the
Division of Medical Sciences

Committee on Virology

have examined a dissertation entitled

Isogenic sets of T. vaginalis harboring different combinations of trichomonasviruses reveal insights into host–virus interactions

presented by Carrie A. Hetzel

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature: Aaron G. Schmidt (Dec 4, 2024 11:37 EST)

Typed Name: Dr. Aaron Schmidt

Signature: Sophie Helaine (Dec 4, 2024 11:19 EST)

Typed Name: Dr. Sophie Helaine

Signature: victoria dsouza (Dec 4, 2024 11:06 EST)

Typed Name: Dr. Victoria D'Souza

Signature: Sebastian Lourido (Dec 5, 2024 12:10 EST)

Typed Name: Dr. Sebastian Lourido

Date: December 02, 2024

Isogenic sets of *T. vaginalis* harboring different combinations of trichomonasviruses reveal insights into host–virus interactions

A Dissertation Presented by

Carrie A. Hetzel

to The Program in Virology

in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

in the subject of

Virology

Harvard University

Cambridge, Massachusetts

December 2024

© 2024 Carrie A. Hetzel

All rights reserved.

Dissertation Advisor: Max L. Nibert Carrie A. Hetzel

Isogenic sets of *T. vaginalis* harboring different combinations of trichomonasviruses reveal insights into host–virus interactions

ABSTRACT

Trichomonasviruses are a genus of double stranded RNA (dsRNA) viruses that persist in the obligate human protozoan parasite *T. vaginalis* (Tvag). Tvag is the causative agent of trichomoniasis, the most common nonviral sexually transmitted infection (STI) worldwide. These viruses, commonly known as TVV1, TVV2, TVV3, TVV4, and TVV5, have been implicated in increasing the severity of trichomoniasis by eliciting greater pelvic inflammation. Unlike many familiar eukaryotic viruses, trichomonasviruses are not thought to have an extracellular lifecycle and instead are believed to be transmitted vertically as the parasite divides. The presence of trichomonasviruses is not thought to be detrimental to the health of the parasite, but rigorous investigation into the effects of trichomonasviruses on their host has to date been challenging. Because uninfected parasites are not thought to be able to be infected with trichomonasviruses, previous studies into the effects of trichomonasviruses on their host have largely relied on comparing virus-positive isolates to virus-negative isolates. However, gene expression differences between these isolates have confounded studies. In 2022, Narayanasamy *et al.* reported that a cytidine nucleoside analog, 2'-C-methylcytidine (2CMC) was effective in clearing Tvag isolates of infection with TVV1, TVV2, and TVV3. In this study, we extended the results of Narayanasamy *et al.* to demonstrate that 2CMC is effective

against in reducing viral RNA abundance of representative strains of all five TVV species. In doing so, we identified differences in species-specific differences in 2CMC susceptibility and harness those susceptibility differences to generate isogenic sets of Tvag clones derived from the same parent isolate that harbored different combinations of trichomonasviruses. Using these newly generated isogenic sets, we investigated effects of viral infection on Tvag growth, survival, and adherence to human cells. While we saw no apparent difference in the growth rate of virus-positive and virus-negative Tvag clones in ideal conditions, our results suggested that virus-positive Tvag clones have an increased ability to survive in harsh conditions and a decreased ability to adhere to human cells. To understand the mechanism underlying these results, we performed differential gene expression analysis using three isogenic Tvag sets and revealed a number of differentially expressed genes between cured and uncured Tvag clones, including the putative transcription factor TvMyb4, which was consistently upregulated across Tvag clones harboring trichomonasviruses. We provide evidence suggesting involvement of TvMyb4 in the persistence of trichomonasviruses and in the survival differences between virus-positive and virus-negative clones. Together, the results presented in this work reveal new insights about differences between trichomonasvirus species and the relationship between trichomonasviruses and their host, providing valuable information about both an important human pathogen and a valuable model system for viral persistence.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First, I would like to acknowledge my thesis advisor, Dr. Max Nibert for supporting me over the past five years. Max fostered my development as an independent scientist from the first day I started in his lab and for that I am grateful. His unwavering faith in me allowed me to take risks while knowing that I had a soft place to land they didn't work out. He also gave me room to explore areas outside of my bench work which allowed me to find my passion for education and outreach. I will always be grateful for his quiet support.

I would also like to thank the other mentors in my scientific career up to this point, including Drs. Bob and Janet Siciliano who welcomed me into their lab as an undergraduate student and Dr. Andrew Timmons, who taught me the very basics of how to be a scientist. Additionally, I am grateful to the members of my dissertation advisory committee— Dr. Aaron Schmidt, Dr. Smita Gopinath, and Dr. Sun Hur— for their feedback and guidance throughout the process.

Within my lab, I would like to acknowledge Dr. Austin Manny and (future Dr.) Adede Appah-Sampong. Austin became like a brother to me over the three years we worked together, making me laugh, listening to me complain, and occasionally driving me up the wall. In the early years of my PhD, everything felt high stakes, but Austin was there to remind me that I'd make it through. Adede was my extremely talented undergraduate student and is responsible for many of the preliminary experiments done in this work. She also taught me how to be a good mentor and teacher, which in turn made me a better scientist.

Outside of the lab, I am grateful to all of my friends who have been by my side before and during my time in grad school. They have struggled alongside me, listened to

me vent, and distracted me when I needed distracting. They have supported me through development as a scientist and as a whole person. Similarly, I must express my deep gratitude to my family who have cheered me on through this entire process. Their unwavering confidence in me has helped me to build confidence in myself. They will forever be my first and loudest cheering squad. Thank you for providing me a place to come home to when I wanted to escape the world of science, and for sending me back to Boston with coolers full of Philadelphia delicacies— pork roll and Wawa soft pretzels.

I am also deeply grateful to my partner, Emily (Emmy) Matthews. Emmy has been my biggest source of reassurance that I was going to make it out the other side. She celebrated with me through the successes, consoled me through the failures, and prodded me to open up about the challenges I was trying (and failing) to ignore. She supported me through the publication of my first lead-author manuscript, bringing me pizza rolls and sodas with lots of ice, and petting my hair when I cried out of exhaustion. She brought a joy to my life that I didn't know was missing.

Finally, I must acknowledge my mom: Dr. Christine Elliott. Although she was not here to see me through my PhD, she is the reason I am here today. She raised me for the first seventeen years of my life and her memory has guided me through the past ten. She taught me to be determined, but also to cut my losses and move on when something wasn't working, like my elementary school quest to master jump roping. She taught me not to take life too seriously and to take everything "five minutes at a time." I often hear her voice in the back of my head, bossy as ever, pushing me to take chances and enjoy life. Most importantly, she showed me how to be brave and keep going even when I really don't feel like it. She is the reason I am here today.

INTRODUCTION

Trichomonasviruses, their host, and the human superhost

Trichomonasviruses are a genus of protozoan viruses comprising of five known species: *Trichomonasvirus vagiprimus*, *Trichomonasvirus vagisecundus, Trichomonasvirus vagitertius*, *Trichomonasvirus vagiquartus*, and *Trichomonasvirus vagiquintus* (Tai and Ip, 1995; Bessarab *et al.*, 2000, 2011; Goodman, Freret, *et al.*, 2011; Manny *et al.*, 2022; Simmonds *et al.*, 2024). Commonly, these viruses are referred to as trichomonas vaginalis viruses (TVVs), and the individual species are referred to as TVV1, TVV2, TVV3, TVV4, and TVV5. These viruses persist long term in the obligate human parasite, *Trichomonas vaginalis* (Goodman, Ghabrial, *et al.*, 2011), which is abbreviated through this text as Tvag. Tvag is the causative agent of trichomoniasis, an infection of the urogenital tract with important relevance to human health. In addition to the human health impacts, which will be discussed in detail in the next section, trichomonasviruses provide an intriguing model to study the impact of not just virus–host biology, but also virus–host–superhost interactions. The following sections will summarize relevant details about each of the key players of this system as well as introduce the work presented in this thesis.

Trichomoniasis and the human superhost

As the most common nonviral sexually transmitted infection (STI) worldwide, trichomoniasis is a major global health concern worldwide (Mercer and Johnson, 2018). In 2020, there were an estimated 156 million new cases in people aged 15 to 49 worldwide (World Health Organization, 2020). Common symptoms of trichomoniasis

include vaginal irritation, increased vaginal discharge, pelvic inflammation, and dysuria (Fichorova, 2009). Complications from untreated trichomoniasis during pregnancy include premature delivery, low birth weight, and increased risk of spontaneous abortion (Mercer and Johnson, 2018). Additionally, trichomoniasis is associated with development of bacterial vaginosis and increased acquisition of HIV and other sexually transmitted infections (Fichorova, 2009).

Trichomoniasis can infect the urogenital tracts of people with any type of genitalia. For brevity, I will refer to the urogenital tract containing a penis to be the male anatomy, and the urogenital tract containing a vagina and uterus to be the female anatomy. However, it is important to remember that not all people with vaginas are indeed female, and not all people with penises are indeed male. Similarly, not all women have vaginas and not all men have penises. With this being said, trichomoniasis infection tends to more often be symptomatic in patients with female urogenital tracts than patients with male urogenital tracts. Approximately 50% of trichomoniasis cases in people with female genitalia are asymptomatic (Sutton *et al.*, 2007), while the greater than 75% of cases in people with male genitalia are estimated to be asymptomatic. (Hirt and Sherrard, 2015). The lack of symptoms increases the risk of transmission of trichomoniasis, as uninfected carriers can spread the infection through sexual intercourse (Sutton *et al.*, 2007).

Despite being largely asymptomatic in people with male genitalia, trichomoniasis has been associated with increased prostate inflammation (Skerk *et al.*, 2004). In one study, infection with the parasite was shown to cause increased prostate inflammation in rats (Jang *et al.*, 2019). Further, trichomonads have been found in prostate tissue of patients with benign prostatic hyperplasia (BPH), a common condition in men over 50

(Mitteregger *et al.*, 2012). Infection with trichomonads has been shown to induce BPH *ex vivo* through induction of NLRP3 inflammasome activation (Gu *et al.*, 2016). Some investigators have suggested a link between trichomoniasis infection and prostate cancer, but results have thus far been conflicting (Yang *et al.*, 2022; Nagata *et al.*, 2023; Sawaya *et al.*, 2024).

Symptomatic trichomoniasis cases have historically been treated with the 5-nitroimidazole drug, metronidazole, or more recently with tinidazole, a derivative with a longer half-life (Hirt and Sherrard, 2015; Leitsch, 2016). However, resistance to 5-nitroimadazole drugs is on the rise (Sobel and Sobel, 2015), increasing the need for alternative therapies. Many other investigators have screened for compounds with trichomonacidal activity; most recently, 5-imidazole carbamates have been shown to be effective at restricting Tvag growth *in vitro* (Martínez-Rosas *et al.*, 2024). However, nitroimidazoles remain the only compounds currently approved by the FDA for treatment of trichomoniasis (Muzny, Van Gerwen and Legendre, 2022).

Despite the existence of these treatments, trichomoniasis remains a public health burden to vulnerable populations. An estimated 85% of new trichomoniasis infections occur in people in developing countries (Secor, 2012). Within the United States, the prevalence of trichomoniasis infection in Black women was estimated to be more than 10 times higher than the prevalence in White women in 2004 (Sutton *et al.*, 2007). More recently, the prevalence of trichomoniasis was estimated to be 8.9% in Black women and 4.2% in Black men. In comparison, the prevalence of trichomoniasis in men and women of other races was estimated to be 0.03% and 0.8%, respectively (Patel *et al.*, 2018). Further, incarcerated individuals and sex workers have increased risks of contracting

trichomoniasis (Kissinger, 2015). Despite these disparities and the total global health burden, trichomoniasis is listed by the Center for Disease Control (CDC) as a neglected parasitic infection (Ibáñez-Escribano and Nogal-Ruiz, 2024). Thus, further research on the parasite that causes this disease, and the endosymbiont viruses harbored by the parasite, is of particular relevance to combatting global health inequities.

T. vaginalis biology

Tvag is a flagellated protozoan the colonizes the human urogenital tract (Leitsch, 2016). In addition to its relevance to human health, its status as one of the earliest branching eukaryotes (Pace, 2009) (Figure I.1) makes it an intriguing model system for studying evolutionary biology, particularly for investigating the emergence of eukaryotic antiviral defenses. Along with other members of the genus *Trichomonas,* Tvag is microaerophilic, meaning it survives best under conditions of low oxygen, and lacks mitochondria, and instead using hydrogenosomes to facilitate cellular respiration (Chapman *et al.*, 1985). Despite the early emergence of its lineage, modern Tvag is considered to be an obligate extracellular parasite that must colonize the human urogenital tract to survive outside of laboratory conditions. The evolutionary relationship between humans and this modern-day *Trichomonas* species is still unknown, but it has been hypothesized that trichomoniasis infection in humans may have first occurred through zoonotic transmission from birds and that the parasite has evolved to survive in the microenvironment of the human urogenital tract. (Maritz *et al.*, 2014).

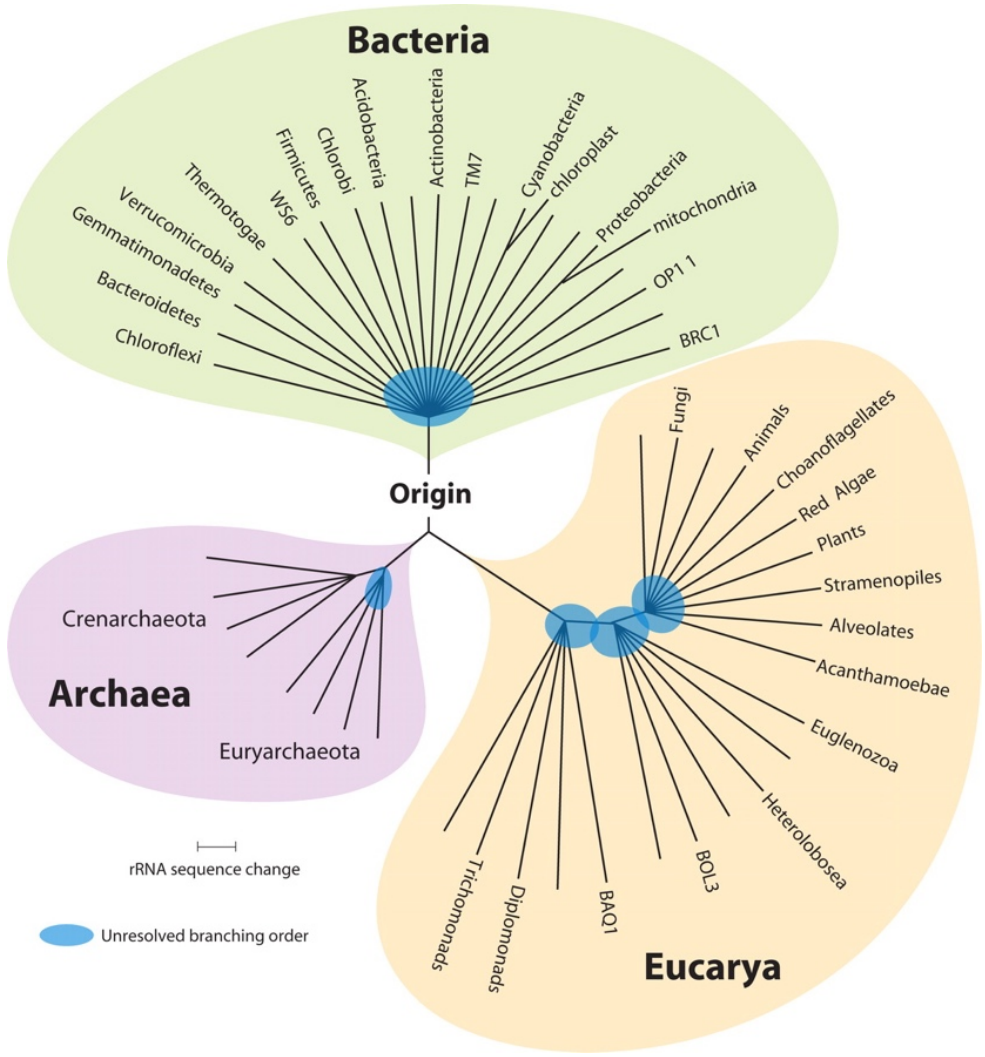


Figure I.1: Simplified Phylogenetic Tree of Life. Trichomonads are an early branching lineage, likely diverging before the acquisition of mitochondria by most eukaryotes. This status makes Tvag an important tool for studying the evolutionary history of host–virus interactions. (American Society for Microbiology, 2021)

To proliferate and establish infection, Tvag attaches to epithelial cells of the human urogenital tract (Mercer and Johnson, 2018). This allows trichomonads to avoid being washed away by gravity and by genital secretions, as well as gain nutrients from feeding off of human cells (Ryan, de Miguel and Johnson, 2011). Trichomonads adhere to human cells through lipoglycans on their surface, which bind to Galectin-1 on host cells (Okumura, Baum and Johnson, 2008). Surface proteins, including TvBAP-1 and TvBAP-2, as well as intracellular proteins TvPP1Y, TvLegu1, and TvROM1 have also been associated with Tvag adherence (Mercer and Johnson, 2018), but a detailed mechanism of Tvag adherence to human cells has not yet been established. Trichomonads also adhere to one another, in a process commonly referred to as “clumping,” which is also thought to prevent washing away of the parasites during the initial establishment of infection (Coceres *et al.*, 2015). Additionally, while a baseline level of adherence is thought to be needed to establish infection, increased adherence is in many cases not correlated with increased host cell lysis (Lustig *et al.*, 2013).

During the course of infection, trichomonads induce contact-dependent lysis of human epithelial cells (Lustig *et al.*, 2013) in order to acquire nutrients (Mercer and Johnson, 2018). Trichomonads have been shown to have both phagocytic (Rendón-Maldonado *et al.*, 1998) and trogocytic (i.e. “nibbling” on human cells) (Midlej and Benchimol, 2010) activities. These processes are believed to result in necrosis of human epithelial cells (Lustig *et al.*, 2013), but the key Tvag effectors of this process have yet to be determined. Some candidate proteases, including TVROM1 (Riestra *et al.*, 2015), TvMP50 (Puente-Rivera *et al.*, 2017) and TvCP2 (Miranda-Ozuna *et al.*, 2019) have been suggested to be involved in host cell killing. Pore forming proteins known as TvSaplips

have been proposed to support the trogocytic activity of trichomonads (Diaz *et al.*, 2020). Additionally, trichomonads can also phagocytose smaller lymphocytes, erythrocytes, and commensal microbes (Rendón-Maldonado *et al.*, 1998) during infection.

Trichomonads can exhibit three distinct morphologies to support their survival and replication in different environments, as well as to promote adherence to human cells (Benchimol, 2008; Dias-Lopes *et al.*, 2018; Beri *et al.*, 2020). Under ideal conditions, trichomonads typically exhibit a trophozoite morphology (Figure I.2A), characterized by a pear-shaped structure and externalized flagella (Wartoña and Honigberg, 1979). This morphology is associated with Tvag replication and motility (Pereira-Neves and Benchimol, 2008). When in stressful conditions, such as high salt or low pH, trichomonads can adopt a pseudocyst morphology (Figure I.2B) by internalizing their flagella and forming a protective outer layer of chitin (Beri *et al.*, 2020). This process is similar to other encysting protozoa, but Tvag lacks a cell wall characteristic of a true cyst structure (Dias-Lopes *et al.*, 2018). Tvag encystation has been suggested to promote short-term survival on surfaces outside of the body, including in swimming pools (Pereira-Neves and Benchimol, 2008). Finally, the amoeboid morphology (Figure I.2C), which is often described as having a pancake-like appearance, is associated with attachment to human epithelial cells (Carlton *et al.*, 2007; Hirt, 2013).

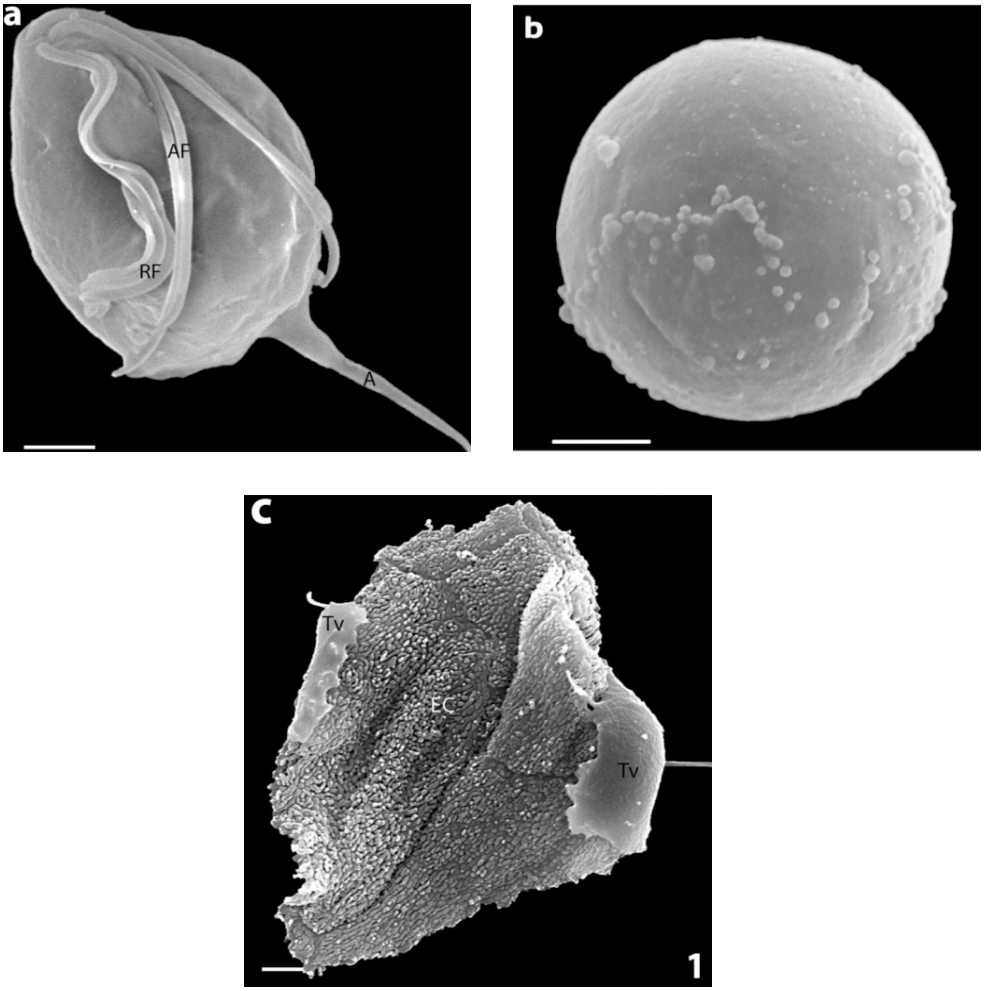


Figure I.2: Morphologies of *T. vaginalis.* Trichomonads can exhibit (A) trophozoite, (B) pseudocyst, or (C) amoeboid morphologies. Ax: Axostyle, AF: anterior flagella, EC: Human vaginal epithelial cell, RF: recurrent flagellum. Tv: *Trichomonas vaginalis.* Bars represent 1µm. (Benchimol et al., 2008).

To facilitate cell-cell communication, trichomonads secrete extracellular vesicles (EVs) (Twu *et al.*, 2013), which can then be internalized either by other trichomonads or by human epithelial cells during infection (Rai and Johnson, 2019). These EVs are similar in size and composition to mammalian exosomes (Twu *et al.*, 2013). EVs have been shown to modulate parasite adherence to host cells (Kochanowsky *et al.*, 2024), due at least in part to increasing the number of cytoneme structures on the surface of

trichomonads (Salas *et al.*, 2023). Crosstalk between parasites is particularly interesting because approximately 11% of clinical isolates are estimated to be mixed isolates containing two or more strains of Tvag (Conrad *et al.*, 2012). Of relevance to this work, TVVs have been shown to both be packaged in and modulate the cargo of Tvag EVs (Govender *et al.*, 2020; Ong *et al.*, 2022). The potential implications of this modulation are described in the discussion chapter of this text.

In addition to trichomonasviruses, which will be discussed at length in the following section, Tvag can be persistently infected with bacterial species *Mycoplasma hominis* (Mh) and *Mycoplasma girerdii* (Mg) with little harmful effects on the protozoan (Fiori *et al.*, 2013; Martin *et al.*, 2013). Infection with Mh-containing Tvag strains has been shown to increase inflammation to the human superhost (Fiori *et al.*, 2013; Mercer *et al.*, 2016). Further, both Mh and Mg have been shown to increase Tvag binding to human cells and resulting hemolysis (Margarita *et al.*, 2022). Additionally, Mh has been demonstrated to promote sensitivity to metronidazole of the protozoan host, but Mg does not appear to affect metronidazole sensitivity (Margarita *et al.*, 2022).

Despite its impact on human health and its relevance to evolutionary biology, Tvag remains largely understudied. To date, only one published genome is available, so little is known about genome variability (Carlton *et al.*, 2007). This is particularly relevant to our work with trichomonasviruses because previous studies looking at effects of trichomonasviruses on their host have relied on comparing Tvag isolates with different genetic backgrounds. Without better information on the genetic variability between isolates, interpreting gene expression data has been difficult. Further, the most recent annotation of this genome reports approximately 37,000 protein coding genes, but most

little to know information is currently known about most of these genes (https://trichdb.org/trichdb/app/record/dataset/DS_e6874bf420). Thus, a main barrier to studying Tvag, and to studying the interplay between Tvag and trichomonasviruses is the lack of information available about individual genes of the Tvag genome and of understanding of the genetic variability across different Tvag isolates.

Trichomonas vaginalis viruses (TVVs)

T. vaginalis isolates can be persistently infected with strains of one or more species of *Trichomonasvirus* (Goodman, Freret, *et al.*, 2011; Manny *et al.*, 2022). Trichomonasviruses are a genus of double stranded RNA (dsRNA) viruses in the family *Pseudototiviridae,* along with other viral genera infecting protozoa or fungi (Simmonds *et al.*, 2024). As previously mentioned, there are currently five known species in the genus *Trichomonasvirus*: *Trichomonasvirus vagiprimus*, *Trichomonasvirus vagisecundus, Trichomonasvirus vagitertius*, *Trichomonasvirus vagiquartus*, and *Trichomonasvirus vagiquintus* (Manny *et al.*, 2022). Commonly, these viruses are referred to as trichomonas vaginalis viruses (TVVs) and individual species referred to as TVV1, TVV2, TVV3, TVV4, and TVV5. Each species has at least five reported strains (See Figure I.3) and the strains are named denoting the Tvag isolate that each strain infects (e.g. TVV1- JH37A#2 is a strain of TVV1 that persists in Tvag isolate JH37A#2). According to phylogenetic analysis performed by Manny et al. (see Appendix 1), strains of TVV2 appear to be more similar in sequence to strains of TVV5 (Figure I.3). Likewise, sequences from strains of TVV3 and TVV4 cluster together, leaving TVV1 as the most divergent from the others (Manny

et al., 2022). It is still largely unknown if these differences in sequence correlate to differences in virus biology or effects on the host.

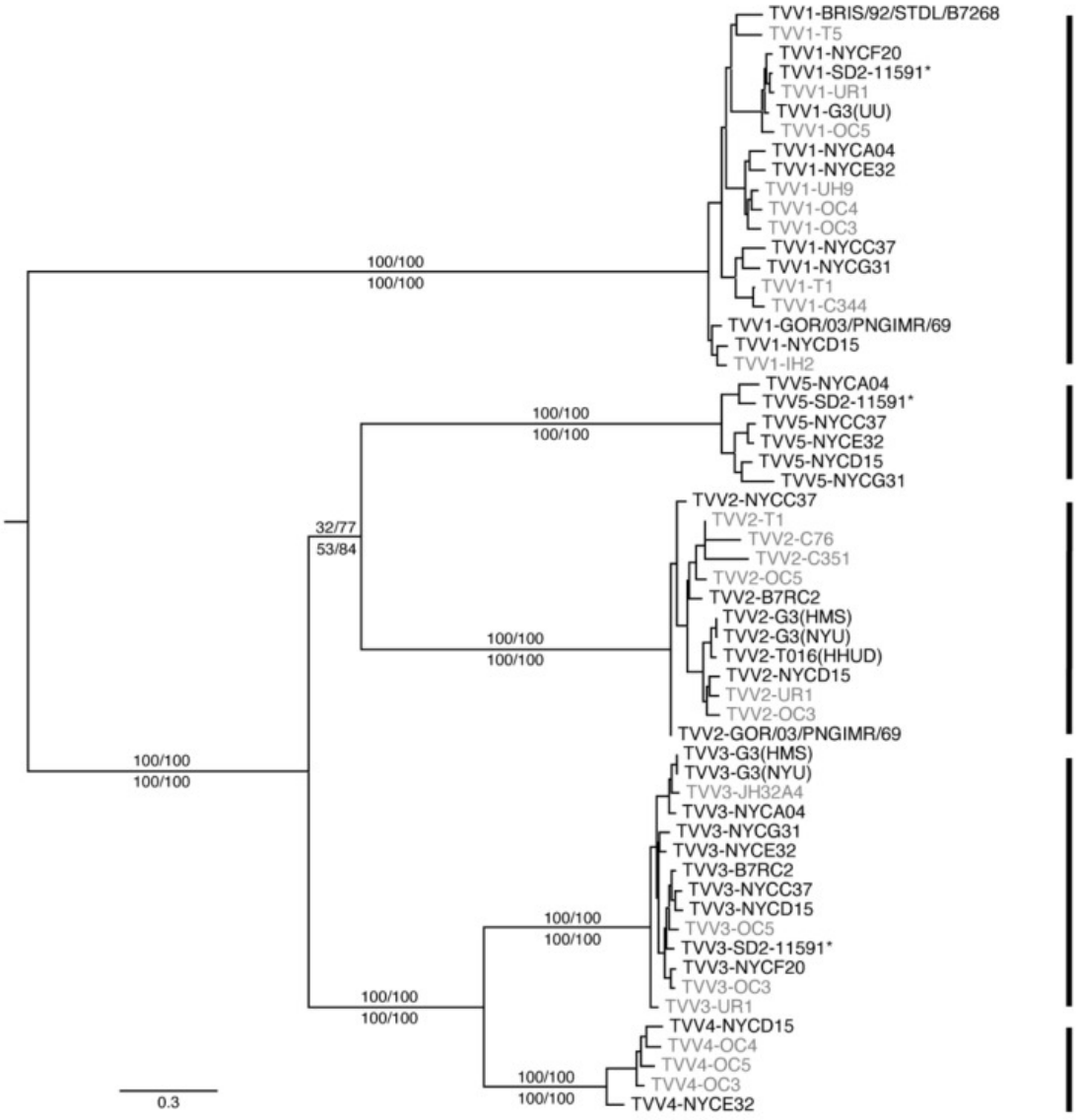


Figure I.3: Phylogenetic Tree of Known Strains of TVV1 – 5. Phylogenetic tree was generated based on amino acid sequences from the CP/RdRp fusion protein. TVV1 strains are the most divergent, while strains from TVV2 and TVV5 cluster together, as do strains from TVV3 and TVV4. See Appendix 1 for more details (Manny et al., 2022).

Trichomonasviruses are relatively simple viruses, making them excellent models for studying host–virus interactions. The trichomonasvirus genome contains only two open reading frames (ORFs) encoding for two proteins: a capsid protein (CP) and RNA-dependent RNA polymerase (RdRp) (Goodman, Ghabrial, *et al.*, 2011). The RdRp is expressed as a fusion protein with CP as a result of a ribosomal frameshifting (Goodman, Freret, *et al.*, 2011). Virion structures of representative strains of TVV1 and TVV2 have thus far been determined (Parent *et al.*, 2013; Stevens *et al.*, 2021). As with other members of the genus *Pseudototiviridae,* both virion structures are comprised of 120 CP subunits (Parent *et al.*, 2013; Stevens *et al.*, 2021).

Unlike most viruses, trichomonasviruses are not thought to have an extracellular phase in their lifecycle and are instead believed to be transmitted vertically from mother to daughter cell as the parasite divides (Goodman, Ghabrial, *et al.*, 2011). These viruses persist long-term in the host, remaining present in parasite cultures even after six months of serial passaging (Goodman, Ghabrial, *et al.*, 2011). Long-term persistence of these viruses suggests that there may be an evolutionary benefit to the protozoan to maintain these viruses at tolerable levels instead of clearing the infection entirely. However, how this détente is achieved and maintained is not yet understood. The Tvag genome does not appear to encode proteins with homology to classical dsRNA sensors such RIG-I or TLR3 (Carlton *et al.*, 2007). RNA interference (RNAi) is one innate immune pathway that does appear to be intact in Tvag (Carlton *et al.*, 2007), but its potential role in limiting TVV replication has not been reported to date. Some preliminary evidence for this pathway will be noted in the discussion chapter of this text, but its role in limiting viral replication is not

conclusive. Thus, a major open area of research in the trichomonasvirus field is in identifying Tvag immune pathways used to control these viruses.

Further, benefits that TVVs might confer to their protozoan host has not yet been established. TVVs have been shown to lead to increased expression of Tvag protein P270 in some isolates, which has been linked to decreased cytotoxicity to HeLa cells (Alderete *et al.*, 1986). It is hypothesized that this may lead evasion of host defenses by Tvag (Goodman, Ghabrial, *et al.*, 2011), although this has not been robustly shown. Similarly, TVVs have been suggested to modulate cysteine protease expression in Tvag (Provenzano, Khoshnan and Alderete, 1997), but whether this altered expression pattern confers a benefit to parasite is still unknown. More research into how trichomonasviruses affect global gene expression is greatly needed in the field. In Chapter 2 of this work, we begin to address this question by performing differential gene expression analysis using infected and uninfected Tvag clones.

In addition to investigation into impacts of TVVs on their host, research into effects of TVVs on the human superhost is also ongoing. Trichomonasviruses are thought to increase pelvic inflammation, thereby increasing the severity of trichomoniasis infection (Fichorova *et al.*, 2012). Because TVVs do not appear to have an extracellular phase in their lifecycle, they are thought to be released by dead and dying parasites, and the free virions then sensed by Toll-like Receptor 3 (TLR3), triggering a proinflammatory cascade in human epithelial cells (Figure I.3) (Fichorova *et al.*, 2012). As an important caveat, these studies were performed using Tvag isolates containing strains of TVV1, either in isolation or in combination with strains of other TVV species, so differences in inflammatory responses elicited by different viral species have yet to be determined.

Further, this study focused on TLR3-meditated inflammatory response, so the impact of TVVs on eliciting inflammation through other immune pathways has also not been shown.

The lack of suitable isolates to investigate effects of TVV species on the host and human superhost has historically been a limiting factor in research progress in the trichomonasvirus field. Because TVVs are not thought to be outwardly infectious (i.e. incubating purified virions with Tvag cultures has not been seen to result in a productive infection), so studying differences between uninfected and infected parasites has been challenging. Similarly, the ability of Tvag isolates to be coinfected with at least four species of trichomonasviruses (Goodman, Freret, *et al.*, 2011; Manny *et al.*, 2022) makes studying these viruses in isolation difficult. Many previous studies have relied on screening Tvag isolates for different combinations of viruses. However, to date, only isolates that are singly infected with strains of TVV1, TVV2, or TVV3 have been reported, and strains of TVV4 and TVV5 have only been seen in combination with strains of other viral species (See Appendix 2). Further, differences in genetic backgrounds have confounded comparisons of virus-positive and virus-negative isolates.

In 2022, a study performed by Narayanasamy *et al.* demonstrated that treatment of Tvag isolates with nucleoside analog, 2'-C-methylcytidine (2CMC) (Figure I.4 A) was effective in reducing abundance of viral RNA from representative strains of TVV1, TVV2, and TVV3 (Figure I.4B). They also used this drug to clear trichomonads of viral infection, providing a strategy for creation of sets of cured and uncured isogenic Tvag clones (Narayanasamy *et al.*, 2022). These results provide a way to overcome one of the biggest problems in the TVV field: the difficulty of creating virus-positive and virus-negative clones derived from the same parent isolate.

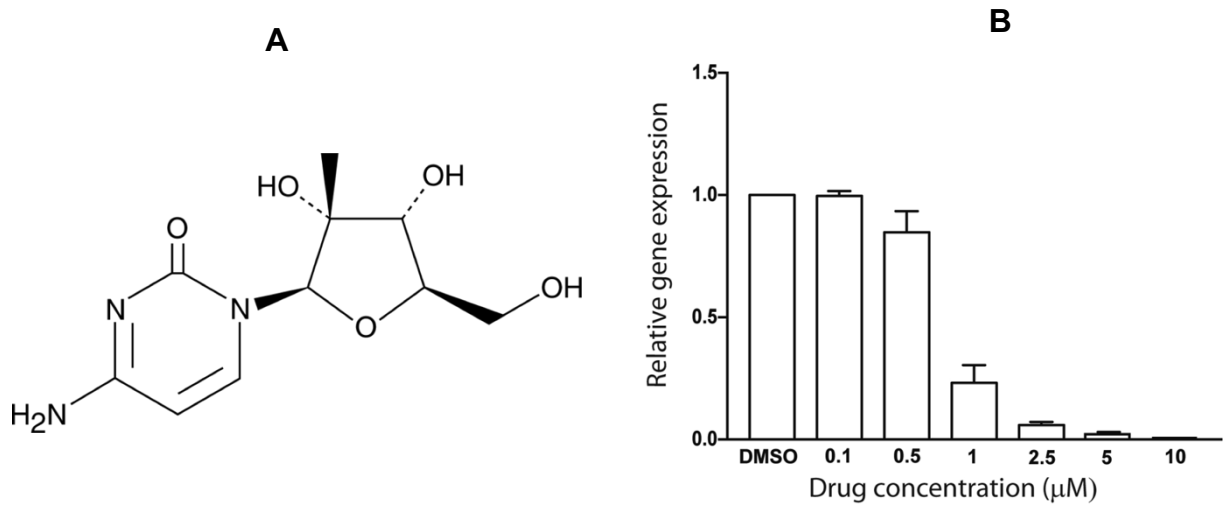


Figure I.4: 2CMC is effective in reducing TVV1 RNA abundance. (A) Structure of nucleoside analog 2'-C-methylcytdine. (B) Data from Narayanasamy *et al.* Trichomonads harboring TVV1 were treated with increasing concentrations of 2CMC, resulting in decreased viral RNA abundance. RT-qPCR was used to measure resulting TVV1 abundance. Full text can be found at:
https://www.sciencedirect.com/science/article/pii/S1684118221001808?via%3Dihub

Using isogenic strains to determine effects of TVVs on the protozoan host

The goals of this thesis project were to (1) use 2CMC to generate isogenic Tvag sets for our own use, and (2) leverage these isolates to investigate effects of TVV infection on Tvag gene expression and phenotype (Figure I.5). In the process of accomplishing the first goal, we also identified differences in susceptibility to 2CMC between representative strains of the five TVV species.

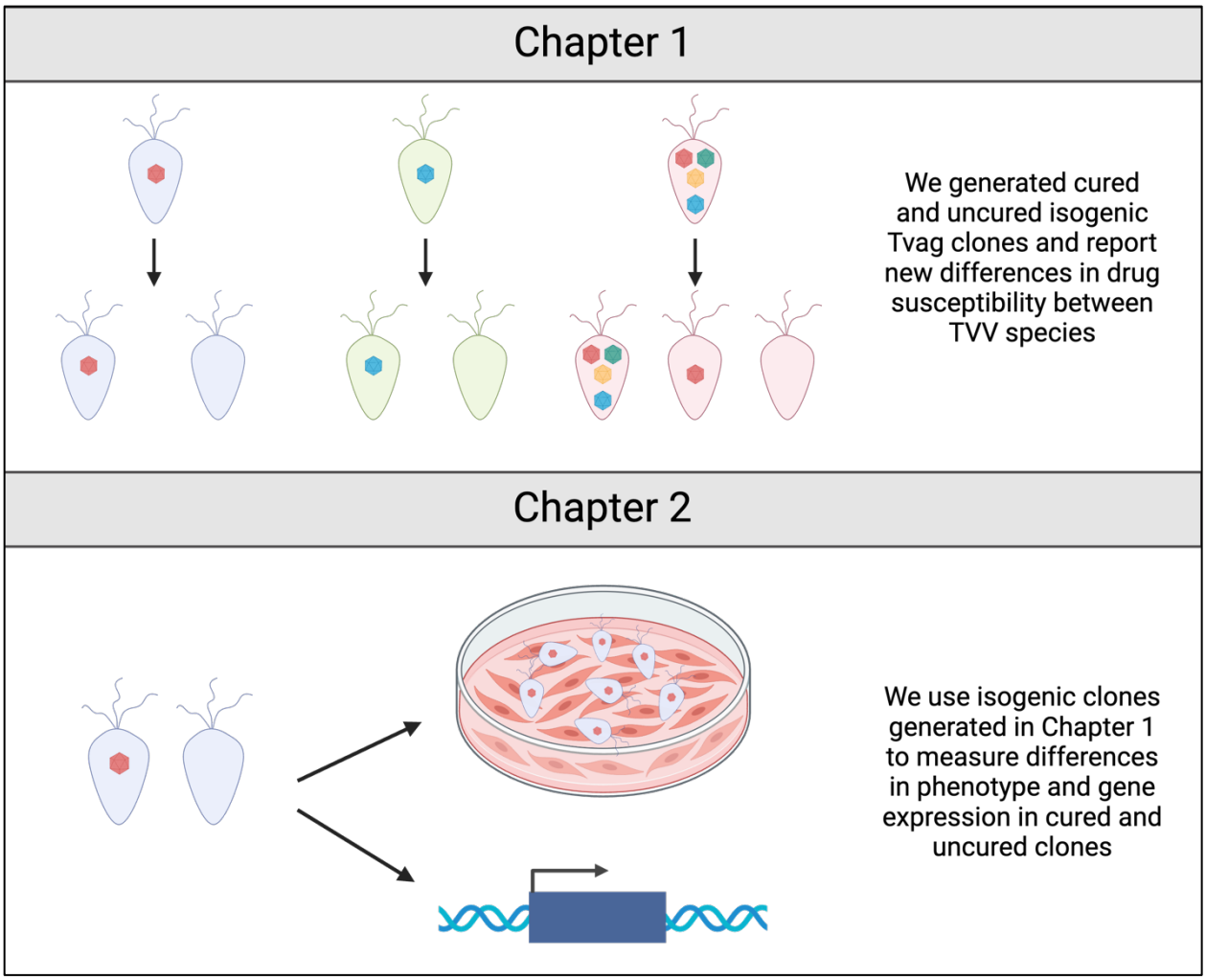


Figure I.5 Graphical Abstract of Research. Isogenic sets generated in Chapter 1 were used to determine differences in phenotype and gene expression between cured and uncured Tvag isolates.

In Chapter 1, we treat a panel of Tvag isolates with 2CMC and corroborate the results of Narayanasamy *et al.* that this drug is able to clear parasites of TVV1, TVV2, and TVV3 infection (Narayanasamy *et al.*, 2022), as well as extend these results to demonstrate 2CMC-mediated inhibition of representative strains of TVV4 and TVV5. We also report newly identified differences in susceptibility to 2CMC treatment between representative strains of the five TVV species. Finally, we utilize these differences in

susceptibility to generate sets of singly infected, multiply infected, and uninfected Tvag clones derived from the same parent isolate.

In Chapter 2, we use three of these isogenic sets to identify differences in Tvag growth and survival, adherence to human cells, and encystation between virus-positive and virus-negative Tvag clones. We report that while the presence of strains TVV1 or TVV3 do not appear to cause any difference in Tvag growth rate compared to cured clones, the presence of these strains does appear to confer a survival advantage to the parasite. Further, isolates infected with representative strains of TVV1 or TVV3 appear to adhere less strongly to human cells. To understand the mechanism behind these differences, we performed RNA sequencing (RNA-seq) using these isogenic sets and identify a number of differentially expressed genes, including a transcription factor, TvMyb4, which was upregulated in virus-positive strains in all three experiments we performed. Further, we show that knockdown of TvMyb4 both decreases viral RNA abundance and limits parasite survival in harsh conditions, indicating a role for this protein in both maintaining viral replication and in modulating the phenotypic differences between cured and uncured Tvag clones.

Implications of this work, as well as future areas of research in the field are discussed in the final section of the main text. Appendices 1 and 2 describe additional work to characterize what viruses are present across commercially available Tvag isolates and expand the known collection of genome sequences from across all five TVV species.

References

Alderete, J.F. *et al.* (1986) 'Phenotypic variation and diversity among *Trichomonas vaginalis* isolates and correlation of phenotype with trichomonal virulence determinants.', *Infection and Immunity*, 53(2), pp. 285–293.

Benchimol, M. (2008) 'The hydrogenosome as a drug target', *Current Pharmaceutical Design*, 14(9), pp. 872–881. Available at: https://doi.org/10.2174/138161208784041114.

Beri, D. *et al.* (2020) 'Demonstration and characterization of cyst-like structures in the life cycle of *Trichomonas vaginalis*', *Frontiers in Cellular and Infection Microbiology*, 9. Available at: https://doi.org/10.3389/fcimb.2019.00430.

Bessarab, I.N. *et al.* (2000) 'The complete cDNA sequence of a type II *Trichomonas vaginalis* virus', *Virology*, 267(2), pp. 350–359. Available at: https://doi.org/10.1006/viro.1999.0129.

Bessarab, I.N. *et al.* (2011) 'Identification and characterization of a type III Trichomonas vaginalis virus in the protozoan pathogen *Trichomonas vaginalis'*, *Archives of Virology*, 156(2), pp. 285–294. Available at: https://doi.org/10.1007/s00705-010-0858-y.

Bokharaei-Salim, F. *et al.* (2020) 'The first detection of co-infection of double-stranded RNA virus 1, 2 and 3 in Iranian isolates of *Trichomonas vaginalis*', *Iranian Journal of Parasitology*, 15(3), pp. 357–363. Available at: https://doi.org/10.18502/ijpa.v15i3.4200.

Carlton, J.M. *et al.* (2007) 'Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis'*, *Science (New York, N.Y.)*, 315(5809), pp. 207–212. Available at: https://doi.org/10.1126/science.1132894.

Chapman, A. *et al.* (1985) 'Energy-dispersive X-ray microanalysis of membrane-associated inclusions in hydrogenosomes isolated from *Trichomonas vaginalis'*, *Journal of General Microbiology*, 131(11), pp. 2933–2939. Available at: https://doi.org/10.1099/00221287-131-11-2933.

Coceres, V.M. *et al.* (2015) 'The C-terminal tail of tetraspanin proteins regulates their intracellular distribution in the parasite *Trichomonas vaginalis*', *Cellular Microbiology*, 17(8), pp. 1217–1229. Available at: https://doi.org/10.1111/cmi.12431.

Conrad, M.D. *et al.* (2012) 'Extensive genetic diversity, unique population structure and evidence of genetic exchange in the sexually transmitted parasite *Trichomonas vaginalis*', *PLoS Neglected Tropical Diseases*, 6(3), p. e1573. Available at: https://doi.org/10.1371/journal.pntd.0001573.

Dias-Lopes, G. *et al.* (2018) 'In-depth quantitative proteomic analysis of trophozoites and pseudocysts of *Trichomonas vaginalis'*, *Journal of Proteome Research*, 17(11), pp. 3704–3718. Available at: https://doi.org/10.1021/acs.jproteome.8b00343.

Diaz, N. *et al.* (2020) 'Production and functional characterization of a recombinant predicted pore-forming protein (TVSAPLIP12) of *Trichomonas vaginalis* in *Nicotiana benthamiana* plants', *Frontiers in Cellular and Infection Microbiology*, 10, p. 581066. Available at: https://doi.org/10.3389/fcimb.2020.581066.

Fichorova, R.N. (2009) 'Impact of *T. vaginalis* infection on innate immune responses and reproductive outcome', *Journal of reproductive immunology*, 83(1–2), pp. 185–189. Available at: https://doi.org/10.1016/j.jri.2009.08.007.

Fichorova, R.N. *et al.* (2012) 'Endobiont viruses sensed by the human host – beyond conventional antiparasitic therapy', *PLOS ONE*, 7(11), p. e48418. Available at: https://doi.org/10.1371/journal.pone.0048418.

Fiori, P.L. *et al.* (2013) 'Association of *Trichomonas vaginalis* with its symbiont *Mycoplasma hominis* synergistically upregulates the in vitro proinflammatory response of human monocytes', *Sexually Transmitted Infections*, 89(6), pp. 449–454. Available at: https://doi.org/10.1136/sextrans-2012-051006.

Goodman, R.P., Freret, T.S., *et al.* (2011) 'Clinical isolates of *Trichomonas vaginalis* concurrently infected by strains of up to four *Trichomonasvirus* species (Family Totiviridae)', *Journal of Virology*, 85(9), pp. 4258–4270. Available at: https://doi.org/10.1128/JVI.00220-11.

Goodman, R.P., Ghabrial, S.A., *et al.* (2011) '*Trichomonasvirus*: a new genus of protozoan viruses in the family Totiviridae', *Archives of Virology*, 156(1), pp. 171–179. Available at: https://doi.org/10.1007/s00705-010-0832-8.

Govender, Y. *et al.* (2020) 'The role of small extracellular vesicles in viral-protozoan symbiosis: lessons from *Trichomonasvirus* in an isogenic host parasite model', *Frontiers in Cellular and Infection Microbiology*, 10, p. 591172. Available at: https://doi.org/10.3389/fcimb.2020.591172.

Gu, N.-Y. *et al.* (2016) '*Trichomonas vaginalis* induces IL-1β production in a human prostate epithelial cell line by activating the NLRP3 inflammasome via reactive oxygen species and potassium ion efflux', *The Prostate*, 76(10), pp. 885–896. Available at: https://doi.org/10.1002/pros.23178.

Hirt, R.P. (2013) '*Trichomonas vaginalis* virulence factors: an integrative overview', *Sexually Transmitted Infections*, 89(6), pp. 439–443. Available at: https://doi.org/10.1136/sextrans-2013-051105.

Hirt, R.P. and Sherrard, J. (2015) '*Trichomonas vaginalis* origins, molecular pathobiology and clinical considerations', *Current Opinion in Infectious Diseases*, 28(1), pp. 72–79. Available at: https://doi.org/10.1097/QCO.0000000000000128.

Ibáñez-Escribano, A. and Nogal-Ruiz, J.J. (2024) 'The past, present, and future in the diagnosis of a neglected sexually transmitted infection: trichomoniasis', *Pathogens*, 13(2), p. 126. Available at: https://doi.org/10.3390/pathogens13020126.

Jang, K.-S. *et al.* (2019) 'Experimental rat prostatitis caused by *Trichomonas vaginalis* infection', *The Prostate*, 79(4), pp. 379–389. Available at: https://doi.org/10.1002/pros.23744.

Kissinger, P. (2015) 'Epidemiology and treatment of trichomoniasis', *Current infectious disease reports*, 17(6), p. 484. Available at: https://doi.org/10.1007/s11908-015-0484-7.

Leitsch, D. (2016) 'Recent advances in the *Trichomonas vaginalis* field', *F1000Research*, 5, p. F1000 Faculty Rev-162. Available at: https://doi.org/10.12688/f1000research.7594.1.

Lustig, G. *et al.* (2013) '*Trichomonas vaginalis* contact-dependent cytolysis of epithelial cells', *Infection and Immunity*, 81(5), pp. 1411–1419. Available at: https://doi.org/10.1128/IAI.01244-12.

Manny, A.R. *et al.* (2022) 'Discovery of a novel species of trichomonasvirus in the human parasite *Trichomonas vaginalis* using transcriptome mining', *Viruses*, 14(3), p. 548. Available at: https://doi.org/10.3390/v14030548.

Margarita, V. *et al.* (2022) 'Effect of the symbiosis with *Mycoplasma hominis* and *Candidatus* Mycoplasma Girerdii on *Trichomonas vaginalis* metronidazole susceptibility', *Antibiotics (Basel, Switzerland)*, 11(6), p. 812. Available at: https://doi.org/10.3390/antibiotics11060812.

Maritz, J.M. *et al.* (2014) 'What is the importance of zoonotic trichomonads for human health?', *Trends in Parasitology*, 30(7), pp. 333–341. Available at: https://doi.org/10.1016/j.pt.2014.05.005.

Martin, D.H. *et al.* (2013) 'Unique vaginal microbiota that includes an unknown mycoplasma-like organism is associated with *Trichomonas vaginalis* infection', *The Journal of Infectious Diseases*, 207(12), pp. 1922–1931. Available at: https://doi.org/10.1093/infdis/jit100.

Martínez-Rosas, V. *et al.* (2024) 'Imidazole carbamates as a promising alternative for treating trichomoniasis: in vitro effects on the growth and gene expression of *Trichomonas vaginalis*', *Molecules (Basel, Switzerland)*, 29(11), p. 2585. Available at: https://doi.org/10.3390/molecules29112585.

Mercer, F. *et al.* (2016) 'Leukocyte lysis and cytokine induction by the human sexually transmitted parasite *Trichomonas vaginalis*', *PLoS Neglected Tropical Diseases*, 10(8), p. e0004913. Available at: https://doi.org/10.1371/journal.pntd.0004913.

Mercer, F. and Johnson, P.J. (2018) '*Trichomonas vaginalis*: pathogenesis, symbiont interactions and host cell immune responses', *Trends in Parasitology*, 34(8), pp. 683–693. Available at: https://doi.org/10.1016/j.pt.2018.05.006.

Meysick, K.C. and Garber, G.E. (1992) 'Interactions between *Trichomonas vaginalis* and vaginal flora in a mouse model', *The Journal of Parasitology*, 78(1), pp. 157–160.

Midlej, V. and Benchimol, M. (2010) '*Trichomonas vaginalis* kills and eats–evidence for phagocytic activity as a cytopathic effect', *Parasitology*, 137(1), pp. 65–76. Available at: https://doi.org/10.1017/S0031182009991041.

Miranda-Ozuna, J.F.T. *et al.* (2019) 'Glucose-restriction increases *Trichomonas vaginalis* cellular damage towards HeLa cells and proteolytic activity of cysteine proteinases (CPs), such as TvCP2', *Parasitology*, 146(9), pp. 1156–1166. Available at: https://doi.org/10.1017/S0031182019000209.

Mitteregger, D. *et al.* (2012) 'High detection rate of *Trichomonas vaginalis* in benign hyperplastic prostatic tissue', *Medical Microbiology and Immunology*, 201(1), pp. 113–116. Available at: https://doi.org/10.1007/s00430-011-0205-2.

Muzny, C.A., Van Gerwen, O.T. and Legendre, D. (2022) 'Secnidazole: a treatment for trichomoniasis in adolescents and adults', *Expert Review of Anti-Infective Therapy*, 20(8), pp. 1067–1076. Available at: https://doi.org/10.1080/14787210.2022.2080656.

Nagata, M. *et al.* (2023) 'No association of *Trichomonas vaginalis* seropositivity with advanced prostate cancer risk in the multiethnic cohort: a nested case-control study', *Cancers*, 15(21), p. 5194. Available at: https://doi.org/10.3390/cancers15215194.

Narayanasamy, R.K. *et al.* (2022) 'Cytidine nucleoside analog is an effective antiviral drug against Trichomonasvirus', *Journal of Microbiology, Immunology, and Infection* 55(2), pp. 191–198. Available at: https://doi.org/10.1016/j.jmii.2021.08.008.

Okumura, C.Y.M., Baum, L.G. and Johnson, P.J. (2008) 'Galectin-1 on cervical epithelial cells is a receptor for the sexually transmitted human parasite *Trichomonas vaginalis*', *Cellular Microbiology*, 10(10), pp. 2078–2090. Available at: https://doi.org/10.1111/j.1462-5822.2008.01190.x.

Ong, S.-C. *et al.* (2022) 'Identification of endosymbiotic virus in small extracellular vesicles derived from *Trichomonas vaginalis'*, *Genes*, 13(3), p. 531. Available at: https://doi.org/10.3390/genes13030531.

Pace, N.R. (2009) 'Mapping the tree of life: progress and prospects', *Microbiology and Molecular Biology Reviews*, 73(4), pp. 565–576. Available at: https://doi.org/10.1128/mmbr.00033-09.

Parent, K.N. *et al.* (2013) 'Structure of a protozoan virus from the human genitourinary parasite *Trichomonas vaginalis'*, *mBio*, 4(2), pp. e00056-13. Available at: https://doi.org/10.1128/mBio.00056-13.

Patel, E.U. *et al.* (2018) 'Prevalence and correlates of *Trichomonas vaginalis* infection among men and women in the United States', *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 67(2), pp. 211–217. Available at: https://doi.org/10.1093/cid/ciy079.

Pereira-Neves, A. and Benchimol, M. (2008) '*Trichomonas vaginalis*: *in vitro* survival in swimming pool water samples', *Experimental Parasitology*, 118(3), pp. 438–441. Available at: https://doi.org/10.1016/j.exppara.2007.09.005.

Provenzano, D., Khoshnan, A. and Alderete, J.F. (1997) 'Involvement of dsRNA virus in the protein composition and growth kinetics of host *Trichomonas vaginalis'*, *Archives of Virology*, 142(5), pp. 939–952. Available at: https://doi.org/10.1007/s007050050130.

Puente-Rivera, J. *et al.* (2017) 'The 50kDa metalloproteinase TvMP50 is a zinc-mediated *Trichomonas vaginalis* virulence factor', *Molecular and Biochemical Parasitology*, 217, pp. 32–41. Available at: https://doi.org/10.1016/j.molbiopara.2017.09.001.

Rai, A.K. and Johnson, P.J. (2019) '*Trichomonas vaginalis* extracellular vesicles are internalized by host cells using proteoglycans and caveolin-dependent endocytosis', *Proceedings of the National Academy of Sciences of the United States of America*, 116(43), pp. 21354–21360. Available at: https://doi.org/10.1073/pnas.1912356116.

Rendón-Maldonado, J.G. *et al.* (1998) '*Trichomonas vaginalis*: in vitro phagocytosis of lactobacilli, vaginal epithelial cells, leukocytes, and erythrocytes', *Experimental Parasitology*, 89(2), pp. 241–250. Available at: https://doi.org/10.1006/expr.1998.4297.

Riestra, A.M. *et al.* (2015) 'A *Trichomonas vaginalis* rhomboid protease and its substrate modulate parasite attachment and cytolysis of host cells', *PLoS Pathogens*, 11(12), p. e1005294. Available at: https://doi.org/10.1371/journal.ppat.1005294.

Ryan, C.M., de Miguel, N. and Johnson, P.J. (2011) '*Trichomonas vaginalis*: current understanding of host–parasite interactions', *Essays in Biochemistry*, 51, pp. 161–175. Available at: https://doi.org/10.1042/bse0510161.

Salas, N. *et al.* (2023) 'Role of cytoneme structures and extracellular vesicles in *Trichomonas vaginalis* parasite-parasite communication', *eLife*, 12, p. e86067. Available at: https://doi.org/10.7554/eLife.86067.

Sawaya, M. *et al.* (2024) 'Sexually and non-sexually transmitted infections and the risk of prostate cancer: Results from the EPICAP study', *Cancer Medicine*, 13(1), p. e6841. Available at: https://doi.org/10.1002/cam4.6841.

Secor, W.E. (2012) '*Trichomonas vaginalis:* treatment questions and challenges', *Expert review of anti-infective therapy*, 10(2), pp. 107–109. Available at: https://doi.org/10.1586/eri.11.159.

Simmonds, P. *et al.* (2024) 'Changes to virus taxonomy and the ICTV Statutes ratified by the International Committee on Taxonomy of Viruses (2024)', *Archives of Virology*, 169(11), p. 236. Available at: https://doi.org/10.1007/s00705-024-06143-y.

Skerk, V. *et al.* (2004) 'The role of unusual pathogens in prostatitis syndrome', *International Journal of Antimicrobial Agents*, 24 Suppl 1, pp. S53-56. Available at: https://doi.org/10.1016/j.ijantimicag.2004.02.010.

Sobel, R. and Sobel, J.D. (2015) 'Metronidazole for the treatment of vaginal infections', *Expert Opinion on Pharmacotherapy*, 16(7), pp. 1109–1115. Available at: https://doi.org/10.1517/14656566.2015.1035255.

Stevens, A. *et al.* (2021) 'Atomic Structure of the *Trichomonas vaginalis* double-stranded RNA virus 2', *mBio*, 12(2), pp. e02924-20. Available at: https://doi.org/10.1128/mBio.02924-20.

Sutton, M. *et al.* (2007) 'The prevalence of *Trichomonas vaginalis* infection among reproductive-age women in the United States, 2001-2004', *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America*, 45(10), pp. 1319–1326. Available at: https://doi.org/10.1086/522532.

Tai, J.H. and Ip, C.F. (1995) 'The cDNA sequence of *Trichomonas vaginalis* virus-T1 double-stranded RNA', *Virology*, 206(1), pp. 773–776. Available at: https://doi.org/10.1016/s0042-6822(95)80008-5.

Twu, O. *et al.* (2013) '*Trichomonas vaginalis* exosomes deliver cargo to host cells and mediate host:parasite interactions', *PLoS Pathogens*, 9(7), p. e1003482. Available at: https://doi.org/10.1371/journal.ppat.1003482.

Wartoña, A. and Honigberg, B.M. (1979) 'Structure of trichomonads as revealed by scanning electron microscopy', *The Journal of Protozoology*, 26(1), pp. 56–62. Available at: https://doi.org/10.1111/j.1550-7408.1979.tb02732.x.

World Health Organization (2020) *Trichomoniasis*. Available at: https://www.who.int/news-room/fact-sheets/detail/trichomoniasis (Accessed: 25 July 2024).

Yang, H.-Y. *et al.* (2022) 'Association between trichomoniasis and prostate and bladder diseases: a population-based case-control study', *Scientific Reports*, 12(1), p. 15358. Available at: https://doi.org/10.1038/s41598-022-19561-2.

Zhang, Z. *et al.* (2023) '*Trichomonas vaginalis* adhesion protein 65 (TvAP65) modulates parasite pathogenicity by interacting with host cell proteins', *Acta Tropica*, 246, p. 106996. Available at: https://doi.org/10.1016/j.actatropica.2023.106996.

CHAPTER 1

Differential Drug Susceptibility Across Trichomonasvirus Species Allows for Generation of Varied Isogenic Clones of *Trichomonas vaginalis*

Carrie A. Hetzel[1,2], Akua A. Appah-Sampong[1,3], Austin R. Hurst-Manny[1,2], Max L. Nibert[1,2,*]

[1]Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

[2]Program in Virology, Division of Medical Sciences, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA

[3]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

Corresponding author

Published manuscript can be accessed at:

https://doi.org/10.3390/pathogens13090733

Abstract

Trichomonas vaginalis (Tvag) is a sexually transmitted human pathogen that is commonly infected with strains of one or more of five known species of Trichomonas vaginalis viruses (TVVs), members of genus *Trichomonasvirus*. TVVs are thought not to have an extracellular phase to their lifecycle and instead to be transmitted vertically from mother to daughter cells. As a result, generation of isogenic virus-positive and virus-negative sets of Tvag clones has been a major barrier to studying interactions between TVVs and their host. Nucleoside analog 2′-C-methylcytidine (2CMC) has been recently reported to clear trichomonads of infections with TVV1, TVV2, and TVV3. We used 2CMC to treat a panel of Tvag isolates that collectively harbor at least one representative strain of each TVV species and thereby provided evidence that infections with TVV4 and TVV5 can also be cleared by 2CMC. Furthermore, our results suggest a newly identified difference in drug susceptibility between TVV species. We took advantage of these susceptibility difference to generate isogenic sets of Tvag clones harboring different combinations of the five TVV species. These results provide both new insight into differences between these species and new avenues for generating tools to study the potential roles of TVVs in Tvag biology.

Introduction

Trichomonasviruses constitute a genus of double-stranded (ds)RNA viruses, genus *Trichomonasvirus*, comprising five currently recognized species: *Trichomonasvirus vagiprimus*, *Trichomonasvirus vagisecundus*, *Trichomonasvirus vagitertius*, *Trichomonasvirus vagiquartus*, and *Trichomonasvirus vagiquintus* (https://ictv.global/taxonomy accessed on 10 August 2024) (Tai and Ip, 1995; Bessarab *et al.*, 2000, 2011; Goodman, Freret, *et al.*, 2011; Manny *et al.*, 2022) These viruses are commonly referred to as Trichomonas vaginalis viruses (TVVs), and members of the five individual species are referred to as strains of TVV1, TVV2, TVV3, TVV4, and TVV5. They were previously classified in the family *Totiviridae* but were recently moved to a new family *Pseudototiviridae*, along with several other genera of dsRNA viruses that infect either fungi or protozoa (https://ictv.global/taxonomy accessed on 10 August 2024). Trichomonasviruses persistently infect the pathogenic protozoan *Trichomonas vaginalis* (abbreviated here as Tvag), an obligate human parasite and the causative agent of trichomoniasis, which tends to be more highly symptomatic in the female urogenital tract (Mercer and Johnson, 2018). Trichomoniasis is the most common nonviral sexually transmitted infection (STI) worldwide (World Health Organization, 2020). In 2020, for example, there were estimated to be 156 million new cases per year in people aged 15–49 (World Health Organization, 2020).

Trichomonasviruses are thought to increase the severity of trichomoniasis by inducing increased pelvic inflammation (Fichorova *et al.*, 2012). Clinically, infection by TVV-containing trichomonads is associated with higher degrees of inflammation and more severe disease progression (Fraga *et al.*, 2012). Cervical epithelial cells show

higher degrees of interferon-mediated inflammation when infected with TVV-containing parasites or purified TVV1 virions ex vivo. These cells can sense TVVs via toll-like receptor 3 (TLR3), which recognizes dsRNA. TLR3 triggers a proinflammatory cascade through interferon regulatory factor 3 (IRF3) and NFkB pathways (Fichorova *et al.*, 2012). The resulting increase in inflammation is thought to contribute to premature delivery during pregnancy, susceptibility to other sexually transmitted pathogens such as HIV and HPV, and other complications (Passmore, Jaspan and Masson, 2016). The inflammatory response may conceivably benefit the parasite by providing a microenvironment in the urogenital tract that promotes its growth, survival, and/or transmission.

Unlike well-known human viruses, TVVs are thought not to have an extracellular phase in their lifecycle and instead are transmitted endogenously from mother to daughter cells as the parasite divides (Goodman, Ghabrial, *et al.*, 2011). Studying the effect of TVVs on the parasite host and the human super host has thus been historically difficult, as an uninfected Tvag isolate cannot simply be infected with virus by exogenous means in the laboratory. Previous studies have relied on comparisons between virus-positive and virus-negative Tvag isolates, but differences in the genetic backgrounds of these host trichomonads could well have confounded the results. Moreover, multiple TVV species can coinfect the same trichomonad, adding difficulty to the study of individual viruses (Khoshnan and Alderete, 1993; Bessarab *et al.*, 2000; Goodman, Ghabrial, *et al.*, 2011; Rivera *et al.*, 2017; Bokharaei-Salim *et al.*, 2020). Most studies to date have been performed using Tvag isolates that were singly infected with TVV1, singly infected with TVV3, or multiply infected with two or more TVV species (Fichorova *et al.*, 2012; Parent

et al., 2013; Ong *et al.*, 2022). To date, no isolate has been reported to be singly infected with either TVV4 or TVV5.

Generation of isogenic clones by clearing trichomonads of TVV infection has also proved challenging, as these viruses persist long term in the parasite, even after six months or more of serial passaging (Goodman, Ghabrial, *et al.*, 2011). Furthermore, trichomonasviruses are thought to have few if any cytopathic effects on the protozoan. Combined, the various findings suggest to us that there is an evolutionary benefit for the protozoan to maintain the virus at tolerable levels. However, how this détente between virus and host is established and maintained, and what benefits these viruses might indeed provide to the host, have yet to be determined. Careful studies of isogenic clones of the parasite with and without the virus should help us to uncover such hypothesized benefits of the virus–host interactions.

A recent study by Narayanasamy et. al. (Narayanasamy *et al.*, 2022) demonstrated that the nucleoside analogue 2′-C-methylcytidine (2CMC) is effective at reducing abundance of representative TVV1, TVV2, and TVV3 strains, and succeeded in using this strategy to generate cured clones of Tvag isolate T1c1, which had previously harbored TVV1. Here, we expand on the work done by those investigators and demonstrate that 2CMC is effective also against representative strains of TVV4 and TVV5. In addition, we elucidate species-specific differences in 2CMC susceptibility and use these differences to generate isogenic sets of singly infected, multiply infected, and uninfected clones of the same Tvag parent isolate.

Materials and Methods

Tvag Isolates and Cell Culture

This study used five Tvag isolates: JH37A#2, JH191A#4, JH32A#4, JH162A#4, and RU357. These isolates were purchased from the American Type Culture Collection (ATCC) and stored at −80 °C until use. Upon thawing, isolates were passaged every 1–3 days in Diamond's Modified Media (DMM) (Fouts and Kraus, 1980). Isolates were screened for the presence of *Mycoplasma hominis* via quantitative PCR (qPCR) as described previously (Fürnkranz, Henrich and Walochnik, 2018), and all isolates used in this study were found to be negative. To ensure a clonal population, approximately 20 trichomonads were plated in 10 mL of 3% agar in DMM a 6-well culture plate. Individual trichomonads were then allowed to form colonies for 3–4 days and three colonies were transferred to liquid media upon becoming visible in soft agar. Resulting clones were labeled as A/B/C (e.g., JH37A#2-A) This process was then repeated to further ensure clonal purity. Resulting clones were designated A1/B1/A2/B2, etc. All experiments were performed using trichomonads in the exponential growth phase, which was confirmed by visual inspection using phase contrast microscopy.

RNA Extraction, cDNA Synthesis, and qPCR

Total RNA was isolated using Trizol Reagent (Thermo Fisher Scientific, Waltham, MA, USA) per manufacturer's instructions. Resulting RNA was then incubated at 95 °C for 1 min before being placed immediately on ice to separate RNA strands. 200 µL of RNA was then used for cDNA synthesis using LunaScript RT SuperMix Kit (New England Biolabs, Ipswich, MA, USA). Resulting cDNA was next used as input for a quantitative

(q)PCR reaction using Luna Universal qPCR Master Mix (New England Biolabs, Ipswich, MA, USA). Cycling conditions were as follows: 95 °C for 60 s, followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. 96-well plates were run in an Applied Biosystems 7500 Fast Real Time PCR System (Invitrogen, Waltham, MA, USA). Detection primers are listed in Table S1. Actin was used as a housekeeping control, and relative RNA expression was quantified using the $2^{-\Delta\Delta Ct}$ method for normalizing gene expression.

Drug Treatment Assays

In all assays, 5×10^5 trichomonads were cultured in 10 mL DMM in a closed 15 mL conical tube. Cells were treated with 0–10 µL of 10 mM 2CMC dissolved in DMSO (final concentration 0–10 µM 2CMC) and supplemented with DMSO as a vehicle control to a final concentration of 0.1% DMSO. For time course experiments, cultures were passaged every 48 h by adding 1 mL of existing culture to 9 mL of fresh DMM and adding 2CMC and DMSO as needed to match starting concentrations. At each passage, 1 mL of culture was pelleted by centrifugation and resuspended in 1 mL of Trizol Reagent for RNA isolation, as described in Section 2.2. For experiments involving varying 2CMC concentration, trichomonads were pelleted and resuspended in Trizol Reagent after 24 h

Generation of Isogenic Clones

After eight days of drug treatment, both treated and control cultures were plated in soft agar to generate clonal populations using the protocol described in Section 2.1. Liquid cultures were then passaged approximately six times before RNA was extracted to

determine if viral RNA was present as described in Section 2.2. Clearance was defined a relative RNA expression of less than 0.1% of starting levels.

Results

Treatment with 2CMC Is Effective at Reducing RNA Expression by Strains of All Five TVV Species, but to Varying Degrees

In this study, we used Tvag isolates JH37A#2, JH191A#4, JH32A#4, JH162A#4, and RU357. Each isolate was clonally purified by picking colonies in soft agar twice serially to ensure a more homogeneous starting population of Tvag cells (clones designated A1, A2, B1, B2, etc.). RT-qPCR was then used to validate that spontaneous clearance of virus had not occurred. In doing so, we showed that JH37A#2-B1 and JH191A#4-B1 are singly infected with TVV1, JH32A#4-A1 is singly infected with TVV3, JH162A#4-A1 is multiply infected with TVV1, TVV2, TVV3, and TVV5, and RU357-B1 is multiply infected with TVV1, TVV2, TVV4, and TVV5 (Figure S1). Strains of the different TVV species present in each isolate are summarized in Table 1.1.

Table 1.1. TVV species present in Tvag isolates, verified by RT-qPCR (Figure S1).

Tvag Isolate	TVV1	TVV2	TVV3	TVV4	TVV5
JH37A#2-B1	+	-	-	-	-
JH191A#4-B1	+	-	-	-	-
JH32A#4-A1	-	-	+	-	-
JH162A#4-A1	+	+	+	-	+
RU357-B1	+	+	-	+	+

To date, 2CMC has been reported to inhibit strains of TVV1, TVV2, and TVV3, but has not been tested on strains of TVV4 and TVV5 (Narayanasamy *et al.*, 2022). To corroborate and expand on these results, we treated Tvag isolates JH37A#2-B1, JH191A#4-B1, JH32A#4-A1, JH162A#4-A1 and RU357-A1 with 10 μM 2CMC or an equivalent amount of DMSO as a vehicle control. After eight days of passaging the trichomonads in media supplemented with 2CMC or DMSO, RNA was extracted, and viral RNA expression was measured by RT-qPCR. Across all of these Tvag isolates, RNA abundance from each TVV strain was reduced by greater than 99.8% relative to the respective control culture, including for TVV4 and TVV5 strains, which had not been shown before (Figure 1.1). Interestingly, reduction in viral RNA abundance seemed to differ between TVV species. In particular, the TVV1 and TVV2 strains in these Tvag isolates appeared to be less susceptible to 2CMC treatment at 48 h post-treatment than did the TVV3, TVV4, and TVV5 strains.

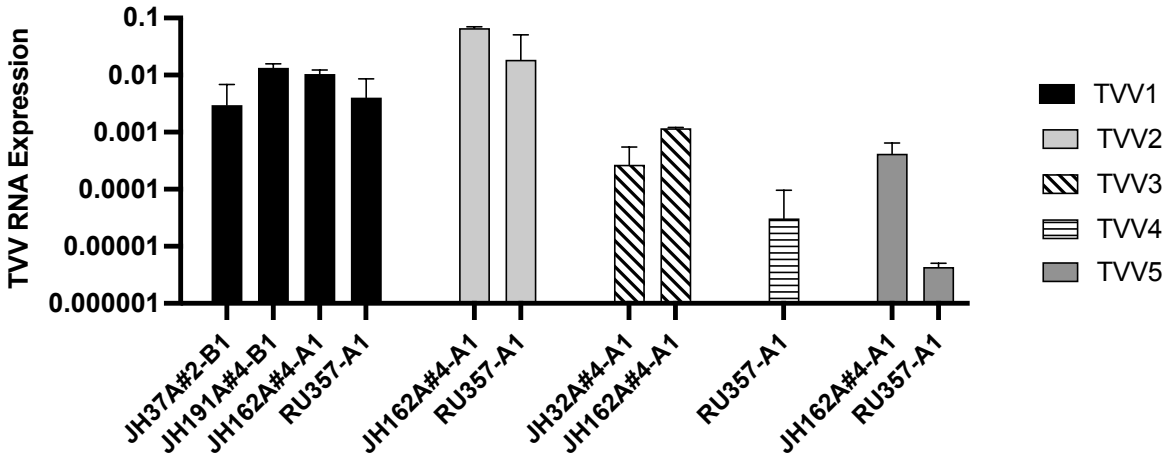


Figure 1.1. Treatment with 2CMC is effective at reducing RNA from strains of all five TVV species. Trichomonads were treated with 10 μM 2CMC for two days and relative viral RNA abundance was quantified using RT-qPCR. Expression values were calculated relative to a DMSO vehicle control and normalized to actin RNA expression as a housekeeping gene. Bars represent the mean of three technical replicates and error bars represent the standard deviation.

Susceptibility to Clearance by 2CMC Varies by TVV Species

To further investigate these apparent differences in 2CMC susceptibility, we passaged two of our singly infected Tvag isolates in 2CMC-containing media for eight days, monitoring viral RNA expression every two days (Figure 1.2A). In doing so, we saw that clearance of TVV3 RNA occurred much more quickly than clearance of TVV1 RNA. Moreover, TVV1 RNA was still detectable after eight days, while TVV3 RNA abundance had been reduced to below our limit of detection (cycle threshold (Ct) value > 35) by day 6. We also performed the same eight-day time course experiment using multiply infected Tvag isolate JH162A#4-A1 and saw that TVV3 RNA and TVV5 RNA reached undetectable levels by day 4, while TVV1 RNA and TVV2 RNA were still detectable at day 8, indicating a lower degree of susceptibility to inhibition by 2CMC on the part of TVV1 and TVV2 (Figure 1.2B). These findings are consistent with those in Figure 1.1 and were further corroborated by experiments with varying 2CMC concentrations (Figure 1.2C). The results in Figure 1.2 with multiply infected isolate JH162A#4-A1 are particularly interesting because monitoring reduction of TVV RNA abundance from different species in the same isolate eliminated the possibility of host differences confounding comparisons between isolates.

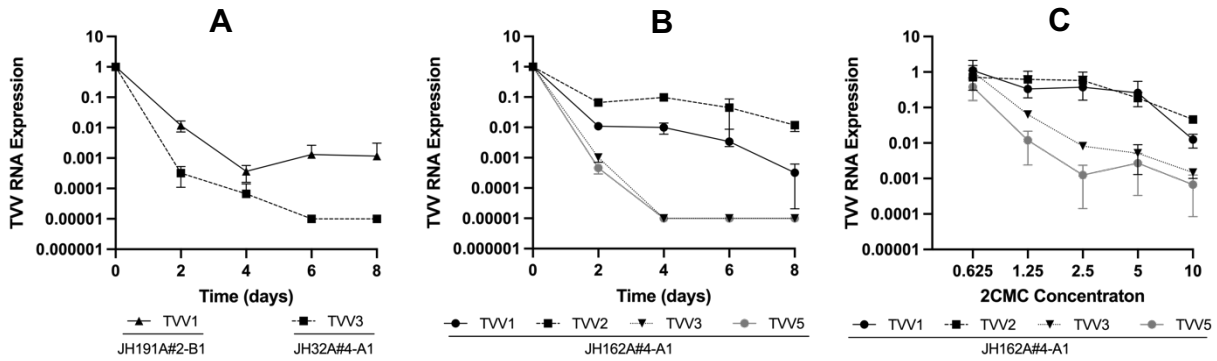


Figure 1.2. Susceptibility to 2CMC treatment varies between TVV species. Cells from (**A**) singly infected Tvag isolates JH191A#4-B1 and JH32A#4-A1 and (**B**) multiply infected isolate JH162A#4-A1 were passaged in media containing 10 μM 2CMC for eight days with timepoints every 48 h. (**C**) Cells from multiply infected isolate JH162A#4-A1 were incubated in media containing 0–10 μM 2CMC for 24 h. Relative viral RNA abundance in each sample was quantified using RT-qPCR. Expression values were calculated relative to a DMSO vehicle control and normalized to actin RNA expression as a housekeeping gene. Datapoints represent the mean of three technical replicates and error bars represent the standard deviation.

To investigate if 2CMC susceptibility was consistent between strains of the same TVV species, we performed eight-day time course experiments with all four of the Tvag isolates that are infected with TVV1 (see Table 1.1). These four strains of TVV1 exhibited similar clearance rates, and TVV1 RNA was still detectable at day 8 for each of them (Figure 1.3A). We also compared the effect of varying 2CMC concentration on the two Tvag isolates that are singly infected with TVV1 (see Table 1.1) and saw little or no apparent difference in 2CMC susceptibility between these two TVV1 strains (Figure 1.3B), providing further evidence that 2CMC susceptibility is largely consistent across different strains of the same TVV species.

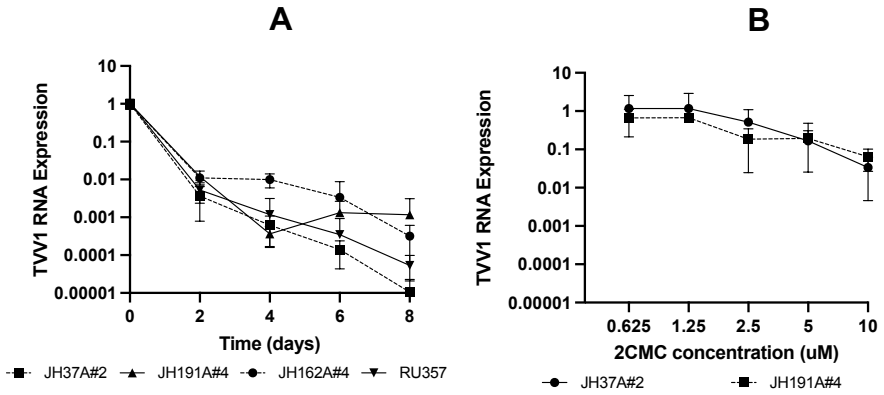


Figure 1.3. Susceptibility to 2CMC is similar between strains of the same TVV species. (**A**) Cells from the TVV1-infected isolates JH37A#2-B1, JH191A#4-B1, JH162A#4-A1, and RU357-B1 were passaged in media containing 10 μM 2CMC for eight days with timepoints every 48 h. (**B**) Cells from the singly TVV1-infected isolates JH37A#2-B1 and JH191A#4-B1 were incubated in media containing 0–10 μM 2CMC for 24 h. Relative viral RNA abundance in each sample was quantified and displayed as described for Figure 1.2.

Because TVV RNA was still present in some Tvag isolates at day 8 post-treatment, we anticipated that viral rebound would occur when 2CMC was removed at day 4. Furthermore, we were interested to see if the apparent differences in susceptibility between strains of different TVV species would affect the magnitude of rebound. When Tvag isolate JH37A#2-B1 (singly infected with TVV1) was tested in this manner, we saw that viral RNA rebounded to starting levels within two days after 2CMC removal (Figure 1.4A), despite viral RNA abundance having been reduced to ~0.2% of starting levels at day 4. On the other hand, when this experiment was performed with Tvag isolate JH32A#4-A1 (singly infected with TVV3), we saw that viral RNA again rebounded within two days after 2CMC removal, but to only ~7% of starting levels (Figure 1.4B), despite viral RNA abundance having been reduced in this case to ~0.02% of starting levels (nearly the limit of detection) at day 4. Thus, consistent with the greater magnitude of TVV3 clearance from Tvag isolate JH32A#4-A1 at day 4 of 2CMC treatment, TVV3 RNA

abundance did not rebound to the same degree in that isolate as did TVV1 RNA in isolate JH37A#2-B1. Sanger sequencing of the coding regions of both virus populations after rebound revealed no mutations, suggesting that rebound was not the result of outgrowth of resistant mutants, but rather the outgrowth of the wild-type viruses that had persisted in the face of 2CMC treatment for four days.

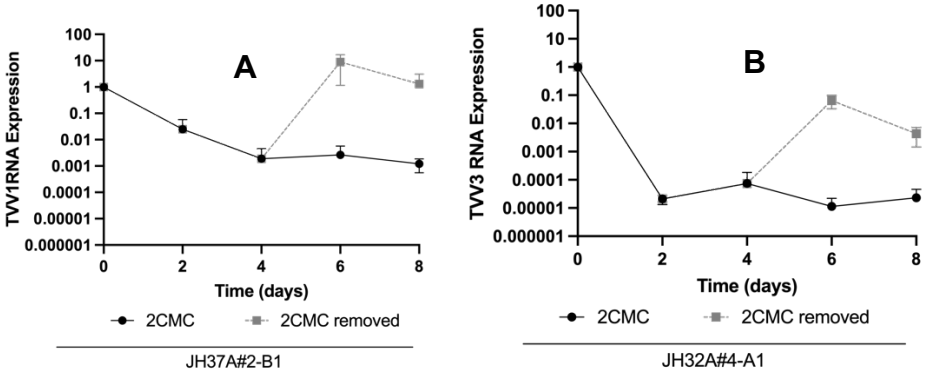


Figure 1.4. TVV infection persists after four days of 10 μM 2CMC treatment and rebounds upon 2CMC removal. Trichomonads from (**A**) singly TVV1-infected Tvag isolate JH37A#2-B1 and (**B**) singly TVV3-infected Tvag isolate JH32A#4-A1 were passaged in media containing 10 μM 2CMC for four days. At day 4 for each isolate, 2CMC was removed from one series of samples and continued for the other. Relative RNA abundance at each timepoint was quantified and displayed as described for Figure 1.2.

Isogenic Clones Generated after 8 Days of 2CMC Treatment

We wanted to generate isogenic uncured and cured clones of Tvag isolates, as was done by Narayanasamy et al. (Narayanasamy *et al.*, 2022) to use for future experiments. However, due to our results regarding time to clearance and viral rebound, we passaged our isolates in 2CMC or DMSO, as a vehicle control, for eight days before plating for Tvag colonies in soft agar to obtain clonal populations. To ensure these clones were cured of TVV, we passaged each approximately six times in culture after selection and monitored viral RNA expression. In doing so, we were able to generate isogenic sets

from singly infected isolates JH37A#2-B1, JH191A#4-B1, and JH32A#4-A1 (Figure 1.5A–C). For these experiments, we defined a cured strain as having viral RNA abundance reduced to 0.1% or less of starting levels, i.e., reduced by three orders of magnitude or more, following six or more passages in culture. All of the clones we derived from 2CMC treatment of isolates JH191A#4-B1 and JH32A#4-A1 were thereby considered cured. Interestingly, one of the four 2CMC-treated clones from JH37A#2-B1 (clone 7 in Figure 1.5A) remained infected with TVV1, providing further evidence that viral RNA can persist in at least some Tvag cells in a culture even after eight days of 2CMC treatment. Sanger sequencing of this clone revealed four synonymous mutations in the coding region of TVV1, but these mutations did not appear to affect 2CMC susceptibility relative to the original virus (Figure S2), suggesting to us that these mutations were not responsible for the persistence of TVV1 in this clone.

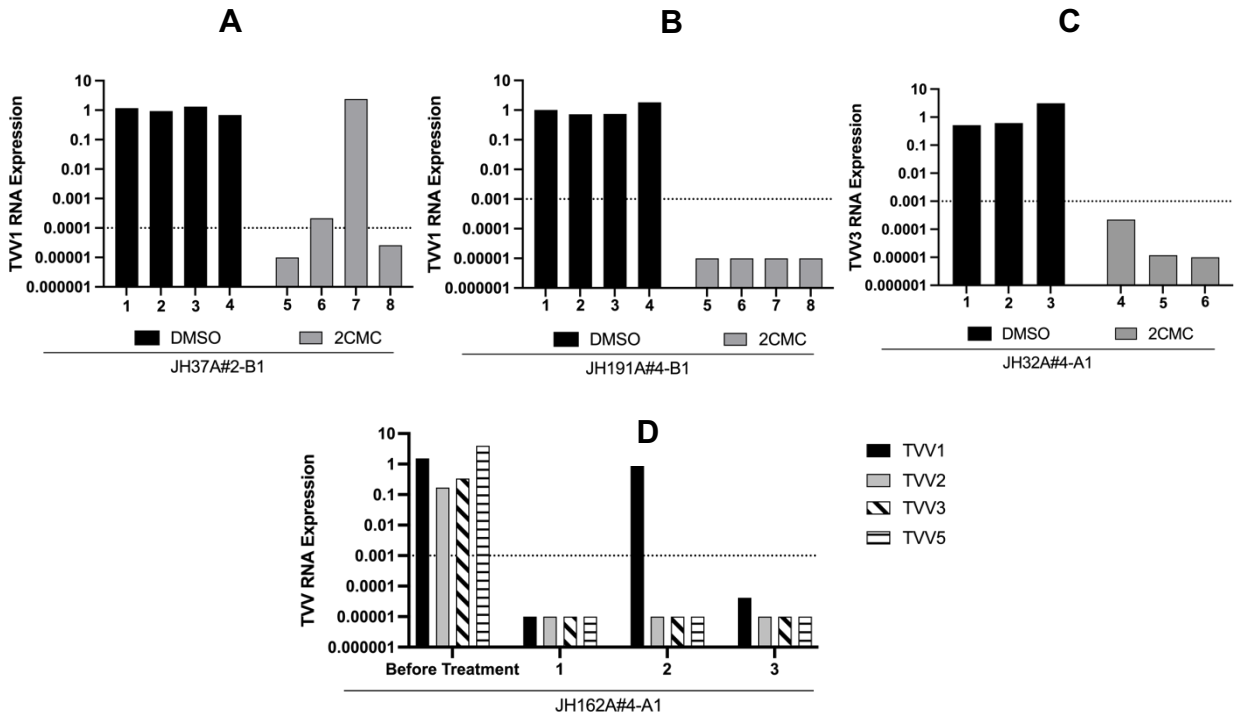


Figure 1.5. Treatment with 10 μM 2CMC allows for clearance of virus from singly and multiply infected isolates. Trichomonads from isolates (**A**) JH37A#2-B1, (**B**) JH191A#4-B1, (**C**) JH32A#4-A1, and (**D**) JH162A#4 were passaged in media containing 10 μM 2CMC or the equivalent amount of DMSO for eight days. At day 8, trichomonads were plated in soft agar and allowed to form colonies. Three or four individual colonies for each isolate and treatment condition were screened for the presence of trichomonasviruses by RT-qPCR. Relative viral RNA abundance in each sample was quantified as described for Figure 1.1. Dotted line represents viral RNA abundance below which clones were considered to be cured. In (D), the samples labeled “Before Treatment” reflect analysis of the parent Tvag isolate (not DMSO treated and cloned in parallel to the 2CMC treatment yielding clones 1–3).

Discussion

Trichomonasviruses constitute an understudied genus, despite their potential as a model system to investigate virus–host interactions in the context of long-term viral persistence in a human pathogen, *T. vaginalis*. This is in part due to the challenge of working with such viruses that appear to lack an extracellular phase in their lifecycle. The generation of virus-positive and virus-negative Tvag clones derived from the same parent isolate, to use for comparative analyses, is thus crucial. Narayanasamy et al. (Narayanasamy *et al.*, 2022) demonstrated the capacity of 2CMC to clear representative strains of TVV1, TVV2, and TVV3 and used this approach to generate uncured and cured isogenic Tvag clones. In this study, we extended those previous results to demonstrate that 2CMC is also effective in clearing representative strains of TVV4 and TVV5 from Tvag isolates, allowing for generation of a wider range of isogenic pairs of Tvag clones. Furthermore, we demonstrated the possibility of using 2CMC treatment to create isogenic sets of Tvag clones containing varied combinations of strains of the five TVV species and generated the first reported isogenic set containing singly infected, multiply infected, and uninfected Tvag clones derived from the same parent isolate (see Figure 1.5). To date, little is known about the relative impact of single vs. multiple TVV infections on the protozoan host, and so such isogenic sets of Tvag clones are essential for allowing careful investigations of these differences.

Isogenic sets of clonally purified host cell cultures are additionally useful for related studies of virus–host biology in the numerous other systems involving protozoan, fungal, or other hosts that are persistently infected by viruses that are not routinely transmitted by extracellular means, i.e., including but extending beyond trichomonasviruses and *T.*

vaginalis. Trichomonasviruses are an example of viruses that provide an intriguing model for studying the virus–host arms race in the context of long-term viral persistence. It is largely unknown how the détente between endosymbiotic viruses and their hosts is achieved and maintained, and how that détente can be broken. Moreover, because Tvag is an early diverging eukaryote (Gunderson *et al.*, 1995), understanding this détente might provide valuable information about the evolution of host defenses, as Tvag lacks many classical innate immune pathways (Carlton *et al.*, 2007) that might be expected to participate in this détente. Isogenic sets of Tvag clones harboring different combinations of trichomonasviruses may, for example, allow for differential gene expression analyses to help to uncover host biochemical pathways that might be involved in controlling TVV replication.

Our results additionally demonstrated the importance of using extended treatment of Tvag cultures with 2CMC in order to achieve viral clearance from all or nearly all Tvag cells in the cultures, followed by clonal purification of cured cells. When we removed 2CMC after four days of treatment, strains of both TVV1 and TVV3 rebounded, despite having been reduced to nearly undetectable levels in the case of TVV3 (see Figure 1.4). Sanger sequencing revealed that the coding regions of the TVV strains present in the rebounded cultures were identical to the starting sequences, indicating to us that rebound was primarily explained by outgrowth of wild-type viruses that had persisted in many of the Tvag cells after four days of 2CMC treatment, and not by outgrowth of any 2CMC-resistant mutants that may have been selected. Based on the relative levels of rebound, we interpreted that after four days of treatment, TVV1 had persisted in most Tvag cells in isolate JH37A#2-B1, whereas TVV3 had persisted in < 10% of Tvag cells in isolate

JH32A#4-A1. Similarly, strains of TVV1 that remained after eight days of 2CMC treatment in one clone each from Tvag isolates JH37A#2-B1 and JH162A#4-A1 (see Figure1.5A, D) provided further evidence for the persistence of virus in at least some Tvag cells after extended treatments with 2CMC.

Currently, there are no reported Tvag isolates that are singly infected with either TVV4 or TVV5. Further investigations might be fruitful for identifying inhibitors to which strains of particular TVV species are uniquely resistant, which would allow for the generation of Tvag clones that are singly infected with strains of each of the different species, including TVV4 and TVV5. Nucleoside analogs in general are well-known inhibitors of viral polymerases (Kamzeeva *et al.*, 2023), and thus further study of differential susceptibility of TVV strains to a larger panel of nucleoside analogs would seem to be especially promising. Because TVV1 and TVV2 strains exhibited the least susceptibility to 2CMC inhibition, they currently present a limitation to being cleared from Tvag cells of a multiply infected parent isolate before strains of the other species in that isolate have been cleared. Thus, inhibitors that are specifically more effective against TVV1 and/or TVV2 strains are especially needed to allow an even greater number of combinations of isogenic clones to be generated.

In addition to further study into the relationships between virus (TVV) and host (Tvag), as well as superhost (human), future directions of this work will be to elucidate why the apparent differences in 2CMC susceptibility between TVV species exist. According to analyses performed by Manny et. al. (Manny *et al.*, 2022), sequences from TVV2 and TVV5 strains are more similar to one another than to strains of the other species, as also are sequences from TVV3 and TVV4 strains, with TVV1 strains being

the most divergent from the others. Intriguingly, susceptibility to 2CMC treatment did not appear to replicate this pattern in our experiments. Instead, TVV1 and TVV2 strains appeared to be less susceptible to inhibition by 2CMC than did TVV3, TVV4, and TVV5 strains. No obvious patterns of sequence differences have been noted in our analyses to date to try to explain this result. We have also performed preliminary structure predictions using AlphaFold2 (Jumper *et al.*, 2021) for the viral polymerase from strains of the different TVV species, but no obvious structural differences to explain these differences in 2CMC susceptibility have yet been recognized. Of further note, apparent susceptibility to 2CMC by strains of different TVV species did not appear to correlate with starting levels of viral RNA in the Tvag parent cultures. For example, in Tvag isolate JH162A#4-A1, TVV3 appeared to have the highest relative starting level of RNA abundance, followed by TVV1, then TVV5, then TVV2 (see Figure S1).

To conclude, further work is needed in a number of important areas identified by this study as suggested above. In addition, structural determinations of the viral polymerase from the different TVV species could allow greater insight into how nucleoside analog 2CMC may interact with this protein, which is essential for viral replication. Lastly, identification of novel strains or clones of the different TVV species, with increased or decreased susceptibility to 2CMC achieved either by further sampling of naturally occurring variation among Tvag clinical isolates or by selecting 2CMC-resistant mutants from Tvag isolates in the laboratory, could also help to elucidate how 2CMC may interact with the viral polymerase and provide additional information about the differences between TVV species. It is even conceivable, though perhaps unlikely, that the effect of

2CMC occurs via a less direct mechanism, such as by affecting nucleotide substrate levels in Tvag cells, which then differentially impact the five TVV species.

Supplementary Materials

The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pathogens13090733/s1, Table S1: Primers used for RT-qPCR; Figure S1: Validation of the presence of trichomonasviruses in isolates tested; Figure S2: Susceptibility to 2CMC is not significantly different between TVV1 in parent isolate JH37A#2-B1 and TVV1 in 2CMC-treated but uncured clone JH37A#2-B1-2.

Author Contributions

Carrie A. Hetzel: Conceptualization of project, methodology design, experimental investigation, validation of results, preparation of initial manuscript, manuscript revision, acquisition of funding

Akua A. Appah-Sampong: Methodology design, experimental investigation, validation of results

Austin R. Manny: Conceptualization of project and initial support on methodology design

Max L. Nibert: Conceptualization of project, manuscript revision, project supervision acquisition of funding

Funding

This research was funded in part by National Institutes of Health (USA) Grants T32AI007245 (Program in Virology at Harvard University, C.A.H. and A.R.H.M.), R01AI132445 (M.L.N), and F31AI176739 (C.A.H.). Additional support was provided by the Program for Research in Science and Engineering (PRISE) at Harvard University (A.A.A.S.).

References

Bessarab, I.N. *et al.* (2000) 'The complete cDNA sequence of a type II *Trichomonas vaginalis* virus', *Virology*, 267(2), pp. 350–359. Available at: https://doi.org/10.1006/viro.1999.0129.

Bessarab, I.N. et al. (2011) 'Identification and characterization of a type III *Trichomonas vaginalis* virus in the protozoan pathogen Trichomonas vaginalis', *Archives of Virology*, 156(2), pp. 285–294. Available at: https://doi.org/10.1007/s00705-010-0858-y.

Bokharaei-Salim, F. *et al.* (2020) 'The first detection of co-infection of double-stranded RNA virus 1, 2 and 3 in Iranian isolates of *Trichomonas vaginalis*', *Iranian Journal of Parasitology*, 15(3), pp. 357–363. Available at: https://doi.org/10.18502/ijpa.v15i3.4200.

Carlton, J.M. *et al.* (2007) 'Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis'*, *Science (New York, N.Y.)*, 315(5809), pp. 207–212. Available at: https://doi.org/10.1126/science.1132894.

Fichorova, R.N. *et al.* (2012) 'Endobiont viruses sensed by the human host – beyond conventional antiparasitic therapy', *PLOS ONE*, 7(11), p. e48418. Available at: https://doi.org/10.1371/journal.pone.0048418.

Fouts, A.C. and Kraus, S.J. (1980) '*Trichomonas vaginalis*: reevaluation of its clinical presentation and laboratory diagnosis', *The Journal of Infectious Diseases,* 141(2), pp. 137–143. Available at: https://doi.org/10.1093/infdis/141.2.137.

Fraga, J. *et al.* (2012) 'Species typing of Cuban *Trichomonas vaginalis* virus by RT-PCR, and association of TVV-2 with high parasite adhesion levels and high pathogenicity in patients', *Archives of Virology,* 157(9), pp. 1789–1795. Available at: https://doi.org/10.1007/s00705-012-1353-4.

Fürnkranz, U., Henrich, B. and Walochnik, J. (2018) '*Mycoplasma hominis* impacts gene expression in *Trichomonas vaginalis'*, *Parasitology Research,* 117(3), pp. 841–847. Available at: https://doi.org/10.1007/s00436-018-5761-6.

Goodman, R.P., Freret, T.S., *et al.* (2011) 'Clinical isolates of *Trichomonas vaginalis* concurrently infected by strains of up to four *Trichomonasvirus* species (Family Totiviridae)', *Journal of Virology*, 85(9), pp. 4258–4270. Available at: https://doi.org/10.1128/JVI.00220-11.

Goodman, R.P., Ghabrial, S.A., et al. (2011) '*Trichomonasvirus*: a new genus of protozoan viruses in the family Totiviridae', *Archives of Virology,* 156(1), pp. 171–179. Available at: https://doi.org/10.1007/s00705-010-0832-8.

Gunderson, J. et al. (1995) 'Phylogeny of trichomonads inferred from small-subunit rRNA sequences', *The Journal of Eukaryotic Microbiology,* 42(4), pp. 411–415. Available at: https://doi.org/10.1111/j.1550-7408.1995.tb01604.x.

Jumper, J. et al. (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

Kamzeeva, P.N. et al. (2023) 'Recent advances in molecular mechanisms of nucleoside antivirals', *Current Issues in Molecular Biology,* 45(8), pp. 6851–6879. Available at: https://doi.org/10.3390/cimb45080433.

Khoshnan, A. and Alderete, J.F. (1993) 'Multiple double-stranded RNA segments are associated with virus particles infecting Trichomonas vaginalis', *Journal of Virology,* 67(12), pp. 6950–6955. Available at: https://doi.org/10.1128/JVI.67.12.6950-6955.1993.

Manny, A.R. et al. (2022) 'Discovery of a novel species of trichomonasvirus in the human parasite *Trichomonas vaginalis* using transcriptome mining', *Viruses*, 14(3), p. 548. Available at: https://doi.org/10.3390/v14030548.

Mercer, F. and Johnson, P.J. (2018) '*Trichomonas vaginalis:* pathogenesis, symbiont interactions and host cell immune responses', *Trends in Parasitology,* 34(8), pp. 683–693. Available at: https://doi.org/10.1016/j.pt.2018.05.006.

Narayanasamy, R.K. et al. (2022) 'Cytidine nucleoside analog is an effective antiviral drug against Trichomonasvirus', *Journal of Microbiology, Immunology, and Infection,* 55(2), pp. 191–198. Available at: https://doi.org/10.1016/j.jmii.2021.08.008.

Ong, S.-C. et al. (2022) 'Identification of endosymbiotic virus in small extracellular vesicles derived from *Trichomonas vaginalis'*, *Genes*, 13(3), p. 531. Available at: https://doi.org/10.3390/genes13030531.

Parent, K.N. et al. (2013) 'Structure of a protozoan virus from the human genitourinary parasite *Trichomonas vaginalis'*, *mBio*, 4(2), pp. e00056-13. Available at: https://doi.org/10.1128/mBio.00056-13.

Passmore, J.-A.S., Jaspan, H.B. and Masson, L. (2016) ‘Genital inflammation, immune activation and risk of sexual HIV acquisition’, *Current opinion in HIV and AIDS*, 11(2), pp. 156–162. Available at: https://doi.org/10.1097/COH.0000000000000232.

Rivera, W.L. et al. (2017) ‘Detection and molecular characterization of double-stranded RNA viruses in Philippine *Trichomonas vaginalis* isolates’, *Journal of Microbiology, Immunology and Infection*, 50(5), pp. 669–676. Available at: https://doi.org/10.1016/j.jmii.2015.07.016.

Tai, J.H. and Ip, C.F. (1995) ‘The cDNA sequence of Tri*chomonas vaginalis* virus-T1 double-stranded RNA’, *Virology*, 206(1), pp. 773–776. Available at: https://doi.org/10.1016/s0042-6822(95)80008-5.

World Health Organization (2020) *Trichomoniasis*. Available at: https://www.who.int/news-room/fact-sheets/detail/trichomoniasis (Accessed: 25 July 2024).

CHAPTER 2

Trichomonasviruses Modulate Survival, Adherence, and Gene Expression of their Protozoan Host

Carrie A. Hetzel[1,2], Akua A. Appah-Sampong[1,3], Austin R. Hurst-Manny[1,2], Max L. Nibert[1,2,*]

[1]*Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA*

[2]*Program in Virology, Division of Medical Sciences, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA*

[3]*Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA*

**Corresponding author*

Abstract

Trichomonasviruses are a genus of double-stranded RNA (dsRNA) viruses that persistently infect the obligate human protozoan parasite *Trichomonas vaginalis* (Tvag). These viruses are not thought to have an extracellular lifecycle and are instead believed to be transmitted vertically from mother to daughter as the parasite divides. The persistent nature of these viruses suggests that there may be an evolutionary benefit to the parasite to maintain these viruses at tolerable levels. However, little is known about beneficial impacts of trichomonasviruses on their host. In this study, we used isogenic virus-positive and virus-negative Tvag clones derived from the same parent isolate to examine effects of TVV infection on parasite growth, survival, and adherence to human cells. In doing so, we provide evidence that TVV+ Tvag clones have an increased ability to survive in harsh conditions and a decreased ability to adhere to human cells. To better understand the mechanism underlying this difference, we performed differential gene expression analysis and identify 179 differentially expressed genes between cured and uncured clones from three isogenic sets. Of these hits, we identified TVAGG3_0365170, which we call TvMyb4, as a consistently upregulated hit in TVV+ clones and demonstrate that knockdown of this gene leads to decreased RNA abundance of representative strains of both TVV1 and TVV3. We also provide evidence that knockdown of TvMyb4 leads to decreased parasite survival in harsh conditions, suggesting that increased TvMyb4 expression is contributing to the survival advantage of TVV+ clones. This work provides new insights into the interplay between trichomonasviruses and their parasite host and suggests a potential evolutionary explanation for long-term viral persistence in the parasite.

Introduction

Trichomonasviruses constitute a genus of double-stranded RNA (dsRNA) viruses that persistently infect the obligate human parasite, *Trichomonas vaginalis* (Tvag) (Goodman, Ghabrial, *et al.*, 2011). Five viral species are currently recognized within this genus: *Trichomonasvirus vagiprimus*, *Trichomonasvirus vagisecundus, Trichomonasvirus vagitertius*, *Trichomonasvirus vagiquintus*, and *Trichomonasvirus vagiquintus* (https://ictv.global/taxonomy) (Tai and Ip, 1995; Bessarab *et al.*, 2000, 2011; Goodman, Freret, *et al.*, 2011; Manny *et al.*, 2022). These viruses are commonly referred to as Trichomonas vaginalis viruses (TVVs) and the members of each of the five species designated as strains of TVV1, TVV2, TVV3, and TVV4, and TVV5, respectively. Genus *Trichomonasvirus* was previously assigned to family *Totiviridae* but recently moved to family *Pseudototiviridae* (https://ictv.global/taxonomy) alongside a subset of other dsRNA viruses infecting protozoa or fungi. More than half of Tvag isolates are estimated to be infected with trichomonasviruses (Goodman, Ghabrial, *et al.*, 2011), and strains of more than one species of trichomonasvirus commonly coinfect the same isolate (Khoshnan and Alderete, 1993; Bessarab *et al.*, 2000; Goodman, Freret, *et al.*, 2011; Rivera *et al.*, 2017; Bokharaei-Salim *et al.*, 2020).

Tvag is the causative agent of trichomoniasis, the most common nonviral sexually transmitted infection (STI) worldwide, with an estimated 156 million cases per year in people aged 15–49 (World Health Organization, 2020). In addition to standard STI symptoms, trichomoniasis is associated with infertility, low birth weight, premature delivery, and increased acquisition of other STIs including HPV and HIV (Fichorova, 2009; Mercer and Johnson, 2018). Trichomonasviruses are thought to increase the severity of

trichomoniasis by increasing pelvic inflammation; the prevailing model is that dsRNA from trichomonasviruses is released by dead and dying parasites, which is then sensed by vaginal epithelial cells via Toll-like receptor 3 (TLR3), triggering an immune cascade through the NFκB and IRF3 pathways (Fichorova *et al.*, 2012). Moreover, infection with trichomonads harboring trichomonasviruses has been shown to lead to more severe disease progression in humans (Fraga *et al.*, 2012). As an important caveat, however, these studies have been performed using Tvag isolates harboring strains of TVV1, either in isolation or in combination with strains of other TVV species, so little is known about the effects of other trichomonasvirus species on the human inflammatory response.

Although TVVs are thought to be largely non-cytopathic to their protozoan host (Goodman, Ghabrial, *et al.*, 2011), they have been linked to changes in certain Tvag phenotypes. For example, TVVs have been shown to increase surface expression of the Tvag P270 protein, which has been linked to decreased cytotoxicity to human cells (Alderete *et al.*, 1986). TVVs have also been shown to modulate cysteine protease expression, which has been hypothesized to promote parasite immune evasion and survival in the human superhost (Provenzano, Khoshnan and Alderete, 1997). The presence of TVVs has been additionally suggested to increase parasite adherence to human cells, particularly in the case of TVV2 (Fraga *et al.*, 2012). Lastly, TVVs have been suggested to modulate composition of extracellular vehicles (EVs) secreted by the protozoan (Rada *et al.*, 2022), and these EVs have been implicated in altering parasite adherence to human cells (Kochanowsky *et al.*, 2024). The detailed molecular mechanisms by which these phenotypic changes occur in association with TVV infection remain an open question in the field.

Further investigation into the effects of trichomonasviruses on their protozoan host has been hindered by limitations of this virus–host system. These viruses persist long term in Tvag and are thought not to have an extracellular phase to their lifecycle. Instead, TVVs are thought to be transmitted vertically as the protozoan divides, or possibly during a yet-to-be-observed means of Tvag mating (Goodman, Ghabrial, *et al.*, 2011). Historically, the lack of an extracellular phase has been a challenge to studying impacts of trichomonasviruses on their host, as generation of isogenic Tvag clones with and without virus has been difficult. Recent work in the field, however, has allowed for generation of isogenic Tvag clones harboring different combinations of virus derived from the same parent isolate (Narayanasamy *et al.*, 2021; Hetzel *et al.*, 2024). In this study, we used isogenic sets of cured and uncured Tvag clones to investigate how representative strains of TVV1 and TVV3 affect parasite growth, survival, and adherence. Furthermore, we performed differential gene expression analyses using isogenic clones derived from three Tvag isolates to identify patterns of gene expression that may be involved in phenotypic differences between the TVV+ and TVV- clones. Finally, we investigated the role of one of these genes in maintaining trichomonasvirus persistence and modulating Tvag survival and adherence phenotypes.

Materials and Methods

Human and Tvag cell culture

Human Vk2/E6E7 vaginal epithelial cells were passaged in keratinocyte-serum free media (KSFM) (Gibco) supplemented with 0.1 ng/mL human recombinant epidermal growth factor (rEGF) (Gibco) and 0.05 mg/mL bovine pituitary extract (BPE) (Gibco), as described previously (Fichorova, Rheinwald and Anderson, 1997). Cells were grown at

37 °C and passaged every 2–3 days. Trichomonads were grown in Diamond's Modified Media (DMM) (Fouts and Kraus, 1980) at 37 °C and passaged every 1–3 days. Tvag isolates used in this study were screened for *Mycoplasma hominis* by quantitative PCR (qPCR) as described previously (Fürnkranz, Henrich and Walochnik, 2018) and determined to be negative for this endosymbiont.

Generation of isogenic sets

Trichomonads were grown in DMM at 37 °C and passaged every 1–3 days. Isogenic sets were generated as described previously (Hetzel *et al.*, 2024). Briefly, Tvag isolates JH37A#2, JH191A#4, and JH32A#4 were thawed from frozen stocks and resuspended in DMM. To generate clonal populations, trichomonads from each isolate were plated in 3% agar in DMM and individual colonies were selected. This process was repeated twice to generate clones JH37A#2-B1, JH191A#4-A1, and JH32A#4-A1. RNA from each clone was extracted and reverse transcription (RT) qPCR was used to verify the presence of strains of TVV1, TVV1, and TVV3, respectively. These clones were then treated with 10 μM 2CMC for 8 days before repeating the soft agar cloning process to generate three cured and three uncured clones from each isolate. Trichomonads from each clone were frozen and aliquots thawed as needed.

Fluorescent labeling of trichomonads

Approximately $1x10^5$ trichomonads were harvested in logarithmic growth phase and resuspended in 1mL PBS. 2 μL of CytoTrace Red CMTPX dye (AAT Bioquest) was added to the resuspended trichomonads. Trichomonads were then incubated for 30 min

at 37 °C and cultures were protected from light during incubation. Cells were washed twice in PBS to remove unabsorbed dye. After washing, trichomonads were resuspended in 200 µL PBS and recounted using a hemocytometer to ensure appropriate cell numbers for downstream applications.

Growth and survival rate measurement

To measure growth in axenic culture, $1x10^4$ trichomonads from clone JH37A#2-A1-1 (TVV1+) and clone JH37A#2-A1-5 (TVV-) were resuspended in 1 mL DMM and incubated in closed microcentrifuge tubes. Cells were counted using a hemocytometer at timepoints over a 96 h period. Trypan blue (Sigma Aldrich) staining was used to distinguish live cells from dead cells. Three independent trials were performed, each with three replicates of each clone. To measure Tvag growth in co-culture with human cells, approximately $1x10^5$ Vk2/E6E7 cells were seeded in 48 well culture plates and allowed to grow to confluence for 2–3 days at 37 °C. Before performing co-culture experiments, media was removed and replaced with 200 µL of fresh KSFM that was prewarmed to 37 °C. Trichomonads were fluorescently labeled as described above and approximately $1x10^5$ trichomonads were added to the wells containing Vk2/E6E7 cells. Plates were imaged by fluorescence microscopy at various times over 72 h, with most measurements taken between 12 h and 24 h to capture the parasite growth after the initial lag period. Cell counts were determined by ImageJ software. To measure parasite survival, trichomonads from three cured and three uncured clones of isolates JH37A#2 and JH32A#4 were harvested in logarithmic growth phase and resuspended in DMM to a concentration of 2.5×10^6 cells per mL. Cells were then diluted 1:10 in Trypan blue dye

(Sigma Aldrich) and placed in a disposable hemocytometer (INCYTO). Three slides were prepared for each clone. The hemocytometer openings were sealed with petroleum jelly to prevent liquid evaporation. Live and dead cell counts were recorded at 0, 2, 4, and 6 h after loading.

Adherence assays

Approximately $1x10^5$ Vk2/E6E7 cells were seeded in 48 well culture plates and allowed to grow to confluence for 2–3 days at 37 °C. Before performing adherence assays, media was removed and replaced with 200 µL of fresh KSFM that was prewarmed to 37 °C. Approximately $1x10^5$ trichomonads were harvested in logarithmic growth phase and labeled with CytoTrace Red CMTPX dye as described above. After washing, trichomonads were resuspended in PBS and approximately $1x10^3$ trichomonads were added to the wells containing Vk2/E6E7 cells. In initial assays, plates were then incubated for 30 min at 37 °C to allow trichomonads to adhere to human cells. Due to decreased culture health after transfection, trichomonads that had undergone transfection of either targeting or control siRNA were incubated for 2 h to promote adherence. After incubation, trichomonads were imaged using fluorescence microscopy. Tvag-human cell layer was then washed twice by removing media, adding 1 mL of PBS and gently pipetting up and down four times before removing PBS. 200 µL of KSFM was added back to wells to keep cells alive and imaging was repeated. Cell counts were determined using ImageJ software.

Differential gene expression analysis

Fresh aliquots of three TVV+ and TVV- clones of JH37A#2-B1, JH191A#2-A1, and JH32A#4-A1 were thawed and passaged 2–3 times ensure parasite health after thawing. Cells were harvested at logarithmic growth phase and total RNA was isolated using TRIzol (Thermo Fisher Scientific) according to manufacturer's instructions. RNA quality and presence/absence of TVV strains in each clone were checked by gel electrophoresis and RT-qPCR, respectively. Remaining RNA was immediately frozen at -20°C until shipping to Quick Biology. Total RNA integrity and quantity was analyzed at this stage using an RNA ScreenTape apparatus (Agilent). Library prep was performed including Poly-A tail selection to enrich for mRNA. Illumina sequencing was performed, generating 150 base pair paired-end reads with a mean depth of 30 million reads per sample. FastQ files were generated and imported in Galaxy (usegalaxy.org) for processing. Reads were trimmed using TrimGalore! and mapped to the 2022 annotated G3 reference genome from TrichDB (https://trichdb.org/common/downloads/release-64/TvaginalisG32022/) using Bowtie2. Bam files from Bowtie2 were imputed into Salmon quant to calculate transcript abundance and DEseq2 was used to performed differential gene expression analysis. Volcano plots displaying the p-adjusted value and log2FC were generated in RStudio using the ggplot2 package.

RT-qPCR quantification of gene expression

RT-qPCR was used to measure both expression of host genes and abundance of viral RNA. Cells were harvested at logarithmic growth phase and total RNA was isolated using TRIzol (Thermo Fisher Scientific) according to manufacturer's instructions. RNA

was then incubated at 95 °C for 1 min to denature viral dsRNA before being placed on ice to prevent strands from reannealing. Approximately 200 ng of this RNA was used for cDNA synthesis using Luna RT Supermix (New England Biolabs), per manufacturers instructions. 1 µL of cDNA was then amplified using the Luna Universal qPCR Master Mix kit (New England Biolabs). Reactions were run in 96-well plates using the Applied Biosystems 7500 Fast Real Time PCR System. The cycling conditions were as follows: 95 °C for 60 s followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. Viral and host gene expression was quantified using the $2^{-\Delta\Delta Ct}$ method. Graphpad Prism software was used to generate graphs and perform statistical analysis for validation and knockdown experiments.

Gene annotation

Significant hits (p-adjusted <0.05) from each experiment for annotation were compiled. Gene lists were first entered intro TrichDB (TrichDB.org) and gene type (i.e. protein coding, pseudogene, rRNA) were recorded. Hits designated as rRNA were discarded. Product description and ortholog group noted by TrichDB was recorded. Amino acid sequences from each hit were run through NCBI BLAST and HHPred to identify potential homologs based on sequence and predicted structure.

siRNA knockdowns

Knockdown of TvMyb4 was achieved by transfection of targeting siRNA using Lipofectamine RNAiMax Reagent (Thermo Fisher Scientific). To do so, $1x10^5$ trichomonads were resuspended in 900uL of prewarmed DMM and added to one well of

a 24-well culture plate. 1.2 µL of Lipofectamine RNAiMax reagent was diluted in 48.8 µL of OptiMEM media (Thermo Fisher scientific). siRNA was resuspended in OptiMEM media to a concentration of 4 µM in 50 µL. Diluted Lipofectamine RNAiMax and siRNA were combined and incubated at room temperature for 10 min before adding to the culture plate. Each knockdown was performed in triplicate in three separate wells. Control reactions were also performed in triplicate with a nontargeting control siRNA. Sequences for both targeting and nontargeting control siRNAs can be found in Table S2. After 4–6 hours, Tvag cultures were transferred into 9 mL of fresh DMM. After 36 h, trichomonads were either harvested for RNA isolation and RT-qPCR or used for survival and adherence assays as described above.

Results

Tvag clones harboring TVVs display a survival advantage in harsh conditions over cured Tvag clones

To investigate effects of trichomonasviruses on their protozoan host, we first examined the growth rates of cured and uncured isogenic Tvag clones derived from the same parent isolate. We first measured growth rate of cured and uncured clones of isolate JH37A#2-A1 in axenic culture and found no apparent difference in growth rate between cured and uncured clones (Figure 2.1A). We then cocultured these clones with human vaginal epithelial cells from the Vk2/E6E7 cell. Again, we found no apparent difference in growth rate between cured and uncured clones of JH37A#2-A1 in this experiment. While we could not conclude that there is no difference in growth rate between cured and uncured clones, these results indicated to us that growth rate may not be a key

evolutionary benefit to the protozoan conferred by the presence of trichomonasviruses. However, these experiments were performed in ideal conditions for Tvag growth, so we could not ascertain whether we would see differences in harsher conditions.

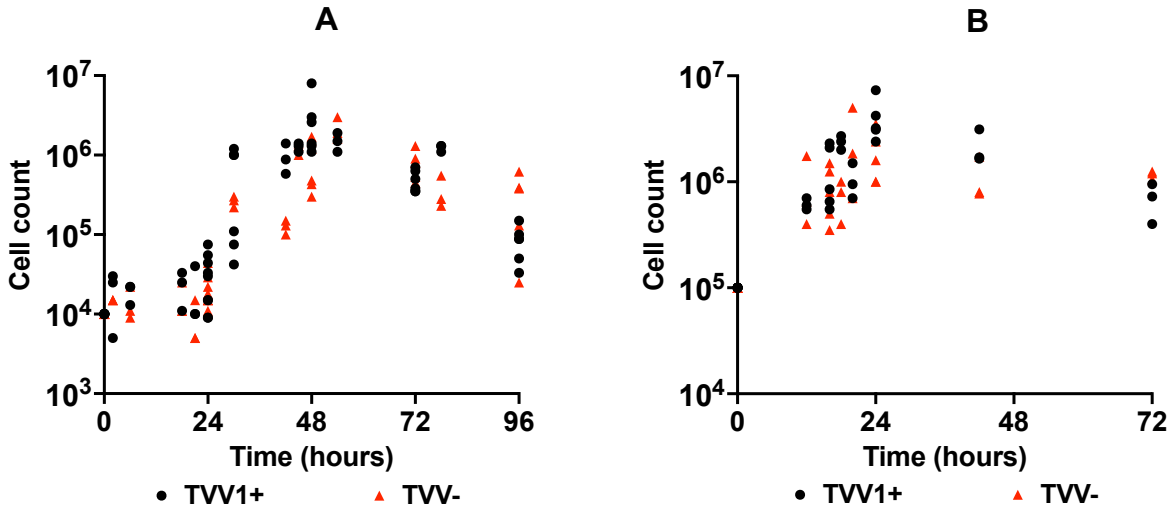


Figure 2.1: Cured and uncured Tvag clones do not demonstrate a clear difference in growth rate in axenic culture or in co-culture with human cells. (A) Trichomonads from cured and uncured clones of isolate JH37A#2-A1 were grown in axenic culture from up to 96 h. Data are compiled from three independent experiments, each with three replicates. (B) Trichomonads from cured and uncured clones of isolate JH37A#2-A1 were co-cultured with human Vk2 cells in axenic culture from up to 72 h. Data are compiled from three independent experiments, each with two to three replicates.

To determine if the presence of trichomonasviruses effects host survival in non-ideal conditions, we examined the survival rate of nutrient-deprived trichomonads. To do so, we resuspended trichomonads in trypan blue dye and measured their viability over 6 h. We first performed this assay with three cured and three uncured clones derived from isolate JH37A#2-B1 and saw that trichomonads harboring TVV1 appeared to die more slowly than trichomonads from cured clones (Figure 2.2A). The use of three clones for each condition allowed us to minimize the chance that this survival difference was a result of differences between individual clones rather than due to the presence of virus. We

repeated this experiment with cured and uncured clones derived from isolate JH32A#4-A1 and saw that trichomonads harboring TVV3 also appeared to die more slowly than cured clones (Figure 2.2B). These results suggest to us that trichomonasviruses may confer a survival advantage to their host when in non-ideal conditions.

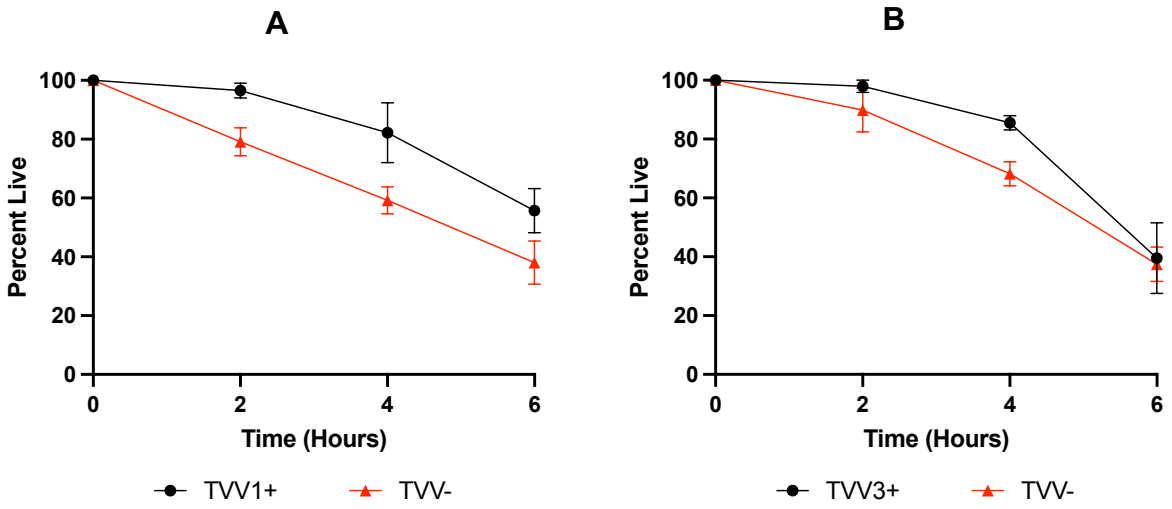
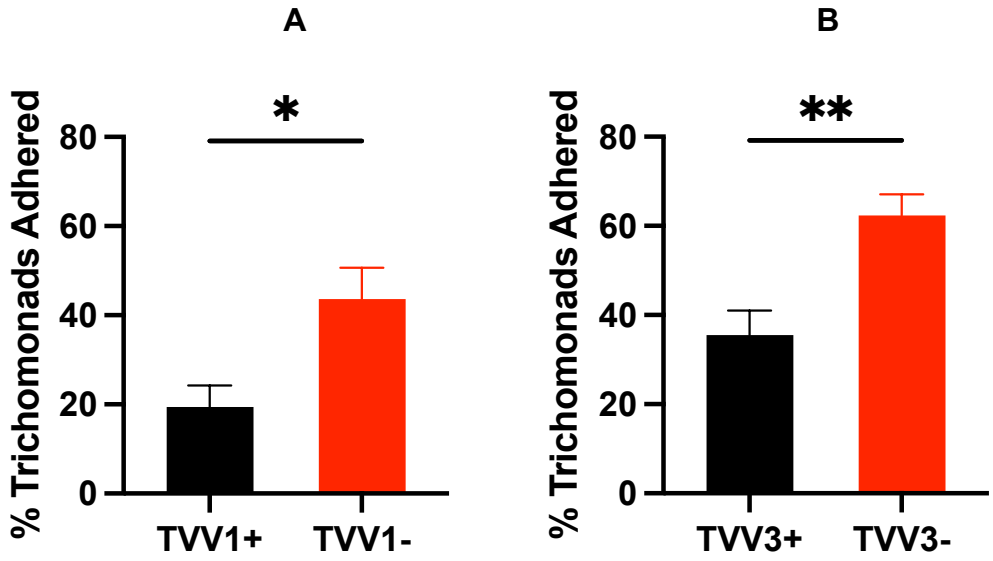


Figure 2.2: Uncured Tvag clones show a survival advantage over cured clones. Trichomonads from cured and uncured clones of Tvag isolates (A) JH37A#2-B1 and (B) JH32A#4-A1 were resuspended in trypan blue dye and placed and observed under phase contrast microscopy in a sealed hemocytometer over six hours. Live/dead percentages were determined using trypan blue staining. Data points represent the mean and SEM from three independent experiments.

Tvag adherence to human cells is increased in cured clones

In addition to growth and survival, we also wanted to investigate the impact of trichomonasviruses on the ability of their host to adhere to human cells. We first incubated fluorescently labeled trichomonads from cured and uncured clones of JH37A#2-B1 with Vk2/E6E7 cells for 30 min. We imaged the culture plates before and after washing with PBS to determine what percentage of trichomonads from each clone were still adhered after washing. Over two independent trials, each with three replicates, an average of 19.5% of trichomonads harboring TVV1 remained attached to the human cells, as

compared to an average of 43.6% of cured trichomonads from the same isolate, indicating that cured clones adhered more strongly to human cells than uncured clones (Figure



2.3A). Under the same experimental setup, an average of 35.6% of trichomonads from isolate JH32A#4-B1 harboring TVV3 remained attached, while an average of 62.3% of cured trichomonads from the same isolate remained attached, corroborating the results from our initial experiment (Figure 2.3B).

Figure 2.3: Cured Tvag clones adhere more strongly to human cells than uncured clones. Trichomonads from cured and uncured Tvag clones of (A) JH37A#2-B1 and (B) JH32A#4-A1 were plated on human Vk2 cells and incubated for 30 minutes before washing with PBS. Images were captured before and after washing and percent adherence was calculated. Data for each isolate are compiled from two independent experiments, each with three replicates. Bars represent mean of all six data points and error bars represent standard error of mean (SEM). Unpaired t-tests were used to determine statistical significance with the following thresholds: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

Isogenic clones with and without TVVs have different patterns of gene expression

To better understand the mechanisms underlying these survival and adherence differences, we performed differential gene expression analysis using RNA from cured

and uncured clones of three isolates, JH37A#2-B1, JH191A#4-A1, and JH32A#4-A1. The uncured clones of these isolates harbored TVV1, TVV1, and TVV3, respectively. We performed bulk RNA sequencing (RNA-seq) on all three sets and performed differential gene expression analysis by mapping reads to the *Trichomonas vaginalis* G3 2022 Genome Sequence and Annotation generated by Carlton et al. (https://trichdb.org/trichdb/app/record/dataset/DS_e6874bf420#Busco), revealing a total of 179 significantly up- and downregulated genes (Figure 2.4). From isogenic clones derived from isolate JH37A#2-B1, we found that 23 genes were upregulated in clones harboring TVV1, while 5 were downregulated in these clones. From isogenic clones derived from isolate JH191A#4-A1, 26 genes were upregulated in clones harboring TVV1, while 8 were downregulated. From isolate JH32A#4-A1, we found that 112 genes were upregulated in clones harboring TVV1, while 5 were downregulated.

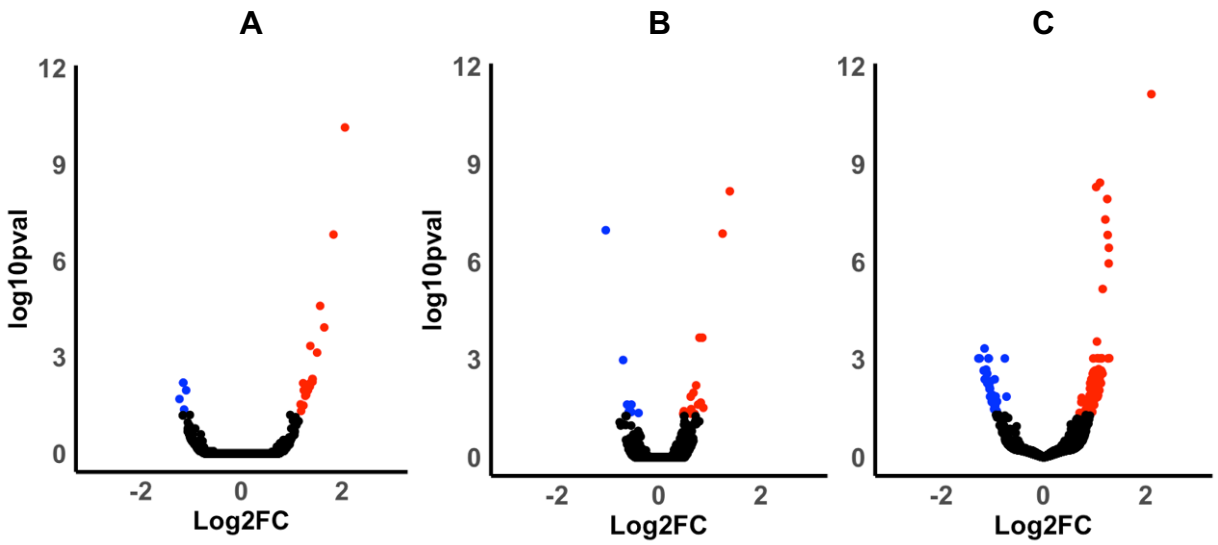


Figure 2.4: RNA-seq reveals a number of differentially expressed genes between cured and uncured Tvag clones. Total RNA was harvested from three cured and three uncured clones of Tvag isolates (A) JH37A#2-B1, (B) JH191A#4-A1, (C) JH32A#4-A1 and bulk RNA-seq was performed. Reads were mapped to the Tvag genome using Bowtie2 and differential gene expression analysis was performed using DESeq2, using the cured clones as the control group. Genes highlighted in red represent the subset that was significantly upregulated in TVV+ isolates. Genes highlighted in blue represent the subset of hits that was significantly upregulated in TVV+ isolates. Hits were considered to be significant if the p-adjusted value was <0.05.

While a complete genome sequence is available for *T. vaginalis*, the vast majority of its genes have little annotation regarding the structure or function of proteins encoded. To characterize our RNA-seq hits, we performed bioinformatic analysis using TrichDB, NCBI-BLAST, and HHPred. Using TrichDB, we were able to classify hits by ortholog group and determined our hits came from 58 different ortholog groups, two of which were represented in more than one of our experiments (Table 2.1). Detailed information from Trich DB can be found in Appendix 3 of this dissertation (Tables A3.1, A3.2, A3.3). Of these groups, only OG6_105664 was represented in all three datasets. Members of this group are largely characterized by the presence of ATPases Associated with Diverse

Cellular Activities (AAA) domains. Three of these hits (TVAGG3_0496300, TVAGG3_0539280, TVAGG3_0539290) are predicted to be endonuclease proteins by TrichDB. Because much of the Tvag genome is still unannotated, little else is available about these hits. An additional ortholog group, OG6_100129 was represented in the hits from our JH37A#2 and JH191A#4 experiments, suggesting a potential difference in gene expression between TVV1 and TVV3 harboring parasites.

Table 2.1: Hits from overlapping ortholog groups. Two ortholog groups were represented in more than one dataset. The lefthand column lists the gene names and the right lists the datasets in which these hits were found.

OG6_105664	
TVAGG3_0496280	JH37A#2, JH32A#4
TVAGG3_0496300	JH37A#2
TVAGG3_0539280	JH37A#2, JH32A#4
TVAGG3_0539290	JH37A#2
TVAGG3_0707390	JH37A#2, JH191A#4
OG6_100129	
TVAGG3_0365170	JH37A#2, JH191A#4
TVAGG3_0192490	JH37A#2

TvMyb4 is upregulated in clones harboring the virus, and knockdown of TvMyb4 leads to reduced viral RNA abundance

Of these hits, none were significantly up- or downregulated in all three sets. However, one gene, TVAGG3_0365170, was significantly upregulated in uncured clones JH37A#2 and JH191A#4. Expression of this gene does appear to be somewhat upregulated in

uncured clones of JH32A#4, although this hit does not reach statistical significance (p-adjusted = 0.11). When we validated this result using RT-qPCR, RNA expression of this gene was indeed upregulated in cured clones from all three isogenic sets (Figure 2.5). Based on the annotations from TrichDB and the homology predictions from NCBI-BLAST and HHPred, we hypothesize that TVAGG3_0365170 encodes a transcription factor of the Myb family. Specifically, it appears to be an R2R3 type transcription factor, due to its two Myb recognition element (MRE) domains (Wang *et al.*, 2020). Three Myb-like proteins in the Tvag genome have thus far been reported (Tsai, Liu and Tai, 2002; Ong *et al.*, 2007; Smith *et al.*, 2011), so we will refer to this protein as TvMyb4, following established nomenclature.

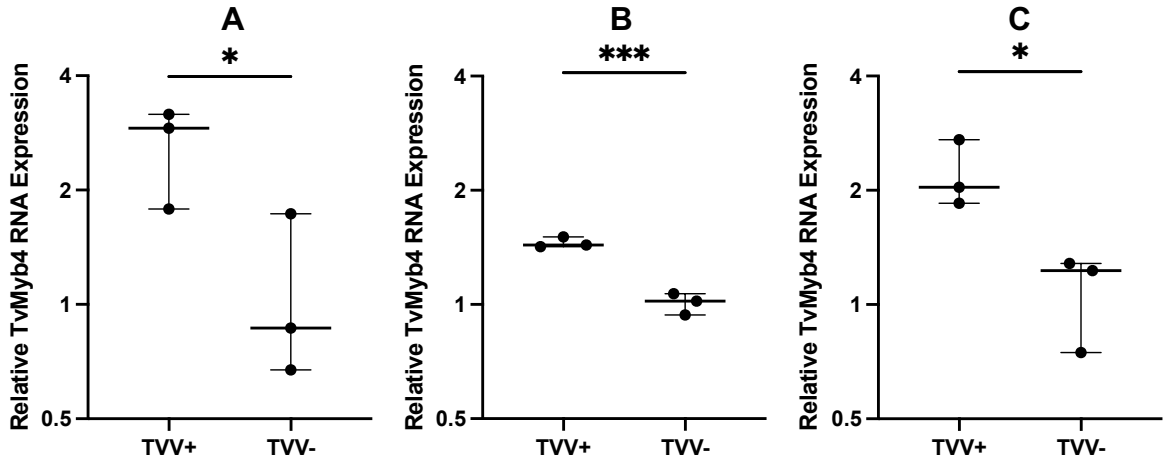


Figure 2.5: Differential Expression of TvMyb4 is confirmed by RT-qPCR. Bulk RNA from three cured and three uncured clones of (A) JH37A#2-B1, (B) JH191A#4-A1, (C) JH32A#4-A1 was harvested and RT-qPCR was used to detect expression of TvMyb4. Relative RNA expression was calculated using the $2^{-\Delta\Delta Ct}$ method using actin as a housekeeping control. Median and 95% confidence interval are displayed. Unpaired t-tests were used to determine statistical significance with the following thresholds: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$.

To further establish the relationship between trichomonasviruses and TvMyb4, we performed knockdowns of TvMyb4 and measured viral RNA abundance in isolates JH37A#2-B1 and JH32A#4-A1. To this end, we delivered siRNA molecules targeting TvMyb4 by lipofectamine transfection and measured the expression of both TvMyb4 and viral RNA (Figure 2.6). In trichomonads from Tvag isolate JH37A#2 treated with the targeting siRNA, TvMyb4 expression was reduced by approximately 36% and TVV1 RNA expression was reduced by approximately 39%, as compared to trichomonads receiving nontargeting siRNA. Similarly, in trichomonads from Tvag isolate JH32A#4 treated with the targeting siRNA, TvMyb4 expression was reduced by approximately 40% and TVV1 RNA expression was reduced by approximately 67%, as compared to trichomonads receiving nontargeting siRNA. The reduction in viral RNA abundance following TvMyb4 knockdown suggests to us that TvMyb4 may have a role in promoting viral replication and maintenance of long-term infection.

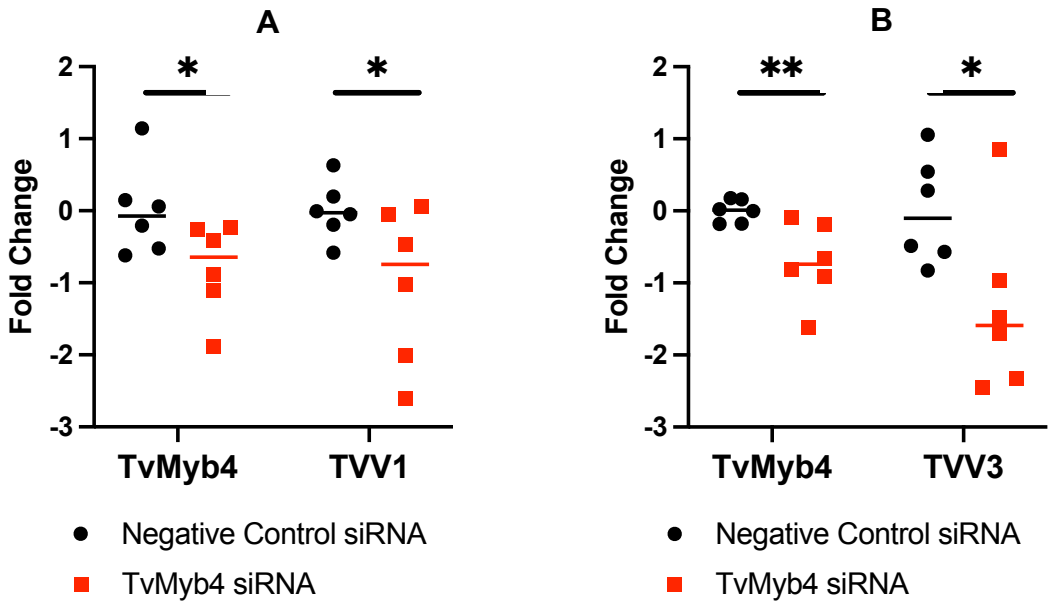


Figure 2.6: Knockdown of TvMyb4 reduces viral RNA expression. Trichomonads from uncured clones of (A) JH37A#2-B1, and (B) JH32A#4-A1 were transfected with either siRNA targeting TvMyb4 or a nontargeting control siRNA. After 40 hours, total RNA was harvested, and RT-qPCR was used to measure viral abundance. Data were combined from two independent experiments, each with three replicates. Relative RNA expression was calculated using the $2^{-\Delta\Delta Ct}$ method using TvActin as a housekeeping control. Individual and median fold changes are represented graphically. Paired t-tests were used to determine statistical significance with the following thresholds: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$. Lines represent the median fold change.

Tvag phenotypic differences are linked to TvMyb4 expression

To investigate the relationship between TvMyb4 and Tvag survival, we performed siRNA knockdowns of TvMyb4 and repeated the assays described above to measure survival rate and adherence. To ensure that any depletion in viral RNA that may be caused by TvMyb4 would not confound our results, we used a cured clone of parent isolate JH32A#4 which had been previously cured of TVV3. In doing so, we saw that trichomonads treated with siRNA targeting TvMyb4 were less able to survive in our experimental setup than trichomonads treated with non-targeting control siRNA (Figure

2.7A). We also repeated our adherence assays following TvMyb4 knockdown, but in this case did not see a statistically significant difference in Tvag adherence between trichomonads that received siRNA targeting TvMyb4 and trichomonads receiving control siRNA, potentially suggesting that these phenotypes may occur through different mechanisms.

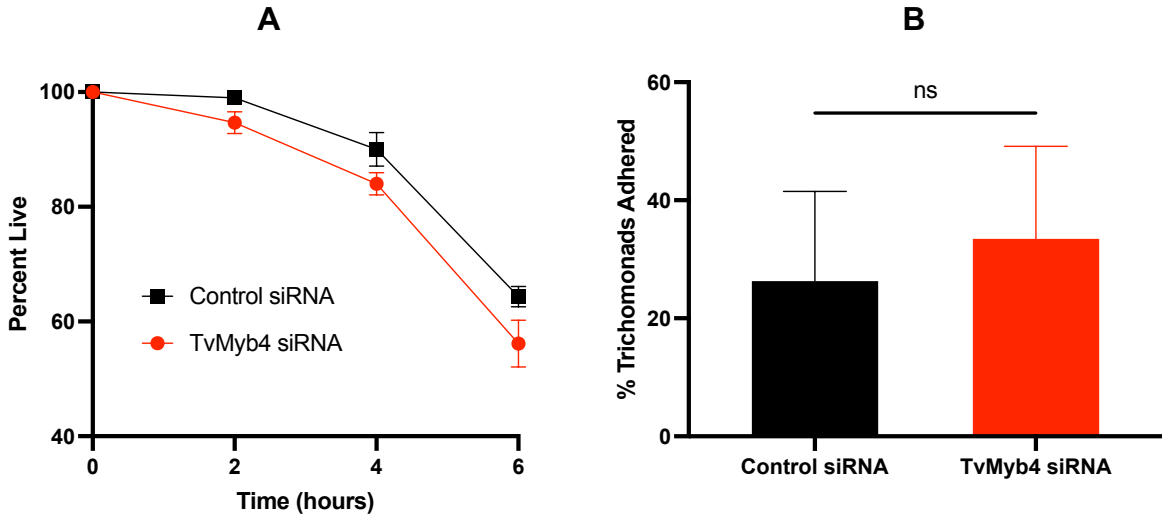


Figure 2.7: Knockdown of TvMyb4 leads to decreased survival, but no significant effect on adherence Trichomonads from a cured clone of JH32A#4-A1 were treated with either siRNA targeting TvMyb4 or a control nontargeting siRNA. After incubation for approximately 40 hours, cell survival and adherence were measured as follows: (A) Parasite viability was observed under phase contrast microscopy over six hours. Live/dead percentages were determined using trypan blue staining. Data points represent the mean and SEM from three independent experiments. (B) Trichomonads were plated on human Vk2 cells and incubated for 2 hours before washing with PBS. Images were taken before and after washing and percent adherence was calculated. Data for each isolate are compiled from two independent experiments, each with three replicates. Mean and SEM are displayed.

Discussion

To date, the impact of trichomonasviruses on Tvag biology has been an open field of investigation. In this paper, we provide evidence that the presence of representative strains of TVV1 and TVV3 is associated with increased parasite survival in harsh conditions and decreased adherence to human cells. Further, by performing differential gene expression analysis using cured and uncured isogenic Tvag clones derived from three parent isolates, we identified TvMyb4 as a candidate hit to explain these differences in phenotype. We demonstrate that knockdown of TvMyb4 expression results in reduced viral RNA abundance of clones of TVV1 and TVV3, suggesting that TvMyb4 may be involved in maintaining trichomonasvirus persistence. Finally, we show that knockdown of TvMyb4 results in increased Tvag survival, suggesting that differential TvMyb4 expression is at least in part responsible for the phenotypic differences between cured and uncured isogenic TVV clones.

The long-term persistence of trichomonasviruses in Tvag suggests that the protozoan benefits from maintaining viral infection at a tolerable level instead of clearing the virus entirely. Our results suggest that the presence of trichomonasviruses may not confer a growth advantage to trichomonads but may support survival of the parasite in non-ideal conditions. As Tvag is transmitted sexually, the ability to withstand harsher conditions may allow for increased transmission of virus-positive trichomonads compared to virus-negative trichomonads. The apparent decrease in the ability of virus-positive trichomonads to adhere to human cells would also be consistent with a possible transmissibility advantage. Interestingly, this decrease in adherence to human cells is in contradiction to a previous report comparing infected and uninfected Tvag isolates (Fraga

et al., 2012), but these differences may be explained by the use of isogenic clones in our experiments, as gene expression differences between isolates may have confounded previous studies. Further work, including *in vivo* experiments in animal models, is needed to elucidate if trichomonasviruses do indeed increase transmissibility of the protozoa.

The mechanism by which trichomonasviruses promote parasite survival and detachment from human cells also remains to be elucidated. One possible mechanism we present is upregulation TvMyb4 expression. Our results suggest that knockdown of TvMyb4 results in decreased parasite survival in harsh conditions. Thus, it is conceivable that increased TvMyb4 expression in uncured clones may be contributing to the increased survival ability of these clones. However, the mechanism by which this happens requires further investigation. Future directions of this work will be to determine if TvMyb4 indeed functions as a transcription factor as we predict and to subsequently investigate the effects of TvMyb4 on transcriptional regulation. Additionally, the involvement of TvMyb4 in parasite adherence remains to be seen. Our results suggest that knockdown of TvMyb4 may not be sufficient to increase parasite adherence to a statistically significant degree. However, it is possible that TvMyb4 may still be involved in Tvag adherence, but that the knockdown of TvMyb4 expression was not strong enough to reveal a significant difference. Other TvMyb proteins have been associated with modulating parasite adherence to human cells (Tsai, Liu and Tai, 2002; Ong *et al.*, 2007; Jiang *et al.*, 2011), so further investigation into the role of TvMyb4 in parasite adherence is warranted. Overexpression of TvMyb4 in cured Tvag clones is a potential avenue for understanding what changes Tvag phenotypes may be attributed to TvMyb4.

Our results also suggest a potential role of TvMyb4 in maintaining long-term persistence of trichomonasviruses. Knockdown of TvMyb4 resulted in decreased viral RNA abundance from representative stains of both TVV1 and TVV3 (see Figure 2.7), suggesting to us that TvMyb4 may promote viral replication and persistence, but the mechanism of this remains to be elucidated. Further, whether or not the differentially expressed genes from OG6_105664 (see Table 2.1) are involved in maintaining trichomonasviruses at tolerable level to the parasite remain to be seen. In our experiment comparing clones of cured and uncured isolate JH37A#2-B1, three of our top four hits (See supplementary table S1) are classified as putative nucleases by TrichDB. Future directions of this work include characterization of these nucleases to determine which, if any, may be effective in cleaving viral RNA and subsequently limiting viral replication.

Differences in gene expression patterns elucidated by different TVV species, or by the different strains of the same viral species, also remain to be seen. We anticipated that patterns of differential gene expression might be more similar in cured and uncured clones of derived from isolates harboring TVV1, but our results did not reveal unique patterns of Tvag gene expression that were triggered by either TVV1 or TVV3. However, overlapping hits in experiments using isolates harboring strains of the same virus species may emerge by expanding this analysis to more isogenic sets. Further, due to isolate availability, we used isogenic sets derived from isolates that were singly infected with strains of either TVV1 or TVV3. Different hits may be identified using isogenic Tvag sets derived from parent isolates harboring strains of TVV2, TVV4, or TVV5. Similarly, these experiments used Tvag clones that were infected with a single TVV species, so repeating

this experiment with cured and uncured stains derived from multiply infected isolates may reveal information about the impact of multiple infections on the parasite host.

To conclude, our results suggest survival and adherence differences between cured and uncured isogenic Tvag clones that may provide a partial explanation for the long-term persistence of trichomonasviruses and suggest the beginnings of a mechanism for the increased survival ability of uncured Tvag clones through the identification of TvMyb4 as a differentially expressed transcription factor. Future directions of this work include elucidating how TvMyb4 may be affecting Tvag survival as well as how TvMyb4 may be involved in maintaining viral persistence. Characterization of other RNA-seq hits, including those in ortholog group OG6_105664, may also provide valuable information on host–virus interactions in this system. Repeating these experiments with additional isogenic Tvag sets may also provide more information about how different TVV species, and different strains of the same TVV species, elicit different patterns of gene expression. Finally, how these changes in gene expression impact the human superhost, including potential increased in transmissibility by Tvag clones harboring trichomonasviruses, remains to be seen.

Supplementary Materials

Significantly significant hits from out differential gene expression analysis are compiled and annotated in Appendix 3 of this text.

Author Contributions

Carrie A. Hetzel: Conceptualization of project, methodology design, experimental investigation, validation of results, data analysis, preparation of initial manuscript, manuscript revision, acquisition of funding

Akua A. Appah-Sampong: Methodology design, experimental investigation, validation of results, data analysis

Austin R. Manny: Conceptualization of project, data analysis

Max L. Nibert: Conceptualization of project, manuscript revision, project supervision acquisition of funding

Funding

This research was funded in part by National Institutes of Health (USA) Grants T32AI007245 (Program in Virology at Harvard University, C.A.H. and A.H.M.), R01AI132445 (M.L.N), and F31AI176739 (C.A.H.). Additional support was provided by the Program for Research in Science and Engineering (PRISE) at Harvard University (A.A.A.S.).

References

Alderete, J.F. *et al.* (1986) 'Phenotypic variation and diversity among *Trichomonas vaginalis* isolates and correlation of phenotype with trichomonal virulence determinants.', *Infection and Immunity*, 53(2), pp. 285–293.

Bessarab, I.N. *et al.* (2000) 'The complete cDNA sequence of a type II *Trichomonas vaginalis* virus', *Virology*, 267(2), pp. 350–359. Available at: https://doi.org/10.1006/viro.1999.0129.

Bessarab, I.N. *et al.* (2011) 'Identification and characterization of a type III *Trichomonas vaginalis* virus in the protozoan pathogen *Trichomonas vaginalis'*, *Archives of Virology*, 156(2), pp. 285–294. Available at: https://doi.org/10.1007/s00705-010-0858-y.

Bokharaei-Salim, F. *et al.* (2020) 'The first detection of co-infection of double-stranded RNA virus 1, 2 and 3 in Iranian isolates of *Trichomonas vaginalis'*, *Iranian Journal of Parasitology*, 15(3), pp. 357–363. Available at: https://doi.org/10.18502/ijpa.v15i3.4200.

Fichorova, R.N. (2009) 'Impact of *T. vaginalis* infection on innate immune responses and reproductive outcome', *Journal of reproductive immunology*, 83(1–2), pp. 185–189. Available at: https://doi.org/10.1016/j.jri.2009.08.007.

Fichorova, R.N. *et al.* (2012) 'Endobiont viruses sensed by the human host – beyond conventional antiparasitic therapy', *PLOS ONE*, 7(11), p. e48418. Available at: https://doi.org/10.1371/journal.pone.0048418.

Fichorova, R.N., Rheinwald, J.G. and Anderson, D.J. (1997) 'Generation of papillomavirus-immortalized cell lines from normal human ectocervical, endocervical, and vaginal epithelium that maintain expression of tissue-specific differentiation proteins', *Biology of Reproduction*, 57(4), pp. 847–855. Available at: https://doi.org/10.1095/biolreprod57.4.847.

Fouts, A.C. and Kraus, S.J. (1980) '*Trichomonas vaginalis*: reevaluation of its clinical presentation and laboratory diagnosis', *The Journal of Infectious Diseases*, 141(2), pp. 137–143. Available at: https://doi.org/10.1093/infdis/141.2.137.

Fraga, J. *et al.* (2012) 'Species typing of Cuban *Trichomonas vaginalis* virus by RT-PCR, and association of TVV-2 with high parasite adhesion levels and high pathogenicity in patients', *Archives of Virology*, 157(9), pp. 1789–1795. Available at: https://doi.org/10.1007/s00705-012-1353-4.

Fürnkranz, U., Henrich, B. and Walochnik, J. (2018) '*Mycoplasma hominis* impacts gene expression in *Trichomonas vaginalis'*, *Parasitology Research*, 117(3), pp. 841–847. Available at: https://doi.org/10.1007/s00436-018-5761-6.

Goodman, R.P., Freret, T.S., *et al.* (2011) 'Clinical isolates of *Trichomonas vaginalis* concurrently infected by strains of up to four *Trichomonasvirus* species (Family Totiviridae)', *Journal of Virology*, 85(9), pp. 4258–4270. Available at: https://doi.org/10.1128/JVI.00220-11.

Goodman, R.P., Ghabrial, S.A., *et al.* (2011) '*Trichomonasvirus:* a new genus of protozoan viruses in the family *Totiviridae'*, *Archives of Virology*, 156(1), pp. 171–179. Available at: https://doi.org/10.1007/s00705-010-0832-8.

Hetzel, C.A. *et al.* (2024) 'Differential drug susceptibility across trichomonasvirus species allows for generation of varied isogenic clones of *Trichomonas vaginalis'*, *Pathogens*, 13(9), p. 733. Available at: https://doi.org/10.3390/pathogens13090733.

Jiang, I. *et al.* (2011) 'Molecular basis of the recognition of the ap65-1 gene transcription promoter elements by a Myb protein from the protozoan parasite *Trichomonas vaginalis'*, *Nucleic Acids Research*, 39(20), pp. 8992–9008. Available at: https://doi.org/10.1093/nar/gkr558.

Khoshnan, A. and Alderete, J.F. (1993) 'Multiple double-stranded RNA segments are associated with virus particles infecting *Trichomonas vaginalis'*, *Journal of Virology*, 67(12), pp. 6950–6955. Available at: https://doi.org/10.1128/JVI.67.12.6950-6955.1993.

Kochanowsky, J.A. *et al.* (2024) '*Trichomonas vaginalis* extracellular vesicles up-regulate and directly transfer adherence factors promoting host cell colonization', *Proceedings of the National Academy of Sciences of the United States of America*, 121(25), p. e2401159121. Available at: https://doi.org/10.1073/pnas.2401159121.

Manny, A.R. *et al.* (2022) 'Discovery of a novel species of trichomonasvirus in the human parasite *Trichomonas vaginalis* using transcriptome mining', *Viruses*, 14(3), p. 548. Available at: https://doi.org/10.3390/v14030548.

Mercer, F. and Johnson, P.J. (2018) '*Trichomonas vaginalis:* pathogenesis, symbiont interactions and host cell immune responses', *Trends in parasitology*, 34(8), pp. 683–693. Available at: https://doi.org/10.1016/j.pt.2018.05.006.

Narayanasamy, R.K. *et al.* (2021) 'Cytidine nucleoside analog is an effective antiviral drug against Trichomonasvirus', *Journal of Microbiology, Immunology, and Infection.* pp. S1684-1182(21)00180–8. Available at: https://doi.org/10.1016/j.jmii.2021.08.008.

Ong, S.-J. *et al.* (2007) 'Activation of multifarious transcription of an adhesion protein ap65-1 gene by a novel Myb2 protein in the protozoan parasite *Trichomonas vaginalis'*, *The Journal of Biological Chemistry*, 282(9), pp. 6716–6725. Available at: https://doi.org/10.1074/jbc.M610484200.

Provenzano, D., Khoshnan, A. and Alderete, J.F. (1997) 'Involvement of dsRNA virus in the protein composition and growth kinetics of host *Trichomonas vaginalis'*, *Archives of Virology*, 142(5), pp. 939–952. Available at: https://doi.org/10.1007/s007050050130.

Rada, P. *et al.* (2022) ‘Double-stranded RNA viruses are released from *Trichomonas vaginalis* inside small extracellular vesicles and modulate the exosomal cargo’, *Frontiers in Microbiology*, 13, p. 893692. Available at: https://doi.org/10.3389/fmicb.2022.893692.

Rivera, W.L. *et al.* (2017) ‘Detection and molecular characterization of double-stranded RNA viruses in Philippine *Trichomonas vaginalis* isolates’, *Journal of Microbiology, Immunology and Infection*, 50(5), pp. 669–676. Available at: https://doi.org/10.1016/j.jmii.2015.07.016.

Smith, A.J. *et al.* (2011) ‘Novel core promoter elements and a cognate transcription factor in the divergent unicellular eukaryote *Trichomonas vaginalis’*, *Molecular and Cellular Biology*, 31(7), pp. 1444–1458. Available at: https://doi.org/10.1128/MCB.00745-10.

Tai, J.H. and Ip, C.F. (1995) ‘The cDNA sequence of *Trichomonas vaginalis* virus-T1 double-stranded RNA’, *Virology*, 206(1), pp. 773–776. Available at: https://doi.org/10.1016/s0042-6822(95)80008-5.

Tsai, C.-D., Liu, H.-W. and Tai, J.-H. (2002) ‘Characterization of an iron-responsive promoter in the protozoan pathogen *Trichomonas vaginalis*’, *Journal of Biological Chemistry*, 277(7), pp. 5153–5162. Available at: https://doi.org/10.1074/jbc.M110234200.

Wang, B. *et al.* (2020) ‘Structural insights into target DNA recognition by R2R3-MYB transcription factors’, *Nucleic Acids Research*, 48(1), pp. 460–471. Available at: https://doi.org/10.1093/nar/gkz1081.

World Health Organization (2020) *Trichomoniasis*. Available at: https://www.who.int/news-room/fact-sheets/detail/trichomoniasis (Accessed: 25 July 2024).

DISCUSSION

Summary of Results

The overall goal of this work is to better understand impacts of trichomonasviruses on their protozoan host, with the aim of uncovering new information about both an important human pathogen and an intriguing model system. We started this project by generating cured and uncured isogenic sets of Tvag clones to use for further experimentation, and in doing so, identified differences in drug susceptibility between representative strains of the five TVV species. We then used these newly generated isogenic sets to uncover differences in gene expression and phenotype between cured and uncured isogenic Tvag clones. The results of both chapters are summarized below, as well as in Figure D.1 at the end of this section.

In Chapter 1, we report our method for using the nucleoside analog 2'-C-methylcytidine (2CMC) to generate isogenic Tvag clones. In doing so, we corroborate the results of Narayanasamy *et al.* that 2CMC is effective against representative strains of TVV1, TVV2, and TVV3 (Narayanasamy *et al.*, 2022) by treating a panel of Tvag isolates containing strains of these species and monitoring viral RNA expression. We also report for the first time that 2CMC is effective in reducing viral RNA abundance of representative strains of TVV4 and TVV5. Additionally, we present evidence for differential susceptibility between TVV species. Our data suggest that representative strains of TVV1 and TVV2 have a greater resistance to 2CMC-mediated clearance than representative strains of TVV3, TVV4, and TVV5. We utilize this apparent difference in resistance to 2CMC to selectively clear strains of specific TVV species from a Tvag isolate that was co-infected with TVV1, TVV2, TVV3, and TVV5. From this isolate, we were able to generate an

isogenic Tvag set containing clones harboring the original combination of viruses, TVV1 alone, or that were cleared of TVV infection entirely. This work demonstrates the utility of 2CMC to create isogenic Tvag sets harboring different combinations of viruses. Further, the differences in 2CMC susceptibility provide new insights into differences between TVV species, posing new questions that will be discussed later in this chapter.

In Chapter 2, we use three of these isogenic sets to examine differences between cured and uncured Tvag clones. In our experiments, we did not see a notable difference in growth rate between TVV+ and TVV- Tvag clones in ideal culture conditions, but did however see an increased survival ability of TVV+ Tvag clones when placed in nutrient deprived conditions. Further, we saw a decreased ability of TVV+ clones to adhere to human cells. To better understand mechanisms underlying these phenotypes, we performed RNA-seq analysis using three cured and uncured Tvag sets and revealed a total of 179 differentially expressed genes. We determined TVAGG3_0365170, which we referred to as TvMyb4, to be our most promising hit due to its consistent upregulation across all three experiments and validated this result using quantitative reverse transcriptase PCR (qRT-PCR). By introducing siRNA against this gene, we were able to knockdown RNA expression of TvMyb4 and demonstrate that viral RNA abundance was also reduced compared to our control reactions, indicating that TvMyb4 may have a role in maintaining viral persistence. We also provide evidence that TvMyb4 knockdown in TVV- trichomonads leads to decreased parasite survival, pointing to TvMyb4 as a potential regulator of the TVV-associated phenotypes we identified.

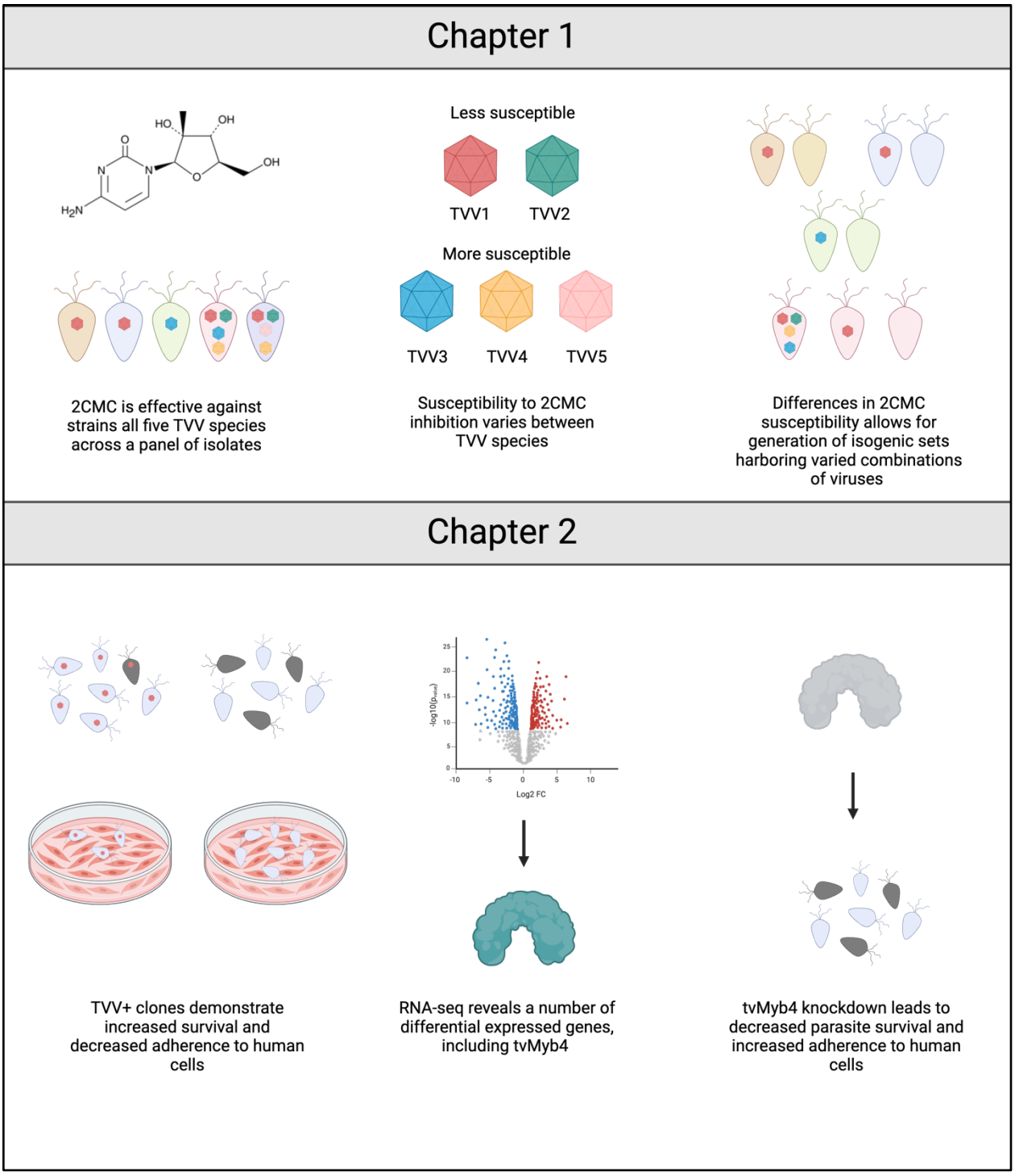


Figure D.1. Graphical Representation of Results. Detailed methods and results are found in Chapters 1 and 2.

Species-specific differences between trichomonasviruses

Differences between trichomonasviruses species, both in the biology of the virus, and in impacts of the parasite host and human superhost, remain an open area of investigation. In Chapter 1, we present evidence that representative strains of TVV1 and TVV2 are better able to resist inhibition by 2CMC than representative strains of TVV3, TVV4, and TVV5. Intriguingly, these patterns do not appear to reflect differences at the sequence level. In our analyses, TVV2 and TVV5 nucleotide sequences appear to cluster together, as do TVV3 and TVV4, leaving TVV1 as the most divergent (Manny *et al.*, 2022) (see Appendix 1). These differences in susceptibility may be explained by structural variations between TVV species, but this information is currently limited. To date, capsid structures of representative strains of TVV1 (Parent *et al.*, 2013) and TVV2 (Stevens *et al.*, 2021) have been published, but similar structures of TVV3, TVV4, and TVV5 capsids are still needed. Further, there are to date no published structures of the RNA dependent RNA polymerase (RdRp) from any TVV species. While little known about 2CMC-mediated inhibition of trichomonasvirus replication, we can speculate about the mechanism of this inhibition based on what is known about 2CMC-mediated inhibition of other viruses. Specifically, 2CMC is thought to induce chain termination of Hepatitis C virus (HCV) through steric hinderance (Eltahla *et al.*, 2015). It is thus conceivable that structural differences in the RdRp proteins between trichomonasvirus species may lead to variable tolerances to steric hinderance by 2CMC.

We performed preliminary structural prediction using AlphaFold2 (Jumper *et al.*, 2021), but no obvious differences between the polymerase structures were revealed. Thus, experimentally determined structures are a crucial next step in understanding the

patterns of 2CMC-mediated inhibition between trichomonasvirus species as well as in uncovering additional differences between viral species. In the absence of structural investigation, an additional line of experimentation may be to passage trichomonads harboring susceptible strains of TVV3 and TVV5 to generate 2CMC-resistance mutants. In the course of this work, we attempted to generate mutant strains of TVV3 by passaging trichomonads from clone JH32A#4-A1 in 1 μM 2CMC but were unable to select Tvag clones harboring mutant TVV3 strains. However, it is possible that different 2CMC concentrations and/or lengths of time may allow for selection of clones with mutant strains of TVV3 or TVV5. If so, the mutated residues may provide key information about the interaction between 2CMC and the viral polymerases.

Understanding the relationship between 2CMC and the viral polymerases across trichomonasvirus species also provides an additional avenue for predicting new species-specific inhibitors of trichomonasviruses. The specificity of 2CMC against TVV1 and TVV2 allowed us to select clones harboring only TVV1 from an isolate that had previously harbored strains of TVV1, TVV2, TVV3, and TVV5 (See Chapter 1). While we did not select any clones from this isolate that either harbored TVV2 only or harbored both TVV1 and TVV2 but were cleared of TVV3 and TVV5, it stands to reason that repeated selection may have generated clones with these combinations of viruses. However, new inhibitors will need to be identified in order to generate Tvag clones with other viral combinations. This is particularly relevant because isolates harboring TVV4 or TVV5 in isolation have not been identified in our experiments (See Appendix 2) or reported in the literature. A combination of one of more inhibitors that could generate clones harboring these species

in isolation, as well as clones harboring new combinations of viruses, would thus be invaluable to understanding differences between trichomonasvirus species.

Additional sets of isogenic Tvag clones would also allow further investigation into host responses to different trichomonasvirus species. In our analysis of phenotypic differences between cured and uncured isogenic Tvag clones, we focused on two isolates harboring strains of either TVV1 or TVV3 (see Chapter 2). These differences between cured and uncured strains appeared to be consistent across the two sets, but repeated experiments using clones harboring various combinations of strains of the five TVV species may reveal more subtle differences between species. Additionally, we performed RNA-seq on three sets of cured and uncured clones, two from parent isolates harboring TVV1 only, and one from a parent isolate harboring TVV3 only. We expected that in addition to overlapping hits from all three experiments, we would see additional gene expression similarities from the two TVV1-containing sets, but upon performing these experiments, we did not identify any patterns of gene expression that were specific to TVV1 cured and uncured sets. However, additional experiments with more sets of isogenic clones might reveal patterns of gene expression that did not appear in these three experiments. Further gene expression comparison between cured and uncured clones is thus needed, both to understand general differences in Tvag phenotype elicited by TVV infection, as well as to elucidate how these differences might vary between TVV species.

Maintenance of host–virus détente

In addition to outstanding questions about species-specific differences between trichomonasviruses, the mechanism of how persistent viral infection of Tvag is achieved and maintained is largely unanswered. However, the long-term persistence of these viruses suggests that the parasite suppresses these viruses to tolerable levels. As discussed in the introduction of this text, Tvag lacks familiar viral dsRNA sensing mechanisms such as RIG-I or TLR3 (Carlton *et al.*, 2007). However, Tvag does appear to have a functional RNA interference (RNAi) pathway that may contribute to suppressing trichomonasvirus replication. RNAi is a major immune factor in plants and other organisms that lack an interferon-based immune system (Muhammad *et al.*, 2019). Typically, functional RNAi pathways include Dicer proteins, which cleave dsRNA into 20-23 nucleotide fragments, Argonaute (AGO) proteins, which bind these dsRNA fragments to target mRNA sequences for interference (Schuster, Miesen and van Rij, 2019). The Tvag genome encodes factors homologous to both human Dicer and human AGO proteins (Carlton *et al.*, 2007). Further, at least two groups have to date used siRNA introduction as a method to successfully knockdown Tvag gene expression (Ravaee *et al.*, 2015; Zhang *et al.*, 2023). In Chapter 2, we use RNAi to knockdown TvMyb4 (See Figure 2.6), providing additional evidence that RNAi is at least somewhat functional in Tvag. Before 2CMC was reported to be an inhibitor of TVV replication, we explored the possibility of using RNAi to generate cured Tvag clones. In doing so, we saw that we were able to reduce TVV1 RNA expression by 90 to 99.9% after 72h (Figure D.2). Due to poor culture health, we were not able to generate cured clones using this method, so we pivoted to using 2CMC to generate cured clones. However, the decrease in viral

abundance following electroporation of siRNA further supports the presence of a functional RNAi system. Whether or not this pathway is used constitutively to limit viral replication without the introduction of exogenous RNA molecules remains to be seen. Potential approaches to answering this question include knockout of Tvag RNAi machinery followed by RT-qPCR to measure viral abundance to determine if RNAi machinery is required to limit viral replication. Similarly, measuring viral RNA abundance after overexpression of RNAi machinery may help reveal if this pathway can be used to limit viral replication.

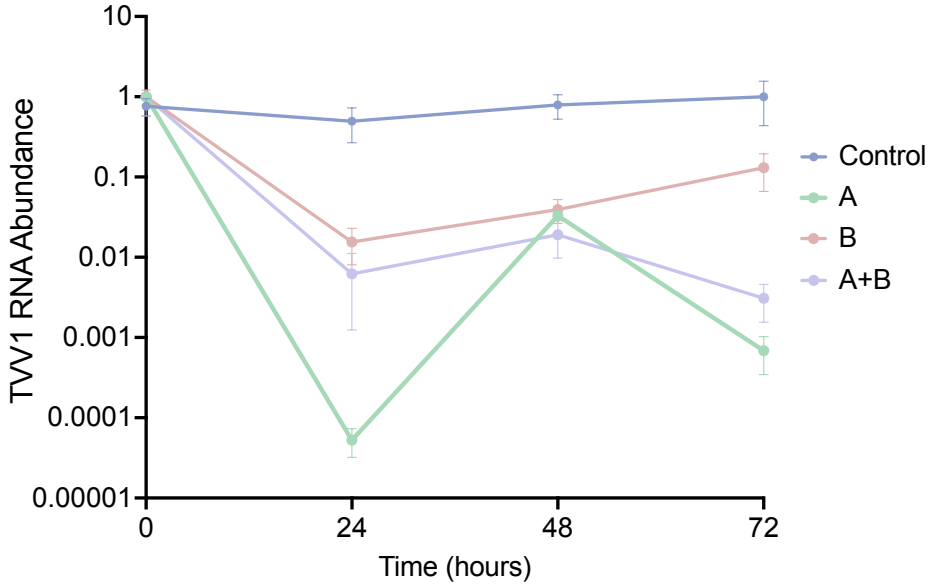


Figure D.2 Introduction of siRNA against TVV1 reduces viral RNA abundance. Electroporation was used to deliver either TVV1-targeting or nontargeting control siRNA molecules to trichomonads from isolate JH37A#2. Total RNA was harvested every 24h and quantified using RT-qPCR using the $2^{-\Delta\Delta Ct}$ method using TvActin as a housekeeping control.

Another potential mechanism that may be used to control trichomonasviruses is through RNA degredation by nucleases. For example, degredation by the 5' to 3' exonuclease XRN1 has been shown to limit replication of ScV-L-A, a related dsRNA virus

to trichomonasviruses (Rowley *et al.*, 2016). There are currently fifteen genes in the 2022 Tvag reference genome that have product annotations classifying them as 5' to 3' exonucleases (https://trichdb.org/trichdb/app/record/dataset/DS_e6874bf420#Busco) that could potentially be involved in a similar mechanism of RNA degradation to yeast XRN1. Additionally, the antiviral activity of XRN1 has also been shown to be involved restriction of other viruses, including flaviviruses (Moon *et al.*, 2012). Flaviviruses can counteract XRN1 degradation via their 5' secondary structure, as XRN1 does not degrade dsRNA, including RNA stem-loop structures (MacFadden *et al.*, 2018). ScV-L-A is unique in that it has a much shorter 5' untranslated region (UTR) than other totiviruses, (Dinman, Icho and Wickner, 1991) which may increase its susceptibility to XRN1 degradation. The length of the TVV 5' UTR likely lends itself to a complicated secondary structure, including a conserved hairpin structure at the extreme 5' end that may be involved in resisting XRN1-mediated degradation (see Appendix 1). To what extent XRN1 may act as an antiviral defense and this hairpin structure serve as an immune evasion strategy remains to be seen. We tried briefly to recombinantly express this protein in *E. coli* with little success, but different systems may allow for expression of XRN1 for biochemical characterization. Additionally, knockdown of XRN1 may allow for investigation into its role in viral replication.

The role of other nucleases, in addition to XRN1, in limiting TVV replication is also unknown. In our differential gene expression analysis (see Chapter 2), we identified five putative nucleases that were significantly upregulated in TVV1+ strains of parent isolate JH37A#2-B1 (Table D.1). Intriguingly, these genes were among the strongest hits in our JH37A#2-B1 experiment (hits 1, 3, 4, and 6 ranked by adjusted p-value). Expression of

one of these genes, TVAGG3_0539280, was also upregulated in TVV3+ strains of JH32A#4-A1. None of the differentially expressed genes in isolate JH191A#4 were predicted to be endonucleases by our analysis. Due to the lack of consistency of expression of these genes, we have not yet investigated these hits further, choosing instead to focus our efforts on validating and characterizing TvMyb4 (see Chapter 2). However, these data do provide future areas of research into the role of nucleases in TVV replication, including the possibility that Tvag may upregulate expression of these genes when TVVs are present to control viral replication. As mentioned above for XRN1, knockdown of these proteins may provide further information about their role in limiting viral replication.

Table D.1 Differentially expressed genes that were predicted to be nucleases by our analysis. Product description entries derived from TrichDB.org

Tvag Gene ID	Product Description	Isolate
TVAGG3_0539280	Endonuclease Protein	JH37A#2, JH32A#4
TVAGG3_0539290	Endonuclease Protein	JH37A#2
TVAGG3_0496300	Endonuclease Protein	JH37A#2
TVAGG3_0585190	3'-5' exonuclease family	JH37A#2

Regardless of what mechanisms Tvag may use to control viral replication, the differential gene expression between virus-positive and virus-negative isogenic Tvag clones suggests that the parasite can recognize to some degree that it is infected with virus. However, the question of how the parasite might sense the virus remains unanswered. One possibility exists that trichomonads can directly sense either viral RNA

or viral proteins. In our analysis, we could not identify any homologs to commonly known viral sensors, but since much of the Tvag genome is still uncharacterized, novel virus-sensing mechanisms may be revealed with further investigation. Another possible explanation is trichomonasviruses compete with their host for resources, and this induces a stress response in the host. Long-term persistence of trichomonasviruses, coupled with our results suggesting that these viruses promote survival in the host (see Chapter 2) indicates that any stress response must be weak enough to not limit host survival. However, it is conceivable that trichomonasviruses would elucidate a mild stress response in the host, leading to changes in transcriptional regulation which may then lead to the changes in adherence and survival phenotypes we have reported. Further characterization of our differentially expressed hits may reveal more information into how Tvag may be sensing these viruses.

Roles for Myb-like proteins in Tvag persistence

The results of our experiments suggest a role for at least one Myb-like transcription factor in TVV persistence and Tvag phenotype. However, the mechanism of how this may happen is currently unknown. Out of the 36,310 protein coding genes listed in the 2022 Tvag reference genome, 486 genes contain at least one Interpro domain containing the term “Myb” (https://trichdb.org/trichdb/app/record/dataset/DS_e6874bf420#Busco).To date, only three Myb-like transcription factors have been identified in Tvag, termed TvMyb1, TvMyb2, and TvMyb3 (Tsai, Liu and Tai, 2002; Ong *et al.*, 2007; Smith *et al.*, 2011). Like these previously identified transcription factors, TvMyb4 contains two Myb recognition element (MRE) domains, characteristic of an R2R3 type transcription factor

(Wang *et al.*, 2020). Further biochemical analysis will be needed to validate the DNA binding activity of TvMyb4, but the predicted structure generated by Alphafold2 (Figure D.3) supports the characterization of this protein as a Myb-family transcription factor. An immediate future direction of this work would be to recombinantly express TvMyb4, which would allow for the structure to be experimentally determined, as well as the DNA binding ability to be assayed by biolayer interferometry.

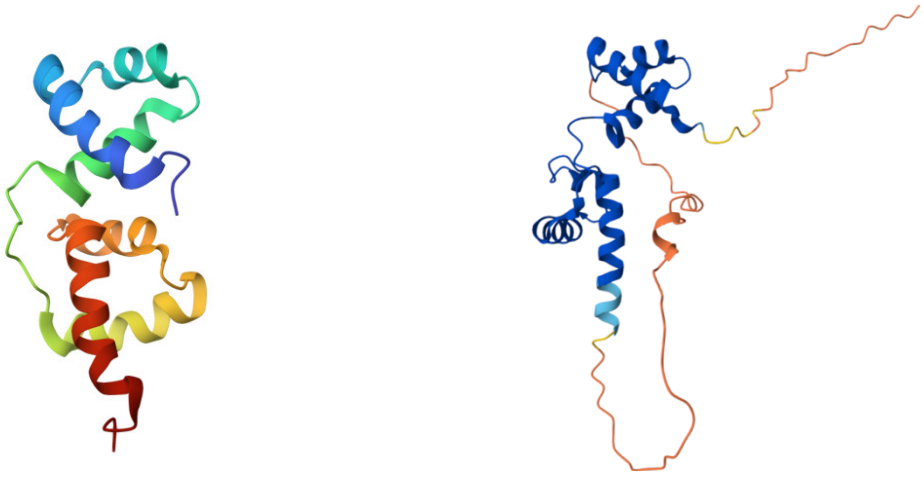


Figure D.3. Predicted structure of TvMyb4 shows similar domains to solved TvMyb1 structure. (Left) Experimentally solved structure of TvMyb1 generated by NMR spectroscopy (Lou *et al.*, 2009). (Right) Predicted structure of TvMyb4 generated by Alphafold2 (Jumper *et al.*, 2021).

If TvMyb4 is indeed shown to have DNA-binding ability, its role in regulating gene expression, and as a result affecting Tvag phenotype, will need to be examined. The three previously identified TvMyb proteins have been implicated in regulating adherence to human cells by modulating expression of Tvag gene TvAP65-1 (Ong *et al.*, 2006, 2007; Hsu *et al.*, 2009). Specifically, TvMyb1 is thought to repress transcription of AP65-1, thereby weakening Tvag adherence (Ong *et al.*, 2006), whereas TvMyb2 and TvMyb3

are thought to increase AP65-1 gene expression, increasing Tvag adherence (Ong *et al.*, 2006, 2007). We could not demonstrate a statistically significant link between TvMyb4 and adherence to human cells and we did not see TvAP65-1, or any apparent homologs, being differentially regulated in our analysis. Thus, further analysis of transcriptional regulation by TvMyb4 is needed. As TvMyb4 expression appears to be upregulated in virus-positive Tvag clones, the most straightforward way to investigate this may be to overexpress TvMyb4 expression in a cured clone and determine what effects this have on gene expression, either globally by RNA-seq or narrowly by RT-qPCR on targeted genes. Another potential avenue to investigate this is to perform RNA-seq of trichomonads following TvMyb4 knockdown, as we have already demonstrated siRNA knockdown of this gene to be possible. Finally, knockout of TvMyb4 followed by RNA-seq to complement these experiments, but it is currently unknown if TvMyb4 is required for cell viability.

In the absence of gene expression studies, previous works regarding TvMyb1, TvMyb2, and TvMyb3 are also useful in investigating the role of Myb-like transcription factors in viral persistence. For instance, our results suggest that TvMyb4 is upregulated in trichomonads harboring trichomonasviruses but does not reveal any information about the post-transcriptional regulation of these protein. Nuclear transport of TvMyb1 (Hsu *et al.*, 2014) and TvMyb3 (Martin *et al.*, 2018) has been shown to be regulated by Tvag cyclophilin proteins, so it is conceivable that the presence of trichomonasviruses may affect the regulation of TvMyb4 beyond at the transcriptional level. Further, because the transport of these proteins is known to be controlled by other cellular proteins, it is possible that these previously characterized Myb-family proteins may be involved in viral

persistence but are not regulated at the transcription level and were thus not identified in our gene expression studies. Future directions of this work will be to determine if the presence of trichomonasviruses impacts nuclear transport of these Myb-family transcription factors to better understand the role of Myb-family transcription factors in TVV persistence.

While TvMyb4 was the only consistent Myb-like hit across all three experiments we performed, we did see one other putative Myb family transcription factor among our differentially expressed hits. Expression of Tvag gene TVAGG3_0192490 was revealed to be significantly downregulated in TVV+ clones derived from parent isolate JH37A#2, despite expression of TvMyb4 being significantly upregulated. We also identified three additional Myb homologs that appear to be downregulated in TVV+ clones of JH37A#2-A1 and two similar hits in our JH32A#4-A1 analysis (Table D.2). While none of these hits reach statistical significance, they point to the possibility of the involvement of additional Myb-family transcription factors in the response to TVVs. Downregulation of these factors in TVV+ clones suggests that these proteins could have opposite effects on gene expression to TvMyb4, or that they are responsible for controlling separate cellular processes entirely. Additionally, the lack of consistency across experiments may imply that there is redundancy in the functions of these proteins. Nevertheless, the potential involvement of Myb-family transcription factors extending beyond TvMyb4 underscores the need for further investigation of these hits.

Table D.2 Putative Myb-like proteins are downregulated in uncured clones of isolates JH37A#2-B1 and JH32A#4-A1. Hits that were classified as being homologs of known Myb-family transcription factors and had an adjusted p-value of less than 0.3 were compiled and reported below. Gene expression of each was reported to be downregulated in uncured isolates as compared to cured isolates.

Gene	JH37A#2	JH191A#4	JH32A#4
TVAGG3_0169690	-	-	0.19
TVAGG3_0192490	0.02	-	-
TVAGG3_0363450	0.26	-	-
TVAGG3_0754380	-	-	0.07
TVAGG3_0858830	0.16	-	-
TVAGG3_0984130	0.21	-	-

Broader effects of trichomonasviruses on Tvag phenotype

Our experiments suggest that trichomonasviruses have an impact on Tvag survival and adherence, but other effects of TVVs on their host have yet to be examined. One potential effect is modulation of parasite morphology. As discussed in the introduction chapter of this dissertation, Tvag can adopt at least three morphologies: trophozoite, ameboid, and pseudocyst (Benchimol, 2008; Dias-Lopes *et al.*, 2018). Trophozoites are thought to encyst in response to stressful conditions (Dias-Lopes *et al.*, 2017; Beri *et al.*, 2020), but little else is known about the mechanism of encystation in Tvag, or if the presence of trichomonasviruses can modulate this process. Our data suggest that the presence of trichomonasviruses leads to increased ability to survive in harsh conditions (See Figure 2.2) and decreased adherence to human cells (See Figure 2.3). Both of these phenotypes are thought to be associated with pseudocyst morphology (Beri *et al.*, 2020). Thus, one possible mechanism to explain these phenotypes is that the presence of trichomonasviruses increases the ability of Tvag to encyst.

We performed some initial experiments to compare the encystation ability of cured and uncured Tvag clones but were unable to robustly distinguish pseudocysts from trophozoites. Much of the literature supporting Tvag encystation has relied upon differential interference contrast microscopy (DIC) or scanning electron microscopy (SEC) to generate detailed images of trichomonad morphologies (Benchimol, 2008; Dias-Lopes *et al.*, 2017). While these images have provided valuable information about the distinct morphologies, they are not ideal for distinguishing trophozoites from pseudocysts in a larger mixed population. One study used fluorescence microscopy and flow cytometry to attempt to quantify pseudocysts in a mixed population using Calcofluor White staining to detect chitin (Beri *et al.*, 2020), but we struggled to detect live pseudocysts using this method. Instead, we detected a large population of dead trichomonads, as determined by FDA viability staining, that were positive for chitin, indicating to us that these may have undergone encystation shortly before dying. More robust methods for detecting pseudocysts, as well as more information regarding the Tvag encystation process may help reveal what, if any, impact trichomonasviruses have on parasite encystation. Understanding this process will help to better understand the mechanism by which trichomonasviruses modulate parasite survival and adherence.

In addition to their role in encystation, the role of trichomonasviruses in parasite transmissibility remains to be seen, but our findings hint at a potential transmissibility advantage for TVV+ trichomonads. The results of our adherence experiments suggest that trichomonads harboring representative strains of TVV1 or TVV3 can be more easily detached from human epithelial cells, indicating a potential increase in their ability to transfer between human hosts during sexual contact. Further, this decrease in adherence

ability does not appear to effect Tvag replication in co-culture with human cells *ex vivo,* indicating that a productive infection would still occur even with this apparent adherence deficit. This transmissibility hypothesis is additionally supported by our results suggesting greater survival of TVV+ trichomonads in harsh conditions, which would allow for survival outside of the urogenital microenvironment during transmission. Transmissibility experiments using animal models of Tvag infection may reveal further information regarding a role for trichomonasviruses in Tvag transmissibility. Robust experiments into transmissibility of trichomonads have been limited in the field due to limitations with existing animal models, which will be discussed later in this chapter.

Impacts of TVVs on human inflammatory response

In addition to effects of TVVs on their parasite host, impacts of TVVs on the human superhost remain an open area of investigation. As pelvic inflammation is a key driver of trichomoniasis symptoms (Fichorova, 2009), understanding the role of TVVs in this inflammation is of particular importance. Further, the microenvironment induced by the inflammatory responses elicited by Tvag infection has also been implicated in the tentative link between Tvag and prostate cancer, (Seo *et al.*, 2014; Kim *et al.*, 2016; Han *et al.*, 2019), so the increased inflammation induced by trichomonasviruses suggests additional clinical relevance to studying these viruses. As discussed in the introduction of this text, TVV virions and free dsRNA are released from dead and dying parasites, which are then recognized by TLR3 in the endosome of human cells, triggering a pro-inflammatory cascade (Fichorova *et al.*, 2012). To what extent other innate immune pathways are engaged in human cells by the presence of TVVs remains to be elucidated.

For instance, it is currently unknown if viral dsRNA is able to enter the cytosol of human cells and be recognized by cytosolic dsRNA sensors. Tvag infection is known to induce NLRP3 inflammasome activation in prostate epithelial cells (Gu *et al.*, 2016) and macrophages, (Riestra *et al.*, 2019) but the effect of TVVs on this inflammatory response has not been examined. The isogenic strains generated as part of this work may be a valuable tool for comparing effects of virus-positive and virus-negative Tvag clones on human inflammatory responses.

In addition to how TVVs are sensed by human cells, the downstream role in eliciting pelvic inflammation has not yet been seen. Treatment of human epithelial cells with either virus-positive Tvag isolates or purified TVV1 virions has been shown to lead to production of pro-inflammatory cytokines *ex vivo* (Fichorova *et al.*, 2012), so it is reasonable to expect that downstream inflammatory processes including immune cell recruitment may be engaged as a result. However, the direct connection between TVV-infection and immune cell recruitment has not been shown. Tvag is known to recruit neutrophils during the course of infection, and a role for TVV-induced inflammation has been hypothesized (Bhakta, Moran and Mercer, 2020), but this link has yet to be demonstrated.

Understanding the role of trichomonasviruses in human inflammation and disease will require a combination of experiments done both in human cells *ex vivo,* as well as *in vitro* in animal models. However, no reliable animal model of trichomoniasis infection has to date been established. A non-human primate (NHP) model has been suggested (Patton *et al.*, 2006), particularly for use in vaccine development, but the cost associated with NHP work makes this model prohibitive for most studies. Differences in the vaginal

pH and microflora between mouse and humans have made establishment of vaginal trichomoniasis infection difficult (Meysick and Garber, 1992). Male mouse models of trichomoniasis using transurethral catheterization have also been proposed and have shown promise in measuring parasite adherence *in vivo* (Molgora *et al.*, 2023), but as symptomatic infections typically occur in the female reproductive tract, a robust female mouse model is greatly needed. Experiments using either subcutaneous or intraperitoneal inoculation have demonstrated robust parasite replication (Galego and Tasca, 2023), but do not accurately reflect human infections. Of particular relevance to this work, pathogenicity comparisons of TVV+ and TVV- trichomonads in animal models have to date relied on measuring abscess formation after subcutaneous infection (Kim *et al.*, 2007). Thus, development of a tractable animal model of trichomoniasis is of crucial importance for studying this parasite and the viruses it harbors.

Attention and Funding for Protozoan Parasite Research

This work reveals a number of unanswered questions and future areas of research into trichomonasvirus biology and their effects on the parasite host and the human superhost. However, a major source of concern in the field is the lack of attention toward the of study protozoa and other eukaryotic parasites. Most recently, VEuPathDB, a major database containing valuable information on protozoan parasites, was recently not renewed by the National Institute of Allergy and Infectious Disease (NIAID), despite the database having 46,000 users in 2023 (Wadman, 2024). TrichDB, an organism-specific resource under the umbrella of VEuPathDB has been invaluable for the work describes in this dissertation (See Chapter 2). Donors from philanthropic organizations have

contributed to reopen this resource with limited functionality, but the long-term future of VEuPathDB is uncertain.

The loss of funding for VEuPathDB reflects a greater trend in the field of parasitology. Of the twenty diseases currently defined by the World Health Organization (WHO) as neglected tropical infections, twelve are caused by parasites (World Health Organization, 2024). In 2014, the Center for Disease Control (CDC) listed five diseases as neglected parasitic infections (NPIs) in the United States: Chagas disease, cysticercosis, toxocariasis, toxoplasmosis, and trichomoniasis (CDC Newsroom, 2014). Despite the efforts of the WHO and CDC to direct research toward these diseases, interest in parasitology has declined. As an example of this, between 1973 and 2024, membership in the American Society for Parasitology has declined by 76% (Richter, 2024). It is thus imperative to recruit new investigators studying parasitology to research institutions.

In addition to their impact on human health, protozoan parasites are an important tool for basic research, as are non-parasitic protozoa. As single cellular eukaryotes, protozoa can help bridge the evolutionary gap between prokaryotes and higher order eukaryotes. With this, the viruses of protozoa also remain an important but understudied area of research. In the wake of the COVID-19 pandemic, much attention was turned toward the study of viruses. However, much of the public attention was and has remained focused on viruses of humans and other animals, as these have the more direct relevance to human health. However, this work underscores the potential for non-human viruses, specifically the viruses of protozoa and other single-celled eukaryotes as a tool for understanding fundamental host–virus interactions. Particularly, trichomonasviruses

provide an intriguing avenue to study how viruses and host cells can develop a mutually beneficial relationship. The future directions discussed in this chapter represent a potential roadmap for new research in this field.

References

Benchimol, M. (2008) 'The hydrogenosome as a drug target', *Current Pharmaceutical Design*, 14(9), pp. 872–881. Available at: https://doi.org/10.2174/138161208784041114.

Beri, D. *et al.* (2020) 'Demonstration and characterization of cyst-like structures in the life cycle of *Trichomonas vaginalis'*, *Frontiers in Cellular and Infection Microbiology*, 9. Available at: https://doi.org/10.3389/fcimb.2019.00430.

Bhakta, S.B., Moran, J.A. and Mercer, F. (2020) 'Neutrophil interactions with the sexually transmitted parasite *Trichomonas vaginalis*: implications for immunity and pathogenesis', *Open Biology*, 10(9), p. 200192. Available at: https://doi.org/10.1098/rsob.200192.

Carlton, J.M. *et al.* (2007) 'Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis'*, *Science (New York, N.Y.)*, 315(5809), pp. 207–212. Available at: https://doi.org/10.1126/science.1132894.

CDC Newsroom (2014) 'Parasitic Infections also occur in the United States'. Available at: https://archive.cdc.gov/www_cdc_gov/media/releases/2014/p0508-npi.html (Accessed:

Dias-Lopes, G. *et al.* (2017) 'Morphologic study of the effect of iron on pseudocyst formation in *Trichomonas vaginalis* and its interaction with human epithelial cells', *Memórias do Instituto Oswaldo Cruz*, 112(10), pp. 664–673. Available at: https://doi.org/10.1590/0074-02760170032.

Dias-Lopes, G. *et al.* (2018) 'In-depth quantitative proteomic analysis of trophozoites and pseudocysts of *Trichomonas vaginalis'*, *Journal of Proteome Research*, 17(11), pp. 3704–3718. Available at: https://doi.org/10.1021/acs.jproteome.8b00343.

Dinman, J.D., Icho, T. and Wickner, R.B. (1991) 'A -1 ribosomal frameshift in a double-stranded RNA virus of yeast forms a gag-pol fusion protein', *Proceedings of the National Academy of Sciences*, 88(1), pp. 174–178. Available at: https://doi.org/10.1073/pnas.88.1.174.

Eltahla, A.A. *et al.* (2015) 'Inhibitors of the Hepatitis C virus polymerase; mode of action and resistance', *Viruses*, 7(10), pp. 5206–5224. Available at: https://doi.org/10.3390/v7102868.

Fichorova, R.N. (2009) ‘Impact of *T. vaginalis* infection on innate immune responses and reproductive outcome’, *Journal of Reproductive immunology*, 83(1–2), pp. 185–189. Available at: https://doi.org/10.1016/j.jri.2009.08.007.

Fichorova, R.N. *et al.* (2012) ‘Endobiont viruses sensed by the human host – beyond conventional antiparasitic therapy’, *PLOS ONE*, 7(11), p. e48418. Available at: https://doi.org/10.1371/journal.pone.0048418.

Galego, G.B. and Tasca, T. (2023) ‘Infinity war: *Trichomonas vaginalis* and interactions with host immune response’, *Microbial Cell*, 10(5), pp. 103–116. Available at: https://doi.org/10.15698/mic2023.05.796.

Gu, N.-Y. *et al.* (2016) ‘*Trichomonas vaginalis* induces IL-1β production in a human prostate epithelial cell line by activating the NLRP3 inflammasome via reactive oxygen species and potassium ion efflux’, *The Prostate*, 76(10), pp. 885–896. Available at: https://doi.org/10.1002/pros.23178.

Han, I.-H. *et al.* (2019) ‘Inflammatory mediators of prostate epithelial cells stimulated with *Trichomonas vaginalis* promote proliferative and invasive properties of prostate cancer cells’, *The Prostate*, 79(10), pp. 1133–1146. Available at: https://doi.org/10.1002/pros.23826.

Hsu, H.-M. *et al.* (2009) ‘Transcriptional regulation of an iron-inducible gene by differential and alternate promoter entries of multiple Myb proteins in the protozoan parasite *Trichomonas vaginalis’*, *Eukaryotic Cell*, 8(3), pp. 362–372. Available at: https://doi.org/10.1128/EC.00317-08.

Hsu, H.-M. *et al.* (2014) ‘Regulation of nuclear translocation of the Myb1 transcription factor by TvCyclophilin 1 in the protozoan parasite *Trichomonas vaginalis’*, *The Journal of Biological Chemistry*, 289(27), pp. 19120–19136. Available at: https://doi.org/10.1074/jbc.M114.549410.

Jumper, J. *et al.* (2021) ‘Highly accurate protein structure prediction with AlphaFold’, *Nature*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

Kim, J.-H. *et al.* (2016) ‘Proliferation of prostate stromal cell induced by benign prostatic hyperplasia epithelial cell stimulated with *Trichomonas vaginalis* via crosstalk with mast cell’, *The Prostate*, 76(15), pp. 1431–1444. Available at: https://doi.org/10.1002/pros.23227.

Kim, J.W. *et al.* (2007) ‘Double-stranded RNA virus in Korean isolate IH-2 of *Trichomonas vaginalis*’, *The Korean Journal of Parasitology*, 45(2), pp. 87–94. Available at: https://doi.org/10.3347/kjp.2007.45.2.87.

Lou, Y.-C. *et al.* (2009) ‘NMR structural analysis of DNA recognition by a novel Myb1 DNA-binding domain in the protozoan parasite *Trichomonas vaginalis’*, *Nucleic Acids Research*, 37(7), p. 2381. Available at: https://doi.org/10.1093/nar/gkp097.

MacFadden, A. *et al.* (2018) 'Mechanism and structural diversity of exoribonuclease-resistant RNA structures in flaviviral RNAs', *Nature Communications*, 9(1), p. 119. Available at: https://doi.org/10.1038/s41467-017-02604-y.

Manny, A.R. *et al.* (2022) 'Discovery of a novel species of trichomonasvirus in the human parasite *Trichomonas vaginalis* using transcriptome mining', *Viruses*, 14(3), p. 548. Available at: https://doi.org/10.3390/v14030548.

Martin, D.H. *et al.* (2013) 'Unique vaginal microbiota that includes an unknown mycoplasma-like organism is associated with *Trichomonas vaginalis* infection', *The Journal of Infectious Diseases*, 207(12), pp. 1922–1931. Available at: https://doi.org/10.1093/infdis/jit100.

Martin, T. *et al.* (2018) '1H, 13C and 15N resonance assignments and secondary structures of cyclophilin 2 from *Trichomonas vaginalis'*, *Biomolecular NMR assignments*, 12(1), pp. 27–30. Available at: https://doi.org/10.1007/s12104-017-9774-3.

Meysick, K.C. and Garber, G.E. (1992) 'Interactions between *Trichomonas vaginalis* and vaginal flora in a mouse model', *The Journal of Parasitology*, 78(1), pp. 157–160.

Molgora, B.M. *et al.* (2023) '*Trichomonas vaginalis* adherence phenotypes and extracellular vesicles impact parasite survival in a novel in vivo model of pathogenesis', *PLoS Neglected Tropical Diseases*, 17(10), p. e0011693. Available at: https://doi.org/10.1371/journal.pntd.0011693.

Moon, S.L. *et al.* (2012) 'A noncoding RNA produced by arthropod-borne flaviviruses inhibits the cellular exoribonuclease XRN1 and alters host mRNA stability', *RNA (New York, N.Y.)*, 18(11), pp. 2029–2040. Available at: https://doi.org/10.1261/rna.034330.112.

Muhammad, T. *et al.* (2019) 'RNA interference: a natural immune system of plants to counteract biotic stressors', *Cells*, 8(1), p. 38. Available at: https://doi.org/10.3390/cells8010038.

Narayanasamy, R.K. *et al.* (2022) 'Cytidine nucleoside analog is an effective antiviral drug against Trichomonasvirus', *Journal of Microbiology, Immunology, and Infection,* 55(2), pp. 191–198. Available at: https://doi.org/10.1016/j.jmii.2021.08.008.

Ong, S.-J. *et al.* (2006) 'Multifarious transcriptional regulation of adhesion protein gene ap65-1 by a novel Myb1 protein in the protozoan parasite *Trichomonas vaginalis'*, *Eukaryotic Cell*, 5(2), pp. 391–399. Available at: https://doi.org/10.1128/EC.5.2.391-399.2006.

Ong, S.-J. *et al.* (2007) 'Activation of multifarious transcription of an adhesion protein ap65-1 gene by a novel Myb2 protein in the protozoan parasite *Trichomonas vaginalis'*, *The Journal of Biological Chemistry*, 282(9), pp. 6716–6725. Available at: https://doi.org/10.1074/jbc.M610484200.

Parent, K.N. *et al.* (2013) 'Structure of a protozoan virus from the human genitourinary parasite *Trichomonas vaginalis'*, *mBio*, 4(2), pp. e00056-13. Available at: https://doi.org/10.1128/mBio.00056-13.

Patton, D.L. *et al.* (2006) 'Development of a nonhuman primate model for *Trichomonas vaginalis* infection', *Sexually Transmitted Diseases*, 33(12), pp. 743–746. Available at: https://doi.org/10.1097/01.olq.0000218871.89901.61.

Ravaee, R. *et al.* (2015) 'Synthetic siRNAs effectively target cystein protease 12 and α-actinin transcripts in *Trichomonas vaginalis'*, *Experimental Parasitology*, 157, pp. 30–34. Available at: https://doi.org/10.1016/j.exppara.2015.06.012.

Richter, H. (2024) 'Parasites are everywhere. Why do so few researchers study them?', *Smithsonian Magazine*, 31 July. Available at: https://www.smithsonianmag.com/science-nature/parasites-are-everywhere-so-why-do-so-few-researchers-study-them-180984753/ (Accessed: 23 September 2024).

Riestra, A.M. *et al.* (2019) '*Trichomonas vaginalis* induces NLRP3 inflammasome activation and pyroptotic cell death in human macrophages', *Journal of Innate Immunity*, 11(1), pp. 86–98. Available at: https://doi.org/10.1159/000493585.

Rivera, W.L. *et al.* (2017) 'Detection and molecular characterization of double-stranded RNA viruses in Philippine *Trichomonas vaginalis* isolates', *Journal of Microbiology, Immunology and Infection*, 50(5), pp. 669–676. Available at: https://doi.org/10.1016/j.jmii.2015.07.016.

Rowley, P.A. *et al.* (2016) 'XRN1 is a species-specific virus restriction factor in yeasts', *PLOS Pathogens*, 12(10), p. e1005890. Available at: https://doi.org/10.1371/journal.ppat.1005890.

Schuster, S., Miesen, P. and van Rij, R.P. (2019) 'Antiviral RNAi in insects and mammals: parallels and differences', *Viruses*, 11(5), p. 448. Available at: https://doi.org/10.3390/v11050448.

Seo, M.-Y. *et al.* (2014) 'Inflammatory response of prostate epithelial cells to stimulation by *Trichomonas vaginalis'*, *The Prostate*, 74(4), pp. 441–449. Available at: https://doi.org/10.1002/pros.22766.

Stevens, A. *et al.* (2021) 'Atomic structure of the *Trichomonas vaginalis* double-stranded RNA virus 2', *mBio*, 12(2), pp. e02924-20. Available at: https://doi.org/10.1128/mBio.02924-20.

Tsai, C.-D., Liu, H.-W. and Tai, J.-H. (2002) 'Characterization of an iron-responsive promoter in the protozoan pathogen *Trichomonas vaginalis*', *Journal of Biological Chemistry*, 277(7), pp. 5153–5162. Available at: https://doi.org/10.1074/jbc.M110234200.

Wadman, H. (2024) ‘Parasitologists up in arms as NIH ends funding for key database’, 11 September. Available at: https://www.science.org/content/article/parasitologists-arms-nih-ends-funding-key-database (Accessed: 23 September 2024).

Wang, B. *et al.* (2020) ‘Structural insights into target DNA recognition by R2R3-MYB transcription factors’, *Nucleic Acids Research*, 48(1), pp. 460–471. Available at: https://doi.org/10.1093/nar/gkz1081.

World Health Organization (2024) *Neglected tropical diseases*. Available at: https://www.who.int/health-topics/neglected-tropical-diseases (Accessed: 23 September 2024).

Zhang, Z. *et al.* (2023) ‘*Trichomonas vaginalis* adhesion protein 65 (TvAP65) modulates parasite pathogenicity by interacting with host cell proteins’, *Acta Tropica*, 246, p. 106996. Available at: https://doi.org/10.1016/j.actatropica.2023.106996.

APPENDIX 1

Discovery of a Novel Species of Trichomonasvirus in the Human Parasite *Trichomonas vaginalis* Using Transcriptome Mining

Austin R. Manny[1,2], Carrie A. Hetzel[1,2], Arshan Mizani[1,3], Max L. Nibert[1,2]

[1]Department of Microbiology, Blavatnik Institute, Harvard Medical School, Boston, MA 02115,

[2]Program in Virology, Division of Medical Sciences, Graduate School of Arts & Sciences, Harvard University, Cambridge, MA 02138

3 Department of Biology, Boston University, Boston, MA 02215

The published manuscript can be accessed as follows:

Manny AR, Hetzel CA, Mizani A, Nibert ML. (2022). Discovery of a novel species of trichomonasvirus in the human parasite *Trichomonas vaginalis* using transcriptome mining. *Viruses, 14,* 548. https://doi.org/10.3390/v14030548

Abstract

Trichomonas vaginalis is the most common non-viral cause of sexually transmitted infections globally. Infection by this protozoan parasite results in the clinical syndrome trichomoniasis, which manifests as an inflammatory disease with acute and chronic consequences. Half or more isolates of this parasite are themselves infected with one or more dsRNA viruses that can exacerbate the inflammatory syndrome. At least four distinct viruses have been identified in *T. vaginalis* to date, constituting species *Trichomonas vaginalis virus 1* through *Trichomonas vaginalis virus 4* in genus *Trichomonasvirus*. Despite the global prevalence of these viruses, few complete coding sequences have been reported. We conducted viral sequence mining in publicly available transcriptomes across 60 RNA-Seq accessions representing at least 13 distinct *T. vaginalis* isolates. The results led to sequence assemblies for 27 novel trichomonasvirus strains across all four recognized species. Using a strategy of de novo sequence assembly followed by taxonomic classification, we additionally discovered six strains of a newly identified fifth species, for which we propose the name *Trichomonas vaginalis virus 5,* also in genus *Trichomonasvirus*. These additional strains exhibit high sequence identity to each other, but low sequence identity to strains of the other four species. Phylogenetic analyses corroborate the species-level designations. These results substantially increase the number of trichomonasvirus genome sequences and demonstrate the utility of mining publicly available transcriptomes for virus discovery in a critical human pathogen.

Introduction

Trichomonas vaginalis viruses (TVVs) are monosegmented (i.e., nonsegmented) dsRNA viruses that infect the parasitic protozoan *Trichomonas vaginalis* (Goodman et al., 2011a). They constitute genus *Trichomonasvirus* in family *Totiviridae* (Goodman et al., 2011a) and are related to monosegmented dsRNA viruses that infect some other parasitic protozoa, namely, viruses that infect *Leishmania* species and constitute genus *Leishmaniavirus* in family *Totiviridae* (Tarr et al., 1988) and viruses that infect *Eimeria* species and constitute proposed genus *Eimeriavirus* in family *Totiviridae* (Wu et al., 2016). Monosegmented dsRNA viruses that infect *Giardia* species and constitute genus *Giardiavirus* (currently classified in family *Totiviridae*, although possibly destined for separation) are more distantly related to trichomonasviruses (Wang & Wang, 1986). TVVs encode two proteins: the viral capsid protein (CP) and a fusion protein, comprising both the CP and the viral RNA-dependent RNA polymerase (RdRp), that results from programmed ribosomal frameshifting (Goodman et al., 2011a).

The prevalence of *T. vaginalis* as a sexually transmitted human pathogen is a major reason for interest in trichomonasviruses. *T. vaginalis* is an extracellular parasite that attaches to epithelial cells of the genitourinary tract of both women and men and is the causative agent of trichomoniasis, the most common non-viral sexually transmitted infection worldwide. Trichomoniasis is associated with perturbed vaginal microbiota, premature delivery, low birth weight, infertility, and the increased transmission and acquisition of other infectious agents including HIV and HPV (Mercer & Johnson, 2018). Currently, the antimicrobial drug metronidazole is used to treat trichomoniasis, but drug resistance is rising, indicating a need for alternative therapies (Leitsch, 2016).

Trichomonasviruses are thought to increase the virulence of *T. vaginalis* by increasing the degree of inflammation during infection. Cervicovaginal epithelial cells sense trichomonasviruses via the recognition of dsRNA by TLR3, which triggers a proinflammatory response through the NF-κB and IRF3 pathways (Fichorova et al., 2012).

A somewhat curious aspect of trichomonasviruses is that they comprise strains of at least four species (*Trichomonas vaginalis virus 1* through *Trichomonas vaginalis virus 4*, abbreviated to TVV1 through TVV4 in strain names) (Bessarab et al., 2000; Bessarab et al., 2010; Goodman et al., 2011b; Tai & Ip, 1995) and that many *T. vaginalis* isolates are concurrently co-infected with different combinations of these species, including some isolates with all four (Goodman et al., 2011b; Rivera et al., 2017). This finding raises the question of the relative roles played by each species and how each may affect the virulence of *T. vaginalis*. It also presents the possibility of biologically relevant interactions among the species, which may give rise to differential effects on the *T. vaginalis* host, and ultimately, the human superhost. The antiviral response described above, for example, has been demonstrated in response to TVV1, but not yet for any of the other species in the absence of TVV1, to the best of our knowledge.

RNA-Seq transcriptome data deposited in public databases, such as sequence reads in the Sequence Read Archive (SRA) database and transcript assemblies in the Transcriptome Shotgun Assembly (TSA) database, both maintained at the National Center for Biotechnology Information (NCBI; Bethesda, MD, USA), are proving a boon for RNA virus discovery (Nibert et al., 2018). For the current study, we decided to screen these databases to discover novel trichomonasvirus strains. We found that the SRA database in particular contains sequence reads from a number of different transcriptome

studies of *T. vaginalis*. From these SRA datasets, we were then able to assemble complete, nearly complete, or partial coding sequences for 27 novel trichomonasvirus strains across all four recognized species. We supplemented this work by determining the complete coding sequences for two other novel strains by de novo sequencing. Notably, we also implemented de novo assembly methods and sensitive homology searches of distantly related sequences to discover six strains of a fifth trichomonasvirus species. We propose the name *Trichomonas vaginalis virus 5* for this newly identified species, abbreviated to TVV5 in strain names. Comparisons of these TVV sequences enhance our understanding of several basic features of these viruses.

Materials and Methods

Analyses of Public Transcriptome Data

RNA-Seq transcriptome datasets in the SRA database at NCBI were screened for TVV- matching sequence reads. This analysis was carried out with a locally implemented BLAST (Camacho et al., 2009) instance to retrieve hits, followed by a deduplication step to exclude erroneous reads cross-mapping to other trichomonasvirus species. Discontiguous megablast ('stand_alone_blast.sh') was first run with multiple species-specific queries against the respective SRA datasets using an e-value threshold of 1e-9. The accessions of these queries obtained from NCBI GenBank are as follows: TVV1: U08999.1, DQ270032.1, HQ607516.1, HQ607513.1, HQ607517.1, JF436869.1; TVV2: NC_003873.1, HQ607514.1, HQ607518.1, HQ607524.1, JF436870.1, JF436871.1; TVV3: NC_004034.1, HQ607515.1, HQ607519.1, HQ607525.1; TVV4 HQ607522.1, HQ607520.1, HQ607526.1. Blastn ('cleanup_blast.sh') was subsequently run on those hits against a local blast database containing 80 TVV sequences retrieved from GenBank,

using an e-value threshold of 10. This confirmed how many of the initial hits indeed best matched the queried trichomonasvirus species. Table A1.1 shows these results for the number of TVV- matching reads for each respective species in each dataset. In addition to this map-to-reference strategy, a de novo assembly approach was used for the discovery of divergent TVV sequences in the SRA datasets. Identification of the initial TVV5 strain was achieved using rnaSPAdes (v1.13.0) (Bushmanova et al., 2018) to assemble sequences. These de novo assemblies were queried against the NCBI nonredundant protein (nr) database using DIAMOND (v0.9.21) (Buchfink et al., 2021) in blastx mode. The DIAMOND parameter '--top 1' was used to force the lowest common ancestor (LCA) algorithm to consider the 99th percentile of the reference sequences to determine the taxonomic origin of each assembly. A custom Python (v3.7.3) script ('diamondToTaxonomy.py') was used to convert NCBI taxonomy IDs to full taxonomic lineages using the JGI-DOE taxonomy server (taxonomy.jgi.doe.gov). Sequence assemblies assigned to family *Totiviridae* were selected for further analysis. These viral assemblies were refined by mapping the reads assigned by BLAST to generate a 50% consensus sequence. This refinement step was performed using a mapping script ('refine_contigs.sh') that implemented BWA-MEM (v0.7.17) (Li, 2013), SAMtools (v1.10) (Li et al., 2009), and BCFtools (v1.10.2) (Danecek et al., 2021). Finalized reads for each novel TVV strain were then assembled into coding-complete, nearly coding-complete, or partial sequences (Supplementary File S1) using CAP3 (v02/10/15; parameters '-o 21 -p 66 -s 300 -z 2') (Huang & Madan, 1999) and CLC Genomics Workbench (v8.0.1; QIAGEN, Redwood City, CA, USA). Normalized coverage values, expressed as RPKM (Supplementary Table S1), were calculated for each assembly. Sequencing depth was

also calculated by mapping each set of TVV-matching reads to a reference strain for that species using BWA-MEM. The SAMtools ‘depth’ function was used to determine the number of mapped reads per nucleotide (nt) position, and a median value was calculated for each virus assembly (Supplementary Table S2) using the ‘median’ function in the statistical programming language R (v4.1.2) (R Core Team). Plots presented throughout this study were developed in R using ggplot2 (v3.3.5) (Wickham, 2016) with extensive use of the tidyverse (v1.3.1) framework (Wickham et al., 2019).

Phylogenetic Tree Construction

Amino acid (aa) sequences for fusion protein CP/RdRp were deduced from newly assembled TVV1 through TVV5 nt sequences and reference TVV1 through TVV4 nucleotide (nt) sequences retrieved from NCBI GenBank (Supplementary Files S1 and S2); partial sequences from GenBank were excluded. These sequences were then aligned with MAFFT (v7.490) (Katoh et al., 2005) using the L-INS-i algorithm (Supplementary File S3). Maximum-likelihood phylogenetic trees were built from this alignment using IQ-TREE (v1.6.11) (Nguyen et al., 2015) on the Los Alamos National Lab webserver (hiv.lanl.gov). The ‘find best and apply’ option (Kalyaanamoorthy et al., 2017) consistently identified JTT+F+I+G4 as the best model. Either standard bootstrapping (Felsenstein; n = 100) or ultrafast bootstrapping (UFBoot2 (Hoang et al., 2018); n = 1000) was conducted in consecutive runs of IQ-TREE, and consensus and bootstrap trees were saved from each run. Each set of trees was then used for transfer analysis by BOOSTER (v0.1.9) (Lemoine et al., 2018) on the Pasteur Institute webserver (booster.pasteur.fr). As expected, an essentially identical consensus tree was obtained from each analysis:

identical in terms of branch topologies and branch lengths limited to two or three significant digits but differing in the support values for some branches. For optimized presentation, the consensus tree was visualized using FigTree (v1.4.4). Support values from both standard and ultrafast bootstrapping, both without and with subsequent transfer analysis, were provided for the main branches, as described in the Figure A1.1 legend. For reference, the original Newick tree files for these four different types of bootstrapping are provided as Supplementary Files S4–S7. The same analyses and presentation steps were additionally performed with the nt sequences of these TVV strains (Figure A1.2), rather than with the CP/RdRp aa sequences, as shown in Figure A1.1.

De novo Analyses of Isolate G3

For polymerase chain reactions, total RNA was isolated from a sample of *T. vaginalis* isolate G3 obtained directly from the American Type Culture Collection (ATCC; Manassas, VA, USA; accession PRA-98) and not subjected to laboratory culturing. cDNA was synthesized using SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA) with random hexamer primers. A 50 µL polymerase chain reaction was carried out for each virus (TVV1, TVV2, TVV3, TVV4, and TVV5) using Taq polymerase (New England Biolabs, Ipswich, MA, USA) for 35 PCR cycles. PCR products were visualized on a 1% agarose gel alongside 1 Kb Plus DNA Ladder (Invitrogen). Primer sequences are listed in Supplementary Table S3.

For high-throughput sequencing, isolate G3 from ATCC was minimally cultured in Diamond's modified Medium (Fouts & Kraus, 1980) for three passages at late-exponential phase. Total RNA was isolated using TRIzol. Contaminating ssRNA was depleted with a

2M LiCl incubation at -20°C overnight. Final enrichment was achieved through nuclease digestion using DNase I and S1 nucleases (Promega Corporation, Madison, WI, USA). Enriched dsRNA was shipped overnight on dry ice to Quick Biology (Pasadena, CA, USA) for sequencing. The RNA-Seq library was prepared according to the KAPA KK8540 RNA HyperPrep kit with 201–300 bp insert size (KAPA Biosystems, Wilmington, MA, USA) using 25–50 ng of total dsRNA as input. Final library quality and quantity were analyzed with the Agilent (Santa Clara, CA, USA) Bioanalyzer 2100 and Invitrogen Qubit 3.0 Fluorometer. Paired-end 150 bp reads were sequenced on an Illumina (San Diego, CA, USA) HiSeq 4000. Sequences were demultiplexed to remove barcodes at the sequencing facility.

For bioinformatic analysis, adapters and other technical sequences were trimmed from the preceding paired-end reads using TrimGalore (v0.6.3) (https://github.com/FelixKrueger/TrimGalore; accessed July 13, 2019) with the following parameters: ‘paired, stringency=5, quality=20’. Bacteriophage ΦX174 positive-control spike-in reads were depleted using BWA-MEM with default parameters. Reads were assigned to their respective trichomonasvirus species using the BLAST workflow above (i.e., initial mapping with discontiguous megablast followed by a clean-up step with blastn). After discarding the reads, species-specific reads were de novo assembled into a draft genome using CAP3 (parameters: ‘- o 21 -p 66 -s 300 -z 2’). In addition to this map-to-reference approach, the de novo assembly strategy described above was also employed. Both approaches converged on the same result, yielding coding-complete sequences from *T. vaginalis* isolate G3 for both TVV2 and TVV3.

Results

Screening for Trichomonasvirus Sequences in Public Transcriptomes

At the time of this study, the SRA database at NCBI contained 60 individual accessions from RNA-Seq transcriptome studies of *T. vaginalis* deposited under five BioProjects from four institutions in the United States, Germany, and Korea (Gould et al., 2013; Woehle et al., 2014; Bradic et al., 2017; Song et al., 2017). These accessions encompass sequence reads from at least 13 distinct *T. vaginalis* isolates. In many cases, the same isolate was represented by several different accessions, which we combined to yield the 16 distinguishable SRA datasets that we subjected to analysis, as listed in Table A1.1. We first screened these datasets for the presence of reads matching any of the four recognized species in genus *Trichomonasvirus* (Bessarab et al., 2000; Bessarab et al., 2010; Goodman et al., 2011a; Goodman et al., 2011b; Tai & Ip, 1995). Briefly, we applied discontiguous megablast at NCBI for performing the database searches, using three to six previously reported nt sequences for each of the four species as queries. Multiple queries were used for each species in an effort to increase the numbers of identified hits from divergent strains of each species that might have been present in these *T. vaginalis* isolates. Results of the screens are summarized in Table A1.1, with numbers that reflect apparently novel TVV strains shown in bold.

Sequence reads matching all four trichomonasvirus species were found, often in large numbers. Only two of the 16 SRA datasets that we distinguished for screening were concluded to be negative for all four species (0 to 13 hits per species): the datasets for *T. vaginalis* isolate T016 from BioProject PRJNA352855 and *T. vaginalis* isolate NYCB20 from BioProject PRJNA280779. The other 14 datasets were concluded to be positive for

at least one species each (67 to 26,085 hits per species): ten for TVV1, seven for TVV2 (including two from the same isolate and institution), nine for TVV3, and two for TVV4. Broken down by *T. vaginalis* isolate, isolates G3 (as reported from BioProject PRJNA345042, University of Utah (UU)) and BRIS/92/STDL/B7268 were positive for TVV1 only; isolate T016 (as reported from BioProjects PRJNA176299 and PRJNA236636, Heinrich Heine University Düsseldorf (HHUD)) was positive for TVV2 only; isolate GOR/03/PNGIMR/69 was positive for TVV1 and TVV2; isolates NYCA04, NYCF20, NYCG31, and SD2 11591* (asterisk a part of isolate name) were positive for TVV1 and TVV3; isolates G3 (as reported from BioProject PRJNA345042, New York University (NYU)) and B7RC2 were positive for TVV2 and TVV3; isolate NYCC37 was positive for TVV1, TVV2, and TVV3; isolate NYCE32 was positive for TVV1, TVV3, and TVV4; and isolate NYCD15 was positive for all four species (Table A1.1). In total, the screening results suggest the identification of 27 novel trichomonasvirus strains, as additionally examined below.

Table A1.1: Screens of SRA datasets at NCBI for TVV-matching sequence reads.

			Sequence read counts from screen for:				
BioProject	**Inst.** [1]	***T. vaginalis* isolate** [2]	**TVV1**	**TVV2**	**TVV3**	**TVV4**	**TVV5**
PRJNA176299	HHUD	T016	0	**18293**	0	0	0
PRJNA236636	HHUD	T016	0	**5254**	0	0	0
PRJNA280779	NYU	BRIS/92/STDL/B7268 [3]	**1017**	4	0	0	17
PRJNA280779	NYU	GOR/03/PNGIMR/69	**1154**	**2447**	2	0	2
PRJNA280779	NYU	G3	2	**385**	**180**	1	4
PRJNA280779	NYU	NYCA04	**2488**	6	**1029**	2	**2783**
PRJNA280779	NYU	NYCB20	4	6	1	0	4
PRJNA280779	NYU	NYCC37	**2114**	**115**	**117**	3	**519**
PRJNA280779	NYU	NYCD15	**2841**	**4702**	**448**	**348**	**7737**
PRJNA280779	NYU	NYCE32	**694**	3	**119**	**589**	**5539**
PRJNA280779	NYU	NYCF20	**2502**	2	**248**	0	6
PRJNA280779	NYU	NYCG31	**422**	5	**66**	0	**6917**
PRJNA280779	NYU	SD2 11591*	**2727**	0	**166**	0	**133**
PRJNA345042	UU	B7RC2	0	**510**	**238**	0	0
PRJNA345042	UU	G3	**26085**	4	10	0	0
PRJNA352855	YU	T016	0	0	0	0	0
Current study	HMS	G3	0	**3114**	**10652**	0	0

[1] Institution: HHUD, Heinrich Heine University Düsseldorf; YU, Yonsei University; UU, University of Utah; NYU, New York University; HMS, Harvard Medical School

[2] As indicated in the metadata for the respective SRA accessions

[3] SRA reads from this *T. vaginalis* isolate and a metronidazole-resistant mutant derived from it were combined for this analysis.

Assembly of Novel Trichomonasvirus Genome Sequences

Having complete coding sequences for virus strains is useful for confirming expected features such as open reading frames and ribosomal frameshifting motifs, for identifying conserved features not previously recognized, for identifying phenotypically important sequence variations, and for allowing robust phylogenetic comparisons. We therefore next assembled the TVV- matching reads into complete coding sequences for as many of the 27 novel trichomonasvirus strains as we could. For this, we used the programs CAP3 and CLC Genomics Workbench to assemble the reads into contigs, and we also separately performed de novo assembly of contigs from the SRA datasets, using the program rnaSPAdes, for corroboration of the results. Through this combination of approaches, we were able to generate and confirm complete coding sequences for thirteen novel strains: ten TVV1 (TVV1-G3(UU), TVV1-BRIS/92/STDL/B7268, TVV1-GOR/03/PNGIMR/69, TVV1-NYCA04, TVV1-NYCC37, TVV1-NYCD15, TVV1-NYCE32, TVV1-NYCF20, TVV1-NYCG31, and TVV1-SD2-11591*) and 3 TVV2 (TVV2-T016, TVV2- GOR/03/PNGIMR/69, and TVV2-NYCD15). In addition, we were able to generate and confirm nearly complete coding sequences for four other novel strains: one TVV2 (TVV2-B7RC2; small 3′ truncation) and three TVV3 (TVV3-B7RC2, TVV3-NYCA04, and TVV3-NYCD15; single small gap in each). Lastly, for the remaining ten novel strains suggested by the findings in Table A1.1 (two TVV2, six TVV3, and two TVV4), we were able to generate and confirm three to six contigs of ≥300 nt in length for each, allowing them also to be included in the subsequent comparisons. RPKM values, using the final full sets of reads used for generating the final assemblies, ranged from 1.2 to 31 for the thirteen coding-complete assemblies (median, 5.2), 0.3 to 2.9 for the four nearly coding-

complete assemblies (median, 0.8), and 0.4 to 1.5 for the longest contig from each of the ten partial assemblies (median, 0.6) (Supplementary Table S1). Median sequencing depth values ranged from 10 to 621 for the thirteen coding-complete assemblies (median, 43), 6 to 16 for the four nearly coding-complete assemblies (median, 10), and 3 to 10 for the longest contig from each of the ten partial assemblies (median, 3) (Supplementary Table S1).

Discovery of a Fifth Trichomonasvirus Species

Following our map-to-reference strategy to characterize additional strains of the four recognized trichomonasvirus species, we next devised an approach that could enable discovery of divergent viruses in these same *T. vaginalis* isolates. A de novo assembly approach was implemented, whereby all adapter-trimmed RNA-Seq reads from a given isolate were assembled using De Bruijn graph-based methods. These assemblies were dynamically translated into the protein space and compared with the NBCI nonredundant protein (nr) database using DIAMOND. Assemblies were compared with the top 1% of matching sequences in the reference database and a taxonomic origin was thereby assigned. Assemblies assigned to viral taxa were retrieved and analyzed.

This assembly first approach successfully reconstructed the TVV sequences obtained from the preceding map-to-reference strategy and also enabled a search for more divergent viruses. Accordingly, *T. vaginalis* isolate NYCE32 was found to carry a virus that could be confidently assigned only to genus *Trichomonasvirus*, with no species-level determination. This 5042 bp assembly exceeded the length of any known TVV strain, although it shared the familiar genomic architecture of two large overlapping reading frames of approximately equal length. Querying this assembly using blastn

against the NCBI nonredundant nucleotide (nt) database indicated its strongest match to be trichomonasvirus TVV2-UR1. Global pairwise alignments were conducted between the 5042 bp assembly and TVV2-UR1 using Clustal Omega (v1.83) (Madeira et al., 2019), demonstrating only 45% identity between these nt sequences and only 36% identity between their deduced CP/RdRp aa sequences. Different species in the family *Totiviridae* are commonly demarcated by a <50% aa-sequence identity (Goodman et al., 2011a), which suggested that isolate NYCE32 harbors a novel trichomonasvirus species.

To test whether other *T. vaginalis* isolates may contain similar viruses to the one in NYCE32, we used the map-to-reference approach with the NYCE32 assembly to identify matching reads in any other dataset. This led us to discover homologous sequences in five other isolates (NYCA04, NYCC37, NYCD15, NYCG31, and SD2-11591*), yielding a total of six additional assemblies of this divergent trichomonasvirus. Two of these assemblies are coding-complete, with a third nearly so (two small gaps); one assembly is partial but encodes a full-length CP; and the remaining two assemblies are partials not fully spanning either viral gene. For the five sequences that cover the predicted CP/RdRp junction, a ribosomal frameshifting site was predicted to involve the heptanucleotide slippery sequence GGGCCCC, which is the same motif as in TVV2. CP/RdRp aa sequences deduced from these six assemblies share >77% pairwise identity with each other, although <50% pairwise identity with any strain of TVV1 through TVV4.

Each of the four recognized trichomonasvirus species represents a distinct monophyletic clade of TVV strains (Goodman et al., 2011b). To examine the relationship of the six divergent assemblies to strains of the recognized species, we constructed maximum-likelihood phylogenetic trees from the aligned CP/RdRp aa sequences of all

new TVV assemblies plus reference TVV genomes retrieved from NCBI GenBank (Figure A1.1). The six divergent assemblies were found to form a monophyletic, discrete clade of their own, with 100% branch support regardless of the different bootstrapping methods we tested. This discrete clade is positioned as sister to the TVV2 clade, consistent with the predicted ribosomal slippery sequence that its members appear to share with TVV2 strains. Given the evidence presented thus far, i.e., a strongly supported discrete clade, high within-clade sequence identity but low sequence identity to TVV1 through TVV4 strains, and a genome length that exceeds that of any previously known TVV strain, we concluded that these six assemblies represent six strains of a novel trichomonasvirus, TVV5, for which we propose the species name *Trichomonas vaginalis virus 5*.

Support values for the branch uniquely shared by the TVV2 and TVV5 clades in Figure A1.1 are lower than those for the other main branches; thus, we recognize that this specific feature of the tree, i.e., the sister relationship between TVV2 and TVV5, might or might not hold up as other novel TVV strains are added in future analyses. It is notable, however, that maximum-likelihood trees we constructed from the nt sequences of these same TVV strains, rather than from their CP/RdRp aa sequences as for Figure A1.1, show discrete TVV2 and TVV5 clades that are again positioned as sisters and have even somewhat stronger support values for their uniquely shared branch (Figure A1.2).

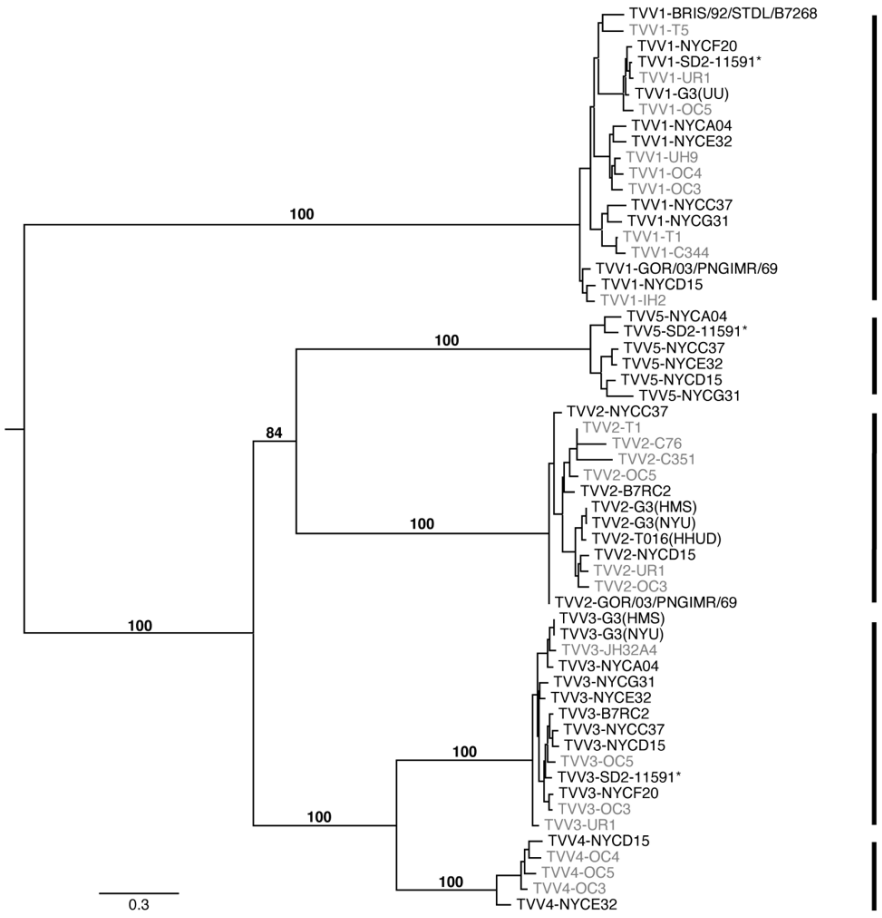


Figure A1.1: CP/RdRp maximum-likelihood phylogenetic tree of TVV1 through TVV5 strains. CP/RdRp aa sequences were deduced from the new TVV assemblies presented in this study (labeled in black) as well as from reference TVV genomes retrieved from NCBI GenBank (labeled in gray). Support values for the main branches are shown as percentages; above the branch is the value from standard bootstrapping with/without subsequent transfer analysis, and below the branch is the value from ultrafast bootstrapping with/without subsequent transfer analysis. The tree is rooted at the midpoint. Bars on the right highlight the five trichomonasvirus species. See Table A1.1 and Supplementary File S2 for explanations of TVV strain names.

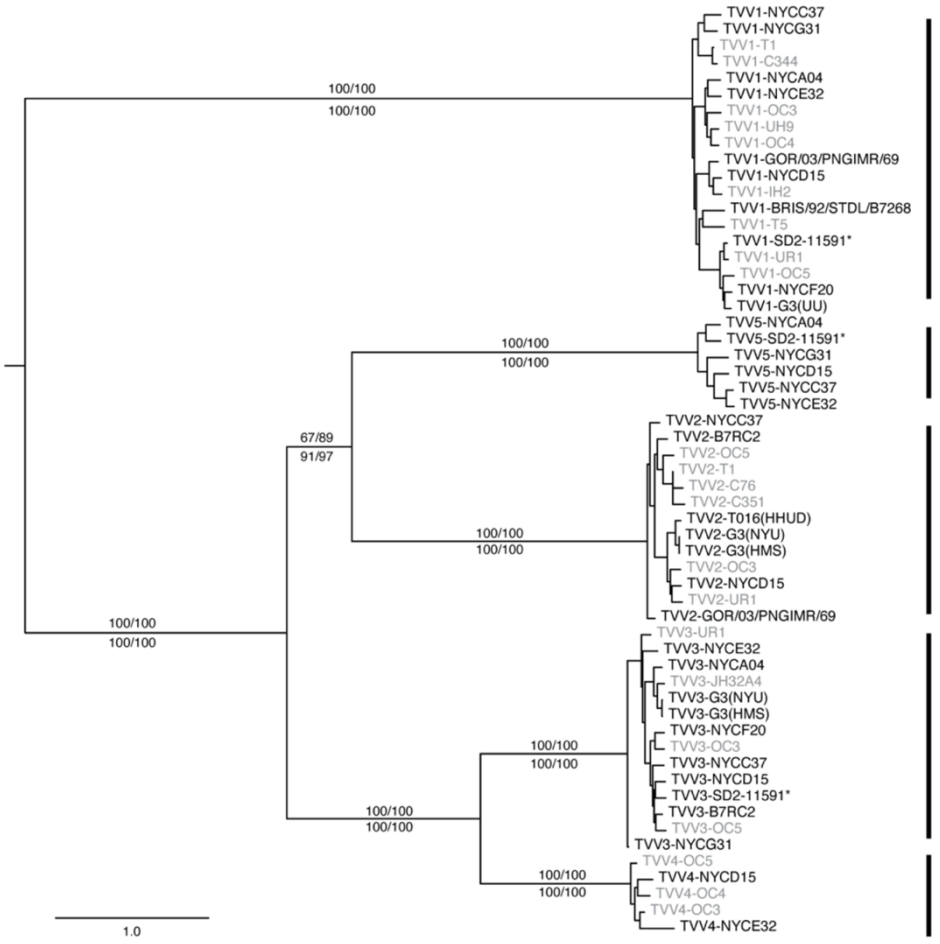


Figure A1.2: Genomic maximum-likelihood phylogenetic tree of TVV1 through TVV5 strains. Nucleotide (nt) sequences of all new TVV assemblies presented in this study (labeled in black) as well as from reference TVV genomes retrieved from NCBI GenBank (labeled in gray) were used for this analysis. Support values for the main branches are shown as percentages; above the branch is the value from standard bootstrapping without/with subsequent transfer analysis, and below the branch is the value from ultrafast bootstrapping without/with subsequent transfer analysis. The tree is rooted at the midpoint. Bars at right highlight the five trichomonasvirus species. This figure parallels Figure A1.1 in the main text, which was instead generated using the deduced CP/RdRp aa sequences of these same TVV strains. As in Figure A1.1, five discrete clades are evident for TVV1 through TVV5, with TVV1 being most divergent and with an apparent sister relationship existing between TVV2 and TVV5 and also between TVV3 and TVV4. The branch uniquely shared by TVV2 and TVV5 has the lowest support values, as also seen in Figure A1.1, though the support values for this branch are higher here than in Figure A1.1.

Reexamination of T. vaginalis Isolate G3

Our results for the presence of TVV sequences in the analyzed SRA datasets (Table A1.1) include inconsistent findings for *T. vaginalis* isolate G3. In particular, G3 from the University of Utah (BioProject PRJNA345042) is positive only for TVV1, whereas G3 from New York University (BioProject PRJNA280779) is positive instead for TVV2 and TVV3. Isolate G3 can be purchased from ATCC and is also the reference isolate for which the *T. vaginalis* whole-genome sequence draft has been reported (Carlton et al., 2007); therefore, we newly acquired and examined this isolate for the presence of trichomonasviruses.

For detecting virus strains that may be carried by isolate G3, we first designed primer pairs based on the TVV1-G3(UU), TVV2-G3(NYU), and TVV3-G3(NYU) sequences described above. Primers for TVV4 and TVV5 were designed against conserved regions of available sequences. RNA was then extracted directly from the cells of *T. vaginalis* isolate G3 that were present in the original sample sent from ATCC, followed by RT-PCR using the respective primer pairs and assay by agarose gel electrophoresis. In this manner, we consistently failed to obtain an amplicon of the expected size using the primer pair based on TVV1-G3(UU) but succeeded in obtaining amplicons of the expected sizes using the primer pairs based on TVV2-G3(NYU) and TVV3-G3(NYU). Moreover, when the latter amplicons were subjected to Sanger sequencing, we found them to be 100% identical to the respective region of TVV2-G3(NYU) (amplicon length excluding primers, 715 nt) and 99.6% identical to the respective region of TVV3-G3(NYU) (amplicon length excluding primers, 798 nt). As

expected, reactions for TVV4 and TVV5 did not yield amplicons. Based on these results, we conclude that the TVV results obtained for *T. vaginalis* isolate G3 reported from New York University are representative of those for G3 isolates currently available from ATCC.

The assembled sequences for TVV2-G3(NYU) and TVV3-G3(NYU) from BioProject PRJNA280779 are not coding-complete; therefore, we also performed an RNA-Seq analysis of *T. vaginalis* isolate G3 that we acquired from ATCC, in an effort to complete the coding sequences of these TVV strains. Following the enrichment of dsRNA from this isolate after limited passage in culture in our laboratory, the sample was submitted for RNA-Seq analysis by a commercial vendor. Results revealed the presence of many sequence reads matching TVV2 and TVV3, but not TVV1, TVV4, or TVV5 (Table A1.1), consistent with the results from BioProject PRJNA280779 described above, as well as with our RT-PCR results. Additionally, reads sufficient in number and coverage were obtained to assemble complete coding sequences for both TVV2-G3(HMS) and TVV3-G3(HMS) (HMS reflecting that this study was performed at Harvard Medical School; see Supplementary Tables S1 and S2 for assembly statistics). These new assemblies exhibit 100% sequence identity with the amplicons for portions of these viruses described above and 99.7% and 99.3% identity, respectively, with the SRA-based partial assemblies for TVV2-G3(NYU) and TVV3-G3(NYU).

Comparisons of Trichomonasvirus Sequences

Untranslated regions (UTRs) at the termini of viral genomes are important sites for replication and/or packaging. Previous reports (Goodman et al., 2011b) have found conserved UTR elements across the four recognized trichomonasvirus species. UTRs

routinely exhibit more sequence flexibility than protein-coding regions, as insertion–deletion (indel) events in the latter can result in deleterious frameshifts and yield defective viral proteins. However, as noted above, UTRs can play vital functional roles too, which can constrain genetic variation in these regions. All TVVs exhibit long 5' UTRs that could contain functional elements such as internal ribosomal entry sites.

To search for functional elements in the UTRs, all new TVV assemblies were compared with the currently available TVV sequences in NCBI GenBank. Per species, these sequences were aligned and any gaps (corresponding to indels between strains) were plotted (Figure A1.3). Most gaps are located in the UTRs, especially the 5' UTR. On the other hand, there are distinct regions in the 5' UTR that lack gaps. As TVV1 is represented by the most sequences, this pattern can be most clearly seen for it: two long gap-free zones where no nucleotides have been inserted or deleted in the available sequences. To investigate whether genetic variation at these nt positions is constrained, conservation plots were also generated for each trichomonasvirus species (Figure A1.4). The conservation plot for TVV1 shows two peaks of high sequence identity at the two gapless regions. The beginning of the CP gene is also highly conserved. This pattern of long 5′ UTRs including distinct zones with constrained sequence variation and no indels points to a functional role for the 5' UTR of TVVs and allows for the possibility of an internal ribosomal entry site in all five trichomonasvirus species.

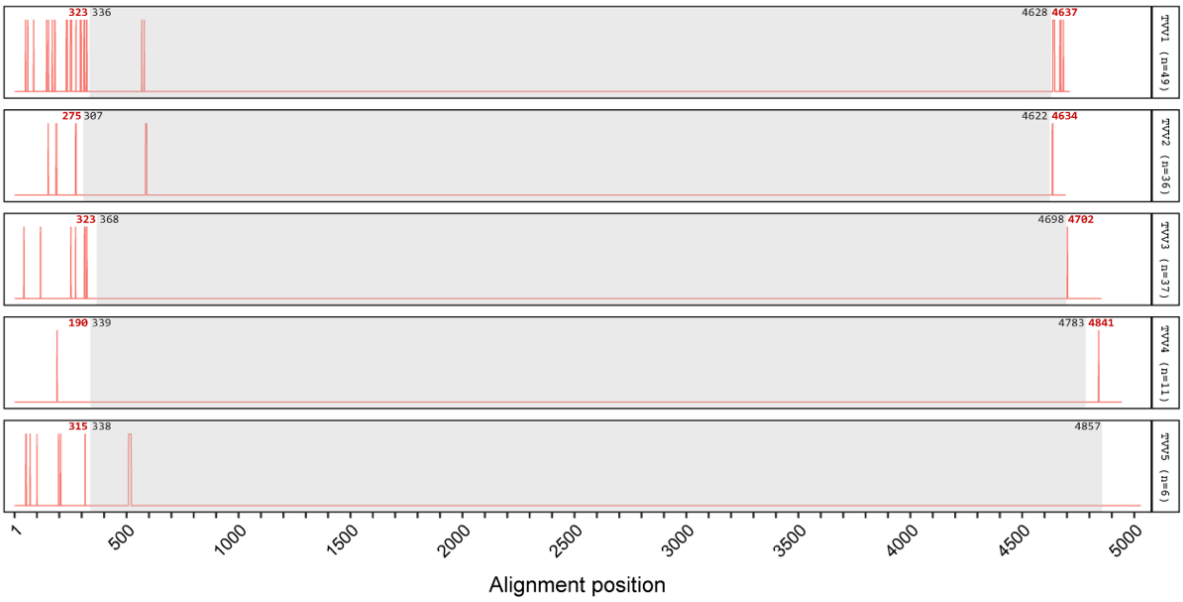


Figure A1.3: Gap plots showing all indels across the aligned nucleotide sequences of TVV1 through TVV5 strains. Indels are concentrated in the 5' and 3' UTRs, although gapless regions also found within the UTRs suggest conserved functional elements. For each trichomonasvirus species, new assemblies were combined with all coding-complete and partial sequences from NCBI GenBank and aligned using MAFFT L-INS-i. The unsequenced ends of partial sequences in the multiple sequence alignment were masked to prevent bias from missing residues. The alignment was analyzed with a custom R script. Gray boxes denote the CDS for the CP/RdRp of each species. Gap positions are indicated by red bars. Red numbers indicate the gap position nearest each CDS boundary; black numbers indicate the CDS boundary positions.

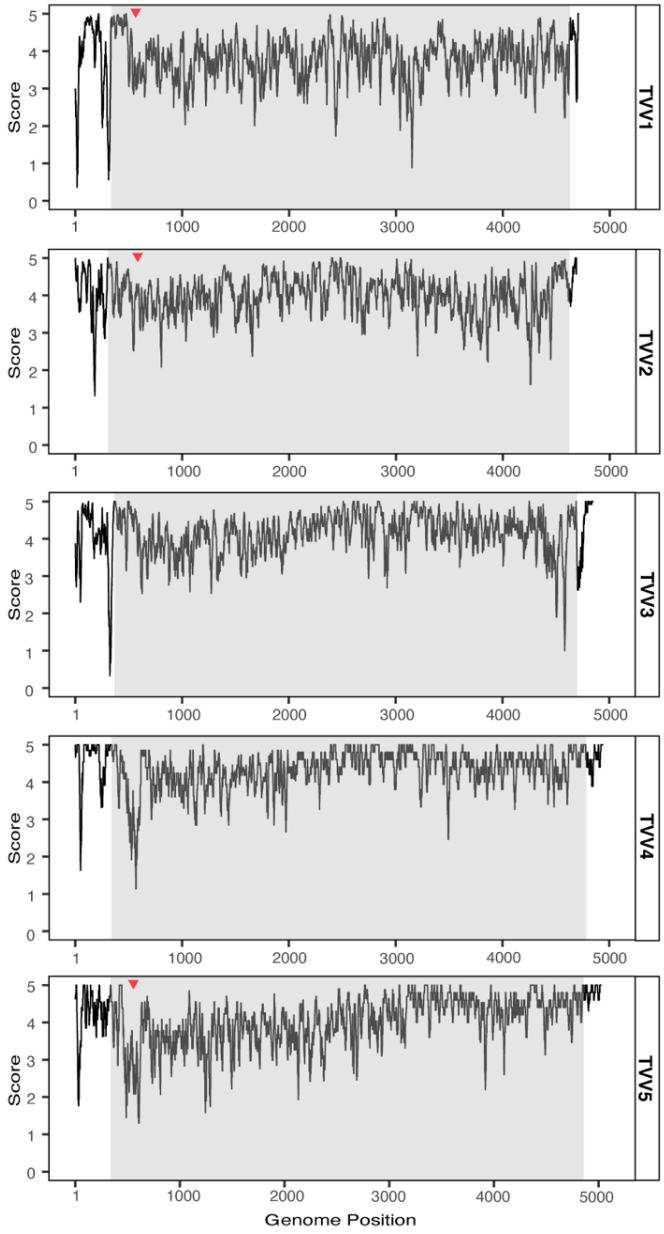


Figure A1.4: Conservation plots for nucleotide sequences of TVV1-TVV5 strains. For each trichomonasvirus species, complete and partial coding sequences were retrieved from NCBI GenBank and combined with new assemblies that were coding-complete or nearly so. A sliding window of 15 nt was chosen for smoothing. The EDNAFULL substitution matrix was used, in which a score of 5 denotes perfect identity at a given position. Gray boxes denote the CDS for the CP/RdRp of each species. Red triangles denote positions of any indels within the CDS (also see Figure A1.3).

Indels were found to be largely excluded from the protein coding sequences (CDSs) of TVV genomes. No strains of TVV3 or TVV4 possess gaps within the CDS; all indels are in either the 5' or 3' UTR. One strain of TVV5 (TVV5-NYCA04) appears to possess a 12 bp deletion in the CDS, near the beginning of the CP gene and corresponding to a skip of 4 aa. TVV1 and TVV2 were found to each have two GenBank sequences with apparent insertions in the CDS. For TVV1, both sequences are for small amplicons from the study of an Iranian patient cohort (Heidary et al., 2013). TVV1-SH8 (GenBank accession AB701566.1) is a 149 bp amplicon sequence and contains an inserted guanosine residue at position 568 relative to the whole genome alignment with all other TVV1 strains. TVV1-SH4 (GenBank accession AB701562.1) is a 142 bp amplicon sequence with an inserted adenosine residue at position 579. In both cases, the inserted residue results in a frameshift that would yield a truncated CP with a divergent C-terminus. Removal of the 1 nt insertion in each of these cases restores the resulting CP sequence to match that in other known TVV1 sequences. For TVV2, both sequences with apparent insertions in the CDS are from the study of a Cuban patient cohort (Fraga et al., 2012). TVV2 strains C76 and C351 (GenBank accessions JF436870.1 and JF436871.1) each contain a multiple nt insertion near the beginning of the CP gene. TVV2-C76 possesses the insertion AAGAAA at positions 585–590, and TVV2- C351 possesses the insertion TAA at positions 588–590. These insertions maintain the coding frame and would result in the respective introduction of two or one extra aa into the CP. The indels identified within the CDS of these few TVV1, TVV2, or TVV5 strains might or might not reflect sequencing or assembly artifacts.

TVV1 genomes were found to possess a conserved CUUUUUGCAC element in the 5’ UTR of full-length sequences. The sole exception is TVV1-NYCC37, which has an extra uracil residue resulting in CUUUUUUGCAC. This uracil-rich element is found in all full-length TVV1 sequences but not in any strains of TVV2, TVV3, TVV4, or TVV5. The 3’ UTRs of each species demonstrate fairly strong sequence conservation within a species, with little similarity across species. These species-specific UTR elements could play a role in the segregation of viral components during prevalent co-infections with multiple trichomonasviruses in a single protozoan host.

Interestingly, our newly assembled TVV1 sequences are all coding-complete, have 5’ termini that match or exceed in length those of previously reported TVV1 genomes, and have 3’ termini that extend within 20 nt of reference TVV1 genomes. Three of the seven TVV2 sequences newly assembled from SRA datasets are coding-complete, with a fourth nearly so. None of the nine TVV3 sequences newly assembled from SRA datasets are coding-complete, but three are nearly so. Neither of the two newly assembled TVV4 sequences is coding-complete or nearly so. Lastly, two of the six newly assembled TVV5 sequences are coding-complete, with a third nearly so. Genome completeness was found to be correlated with coverage depth (p = 0.001; Figure A1.5). To evaluate whether the number of mapped reads is related to trichomonasvirus species, a nonparametric Kruskal–Wallis test was conducted to evaluate whether each species had equivalent numbers of mapped reads per sample (normalized as RPKM values). RPKM values were found to vary significantly across species (p = 0.029), presenting the possibility of different levels of each TVV within a host protozoan (Figure A1.6). To evaluate sequence conservation across trichomonasvirus species, CP/RdRp aa

sequences were deduced from newly assembled sequences and globally aligned per species using Clustal Omega. Pairwise identity matrices for each trichomonasvirus species are provided in Supplementary Tables S4–S8, which show that the analyzed strains share ≥77.9% pairwise identity within each species.

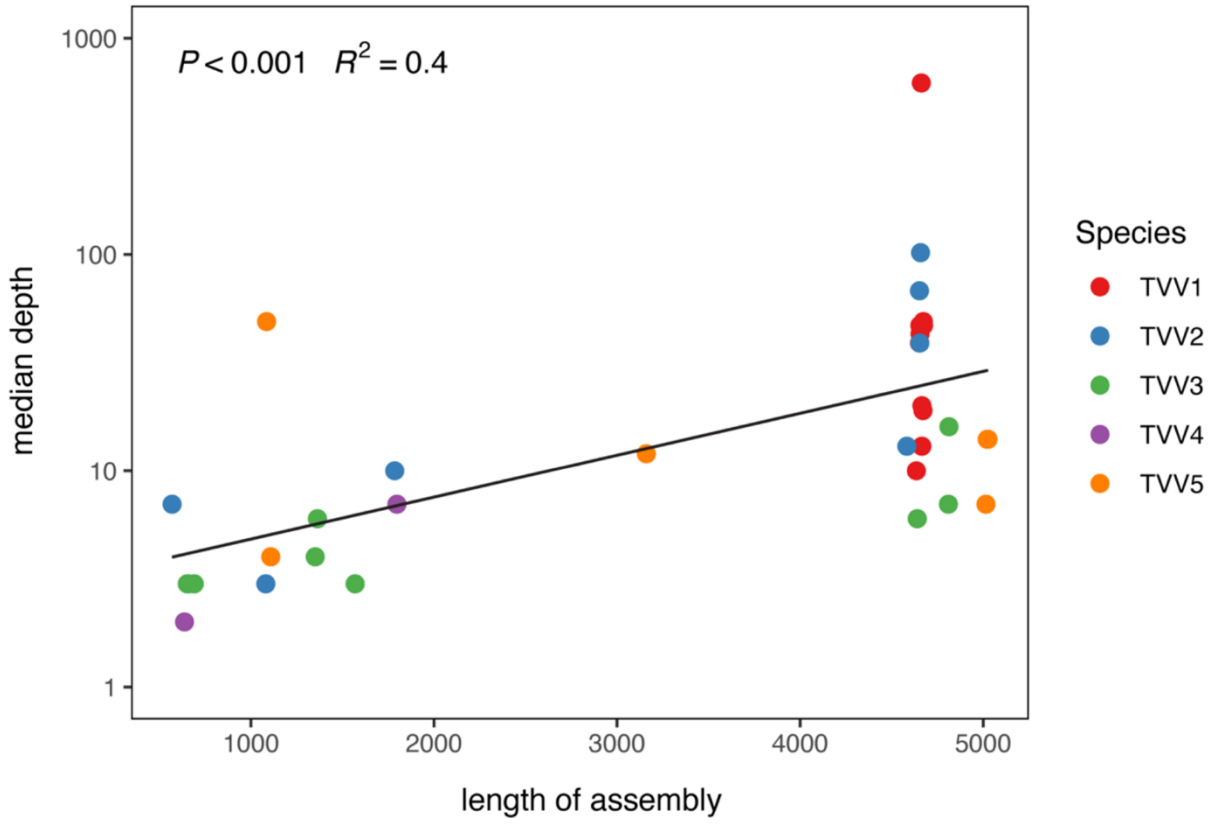


Figure A1.5: Distribution plot of TVV assembly length versus coverage depth. Sequence lengths (in nt) are plotted against median coverage depth (reads per assembly position) for each new assembly, color-coded per trichomonasvirus species. All new assemblies presented in this study were included in this analysis. The line represents the results of a simple linear regression analysis performed on these data, with the resulting *r2* value and *p*-value as shown. This analysis was performed to assess whether assembly length and coverage depth are correlated. The regression analysis indeed supports this correlation, with higher coverage depths favoring longer (coding-complete or nearly so) assemblies. The fact that many TVV genomes exhibited low coverage depths is likely explained by the fact that these RNA-Seq datasets represent metatranscriptomes from *T. vaginalis* samples not enriched for TVV RNAs.

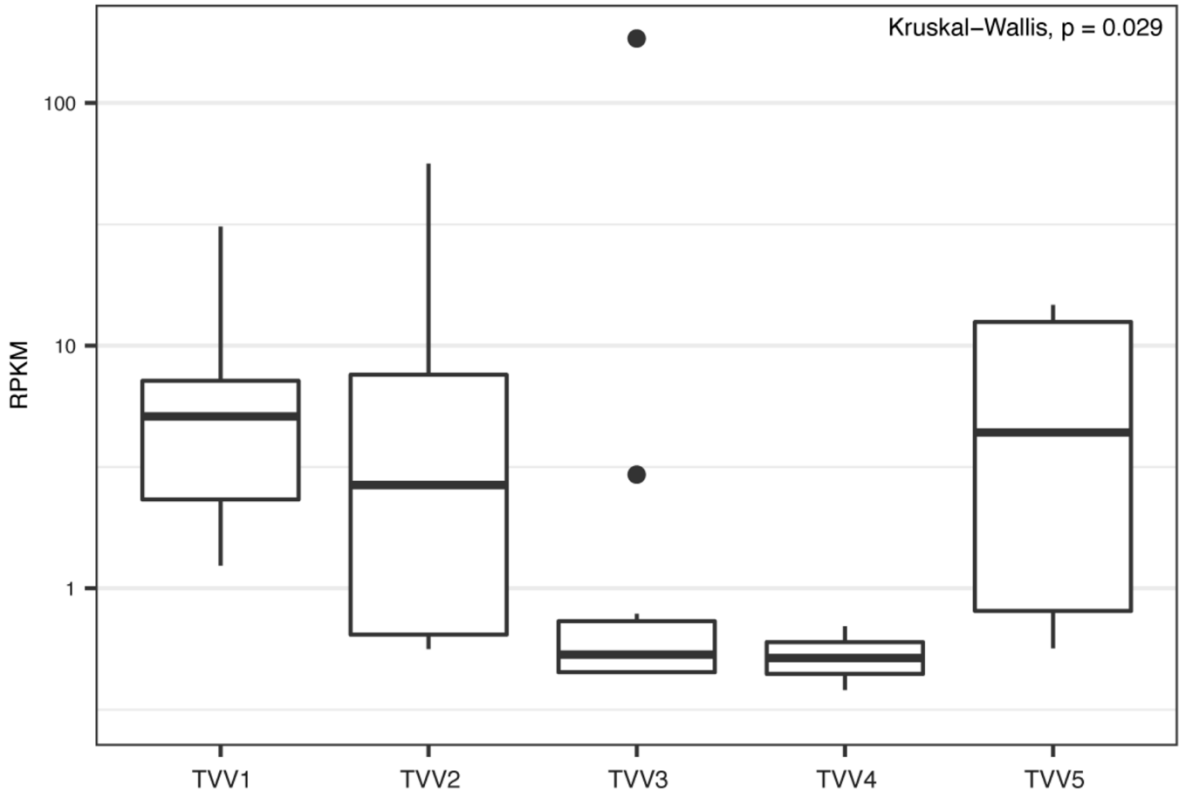


Figure A1.6: Box-and-whisker plots of mapped reads, normalized by RPKM, per trichomonasvirus species. All new assemblies presented in this study were included in this analysis. The median RPKM value for each species is presented as a bold horizontal line. Upper and lower quartiles are represented respectively by the top and bottom of each box, and upper and lower extremes are indicated by the whiskers (vertical lines). Two outliers interpreted to be present in the values for TVV3 are shown as black dots. A nonparametric Kruskal-Wallis one-way ANOVA test was performed, with the resulting *p*-value as shown. This test was used to assess whether the abundance of viral reads was equivalent between species. Instead, the numbers of TVV-matching reads were found to vary across species, suggesting different levels of transcription or replication for the different species. The fact that many of the TVV genomes exhibited low RPKM values is likely explained by the fact that these RNA-Seq datasets represent metatranscriptomes from *T. vaginalis* samples not enriched for TVV RNAs.

All TVV1 genomes assembled in this study are coding-complete, which allowed for the most detailed analysis of a trichomonasvirus species. Almost all TVV1 CP/RdRp sequences are 1430 aa in length, based on predicted initiation at the first in-frame AUG codon. However, both TVV1- UR1-1 and SD2-11591* contain an MGIP N-terminal extension, bringing their apparent lengths to 1434 aa. All TVV1 sequences share a triple serine at their C-termini, except for TVV1-NYCC37, which has STS. Average pairwise identity across the TVV1 CP/RdRp aa sequences is 85.9%. Scanning across the CP/RdRp sequence with a window of 10 aa reveals a single area of the genome with <50% conservation. This region occurs halfway through the CP/RdRp aa sequences, centered on position 701 (705 for TVV1-UR1-1 and SD2-11591*). Interestingly, absolute conservation is observed 20 aa upstream, the site of the ribosomal frameshift.

For TVV2, average pairwise identity for the CP/RdRp aa sequences is 88.3%. The CP/RdRp sequences deduced from coding-complete TVV2 genomes are 1438 aa in length. The N-terminal motif MASTL is found in all these sequences, except in TVV2-T016(HHUD), where it is MAATL. The C-termini of all full-length TVV2 fusion proteins end with PVYV. For TVV3, average pairwise identity for the CP/RdRp aa sequences is 92.1%. The N-termini of all TVV3 sequences that extend to this region begin with MSAPEPLNTEVR, and the C-termini of all TVV3 fusion proteins that extend to this region end with GHGLRSG. Analysis of the TVV4 CP/RdRp aa sequences was impeded by the fact that no TVV4 genomes that are coding-complete or nearly so could be newly assembled from the SRA datasets. The TVV4-NYCD15 and TVV4-NYCE32 assemblies contain partial CP-coding sequences, which have a pairwise identity score of 90.3%. These sequences were aligned to the CP-coding sequences of the three TVV4 genomes

in GenBank: TVV4-OC3, TVV4-OC4, and TVV4-OC5. TVV4-NYCD15 shares the N-terminal motif MSAI with TVV4-OC3 through TVV-OC5. Neither new TVV4 assembly reaches the CP C-terminus.

TVV5 CP/RdRp proves to be the largest of all trichomonasvirus proteins. The three assemblies that are coding-complete or nearly so encode a fusion protein of 1506 aa. Examination of all six TVV5 strains shows a largely conserved N-terminus. The exception is TVV5-NYCA04, which possesses a guanosine at genome position 283 (instead of the adenosine found in all other sequences with coverage at that position), yielding a different proposed start codon. Along with a nearby single nt deletion, this results in an N-terminal extension of 18 aa in the CP and CP/RdRp of this strain. Three TVV5 sequences maintain coverage through the CP/RdRp C-terminus, which ends in PAVPIAT in all three.

Programmed ribosomal frameshifts are instrumental in TVV biology, giving rise to the catalytic CP/RdRp fusion protein. All strains sequenced in this report demonstrate absolute conservation of the known or predicted heptanucleotide slippery sequence in each respective species: TVV1, CCUUUUU; TVV2/TVV5, GGGCCCC; and TVV3/TVV4, GGGCCCU. No variation in these motifs was observed for any strain of any species. A common observation across all TVV species is that although the aa sequences deduced from the region surrounding the ribosomal frameshift are strictly conserved, 10–30 aa downstream is a region of extreme divergence. This observation is described above for TVV1. For another example, the TVV2 ribosomal frameshift site represents position 699 in the deduced aa sequence, and positions 693–702 are perfectly conserved across all strains. However, shortly downstream is a region of low conservation, dropping below 50% at positions 737–738, after which higher sequence identity (80–95%) is restored.

The evolutionary constraint of the ribosomal frameshift might cause higher diversifying selection pressure in the neighboring genomic space, or perhaps this region of aa sequence shortly downstream of the frameshifting site is less subject to structural constraints, possibly representing a “hinge” between CP and RdRp portions of the fusion protein.

To evaluate conserved, terminal RNA structures across trichomonasviruses, secondary- structure predictions were performed for strains of all five species. Specifically, the first and last 100 nt from all TVV strains whose available sequences closely approach one or both expected termini of the TVV genome were analyzed for RNA secondary structures at the 5’ or 3’ end of their coding strands using RNAfold (v2.4.18) (Lorenz et al., 2011). As expected from previous reports (Goodman et al., 2011b), a long stem–loop structure was again predicted at the 5’ terminus of each analyzed sequence, regardless of species and including for TVV5 strains. In addition, strains of TVV2, TVV3, TVV4, and TVV5 were predicted to possess a double stem–loop structure at the 3’ terminus of each analyzed sequence (Figure A1.7). This structure consists of two adjoining stem–loops with no intervening nt residues. These structures extend to within 5 nt of the expected 3’ terminus in TVV2, TVV3, and TVV5 strains, and to within 21 nt of the expected 3’ terminus in TVV4 strains. This conserved RNA feature could play a role in packaging the genome into virions and/or recognition by the viral RdRp.

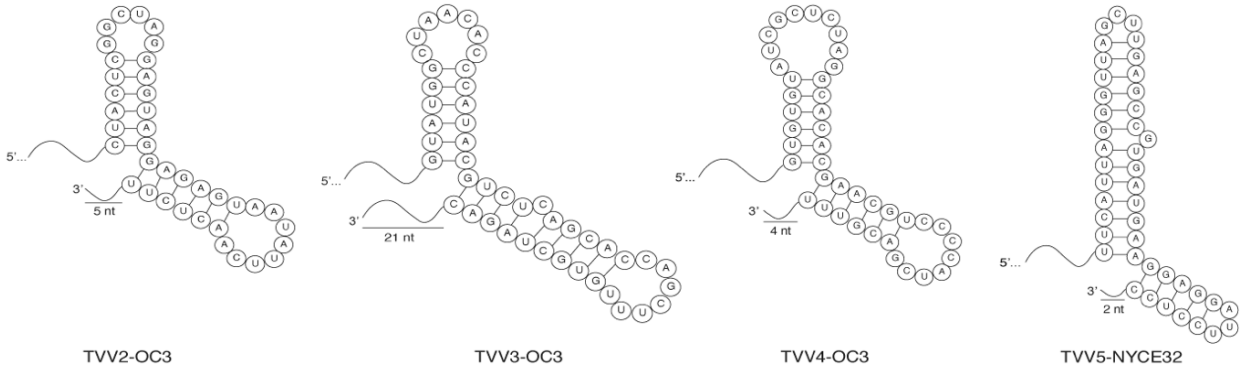


Figure A1.7: Double stem–loop structures near the coding-strand 3' termini of TVV2 through TVV5 strains. Secondary-structure predictions identified this conserved feature, shown here for representative strains TVV2-OC3, TVV3-OC3, TVV4-OC3, and TVV5-NYCE32. This feature extends to within 21 nt of the coding-strand 3' terminus of each virus and consists of two adjoining stem–loops with no intervening nt residues.

Discussion

Discovery of a fifth species in genus *Trichomonasvirus* emphasizes the utility of mining viral sequences from publicly available transcriptomes. A map-reads-to-reference strategy was initially useful for characterizing novel strains of the four recognized trichomonasvirus species. An orthogonal 'assemble-first, classify-later' approach then allowed us to uncover the newly identified fifth species infecting *T. vaginalis*. These complimentary methods enable a robust exploration of sequence data while incorporating the flexibility needed for finding new viruses. Each new TVV assembly held up to rigorous scrutiny, corroborated in all cases by both reference-based and de novo assembly methods. Supplementary Table S1 shows the normalized read count per virus assembly presented in this report, expressed as RPKM values. Most assemblies have an RPKM of <5. Relatedly, most of these assemblies have less than 20x coverage (Supplementary Table S2). These values are high enough to allow the assembly of viral sequences and yielded the assembly of several coding-complete TVV genomes, as presented in this report. However, the fact that these values are relatively low might also explain the number of partial assemblies we obtained. For example, for the two newly assembled TVV4 sequences, both of which are partials, each RPKM was <1 and the median coverage was <10x. Although viral sequence mining was demonstrably successful using these untargeted datasets, methods for enriching viral genomes or transcripts in RNA-Seq libraries may be considered for the reliable determination of more coding-complete viral genomes.

Co-infection of many *T. vaginalis* isolates with strains of two or more trichomonasvirus species, as seen previously (Goodman et al., 2011b) and again in this

study, raises many interesting questions, including whether the different species might be subject to recombination. Although some other RNA viruses (e.g., coronaviruses) exhibit frequent detectable recombination events (Zhu et al., 2020; Gribble et al., 2021), the extent of recombination among trichomonasviruses appears to be low. Pairwise alignment of representative strains of the five species failed to reveal any obvious chimeric breakpoints indicative of recombination events. Furthermore, detailed analysis with the DualBrothers package (v1.1.5) (Minin et al., 2005) failed to detect recombination between TVV strains. This is perhaps not surprising, because most dsRNA viruses are thought to replicate their individual genomes only within the confines of their protein capsids, in which their RdRp or CP/RdRp molecules are also packaged (Tao et al., 2002). This capsid barrier may sequester the replicating RNAs from the potentially distinct RNAs inside other capsids, reducing the chance of interspecies recombination and replication of chimeric molecules. Moreover, the subcellular localizations of different trichomonasviruses within a single protozoan remain unclear. It is possible that during viral co-infections of a shared protozoan cell, distinct TVV strains are additionally secluded from one another in some manner, further reducing the chance of interspecies recombination.

Goodman et al. (Goodman et al., 2011b) reported a conserved, short 5′-terminal region of sequences (36 or 37 nt) in strains TVV1-UH9, TVV1-UR1, TVV1-OC3, TVV1-OC4, and TVV1- OC5 (GenBank accessions HQ607516.1, HQ607513.1, HQ607517.1, HQ607521.1, and HQ607523.1), which was missing from previously reported sequences for strains TVV1-T1, TVV1-T5, and TVV1-IH2 (GenBank accessions U08999.1, U57898.1, and DQ270032.1) (Tai & Ip, 1995; Kim et al., 2007; Su & Tai, 1996) and is

also missing from the subsequently reported sequence for strain TVV1-C344 (GenBank accession JF436869.1) (Madeira et al., 2019). This short sequence extension is notable because it allows the formation of a long 5′-terminal stem– loop structure, which seems likely to be involved in RNA stability and/or other functions in TVV1 replication. All of the new TVV1 sequences reported here extend into this conserved 5′-terminal region, providing further evidence that the sequences for TVV1-T1, TVV1-T5, TVV1-IH2, and TVV1-C344 are likely truncated at their 5′ ends. The fact that the new TVV1 sequences appear to be less but still partially truncated themselves is not surprising, given that RNA-Seq-derived transcript assemblies, in our experience, are often missing a few residues at each end of respective transcripts. Strains of all five trichomonasvirus species have long 5′ UTRs, >280 nt each, as well as numerous AUG codons preceding the first in-frame AUG codon in the CP gene (Goodman et al., 2011a). These features, also evident in the novel strains reported here, suggest to us that at least some of these 5′-UTR sequences contribute to forming an internal ribosome entry site/structure important for viral translation, as shown or suggested for several other members of family *Totiviridae* or related dsRNA viruses (Garlapati & Wang, 2005; Chiba et al., 2018). The increased number of complete coding sequences now available for alignment as a consequence of the current study additionally reveals that indels between TVV strains are found in the 5′ UTR of each trichomonasvirus species, as shown in Figure A1.3. These indels may mark sequence locations that are not directly involved in essential functions at the RNA level or in forming essential RNA structures. Indels between strains have also been found in the 3′ UTRs of four trichomonasvirus species but are generally not found in the long CDS that occupies most of each genome.

Regarding TVV strains in *T. vaginalis* isolate G3, our de novo sequencing results and those from the study by Bradic et al. (Bradic et al., 2017) based at New York University (BioProject PRJNA280779) concur in identifying TVV2 and TVV3, but not TVV1, TVV4, or TVV5 in this isolate. Moreover, the high levels of nt sequence identity (≥99.3%) between TVV2-G3(HMS) and TVV2- G3(NYU) and also between TVV3-G3(HMS) and TVV3-G3(NYU) are consistent with the limited divergence of these viruses, although the *T. vaginalis* isolate was cultured first at ATCC and then separately at the two recipient institutions. On the other hand, the discrepancies found for the viruses in this isolate from the study at University of Utah (BioProject PRJNA345042; presence of TVV1, but not TVV2, TVV3, TVV4, or TVV5) are harder to explain, especially given that this isolate was again obtained from ATCC according to BioSample metadata from that study. One possibility would seem to be that the SRA data for isolates G3 and B7RC2 from the University of Utah study might have been transposed in the database, because the current study identified both TVV2 and TVV3 in isolate B7RC2, as expected instead for isolate G3 based on other results. However, TVV2-B7RC2 and TVV3-B7RC2 are substantially divergent from TVV2-G3(HMS) and

TVV3-G3(HMS) (≤85.4% nt sequence identity), which makes this explanation seem unlikely. At this stage, then, these discrepancies remain unexplained, but they highlight the need for investigators to confirm the virus content of *T. vaginalis* isolates used in each new study that is focused on these viruses.

The current study also identifies discrepancies in the TVV content of *T. vaginalis* isolate T016. From two studies at Heinrich Heine University Düsseldorf (BioProjects PRJNA176299 and PRJNA236636) (Gould et al., 2013; Woehle et al., 2014), isolate T016

is found to contain TVV2, but not TVV1, TVV3, TVV4, or TVV5; the TVV2 sequences derived from those two studies are found to be identical. In contrast, isolate T016 from Yonsei University (BioProject PRJNA352855) (Song et al., 2017) is found to be negative for all five trichomonasvirus species. In this case, because the discrepancies involve the presence or absence of a single TVV strain, the *T. vaginalis* isolate T016 from Yonsei University might have simply been cured of TVV2 during culture, as has been reported to occur in other cases (Wang & Wang, 1987; Men-Fang et al., 1993)

The double stem–loop structure predicted near the coding-strand 3′ terminus of each analyzed TVV2, TVV3, TVV4, and TVV5 sequence is a striking feature. Conserved secondary structures have been found in other RNA viruses to be involved in replication or packaging. Although this double stem–loop varies in length somewhat between trichomonasviruses, the maintenance of its overall shape in TVV2 through TVV5 suggests a functional role in the viral life cycle. This feature was not found in TVV1, which is not entirely surprising. The evolutionary distance is considerable between TVV1 strains and those of the other four species, as deduced from phylogenetic analyses (see Figure A1.1) as well as facets of viral biology such as translation strategies. Strains from TVV2 to TVV5 seem to employ a −1 ribosomal frameshifting mechanism for translation of the CP/RdRp fusion protein, whereas TVV1 strains use a −2 frameshifting mechanism. TVV1 may thus possess its own characteristic RNA structures to fulfill vital roles in its life cycle. Future biochemical studies dissecting the functional potential of TVV RNA sequences appear especially warranted.

Supplementary Materials

Table S1: Sequence read counts for the new TVV assemblies, Table S2: Sequencing depth values for the new TVV assemblies, Table S3: Primers used for polymerase chain reactions for virus detection in *T. vaginalis* isolate G3. Table S4: Pairwise identity matrix of TVV1 CP/RdRp aa sequences, Table S5: Pairwise identity matrix of TVV2 CP/RdRp aa sequences, Table S6: Pairwise identity matrix of TVV3 CP/RdRp aa sequences, Table S7: Pairwise identity matrix of TVV4 CP/RdRp aa sequences, Table S8: Pairwise identity matrix of TVV5 CP/RdRp aa sequences, File S1: TVV genome nt sequences newly assembled from SRA datasets, File S2: GenBank accession numbers of other sequences used for phylogenetic analyses, File S3: Alignment of CP/RdRp aa sequences used for the phylogenetic analyses represented in Figure A1.1, File S4: Newick tree file represented in Figure A1.1 for 100 standard bootstraps, File S5: Newick tree file represented in Figure A1.1 for 100 standard+transfer bootstraps, File S6: Newick tree file represented in Figure A1.1 for 1000 ultrafast bootstraps, File S7: Newick tree file represented in Figure A1.1 for 1000 ultrafast+transfer bootstraps.

Acknowledgments

We wish to acknowledge and thank those scientists who made this study possible by depositing their sequence reads in the SRA database. Reviewer's and editor's comments were also very helpful for improving our manuscript.

Author Contributions

Austin R. Manny: Conceptualized project, designed methodology, conducted reference-based and *de novo* transcriptome mining, cultured *T. vaginalis* isolate G3, isolated RNA and enriched for dsRNA, performed data analysis, prepared original manuscript, prepared figures, edited manuscript

Carrie A. Hetzel: Performed data analysis, validated findings, carried out RNA structure prediction investigations, prepared figures, edited manuscript

Arshan Mizani: Performed data analysis, validated findings, edited manuscript

Max L. Nibert: Conceptualized project, designed methodology, conducted reference-based sequence mining, prepared original manuscript, edited manuscript, supervision, acquired funding

References

Bessarab, I.N.; Liu, H.-W.; Ip, C.-F.; Tai, J.-H. The complete cDNA sequence of a type II Trichomonas vaginalis virus. *Virology* 2000, 267, 350–359

Bessarab, I.N.; Nakajima, R.; Liu, H.-W.; Tai, J.-H. Identification and characterization of a type III Trichomonas vaginalis virus in the protozoan pathogen *Trichomonas vaginalis*. *Arch. Virol.* 2010, 156, 285–294

Bradic, M.; Warring, S.D.; Tooley, G.E.; Scheid, P.; Secor, W.E.; Land, K.M.; Huang, P.-J.; Chen, T.-W.; Lee, C.-C.; Tang, P.; et al. Genetic indicators of drug resistance in the highly repetitive genome of *Trichomonas vaginalis*. *Genome Biol. Evol.* 2017, 9, 1658–1672

Buchfink, B.; Reuter, K.; Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* 2021, 18, 366–368

Bushmanova, E.; Antipov, D.; Lapidus, A.; Przhibelskiy, A.D. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *bioRxiv* 2018, 420208

Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: architecture and applications. *BMC Bioinform.* 2009, 10, 421

Carlton, J.M.; Hirt, R.P.; Silva, J.C.; Delcher, A.L.; Schatz, M.; Zhao, Q.; Wortman, J.R.; Bidwell, S.L.; Alsmark, U.C.M.; Besteiro, S.; et al. Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* 2007, 315, 207–212

Chiba, S.; Jamal, A.; Suzuki, N. First evidence for internal ribosomal entry sites in diverse fungal virus genomes. *mBio* 2018, 9, e02350-17

Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al., Twelve years of SAMtools and BCFtools, *Gigascience* 2021, 10, giab008

Fichorova, R.N.; Lee, Y.; Yamamoto, H.S.; Takagi, Y.; Hayes, G.R.; Goodman, R.P.; Chepa- Lotrea, X.; Buck, O.R.; Murray, R.; Kula, T.; et al. Endobiont viruses sensed by the human host— beyond conventional antiparasitic therapy. *PLoS ONE* 2012, 7, e48418

Fraga, J.; Rojas, L.; Sariego, I.; Fernández-Calienes, A. Genetic characterization of three Cuban Trichomonas vaginalis virus. Phylogeny of *Totiviridae* family. *Infect. Genet. Evol.* 2012, 12, 113– 120

Fouts, A.C.; Kraus, S.J. *Trichomonas vaginalis*: reevaluation of its clinical presentation and laboratory diagnosis. *J. Infect. Dis.* 1980, 141, 137–143

Garlapati, S.; Wang, C.C. Structural elements in the 5′-untranslated region of giardiavirus transcript essential for internal ribosome entry site-mediated translation initiation. *Eukaryot. Cell* 2005, 4, 742–754

54

Goodman, R.P.; Ghabrial, S.A.; Fichorova, R.N.; Nibert, M.L. *Trichomonasvirus*: a new genus of protozoan viruses in the family *Totiviridae*. *Arch. Virol.* 2011(a), 156, 171–179

Goodman, R.P.; Freret, T.S.; Kula, T.; Geller, A.M.; Talkington, M.W.T.; Tang-Fernandez, V.; Suciu, O.; Demidenko, A.A.; Ghabrial, S.A.; Beach, D.H.; et al. Clinical Isolates of *Trichomonas vaginalis* concurrently infected by strains of up to four trichomonasvirus species (family *Totiviridae*). *J. Virol.* 2011(b), 85, 4258–4270

Gould, S.B.; Woehle, C.; Kusdian, G.; Landan, G.; Tachezy, J.; Zimorski, V.; Martin, W.F. Deep sequencing of *Trichomonas vaginalis* during the early infection of vaginal epithelial cells and amoeboid transition. *Int. J. Parasitol.* 2013, 43, 707–719

Gribble, J.; Stevens, L.J.; Agostini, M.L.; Anderson-Daniels, J.; Chappell, J.D.; Lu, X.; Pruijssers, A.J.; Routh, A.L.; Denison, M.R. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. *PLOS Pathog.* 2021, 17, e1009226

Heidary, S.; Bandehpour, M.; Valadkhani, Z.; Seyyed–Tabaee, S.; Haghighi, A.; Abadi, A.; Kazemi, B. Double-stranded RNA viral infection in Tehran Trichomonas vaginalis isolates. *Iran. J. Parasitol.* 2013, 8, 60–64.

Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 2018, 35, 518–522

Huang, X.; Madan, A. CAP3: a DNA sequence assembly program. *Genome Res.* 1999, 9, 868– 877.

Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 2017, 14, 587–589

Katoh, K.; Kuma, K.-i.; Toh, H.; Miyata, T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 2005, 33, 511–518. https://doi.org/10.1093/nar/gki198.

Kim, J.W.; Chung, P.-R.; Hwang, M.-K.; Choi, E.Y. Double-stranded RNA virus in Korean isolate IH-2 of *Trichomonas vaginalis*. *Korean J. Parasitol.* 2007, 45, 87

Leitsch, D. Recent advances in the *Trichomonas vaginalis* field. *F1000Research* 2016, 5, 162

Lemoine, F.; Domelevo Entfellner, J.-B.; Wilkinson, E.; Correia, D.; Dávila Felipe, M.; De Oliveira, T.; Gascuel, O. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* 2018, 556, 452–456

Lorenz, R.; Bernhart, S.H.; Höner zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 2011, 6, 26

Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 2009, 25, 2078–2079

Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013, arXiv:1303.3997.

55

Madeira, F.; Park, Y.M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A.R.N.; Potter, S.C.; Finn, R.D.; et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.* 2019, 47, W636–W641

Men-Fang, S.; Pey-Ru, L.; Chung-Shinn, L. Killing of *Trichomonas vaginalis* by complement- mediated lysis is not associated with the presence of Trichomonas vaginalis virus. *Int. J. Parasitol.* 1993, 23, 675–680

Mercer, F.; Johnson, P.J. *Trichomonas vaginalis*: pathogenesis, symbiont interactions, and host cell immune responses. *Trends Parasitol.* 2018, 34, 683–693

Minin, V.N.; Dorman, K.S.; Fang, F.; Suchard, M.A. Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* 2005, 21, 3034–3042

Nibert, M.L.; Vong, M.; Fugate, K.K.; Debat, H.J. Evidence for contemporary plant mitoviruses. *Virology* 2018, 518, 14–24

Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 2015, 32, 268–274

R Core Team. *R Foundation for Statistical Computing.* Available online: https://www.r-project.org/ (accessed on 28 February 2022)

Rivera, W.L.; Justo, C.A.C.; Relucio-San Diego, M.A.C.V.; Loyola, L.M. Detection and molecular characterization of double-stranded RNA viruses in Philippine *Trichomonas vaginalis* isolates. *J. Microbiol. Immunol. Infect.* 2017, 50, 669–676

Song, M.-J.; Kim, M.; Choi, Y.; Yi, M.; Kim, J.; Park, S.-J.; Yong, T.-S.; Kim, H.-P. Epigenome mapping highlights chromatin-mediated gene regulation in the protozoan parasite *Trichomonas vaginalis*. *Sci. Rep.* 2017, 7, 45365

Su, H.-M.; Tai, J.-H. Genomic organization and sequence conservation in type I Trichomonas vaginalis viruses. *Virology* 1996, 222, 470–473

Tai, J.-H.; Ip, C.-F. The cDNA sequence of Trichomonas vaginalis virus-T1 double-stranded RNA. *Virology* 1995, 206, 773–776

Tao, Y.; Farsetta, D.L.; Nibert, M.L.; Harrison, S.C. RNA synthesis in a cage — structural studies of reovirus polymerase λ3. *Cell* 2002, 111, 733–745

Tarr, P.I.; Aline, R.F.; Smiley, B.L.; Scholler, J.; Keithly, J.; Stuart, K. LR1: A candidate RNA virus of *Leishmania*. *Proc. Natl. Acad. Sci. USA* 1988, 85, 9572–9575

Wang, A.L.; Wang, C.C. The double-stranded RNA in *Trichomonas vaginalis* may originate from virus-like particles. *Proc. Natl. Acad. Sci. USA* 1986, 83, 7956–7960

Wang, A.; Wang, C.C.; Alderete, J.F. *Trichomonas vaginalis* phenotypic variation occurs only among trichomonads infected with the double-stranded RNA virus. *J. Exp. Med.* 1987, 166, 142– 150

Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: New York, NY, USA, 2016.

Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the tidyverse. *J. Open Source Softw.* 2019, 4, 1686

Woehle, C.; Kusdian, G.; Radine, C.; Graur, D.; Landan, G.; Gould, S.B. The Parasite *Trichomonas vaginalis* expresses thousands of pseudogenes and long non-coding RNAs independently from functional neighbouring genes. *BMC Genomics* 2014, 15, 906

Wu, B.; Zhang, X.; Gong, P.; Li, M.; Ding, H.; Xin, C.; Zhao, N.; Li, J. *Eimeria tenella*: a novel dsRNA virus in *E. tenella* and its complete genome sequence analysis. *Virus Genes* 2016, 52, 244–252

Zhu, Z.; Meng, K.; Meng, G. Genomic recombination events may reveal the evolution of coronavirus and the origin of SARS-CoV-2. *Sci. Rep.* 2020, 10, 21617

APPENDIX 2

Systematic characterization of dsRNA viruses in commercially available isolates of the human parasite *Trichomonas vaginalis*

Austin R. Manny[1,2], Carrie A. Hetzel[1,2], Max L. Nibert[1,2]

[1]Department of Microbiology, Blavatnik Institute, Harvard Medical School, Boston, MA 02115,

[2]Program in Virology, Division of Medical Sciences, Graduate School of Arts & Sciences, Harvard University, Cambridge, MA 02138

The work in this chapter represents a preliminary draft of a manuscript that will be submitted for journal publication in the near future.

Abstract

Trichomonas vaginalis is a widespread protozoan parasite that infects upwards of 250 million people per year globally. *T. vaginalis* is the most common non-viral sexually transmitted infection which presents clinically as trichomoniasis. This inflammatory syndrome can manifest with both acute and chronic consequences. Furthermore, *T. vaginalis* can be infected by up to five distinct dsRNA viruses, *Trichomonas vaginalis virus 1* through *Trichomonas vaginalis virus 5* (or, TVVs). These viruses comprise the genus *Trichomonasvirus* within family *Totiviridae*. One of these viruses, namely TVV1, has been shown to exacerbate *T. vaginalis*-mediated disease. Viral co-infections are exceedingly common in *T. vaginalis* and preclude reductionist study of the impact of each virus. Inconsistencies in the literature about which *T. vaginalis* isolates harbor which trichomonasviruses further impede our understanding of these viruses. We set out to systematically characterize the viruses infecting every commercially available isolate of *T. vaginalis* from the ATCC using an approach of high-throughput sequencing of RNA enriched for dsRNA molecules (dsRNA-seq). This resulted in the assembly of 39 new coding-complete TVV genomes from 16 isolates of *T. vaginalis*. Excitingly, five of these commercially available isolates were found to harbor TVV5. A subset of the isolates was subjected to 3' Rapid Amplification of cDNA Ends (3'RACE), resulting in the determination of the full-length TVV5 genome. In doing so, we discovered an unexpected terminal nucleotide at each end of the TVV5 dsRNA genome. Strains of TVV1-4 were then subjected to the same line of experimentation, revealing that the additional terminal residues were present in all other TVV species as well, thus redefining the genomes of all trichomonasviruses. Taken together, this work vastly expands the number of coding-

complete trichomonasvirus genome sequences, redefines the genomes of all trichomonasviruses, and for the first time identifies commercially available isolates of *T. vaginalis* that harbor a recently discovered fifth species of trichomonasvirus.

Introduction

Trichomonas vaginalis is the leading non-viral sexually transmitted infection around the world (Mercer & Johnson, 2018; World Health Organization, 2011). Upon a person's exposure to *T. vaginalis*, the protozoan parasite colonizes the urogenital epithelium where it can form an acute or chronic infection. Infections with *T. vaginalis* can be symptomatic or asymptomatic, presenting as the inflammatory syndrome trichomoniasis. Infection with *T. vaginalis* can also yield higher- order, long-term complications such as increased risk of acquisition of HPV and HIV, heightened chance of preterm labor, and pelvic inflammatory disease. (Fichorova et al., 2009; Mielczarek & Blaszkowska, 2016; Ghosh et al., 2017; Nakubulwa et al., 2015). While the parasite is generally sensitive to treatment with the oral antiparasitic metronidazole, the emergence of drug-resistant isolates of *T. vaginalis* is on the rise (Leitsch, 2016).

T. vaginalis can itself be infected by one or more dsRNA viruses (Goodman et al., 2011a). These viruses, *Trichomonas vaginalis virus 1* through *Trichomonas vaginalis virus 5* (or, TVVs) comprise the genus *Trichomonasvirus* within family *Totiviridae* (Goodman et al., 2011b; Manny et al., 2022). These dsRNA viruses form persistent infections of their host parasite. One of these viruses, TVV1, was shown to exacerbate the inflammatory disease caused by its host parasite (Fichorova et al., 2012). TVV1 particles can be released from its parasite host and be taken up by human epithelial cells, where the viral dsRNA genome is detected by the innate immune sensor TLR3. A proinflammatory antiviral immune cascade is then launched, which heightens the inflammatory environment elicited by *T. vaginalis* (Fichorova et al., 2012). In this way, these viral symbionts can contribute to *T. vaginalis*-mediated disease. Contributions to

pathogenesis by the other trichomonasviruses remain undetermined. It was previously reported that a single clonally purified strain of *T. vaginalis* could be simultaneously infected with TVV1-4, highlighting the co- infections that are common with TVVs (Goodman et al., 2011a). Recently, a fifth species of trichomonasvirus was discovered using a transcriptome mining approach whereby publicly available RNA-seq datasets were mined for trichomonasvirus sequences (Manny et al., 2022). In doing so, a distantly related virus was identified in six isolates of *T. vaginalis*. This novel virus, TVV5, contains the largest genome of any trichomonasvirus yet identified, at over 5 kilobases while TVV1-4 possess genomes of 4.6-4.9 kb (Goodman et al., 2011b; Manny et al., 2022). While TVV5 was found in the transcriptomes of six different patient isolates, no readily accessible isolate of *T. vaginalis* harboring TVV5 has been reported to date.

In the three decades since the first trichomonasvirus was identified, many groups have undertaken surveys to determine the prevalence of these viruses in clinical isolates of *T. vaginalis* (Wendel et al., 2002; Weber et al., 2003; Kim et al., 2007; Fraga et al., 2012; Heidary et al., 2013; El-Gayar et al., 2016; Rivera et al., 2017; Margarita et al., 2019). While the proportion of virus- infected isolates within a single study does vary, in aggregate approximately half of *T. vaginalis* isolates have been found to harbor one or more trichomonasviruses. These studies are helpful in providing a baseline understanding of the prevalence of the different TVVs. However, as the samples are typically collected in a clinical setting, these isolates are not readily accessible for further investigation. Furthermore, these clinical survey papers typically rely on polymerase chain reaction methodologies for detection; if viral sequences are reported, they are generally only a small amplicon of approximately 500 bp within a conserved region of the viral

genome. As a result, the number of TVV genomes deposited in public databases remains limited. Recent studies to mine viral sequences from transcriptome data have been successful at generating new virus sequences for study (e.g., Manny et al. 2022, Nibert et al. 2018). We sought to build upon this and determine trichomonasvirus sequences from commercially available isolates of *T. vaginalis* using high-throughput sequencing of RNA enriched for dsRNA molecules (dsRNA-seq). In this way, the genomes of 39 new strains of trichomonasvirus were determined, including five strains of the recently discovered *Trichomonas vaginalis virus 5*.

Materials and Methods

Selection Criteria, Ordering, and Culturing T. vaginalis Isolates

All readily available isolates of T. vaginalis were purchased from the American Type Culture Collection (ATCC), totaling 34 isolates. Eight additional isolates belonging to limited or special collections not readily available were not purchased. Of the 34 isolates obtained, 2 were excluded due to insufficient sample-associated metadata. One of these isolates, T. vaginalis G3, was of interest for, and reported in, our previous study (Manny et al., 2022). The analyses in this current study include viruses from this isolate. 32 isolates were thus investigated in the present study. All isolates were shipped on dry ice and maintained in liquid nitrogen storage until ready to use. Each isolate was minimally cultured in 15 mL of Diamond's Modified Medium (Fouts et al., 1980) for 2-4 days until late-logarithmic growth was reached. Half of each sample was aliquoted and moved to liquid nitrogen storage. The remaining half was used for immediately for RNA isolation.

RNA Isolation and dsRNA Enrichment

After minimal culture, each sample was resuspended in 10 mL of TRIzol reagent (Invitrogen, Waltham, MA, USA) to carry out RNA isolation using the standard phenol-chloroform extraction with RNA ultimately resuspended in ultrapure water. ssRNA was depleted via incubation with 2M LiCl at -20℃ overnight. The sample was centrifuged at 16100 x g for 20 minutes at 4℃ to pellet ssRNA. The resulting supernatant was then further enriched for dsRNA using cellulose affinity chromatography. 175 mg of cellulose MN 301 (Macherey-Nagel, Düren, Germany) was equilibrated in NaCl-Tris-EDTA (STE) buffer with 16% ethanol. The ssRNA-depleted sample was then added to the cellulose solution, applied to a Wizard Plus Midicolumn (Promega, Madison, WI, USA), washed several times with STE buffer containing ethanol, and purified dsRNA was eluted with the addition of 400 µL of STE buffer without ethanol. Enriched dsRNA was concentrated using isopropanol and resuspended in 15 µL ultrapure water. Presence of one or more trichomonasviruses was assayed by agarose gel electrophoresis for detection of a distinct 5 kb band corresponding to the viral dsRNA genome. dsRNA concentrations were determined using the Qubit RNA HS assay (Invitrogen, Waltham, MA, USA) to ensure sufficient RNA for sequencing. Isolates screened as virus-negative were confirmed by running total RNA on a 1% agarose gel and corroborating the absence of viral genomes.

High-throughput RNA Sequencing

The enriched dsRNA samples from all virus-infected isolates were shipped overnight on dry ice to Quick Biology (Pasadena, CA, USA) for high-throughput sequencing. The RNA-Seq library was prepared according to the KAPA KK8540 RNA

HyperPrep kit with an insert size of 201–300 bp (KAPA Biosystems, Wilmington, MA, USA) using 25–50 ng of total dsRNA as input. The quality and quantity of each prepared sequencing library was analyzed with the Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) and Qubit 3.0 Fluorometer assay (Invitrogen, Waltham, MA, USA). Paired-end 150 bp reads were sequenced on an Illumina (San Diego, CA, USA) HiSeq 4000, targeting 10 million reads per sample. Sequences were demultiplexed to remove barcodes at the sequencing facility.

Bioinformatic Analysis of dsRNA-sequencing Data

The sequencing reads were analyzed using a custom-built pipeline constructed using the Nextflow framework: tvv-nf. Briefly, demultiplexed reads are taken as input and the pipeline filters low-quality reads, trims off adapter sequences, depletes reads mapping to the genome of the bacteriophage ΦX174 (spiked in to ensure library complexity), and maps the cleaned sequencing reads to reference genomes of TVV1, TVV2, TVV3, TVV4, and TVV5. After running the pipeline, binned reads were then separately de novo assembled into coding-compete viral genomes and refined using Geneious Prime version 2021.1 (Biomatters Ltd., Auckland, New Zealand). Filtering and trimming steps were performed using TrimGalore v0.6.3 (https://github.com/ FelixKrueger/TrimGalore; accessed on 13 July 2019). Depletion of ΦX174 reads was achieved using BWA-MEM (Li, 2013) in paired-end mode. Mapping and binning to TVV reference genomes was performed using BBSplit v38.57 (https://sourceforge.net/projects/bbmap/; accessed 16 July 2019) . To clean up the binned reads, each set of binned reads was analyzed using a local instance of BLAST (Camacho et al., 2009), using blastn with an e-value threshold

of 1e-3. Binned reads were queried against a database of the reference TVV sequences. Only reads initially binned to a specific TVV species and again assigned by BLAST to that same species were retained. Reference genomes used by BBSplit and BLAST for TVV1-4 were obtained from NCBI GenBank with the following accessions: TVV1: U08999.1, DQ270032.1, HQ607516.1, HQ607513.1, HQ607517.1, JF436869.1; TVV2: NC_003873.1, HQ607514.1, HQ607518.1, HQ607524.1, JF436870.1, JF436871.1; TVV3: NC_004034.1, HQ607515.1, HQ607519.1, HQ607525.1; TVV4 HQ607522.1, HQ607520.1, HQ607526.1. Reference sequences for TVV5 were TVV5-NYCA04, TVV5-NYCC37, TVV5-NYCD15, and TVV5-NYCE32 determined by Manny et al. 2022.

Assembly and refinement of viral genomes was conducted using Geneious. Assembly was performed on the reads binned to each virus using the 'De Novo Assemble' function with the Geneious assembler with sensitivity set to 'High Sensitivity / Medium'. Refinement was achieved by iteratively mapping the binned reads to the assembled genome using Bowtie2 (Langmead & Salzberg, 2012) and producing a consensus genome using the 'Generate consensus sequence' function with the threshold set to 'Highest Quality (50%)'. Unmapped reads not corresponding to any TVV were analyzed by the pipeline for detection of other viruses. Unassembled unmapped reads were compared to other viruses at the nucleotide level using Kraken2 v2.0.9-beta (Wood et al., 2019) against a database constructed from the genomes of all viruses in the NCBI Viruses portal, along with the reference sequences for TVV1-5 and the transcriptome of *T. vaginalis* (TrichDB v46, retrieved from www.trichdb.org, (Carlton et al., 2007; Amos et al., 2021)). Unmapped reads were also de novo assembled using rnaSPAdes v3.14.0 (Bushmanova et al., 2018), and any viral assemblies were identified with DIAMOND

v2.0.2 (Buchfink et al., 2021) in ‘blastx’ mode against the NCBI nonredundant (‘nr’) protein database.

All code for the tvv-nf pipeline detailing the software and parameters, and well as reference sequences, have been made freely available at the tvv-nf repository at www.github.com/austinreidmanny/tvv-nf (last accessed April 20, 2022). A file named ‘conda.yml’ is available within the repository that can be used by the Conda package manager (Anaconda Inc., Austin, TX, USA) to install all software used in the tvv-nf pipeline. A schematic of the tvv-nf pipeline can be seen in Figure A2.1 and was made using Whimsical (http://whimsical.com, last accessed May 3, 2022).

Sequencing of Genomic Termini Using 3’ Rapid Amplification of cDNA Ends (3’RACE)

A subset of dsRNA samples were subjected to dedicated genomic terminal sequencing using 3’RACE. DNA primers and RNA oligonucleotide adapters (Table S1) were ordered from IDT (Integrated DNA Technologies Inc., Coralville, IA, USA). The adapter sequence was designed with a modified 5’ end containing a terminal phosphate group to facilitate the enzymatic reaction and a modified 3’ end containing a terminal amino group to ensure addition of only a single adapter onto a given RNA molecule. dsRNA was denatured to yield ssRNA by heating to 95℃ for 5 minutes followed by immediately placing on ice for 2 minutes. The adapter was then ligated to the viral RNA using T4 RNA ligase enzyme (New England Biolabs, Ipswitch, MA, USA). Polymerase chain reaction was then carried out using an adapter-complementary terminal primer and an internal primer designed against a previously determined region of the viral genome. PCR products were subjected to agarose gel electrophoresis and amplicons were excised

from the 1% agarose gel using the Monarch DNA Gel Excision Kit (New England Biolabs, Ipswitch, MA, USA). The sequence of the resulting amplicon was determined using Sanger sequencing with the genome-internal primer by Azenta Life Sciences (South Plainfield, NJ, USA). In later confirmatory rounds of 3'RACE experimentation, T4 RNA ligase was freshly sourced from an alternative vendor, Promega Corporation (Madison, WI, USA) and the 3'RACE adapter was reordered from IDT (Coralville, IA, USA).

Transcriptome Mining for Additional TVV5 Sequences

The virus discovery platform Serratus (Edgar et al., 2022) was used to search for transcriptome datasets that could harbor additional strains of TVV5. Serratus systematically catalogues viral sequences contained in transcriptomes in the NCBI Sequence Read Archive and provides an online dashboard to query these results. Datasets were screened as possibly possessing TVV5 sequences if they contained sequences matching TVV1-4. To conduct this screen, Serratus was used in RdRp mode to search for datasets with reads assigned to TVV1, TVV2, TVV3, or TVV4 with a sequence identity of 75-100% and a score of 50-100. Datasets fitting these criteria were analyzed. Datasets that were previously analyzed in Manny et al. 2022 were excluded. One dataset, BioProject PRJNA473698, was selected for further investigation. Transcriptomes from the same patient and sample type were combined to mine virus sequences. Sequencing reads were mapped to all TVV1-5 genomes determined in this study using a local instance of nucleotide blast with task set to 'discontiguous-megablast' and an e-value threshold of 1e-9. These hits were then cleaned up by remapping to all TVV dsRNA-seq genomes using blastn with an e-value threshold of 10, as described

above. Each set of TVV reads was retrieved, de novo assembled, and refined to yield viral sequence assemblies. These assembly and refinement steps were conducted using Geneious Prime, by mapping to a reference TVV strain, generating a 50% consensus sequence, and reiteratively mapping the read set to yield a final coding-complete, gapless viral genome.

Results

Determination of Trichomonasvirus Genomes from Readily Available Isolates of T. vaginalis

All readily purchasable isolate of *Trichomonas vaginalis* were obtained directly from the American Type Culture Collection (ATCC), a publicly accessible repository of biological reagents. Two of these isolates were rejected due to insufficient sample information, resulting in a sample set of 32 isolates. Each isolate was minimally cultured to late-logarithmic phase before isolating total RNA. Single-stranded RNA was depleted by overnight incubation with 2M lithium chloride, and dsRNA was further enriched using cellulose affinity chromatography. Purified dsRNA was assayed for presence of one or more trichomonasvirus by detection of a characteristic 5 kb dsRNA genome band using agarose gel electrophoresis. Of the 32 isolates of *T. vaginalis,* 16 isolates were screened as virus-infected while the other half were determined to be virus-negative. This proportion of isolates harboring TVV is consistent with previous reports (Weber et al., 2003; Kim et al., 2007; Fraga et al., 2012; Heidary et al., 2013; El-Gayar et al., 2016; Rivera et al., 2017; Margarita et al., 2019). The dsRNA from each virus-infected isolate was subjected to high- throughput sequencing, yielding on average 10 million read pairs

per sample. The resulting dsRNA-seq data was processed through a custom-built virus-discovery pipeline, tvv-nf (Figure A2.1). Briefly, this pipeline takes in the dsRNA-seq reads, performs a series of filtering and trimming steps, and maps the reads to TVV genomes. These binned reads were then separately assembled de novo into viral genomes for downstream analysis.

In total, coding-complete genomes were determined for 39 new strains of trichomonasviruses. Top-level characteristics of these sequences are provided in Table A2.1. This represents 13 new genome sequences of TVV1, 8 new genome sequences of TVV2, 10 new genome sequences of TVV3, and 3 new genomes of TVV4. Excitingly, five new genome sequences were also determined for the recently discovered TVV5, denoting the first set of commercially available isolates to harbor the novel virus. Coding-complete genomes were assembled for all viruses identified; no partial assemblies were generated for any virus. No viruses beyond those belonging to genus *Trichomonasvirus* were found.

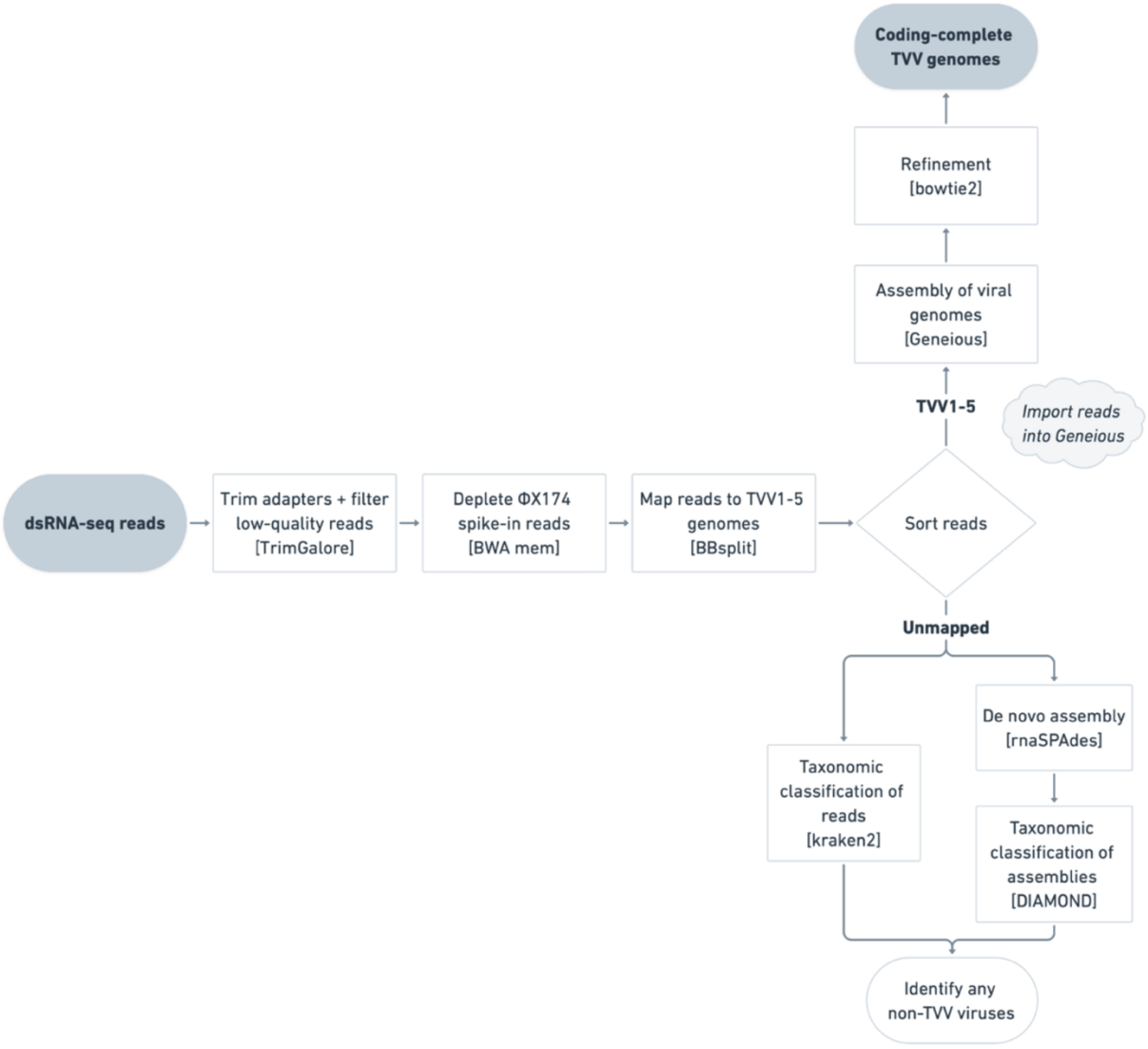


Figure A2.1: The dsRNA-seq analysis pipeline, tvv-nf. This schematic delineates the sequencing analysis pipeline, tvv-nf, used to transform the initial dsRNA-seq reads into coding- complete TVV genomes. This pipeline takes in the raw sequencing reads, trims adapter sequences, filters low-quality reads, depletes reads corresponding to spiked-in ΦX174 DNA, maps reads to reference sequences of TVV1-5, exports TVV reads for downstream assembly and refinement, and assesses unmapped reads for presence of viruses beyond trichomonasviruses.

Table A2.1: Novel trichomonasvirus genomes determined in this study and key characteristics of the sequences. Virus strains subjected to 3'RACE are bolded. CP: capsid protein, CP/RdRp: capsid/RNA-dependent RNA polymerase protein, UTR: untranslated region.

Virus	Genome (bp)	CP (aa)	CP/RdRp (aa)	5'-UTR (bp)	3'-UTR (bp)
TVV1-165307-1	**4689**	**678**	**1430**	**327**	**71**
TVV1-RU357	**4681**	**678**	**1430**	**323**	**67**
TVV1-123414	4674	678	1430	315	68
TVV1-CDC-337	4673	678	1430	316	66
TVV1-JH31A4	4673	678	1430	317	65
TVV1-JH37A2	4672	678	1430	314	67
TVV1-JH162A4	4674	678	1430	316	67
TVV1-JH191A4	4671	678	1430	317	63
TVV1-JRS-TV-120	4670	678	1430	319	60
TVV1-NYH209	4664	678	1430	312	61
TVV1-NYH272	4673	678	1430	316	66
TVV1-NYH286	4668	678	1430	318	59
TVV1-RU393	4677	678	1430	317	69
TVV2-165307-1	**4675**	**709**	**1436**	**298**	**67**
TVV2-RU357	**4677**	**709**	**1436**	**299**	**68**
TVV2-123414	4671	709	1436	294	67
TVV2-C1NIH	4664	709	1436	288	66
TVV2-G3-HMS	4656	709	1436	287	59
TVV2-JH162A4	4671	709	1436	297	64
TVV2-NYH209	4659	709	1436	284	65
TVV2-NYH286	4664	709	1436	288	66
TVV3-165307-1	**4850**	**708**	**1443**	**364**	**155**
TVV3-C1NIH	4831	708	1443	354	146
TVV3-CDC-337	4835	708	1443	350	154
TVV3-G3-HMS	4820	708	1443	344	145
TVV3-JH32A4	4829	708	1443	352	146
TVV3-JH162A4	4828	708	1443	348	149
TVV3-JRS-TV-120	4822	708	1443	347	144
TVV3-NYH209	4836	708	1443	360	145
TVV3-NYH286	4828	708	1443	352	145
TVV3-RU393	4834	708	1443	356	147
TVV4-RU357	**4908**	**734**	**1469**	**338**	**161**
TVV4-JRS-TV-120	4929	746	1481	329	155
TVV4-NYH272	4848	726	1461	309	154
TVV5-165307-1	**5031**	**763**	**1500**	**348**	**181**
TVV5-JRS-TV-120	**5049**	**769**	**1506**	**349**	**180**
TVV5-RU357	**5034**	**783**	**1520**	**292**	**180**
TVV5-JH162A4	5041	769	1506	343	178
TVV5-NYH209	5028	784	1521	291	172

Defining the First Full-length Genomes of TVV5 Using 3'RACE

With the identification of five isolates harboring TVV5 in our sample set, we sought to determine the full-length genome of *Trichomonas vaginalis virus 5* using dedicated sequencing of genomic termini. The extreme ends of dsRNA viruses can play critical roles in the viral lifecycle, often containing functional elements such as packaging signals or secondary structures important for recognition by the viral RNA-dependent RNA polymerase (RdRp). One technique, 3' Rapid Amplification of cDNA Ends (or 3'RACE), has been used by several groups, including ours, to determine the full-length genome of dsRNA viruses including the extreme terminal ends which can be missed by high-throughput sequencing. In this approach, RNA oligonucleotide adapters of known sequence are ligated to the 3' end of each strand of the double-stranded viral genome using T4 RNA ligase. Polymerase chain reaction is then carried out using an adapter-complementary terminal primer and an internal primer designed against the previously determined region of the viral genome. The sequence of the resulting amplicon is finally determined using Sanger sequencing. In this way, both ends of the dsRNA viral genome can be determined with high precision using the same technique. For simplicity, we refer to the 5' and 3' termini of the coding-strand of the viral genome rather than the 3' ends of each strand. Previous reports identified conserved 5'-GC and UC-3' dinucleotide terminal motifs for all known trichomonasviruses up to that point in time (i.e., TVV1-4), and TVV5 was expected to share those conserved sequences (Goodman et al., 2011a).

We subjected dsRNA from *T. vaginalis* isolate 165307-1 to 3'RACE sequencing, enabling the determination of the first full-length genome of TVV5. The genome size of TVV5 was reported as 5026 bp for the prototypical TVV5-NYCE32. This 3'RACE

approach uncovered an additional 12 bp at the 5’ terminus and additional 19 bp at the 3’ terminus, thus yielding a TVV5 genome size of 5057 bp, the longest trichomonasvirus genome reported to date. The genomic termini of TVV5 were expected to be comprised of the conserved 5’-GC and UC-3’ nucleotide motifs. However, one additional nucleotide was observed at each strand of the viral genome. Indeed, the newly determined coding strand of the genome was found to begin with 5’-UGC and end with UCA-3’. To preclude any artifactual results, this experiment was repeated using freshly ordered primers and RNA adapters, and with T4 RNA ligase purchased from another supplier; we initially used enzyme from New England Biolabs (Ipswitch, MA, USA) and then repeated the experiment with enzyme from Promega (Madison, WI, USA). The characteristic trinucleotide motifs were observed once again. To further substantiate this finding, the strain of TVV5 from isolates JRS- TV-120 and RU357 were likewise subjected to this line of experimentation. TVV5-JRS-TV-120 was found to possess both trinucleotide motifs as well. TVV5-RU357 was found to possess the UCA-3’ motif. However, the 5’ terminus was found to begin with the trinucleotide 5’-CGC, maintaining an additional terminal residue—and that residue being a pyrimidine—but with a cytosine in place of uracil.

Redefining the genomic termini of all trichomonasviruses using 3’RACE

Given the unexpected findings from determining the full-length genomes of multiple TVV5 strains, this approach was extended to strains of the four other trichomonasvirus species. Two strains of TVV1, two strains of TVV2, and one strain each of TVV3 and TVV4 were subjected to 3’RACE experimentation (Figure A2.2). All viruses (TVV1-165307-1, TVV1-RU357, TVV2-165307- 1, TVV2-RU357, TVV3-165307-1, and TVV4-RU357) were found to possess not the previously reported dinucleotide motifs but

instead the newly discovered 5'-UGC and UCA-3' motifs. Interestingly, no virus was observed to have the 5'-CGC sequence aside from TVV5-RU357. Given the same repeated observation in viral strains from all TVV species and from multiple isolates of *T. vaginalis*, we conclude that the newly identified trinucleotide terminal motifs of 5'- YGC (where Y is a pyrimidine, namely cytosine or ucacil) and UCA-3' are likely to be conserved across trichomonasviruses generally and that previously reported trichomonasvirus genomes are likely incomplete.

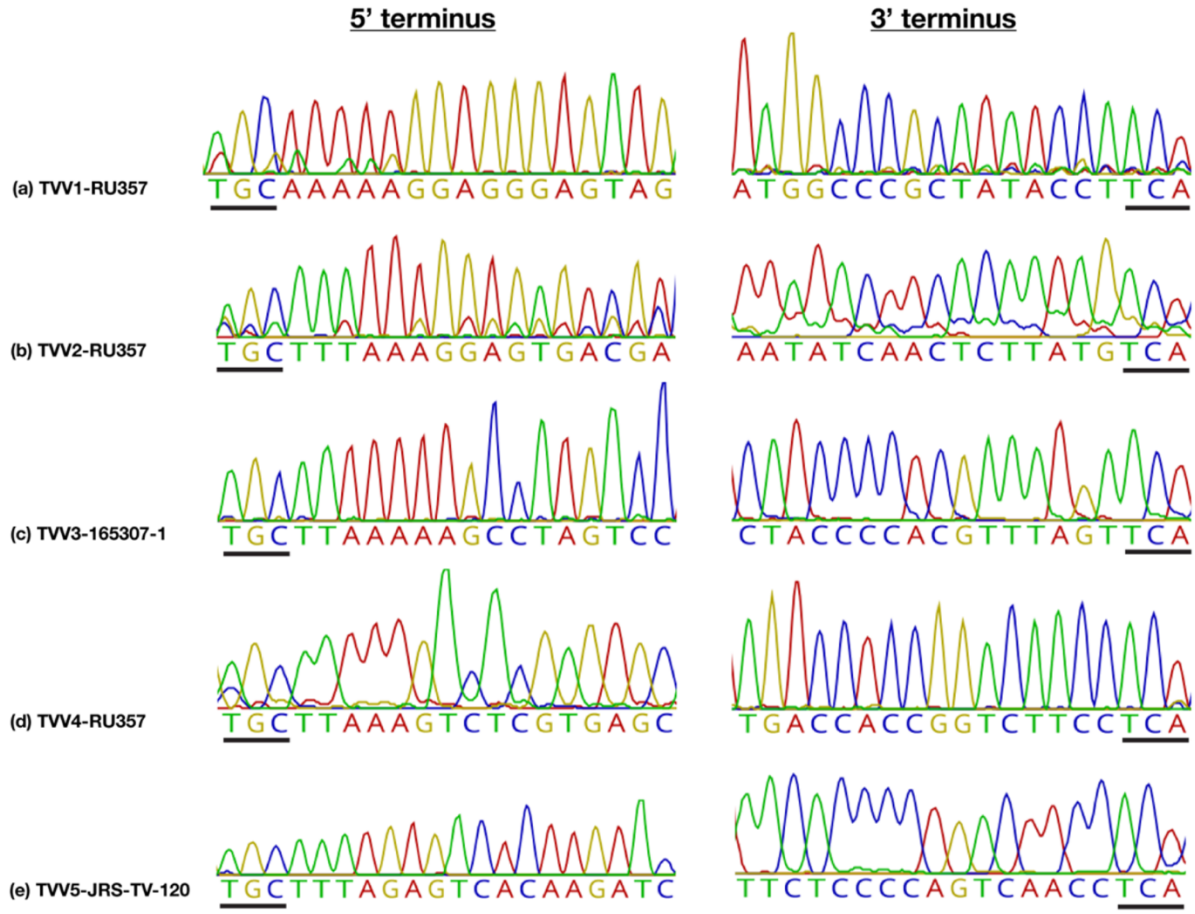


Figure A2.2: Chromatograms of previously undescribed genomic terminal trinucleotide motifs in trichomonasviruses. Sequences of one strain per species are presented for (a) TVV1, (b) TVV2, (c) TVV3, (d) TVV4, and (e) TVV5, corresponding to the 5' and 3' termini of the coding strand of the viral genome. Conserved trinucleotide motifs at genomic termini are indicated by black bars.

Analysis and characterization of new trichomonasvirus sequences

As listed above, 39 coding-complete trichomonasvirus genomes were determined in the course of this study. Patterns observed by previous studies of TVV genomes are maintained here. All of the genomes exhibit the characteristic trichomonasvirus protein coding strategy of two long open reading frames with a small overlap which harbors the programmed ribosomal frameshifting site. The lengths of each strain's genome, protein products, 5'-UTR, and 3'-UTR are provided in Table A2.1. TVV5 strains were found to possess the longest genomes, all in excess of 5 kb, as well as the longest CP/RdRp fusion proteins. Determination of at least one full-length genome per viral species using 3'RACE enabled the examination of 5'-UTR and 3'-UTR lengths. When compared to the genomic boundaries determined by 3'RACE, the dsRNA-seq genomes are all nearly full-length, with all sequences determined to within 17 bp of the 5'-terminus and to within 10 bp of the 3'-terminus (Figure A2.3). TVV3 strains possess the longest 5'-UTRs, and TVV5 strains possess the longest 3'-UTRs. TVV3 and TVV4 exhibit UTRs of similar sizes, potentially attributable to their phylogenetic relationship as sister clades, but the UTRs of TVV2 and TVV5 are appreciably different despite sharing a similar phylogenetic association.

All strains of TVV1, TVV2, and TVV3 were found to possess a singular, signature length for the capsid (CP) and polymerase (CP/RdRp). TVV1 strains were found to encode a CP of 678 aa and CP/RdRp of 1430 aa. TVV2 strains encoded a CP of 709 aa and CP/RdRp of 1436 aa. TVV3 strains encoded a CP of 708 aa and CP/RdRp of 1443 aa. In contrast, the protein products of TVV4 and TVV5 did not exhibit a singular length within each species, with the lengths of TVV4 proteins varying up to 20 aa between

strains. The same 20 aa difference in protein products is observed between two strains of TVV5. The differential variances in the distribution of CP/RdRp lengths between TVV1-3 and TVV4-5 is easily detected (Figure A2.4D). Given the small genome size of these viruses, the fraction of the genome dedicated to encoding protein—and the reciprocal fraction making up the untranslated regions—was investigated (Figure A2.4B). TVV1 and TVV2 have the highest proportions of their genome devoted to encoding CP and CP/RdRp proteins, while an appreciably larger portion of the TVV3, TVV4, and TVV5 genomes serve as UTRs (Figure A2.4B). Despite TVV3-TVV5 possessing genomes with smaller CDS fractions, the polymerase proteins of these viruses are larger than those of TVV1 and TVV2 (Table A2.1, Figure A2.4D). This can be explained by the larger genome sizes of TVV3-TVV5 versus TVV1 and TVV2. The larger genomes of TVV3-TVV5 are accompanied by larger 3'-UTRs as well (Figure A2.4F). The 5'-UTR lengths tend to cluster by virus; however, these groupings are less stark than 3'-UTRs (Figure A2.4E).

As previous studies have used a 50% amino acid identity threshold to delineate individual species within genus *Trichomonasvirus* (a threshold that is supported by phylogenetic analysis) (Goodman et al. 2011b), CP/RdRp sequences from the newly determined genomes were examined. Average pairwise identities of strains within a given trichomonasvirus species showed strong within-species similarity: TVV1, 86.4%; TVV2, 91.2%; TVV3, 92.5%; TVV4, 83.0%; TVV5: 85.3%. In contrast, the sequence identity drops precipitously when crossing species boundaries. When comparing pairwise identities for each species against strains of all other species, the average cross-species pairwise identity of other species to TVV1 strains was 22.7%. Repeating this analysis for TVV2-5 yields averages of 31% (TVV2), 31.7% (TVV3), 35.6% (TVV4), and 30.1%

(TVV5). These results of full CP/RdRp fusion proteins supports previous findings of strong TVV species-specific clustering.

The RdRp component of the CP/RdRp fusion proteins of TVV1-5 are comprised of three discrete domains: an N-terminal domain, a core polymerase domain, and a C-terminal domain. The central polymerase domain maintains a much higher amino acid similarity score (65%) compared to the NTD or CTD (40 and 47%, respectively) using the standard BLSM62 substitution matrix across a TVV1-5 CP/RdRp protein alignment using MAFFT L-ins-i. This sequence divergence could indicate that the conserved polymerase core remains under strong selective pressure to maintain function, while the NTD and CTD portions of the enzyme are freer to change and perhaps evolve species-specific functionalities.

	5’-terminus	3’-terminus
TVV1-RU357*	**TGCAAAAAGGAGGGAGTAGT**	**ATGGCC-CGCTATACCTTCA**
TVV1-165307-1*	**TGCAAAA-AGAGGGAGTGAT**	**ATGGTCTCACTATACCTTCA**
TVV1-CDC337	AGAGGGAGTAGT	TTCGTCTCACTATACCTTCA
TVV1-JH191A4	AGAGGGAGTGAT	TTGGTCTCACTATA
TVV1-NYH286	AGAGGGAGTGAT	TTGGTCTCACTATA
TVV1-JH37A2	GGAGGGAGTAGT	ATGGCCCCGCTATACCTTC
TVV1-123414	GGAGGGAGTAGT	ATGGCCCCGCTATACCTT
TVV1-JRS-TV-120	GGAGGGAGTGAT	TTGGTCTCGTTATACCT
TVV1-JH162A4	GAGGGAGTGAT	TTGGTCTCACTATACCTTC
TVV1-JH31A4	GAGGGGGTTAC	ATGGTCTCACTATACC
TVV1-RU393	GAGGGAGTGAT	TTGGCCTCACTATACCTTC
TVV1-NYH209	GAGGGAGTAGT	ATGGCC-CGCTATA
TVV1-NYH272	GAGGGAGTCAT	ATGGTCTCACTATACC
TVV2-165307-1*	**TGCTTTAAAAGGAGTGACGA**	**TAATACCAACTCTTACGTCA**
TVV2-RU357*	**TGCTTT-AAAGGAGTGACGA**	**TAATATCAACTCTTATGTC**
TVV2-123414	CTTT-GAAGGAGTGACGA	TAATATCAACTCTTACGTCA
TVV2-JH162A4	CTTT-AAAGGAGTGACGA	TAATATCAACTCTTA
TVV2-C1NIH	AGGAGTGACGA	TAATATCAACTCTTACGTC
TVV2-G3(HMS)	AGTGACGA	TAATATCAAC
TVV2-NYH209	GTGACGA	TAATATCAACTCTTACG
TVV2-NYH286	GGAGTGACGA	TAATATCAACTCTTACGTC
TVV3-165307-1*	**TGCTTAAAAAGCCTAGTCCA**	**ACTACCCCACGTTTAGTTCA**
TVV3-NYH209	GCTTAAAAAGCCTAGTCCA	ACTACCCCACGTT
TVV3-RU393	AAGGCCTAGTCCA	ACTACCCCACGTT
TVV3-C1NIH	AAGCCTAGTCCG	ACTACCCCACGTT
TVV3-JH32A4	AGCCTAGTCCA	ACTACCCCACGT
TVV3-NYH286	GCCTAGTCCA	ACTACCCCACGT
TVV3-CDC337	GCCTAGTCCA	ACTACCCCACGTGTAGTTC
TVV3-JH162A4	AGTCCA	ACTACCCCACGTTTAG
TVV3-JRS-TV-120	GTCCG	ACTACCCCACG
TVV3-G3(HMS)	CCA	ACTACCCCACG
TVV4-RU357*	**TGCTTAAAGTCTCGTGAGCT**	**ATGACCACCGGTCTTCCTCA**
TVV4-JRS-TV-120	TCCCAGTGAGCT	ATGACTCCCGGTC
TVV4-NYH272	CCCGTGAGCT	ATGAATCCCGGTC
TVV5-165307-1*	**TGCTTTAGAATCGAAAGATC**	**ATTCTCCCCAGTCATCCTCA**
TVV5-JRS-TV-120*	**TGCTTTAGAGTCACAAGATC**	**ATTCTCCCCAGTCAACCTCA**
TVV5-RU357*	**CGCTTTAGAATCGCAAGATC**	**ATTCTCCCCAGTCATCCTCA**
TVV5-NYH209	CTTTAGAATCGCAAGATC	ATTCTCCCCAG
TVV5-JH162A4	TTCCGAATCGCAAGATC	ATCCTCCCCAGTCATCCT

Figure A2.3: Alignments of dsRNA-seq determined TVV genome assemblies with focus on the 5’ and 3’ termini. First and last twenty nucleotide positions are shown for each terminus with respect to the coding-strand of the viral genome (positions according to longest genome). Full- length TVV genomes, as determined by 3’RACE experimentation to define the genomic termini, are indicated by an asterisk and are bolded.

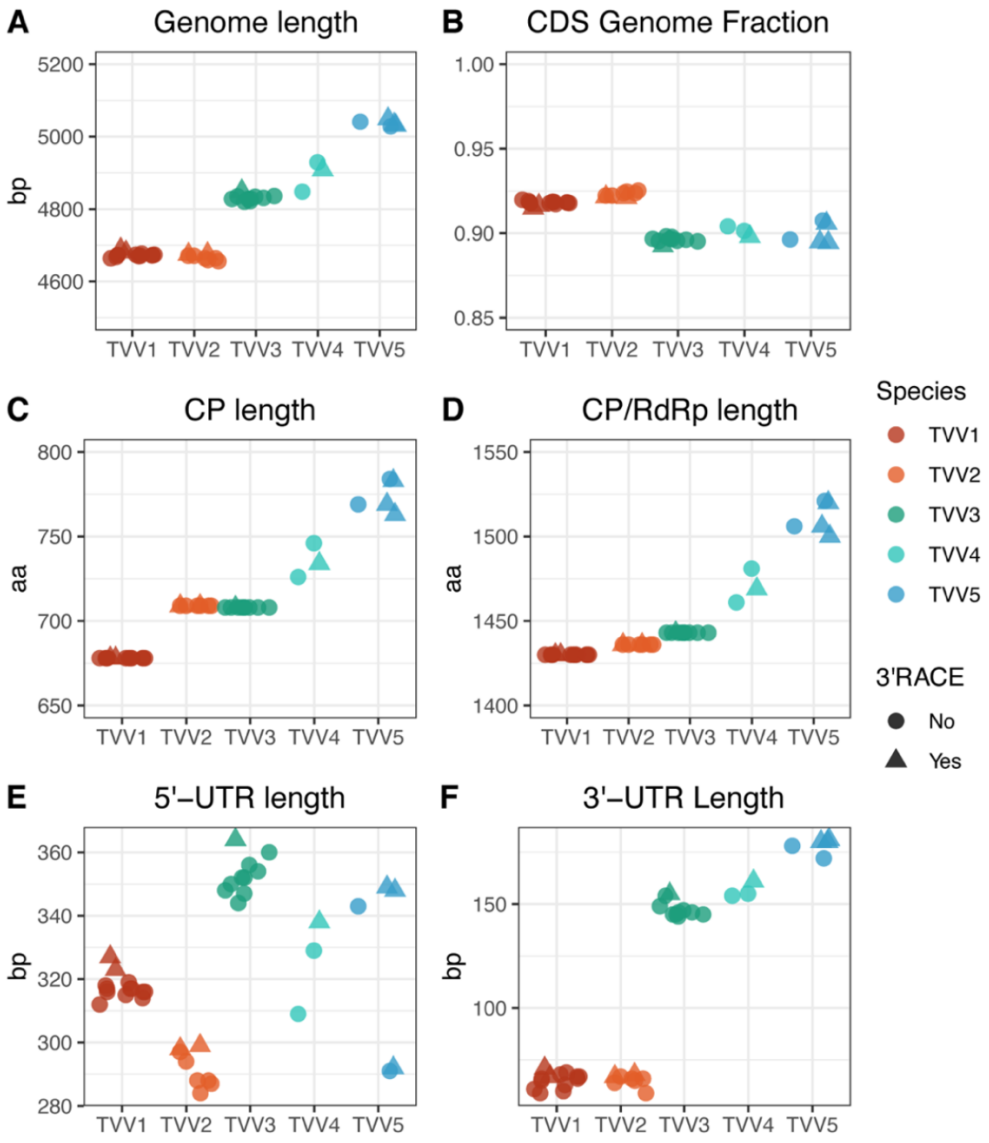


Figure A2.4: Visualization of key features of newly determined trichomonasvirus genomes.

(a) Genome lengths of all new virus strains, in base-pairs (bp); (b) the proportion of each viral genome comprised of the protein coding sequence (CDS); (c) lengths of the CP protein for each virus, in amino acids (aa); (d) lengths of the CP/RdRp fusion protein inferred for each virus based on a -1 or -2 programmed ribosomal frameshift, in aa; (e) The size in bp of the 5' untranslated region (5'-UTR) for all new viruses; (f) The size in bp of the 3' untranslated region (3'-UTR) for all new viruses. The viruses are colored according to species. Strains subjected to 3'RACE experiments to construct a full-length genome are indicated by triangles.

Identification of TVV5 Outside of the United States

A new virus discovery platform, Serratus (Edgar et al., 2022), streamlines the process of identifying viruses in public transcriptomes. The researchers behind Serratus used viral sequences in NCBI GenBank in 2021 as reference to profile all RNA-seq data in the NCBI Sequence Read Archive (SRA) database. We sought to use this tool to expand the TVV5 sequences available for study, especially as all TVV5 strains identified to date (including this current study) are from *T. vaginalis* originally isolated from patients in the United States. As TVV5 was only recently discovered, this virus was not indexed by Serratus. We thus used this tool to identify target datasets containing sequences for TVV1 through TVV4 with the premise that some *T. vaginalis* isolates could be coinfected with TVV5. In this way, we identified a set of TVV+ transcriptomes from a study tracing the evolution of HIV in a cohort of patients in South Africa (BioProject PRJNA473698). Serratus initially identified four RNA-seq libraries from this study: two libraries each from patients designated FR004 and FR006. To mine these datasets for sequences corresponding to TVV5, we used a map-to-reference strategy, aligning the transcriptome sequences to our newly determined TVV5 genomes. Coding-complete genome assemblies were generated for both patients, yielding TVV5-FR004 (5020 bp) and TVV5-FR006 (4977 bp).

Examining this study more closely, RNA-seq datasets were available for four patients. We proceeded to survey all datasets from this study, systematically characterizing all trichomonasvirus sequences from the 11 accessions comprising this study. Two sequencing libraries were available for patient FR014, one from a cervicovaginal lavage sample and one from a plasma sample. Patients FR004 and FR006

each had four corresponding RNA-seq libraries: two from cervicovaginal lavage samples and two from plasma samples. A final transcriptome was available for a patient designated USA, with the sequencing library generated from a plasma sample. Transcriptomes corresponding to patient FR014 did not have sufficient read counts for generation of any trichomonasvirus sequence assemblies (range: 0 reads from plasma, 0-16 reads from lavage; Table S2). The plasma transcriptomes from patients FR004 and FR006 had similarly low TVV read counts (ranges: FR004, 1-20 reads; FR006, 0-13 reads; Table S2). The remaining plasma sample, belonging to the USA patient, had no reads mapping to any trichomonasvirus. In contrast, the cervicovaginal lavage transcriptomes from patients FR004 and FR006 were rich with trichomonasvirus sequences, facilitating the assembly of additional coding- complete viral genomes: TVV1-FR004 (4683 bp), TVV3-FR004 (4822 bp), TVV4-FR004 (4867 bp), TVV1-FR006 (4664 bp), TVV2-FR006 (4655 bp), and TVV3-FR006 (4788 bp). Characteristics of these virus sequences are summarized in Table A2.2.

Taken together, these findings represent the first identification of TVV5 in patients outside of the United States. As strains of each of the other TVVs have been detected in isolates around the world, this brings the geographic range of TVV5 closer in line to the other trichomonasviruses (Weber et al., 2003; Kim et al., 2007; Fraga et al., 2012; Heidary et al., 2013; El-Gayar et al., 2016; Rivera et al., 2017; Margarita et al., 2019). Co-infection with TVV5 and other trichomonasviruses in these patient samples mirrors what is seen with our dsRNA-seq data and previous reports of widespread TVV co-infections (Goodman et al., 2011a).

Table A2.2: Trichomonasvirus genomes assembled from newly identified transcriptome datasets (NCBI BioProject: PRJNA473698). Key characteristics of these sequences are listed.

Virus	Genome (bp)	CP (aa)	CP/RdRp (aa)	5'-UTR (bp)	3'-UTR (bp)
TVV1-FR004	4683	678	1430	324	68
TVV3-FR004	4822	708	1443	363	128
TVV4-FR004	4867	732	1467	320	144
TVV5-FR004	5020	769	1506	344	156
TVV1-FR006	4664	678	1430	324	49
TVV2-FR006	4655	709	1436	293	52
TVV3-FR006	4788	708	1433	328	129
TVV5-FR006	4977	770	1507	313	141

Discussion

The findings presented in the current study contribute a sizable increase in the number of coding-complete genomes of trichomonasviruses now available for study. Prior to this work, the genome counts were especially limited for TVV3, TVV4, and TVV5. The combined number of TVV genomes from NCBI GenBank and transcriptome-derived genomes from our previous study (Manny et al., 2022) put the number of coding-complete genomes for each trichomonasvirus as follows: TVV1, 21; TVV2, 11; TVV3, 8; TVV4, 3; and TVV5, 2. This dearth of genome sequences, most acutely for TVV3-5, precludes serious investigation into the biology of trichomonasviruses. The work presented in this study helps remedy this problem by substantially increasing the number of genomes for TVV1 and TVV2, and at least doubling the number of genomes per species for TVV3, TVV4, and TVV5. Notably, the viral genomes were determined from isolates readily accessible from a common commercial supplier. This acts as catalogue for which commercial isolates of *T. vaginalis* are infected with which species of TVV. We hope that this resource can lower the barrier to entry for studying the molecular biology of

trichomonasviruses, as research is not limited by availability, provenance, or viral status of starting reagents. The identification that five of these readily available isolates are infected with TVV5 was regarded as a welcome finding. This finding of TVV5 in multiple isolates within our sample set allowed us to conduct full-length sequencing on several strains of TVV5 using 3'RACE. This enabled the determination of the first full-length genomes of TVV5 and revealed unexpected genomic terminal residues. These results revealed that the genomes of trichomonasviruses are not comprised of previously reported conserved dinucleotide motifs (Goodman et al., 2011a), but instead feature characteristic trinucleotide motifs missed by previous studies.

Our dsRNA-seq approach demonstrated utility and efficiency for assembling coding- complete genomes for trichomonasviruses. In contrast to our previous study which relied on sequencing data from transcriptomes not enriched for virus (Manny et al., 2022), this study's dsRNA-seq approach was able to yield no partial assemblies. The assembled TVV genomes were almost full-length, in every case within 20 bp of the expected 5' and 3' terminus. Conceptually, one drawback of our dsRNA-seq approach was that it was designed around the detection of trichomonasviruses specifically and dsRNA viruses more generally. While some groups (e.g., Castillo et al., 2011) have reported success with using dsRNA selection to pick up ssRNA viruses (assumedly through the dsRNA replication intermediate), we did not see that in our results. While our approach was clearly effective for determining trichomonasvirus sequences, some RNA viruses (notably those with a ssRNA genome) and DNA viruses could be missed by this strategy.

Acknowledgements

The authors would like to thank the Harvard Medical School Research Computing Group for their work in supporting the O2 High Performance Compute (HPC) cluster. Much of the computational work detailed in this study was conducted using the O2 HPC cluster.

Author Contributions

Austin R. Manny: Conceptualized project, designed methodology, cultured *T. vaginalis*, conducted RNA isolation and dsRNA enrichment experiments, carried out 3'RACE genomic terminal sequencing experiments, developed dsRNA-seq data analysis pipeline, analyzed data, conducted transcriptome mining, prepared original manuscript, prepared figures, edited manuscript

Carrie A. Hetzel: Cultured *T. vaginalis,* conducted RNA isolation and dsRNA enrichment experiments, performed data analysis

Max L. Nibert: Conceptualized project, designed methodology, supervision, acquired funding

References

Buchfink, B., et al. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods***,** 18, 366–368

Bushmanova, E., et al. (2018). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *bioRxiv*, 420208

Camacho, C., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.***,** 10, 421 Castillo A., et al. (2011). Rapid isolation of mycoviral double-stranded RNA from *Botrytis cinerea* and *Saccharomyces cerevisiae*. *Virol. J.* 8, 38 Edgar, R.C., et al. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature* 602**,** 142–147

El-Gayar, E.K., et al. (2016). Molecular characterization of double-stranded RNA virus in *Trichomonas vaginalis* Egyptian isolates and its association with pathogenicity. *Parasitology Research* 115, 4027–4036

Fichorova, R.N., et al. (2009). Impact of T. vaginalis infection on innate immune responses and reproductive outcome. *J. Reprod. Immunol*. 83, 185–189

Fichorova, R.N., et al. (2012). Endobiont viruses sensed by the human host—beyond conventional antiparasitic therapy. *PLoS ONE***,** 7, e48418

Fouts, A.C. & Kraus, S.J. (1980). *Trichomonas vaginalis*: reevaluation of its clinical presentation and laboratory diagnosis. *J. Infect. Dis.***,** 141, 137–143 Fraga, J., et al. (2012). Species typing of Cuban Trichomonas vaginalis virus by RT-PCR, and association of TVV-2 with high parasite adhesion levels and high pathogenicity in patients. *Archives of Virology*, 157, 1789—1795

Ghosh, I. et al. (2017). Association between high risk human papillomavirus infection and co- infection with *Candida* spp. and *Trichomonas vaginalis* in women with cervical premalignant and malignant lesions. *J. Clin. Virol.* 87, 43–48

Goodman, R.P, et al. (2011a). Clinical Isolates of *Trichomonas vaginalis* concurrently infected by strains of up to four trichomonasvirus species (family *Totiviridae*). *J. Virol.***,** 85, 4258–4270

Goodman, R.P., et al. (2011b). *Trichomonasvirus*: a new genus of protozoan viruses in the family *Totiviridae*. *Arch. Virol.,* 156, 171–179

Heidary, S., et al. (2013). Double-Stranded RNA Viral Infection in Tehran Trichomonas vaginalis Isolates. *Iran J Parasitol.* 8, 60–64

Kim, J.W., et al. (2007). Double-stranded RNA virus in Korean Isolate IH-2 of Trichomonas vaginalis. *Korean Journal of Parasitology* 45, 87

Leitsch, D. (2016). Recent advances in the Trichomonas vaginalis field. *F1000 Research,* 5, 162

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2*. Nat Methods* 9, 357–359

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv,* 1303.3997

Manny, A.R., Hetzel, C.A., Mizani, A., and Nibert, M.L. (2022). Discovery of a Novel Species of Trichomonasvirus in the Human Parasite *Trichomonas vaginalis* Using Transcriptome Mining. Viruses *14*, 17

Margarita, V., et al. (2019). Prevalence of double-stranded RNA virus in *Trichomonas vaginalis* isolated in Italy and association with the symbiont *Mycoplasma hominis*. *Parasitology Research*, 118, 3565–3570

Mercer, F. and Johnson, P.J. (2018). *Trichomonas vaginalis*: pathogenesis, symbiont interactions, and host cell immune responses. *Trends Parasitol.*, 34, 683–693

Mielczarek, E. and Blaszkowska, J. (2016). Trichomonas vaginalis: pathogenicity and potential role in human reproductive failure. *Infection* 44, 447–458

Nibert, M.L., et al. (2018). Evidence for contemporary plant mitoviruses. *Virology*, 518, 14–24

Rivera, W.L., et al. (2017). Detection and molecular characterization of double-stranded RNA viruses in Philippine Trichomonas vaginalis isolates. *Journal of Microbiology, Immunology and Infection* 50, 669–676

Weber, B., et al. (2003). Double stranded RNA virus in South African *Trichomonas vaginalis* isolates. *Journal of Clinical Pathology* 56, 542–543

Wendel, K.A., et al. (2002). Double-Stranded RNA Viral Infection of *Trichomonas vaginalis* Infecting Patients Attending a Sexually Transmitted Diseases Clinic. *J Infect Dis* 186*,* 558–561

Wood, D.E., et al. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol*. 28;20(1):257

World Health Organization. (2011). Prevalence and Incidence of Selected Sexually Transmitted Infections, *Chlamydia trachomatis, Neisseria gonorrhoeae, Syphilis,* and *Trichomonas vaginalis*: Methods and Results Used by the WHO to Generate 2005 Estimates. WHO

APPENDIX 3

Supplementary Information from Chapter 2: Trichomonasviruses Modulate Survival, Adherence, and Gene Expression of their Protozoan Host

Carrie A. Hetzel[1,2], Akua A. Appah-Sampong[1,3], Austin R. Hurst-Manny[1,2], Max L. Nibert[1,2,*]

[1]Department of Microbiology, Harvard Medical School, Boston, MA 02115, USA

[2]Program in Virology, Division of Medical Sciences, Graduate School of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA

[3]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA

Corresponding author

Table A3.1. Significant hits between TVV1+ and TVV1- clones of isolate JH37A#2-B1. Hits were designated as significant if p-adjusted < 0.05. Hits with gene IDs colored red were upregulated in TVV1+ clones, and hits with gene IDs colored blue were downregulated in TVV1+ clones. Product description, ortholog group, and InterPro description are taken from Trich DB.

Gene ID	Gene Type	Product Description	Ortholog Group	InterPro Description
TVAGG3_0539280	protein coding gene	endonuclease protein	OG6_105664	N/A
TVAGG3_0707390	protein coding gene	midasin-related family	OG6_105664	AAA+ ATPase domain;ATPase, dynein-related, AAA domain;P-loop containing nucleoside triphosphate hydrolase;von Willebrand factor A-like domain superfamily
TVAGG3_0539290	protein coding gene	endonuclease protein	OG6_105664	von Willebrand factor A-like domain superfamily
TVAGG3_0496300	protein coding gene	endonuclease protein	OG6_105664	N/A
TVAGG3_0741070	protein coding gene	protein polyubiquitination	OG6_112186	Zinc finger C2H2-type
TVAGG3_0585190	protein coding gene	3'-5' exonuclease family	OG6_148709	3'-5' exonuclease domain;Ribonuclease H-like superfamily
TVAGG3_0741040	protein coding gene	tissue development protein family	OG6_112186	Zinc finger C2H2-type
TVAGG3_0996410	pseudogene	glutenin, high molecular weight subunit PW212-related protein family	N/A	N/A
TVAGG3_0496280	protein coding gene	glutenin, high molecular weight subunit PW212-related protein family	OG6_105664	N/A
TVAGG3_1065710	pseudogene	protein polyubiquitination	N/A	N/A
TVAGG3_0716620	protein coding gene	ATP-dependent RNA helicase protein	OG6_112186	DEAD/DEAH box helicase domain;Helicase superfamily 1/2, ATP-binding domain;P-loop containing nucleoside triphosphate hydrolase

TVAGG3_0552600	protein coding gene	GTPase, IMAP family member-related family	OG6_104495	AIG1-type guanine nucleotide-binding (G) domain;P-loop containing nucleoside triphosphate hydrolase;GTPase GIMA/IAN/Toc
TVAGG3_0878700	protein coding gene	spectrin binding	OG6_117387	Ankyrin repeat;Domain of unknown function DUF3447;Ankyrin repeat-containing domain superfamily
TVAGG3_0986850	protein coding gene	aminopeptidase family	OG6_100598	Peptidase M24;Aminopeptidase P, N-terminal;Creatinase/Aminopeptidase P/Spt16, N-terminal;Creatinase/aminopeptidase-like
TVAGG3_0716610	pseudogene	helicase activity	N/A	N/A
TVAGG3_0496290	protein coding gene	glutenin, high molecular weight subunit PW212-related protein family	OG6_527651	N/A
TVAGG3_0365170	protein coding gene	RNA polymerase II transcription regulator recruiting protein	OG6_100129	SANT/Myb domain;Homeobox-like domain superfamily;Myb domain
TVAGG3_0609820	protein coding gene	amino acid transmembrane transporter protein	OG6_100459	Amino acid transporter, transmembrane domain
TVAGG3_0649380	protein coding gene	spectrin binding	OG6_117387	Ankyrin repeat;Ankyrin repeat-containing domain superfamily
TVAGG3_0964910	protein coding gene	Secreted RxLR effector peptide protein [Source:UniProtKB/TrEMBL;Acc:A2E2A3]	OG6_187668	N/A
TVAGG3_0716630	protein coding gene	tissue development protein family	OG6_112186	N/A
TVAGG3_0506690	protein coding gene	Coiled-coil domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A2F9W3]	OG6_109144	N/A
TVAGG3_0192490	protein coding gene	RNA polymerase II transcription regulator recruiting protein	OG6_100129	SANT/Myb domain;Homeobox-like domain superfamily;SANT domain;Myb domain
TVAGG3_1065730	protein coding gene	helicase protein	OG6_112186	N/A

TVAGG3_0741050	pseudogene	ATP-dependent RNA helicase activity	N/A	N/A
TVAGG3_0369320	protein coding gene	protein of unknown function (DUF814)	OG6_110486	NFACT, RNA-binding domain;Jlp2/Ccd25
TVAGG3_0564820	protein coding gene	unspecified product	OG6_150454	N/A
TVAGG3_0189810	protein coding gene	regulation of centriole replication	OG6_103537	Protein kinase domain;Serine/threonine-protein kinase, active site;Protein kinase-like domain superfamily

Table A3.2. Significant hits between TVV1+ and TVV1- clones of isolate JH191A#4-A1. Hits were designated as significant if p-adjusted < 0.05. Hits with gene IDs colored red were upregulated in TVV1+ clones, and hits with gene IDs colored blue were downregulated in TVV1+ clones. Product description, ortholog group, and InterPro description are taken from Trich DB.

Gene ID	Gene Type	Product Description	Ortholog Group	InterPro Description
TVAGG3_0471700	protein coding gene	glucokinase family	OG6_114114	Glucokinase;ATPase, nucleotide binding domain
TVAGG3_0175870	protein coding gene	Listeria-Bacteroides repeat domain (List Bact rpt) family	OG6_108058	N/A
TVAGG3_0707390	protein coding gene	midasin-related family	OG6_105664	AAA+ ATPase domain;ATPase, dynein-related, AAA domain;P-loop containing nucleoside triphosphate hydrolase;von Willebrand factor A-like domain superfamily
TVAGG3_1047160	protein coding gene	leucine-rich repeats (6 copies)-containing protein	OG6_110117	BspA type Leucine rich repeat region
TVAGG3_1047170	protein coding gene	leucine-rich repeats (6 copies)-containing protein	OG6_110117	BspA type Leucine rich repeat region
TVAGG3_0751360	protein coding gene	pyruvate-flavodoxin oxidoreductase family	OG6_104405	Pyruvate-flavodoxin oxidoreductase, central domain;Pyruvate flavodoxin/ferredoxin oxidoreductase, pyrimidine binding domain;Transketolase C-terminal/Pyruvate-ferredoxin oxidoreductase domain II;Thiamine pyrophosphate enzyme, TPP-binding;Pyruvate-flavodoxin oxidoreductase;4Fe-4S ferredoxin-type, iron-sulphur binding domain;4Fe-4S ferredoxin, iron-sulphur binding, conserved site;Pyruvate-flavodoxin oxidoreductase, EKR domain;Pyruvate/ketoisovalerate oxidoreductase, catalytic domain;Thiamin diphosphate-binding fold;Pyruvate:ferredoxin oxidoreductase, core domain II

TVAGG3_0315460	protein coding gene	ATPase activity, coupled to transmembrane movement of substances	OG6_100045	ABC transporter-like, ATP-binding domain;AAA+ ATPase domain;ABC-2 type transporter, transmembrane domain;ABC transporter A;P-loop containing nucleoside triphosphate hydrolase
TVAGG3_0720960	protein coding gene	guanylate cyclase protein	OG6_101146	PAS domain;Adenylyl cyclase class-3/4/guanylyl cyclase;Nucleotide cyclase;PAS domain superfamily
TVAGG3_0662820	protein coding gene	Virion structural protein [Source:UniProtKB/TrEMBL;Acc:A2DUM3]	OG6_161890	N/A
TVAGG3_0720940	protein coding gene	guanylate cyclase protein	OG6_101146	N/A
TVAGG3_0058000	protein coding gene	protein of unknown function (DUF3447)	OG6_100012	Ankyrin repeat;Domain of unknown function DUF3447;Ankyrin repeat-containing domain superfamily
TVAGG3_0196800	protein coding gene	ZZ-type domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A2G9Y6]	OG6_225474	N/A
TVAGG3_0608330	protein coding gene	regulation of pentose-phosphate shunt	OG6_101500	Phosphoglycerate/bisphosphoglycerate mutase, active site;Phosphoglycerate mutase 1;Histidine phosphatase superfamily, clade-1;Histidine phosphatase superfamily
TVAGG3_0848720	protein coding gene	bifunctional inhibitor/lipid-transfer protein/seed storage 2s albumin superfamily protein family	OG6_106613	N/A
TVAGG3_0365170	protein coding gene	RNA polymerase II transcription regulator recruiting protein	OG6_100129	SANT/Myb domain;Homeobox-like domain superfamily;Myb domain
TVAGG3_0281860	protein coding gene	Lipoprotein [Source:UniProtKB/TrEMBL;Acc:A2DYC6]	OG6_132165	N/A
TVAGG3_0059580	transposable element gene	Secreted protein [Source:UniProtKB/TrEMBL;Acc:A2DMQ9]	OG6_225270	N/A
TVAGG3_0933980	protein coding gene	ethanolamine kinase protein	OG6_101726	Protein kinase-like domain superfamily
TVAGG3_0383980	protein coding gene	COA binding domain containing protein family	OG6_152721	CoA-binding;Succinyl-CoA synthetase-like;Succinyl-CoA synthetase-like,

				flavodoxin domain;NAD(P)-binding domain superfamily
TVAGG3_0040170	protein coding gene	regulation of choline O-acetyltransferase protein	OG6_101184	Peptidase M8, leishmanolysin
TVAGG3_0306240	protein coding gene	winged helix DNA-binding domain family	OG6_110237	E2F/DP family, winged-helix DNA-binding domain;E2F Family;Winged helix DNA-binding domain superfamily
TVAGG3_0795510	protein coding gene	MAP kinase kinase kinase protein	OG6_141673	Protein kinase domain;Protein kinase-like domain superfamily
TVAGG3_1024660	protein coding gene	spectrin binding	OG6_100012	N/A

Table A3.3. Significant hits between TVV3+ and TVV3- clones of isolate JH32AA#4-A1. Hits were designated as significant if p-adjusted < 0.05. Hits with gene IDs colored red were upregulated in TVV3+ clones, and hits with gene IDs colored blue were downregulated in TVV3+ clones. Product description, ortholog group, and InterPro description are taken from Trich DB.

Gene ID	Gene Type	Product Description	Ortholog Group	InterPro Description
TVAGG3_0322610	protein coding gene	glycoprotein 38 family	OG6_106291	N/A
TVAGG3_0400780	protein coding gene	cysteine-type peptidase protein	OG6_101407	Peptidase C14, caspase domain;Caspase-like domain superfamily
TVAGG3_0400790	protein coding gene	cysteine-type peptidase protein	OG6_101407	Peptidase C14, caspase domain;Caspase-like domain superfamily
TVAGG3_0399840	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0399370	pseudogene	pseudogene	N/A	N/A
TVAGG3_0400340	protein coding gene	Fusion glycoprotein F0	OG6_106173	N/A
TVAGG3_0400540	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0399570	pseudogene	pseudogene	N/A	N/A
TVAGG3_0399820	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0802710	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA-directed DNA polymerase, family B;Ribonuclease H-like superfamily;DNA/RNA polymerase superfamily
TVAGG3_0055530	protein coding gene	guanylate cyclase protein	OG6_101752	N/A
TVAGG3_0033670	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA-directed DNA polymerase, family B;Ribonuclease H-like superfamily;DNA/RNA polymerase superfamily

TVAGG3_0693750	protein coding gene	glucose import	OG6_100060	Sugar/inositol transporter;Major facilitator, sugar transporter-like;Major facilitator superfamily domain;MFS transporter superfamily
TVAGG3_0784460	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;Ribonuclease H-like superfamily
TVAGG3_0055540	protein coding gene	guanylate cyclase protein	OG6_101752	N/A
TVAGG3_0128710	transposable element pseudogene	Mariner transposase	N/A	N/A
TVAGG3_0293220	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0399600	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0526710	protein coding gene	photosystem I iron-sulfur center-related family	OG6_122732	2Fe-2S ferredoxin-type iron-sulfur binding domain;Iron hydrogenase, large subunit, C-terminal;Iron hydrogenase;4Fe-4S ferredoxin-type, iron-sulphur binding domain;4Fe-4S ferredoxin, iron-sulphur binding, conserved site
TVAGG3_0539210	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0693780	protein coding gene	glucose import	OG6_100060	Sugar/inositol transporter;Major facilitator, sugar transporter-like;Major facilitator superfamily domain;MFS transporter superfamily
TVAGG3_0769420	transposable element gene	unspecified product	OG6_103462	N/A
TVAGG3_0802460	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	N/A

TVAGG3_0790510	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;Ribonuclease H-like superfamily;DNA/RNA polymerase superfamily
TVAGG3_0178180	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0399310	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_1093610	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_1080200	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_0398550	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0398540	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0002460	transposable element pseudogene	pseudogene	N/A (orthology not determined for pseudogenes)	N/A
TVAGG3_0089970	protein coding gene	lactate dehydrogenase isozyme 2	OG6_101066	Lactate/malate dehydrogenase, N-terminal;L-lactate/malate dehydrogenase;Malate dehydrogenase, type 2;Lactate dehydrogenase, protist;Lactate dehydrogenase/glycoside hydrolase, family 4, C-terminal;Lactate/malate dehydrogenase, C-terminal;NAD(P)-binding domain superfamily
TVAGG3_0153060	protein coding gene	Chromo domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A2EYW9]	OG6_525947	N/A
TVAGG3_0399590	protein coding gene	unspecified product	OG6_106173	N/A

TVAGG3_0956480	protein coding gene	Glycoprotein [Source:UniProtKB/TrEMBL;Acc:A2DCJ4]	OG6_530919	N/A
TVAGG3_0019540	transposable element pseudogene	pseudogene	N/A (orthology not determined for pseudogenes)	N/A
TVAGG3_0465450	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;Ribonuclease H-like superfamily
TVAGG3_0046390	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA/RNA polymerase superfamily
TVAGG3_0757580	transposable element pseudogene	pseudogene	N/A (orthology not determined for pseudogenes)	N/A
TVAGG3_0242460	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA-directed DNA polymerase, family B;Ribonuclease H-like superfamily;DNA/RNA polymerase superfamily
TVAGG3_1077150	transposable element pseudogene	pseudogene	N/A (orthology not determined for pseudogenes)	N/A
TVAGG3_0399340	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0548650	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0209230	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA-directed DNA polymerase, family B;Ribonuclease H-like superfamily;DNA/RNA polymerase superfamily

TVAGG3_0240870	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA-directed DNA polymerase, family B;DNA/RNA polymerase superfamily
TVAGG3_0691350	protein coding gene	Por_Secre_tail domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A2EEF4]	OG6_128659	N/A
TVAGG3_0904070	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0055550	protein coding gene	guanylate cyclase protein	OG6_101752	N/A
TVAGG3_0649160	protein coding gene	EF-hand family	OG6_150388	EF-hand domain;EF-hand domain pair;EF-Hand 1, calcium-binding site
TVAGG3_0790520	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	N/A
TVAGG3_0400130	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0399560	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0114470	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0693770	protein coding gene	glucose import	OG6_100060	Major facilitator, sugar transporter-like;Major facilitator superfamily domain;MFS transporter superfamily
TVAGG3_0062820	protein coding gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA-directed DNA polymerase, family B;Ribonuclease H-like superfamily;DNA/RNA polymerase superfamily
TVAGG3_0399300	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0398520	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_1088880	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A

TVAGG3_0046380	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	Ribonuclease H-like superfamily
TVAGG3_0784450	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	N/A
TVAGG3_0925730	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0465460	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	N/A
TVAGG3_0399330	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0802450	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;Ribonuclease H-like superfamily;DNA/RNA polymerase superfamily
TVAGG3_0733680	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0817410	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA/RNA polymerase superfamily
TVAGG3_0829180	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	N/A
TVAGG3_0493750	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0693760	protein coding gene	glucose import	OG6_100060	Major facilitator, sugar transporter-like;Major facilitator superfamily domain;MFS transporter superfamily
TVAGG3_0400110	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_1075970	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A

TVAGG3_0268600	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0398530	pseudogene	pseudogene	N/A	N/A
TVAGG3_1040350	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;Ribonuclease H-like superfamily;DNA/RNA polymerase superfamily
TVAGG3_0033540	transposable element gene	integrase core domain containing protein family	OG6_100027	Integrase, catalytic core;Ribonuclease H-like superfamily
TVAGG3_0399420	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0817420	transposable element pseudogene	pseudogene	N/A	N/A
TVAGG3_0967650	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	Ribonuclease H-like superfamily
TVAGG3_0829190	transposable element pseudogene	pseudogene	N/A	N/A
TVAGG3_0229900	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0266240	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0122970	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0293230	protein coding gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0325090	protein coding gene	NAD(P)H oxidoreductase-related family	OG6_112987	Flavodoxin-like fold;Flavoprotein-like superfamily
TVAGG3_0465440	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA/RNA polymerase superfamily

TVAGG3_1080460	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_0063300	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	N/A
TVAGG3_0496280	protein coding gene	glutenin, high molecular weight subunit PW212-related protein family	OG6_105664	N/A
TVAGG3_0693810	pseudogene	glucose import	N/A	N/A
TVAGG3_0967660	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA-directed DNA polymerase, family B;DNA/RNA polymerase superfamily
TVAGG3_0653560	protein coding gene	mitochondrial ATP-dependent permease MDL1 family	OG6_100059	ABC transporter-like, ATP-binding domain;AAA+ ATPase domain;ABC transporter type 1, transmembrane domain;ABC transporter-like, conserved site;P-loop containing nucleoside triphosphate hydrolase;ABC transporter type 1, transmembrane domain superfamily;Type 1 protein exporter
TVAGG3_0790500	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA/RNA polymerase superfamily
TVAGG3_0074840	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	Ribonuclease H-like superfamily
TVAGG3_1080190	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_1079180	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_0400090	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0689290	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A

TVAGG3_0749400	protein coding gene	glycerol-3-phosphate catabolic process	OG6_100385	Glycerol-3-phosphate dehydrogenase, NAD-dependent, C-terminal;Glycerol-3-phosphate dehydrogenase, NAD-dependent;6-phosphogluconate dehydrogenase-like, C-terminal domain superfamily;Glycerol-3-phosphate dehydrogenase, NAD-dependent, N-terminal;Glycerol-3-phosphate dehydrogenase, NAD-dependent, eukaryotic;NAD(P)-binding domain superfamily
TVAGG3_0955210	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0983760	protein coding gene	PLAC8 family	OG6_100784	PLAC8 motif-containing protein
TVAGG3_1093600	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_0063620	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	Ribonuclease H-like superfamily
TVAGG3_0818470	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA/RNA polymerase superfamily
TVAGG3_0031360	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_0399630	protein coding gene	Ig-like_bact domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A2HXV9]	OG6_106173	N/A
TVAGG3_0693550	protein coding gene	cysteine synthase family protein domain-containing protein	OG6_100154	Cysteine synthase/cystathionine beta-synthase, pyridoxal-phosphate attachment site;Tryptophan synthase beta chain-like, PALP domain;Cysteine synthase;Tryptophan synthase beta chain-like, PALP domain superfamily
TVAGG3_0400140	protein coding gene	unspecified product	OG6_106173	N/A

TVAGG3_0818480	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;Ribonuclease H-like superfamily
TVAGG3_0336550	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0594600	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0399440	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_1076120	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_1088890	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_0774840	protein coding gene	organellar and viral DNA polymerase type B	OG6_100007	N/A
TVAGG3_0826680	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;Ribonuclease H-like superfamily
TVAGG3_0240880	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	Ribonuclease H-like superfamily
TVAGG3_0826690	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA/RNA polymerase superfamily
TVAGG3_0938510	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	N/A
TVAGG3_1075990	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_0398500	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0402100	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A

TVAGG3_0911810	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_1080450	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_0166720	protein coding gene	long chain fatty acid--CoA ligase family	OG6_100152	AMP-dependent synthetase/ligase domain;AMP-binding, conserved site
TVAGG3_0167040	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0539280	protein coding gene	endonuclease protein	OG6_105664	N/A
TVAGG3_0181220	protein coding gene	organellar and viral DNA polymerase type B	OG6_100007	N/A
TVAGG3_0496290	protein coding gene	glutenin, high molecular weight subunit PW212-related protein family	OG6_527651	N/A
TVAGG3_0973690	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA/RNA polymerase superfamily
TVAGG3_0399290	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0802440	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA/RNA polymerase superfamily
TVAGG3_0211730	transposable element gene	zebrafish ch211-108c17.2-related family	OG6_103462	N/A
TVAGG3_0398510	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0192920	protein coding gene	methionine adenosyltransferase (MAT) family	OG6_100388	S-adenosylmethionine synthetase;S-adenosylmethionine synthetase, N-terminal;S-adenosylmethionine synthetase, central domain;S-adenosylmethionine synthetase, C-terminal;S-adenosylmethionine synthetase, conserved site;S-adenosylmethionine synthetase superfamily

TVAGG3_0046370	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	N/A
TVAGG3_0711540	protein coding gene	putative phospholipid-transporting ATPase family	OG6_156384	P-type ATPase;P-type ATPase, subfamily IV;P-type ATPase, A domain superfamily;P-type ATPase, phosphorylation site;P-type ATPase, transmembrane domain superfamily;P-type ATPase, cytoplasmic domain N;P-type ATPase, C-terminal;P-type ATPase, N-terminal;HAD-like superfamily;P-type ATPase, haloacid dehalogenase domain
TVAGG3_0974480	protein coding gene	unspecified product	OG6_124115	N/A
TVAGG3_1079200	transposable element pseudogene	organellar and viral DNA polymerase type B	N/A	N/A
TVAGG3_0063280	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA-directed DNA polymerase, family B, mitochondria/virus;DNA/RNA polymerase superfamily
TVAGG3_0659270	protein coding gene	unspecified product	OG6_107178	N/A
TVAGG3_0399450	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0399830	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_0399610	protein coding gene	unspecified product	OG6_106173	N/A
TVAGG3_1075960	transposable element gene	organellar and viral DNA polymerase type B	OG6_100007	DNA/RNA polymerase superfamily