Transcriptome sequencing of *Trichonympha* from *Reticulitermes hesperus* and multiprotein phylogenetic analysis of selected Spirotrichonymphids, Cristamonads, and

Trichonymphids

by

Caryn Cooper

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2022

© Caryn Cooper, 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate

and Postdoctoral Studies for acceptance, the thesis entitled:

Transcriptome Sequencing of *Trichonympha* from *Reticulitermes hesperus* and Multi-Protein Phylogenetic Analysis of Selected Spirotrichonymphids, Cristamonads, and Trichonymphids

submitted by Caryn Cooper in partial fulfillment of the requirements for

the degree of Master of Science

in <u>Botany</u>

Examining Committee:

Patrick Keeling, Professor, Department of Botany, UBC

Supervisor

Naomi Fast, Professor, Department of Botany, UBC

Supervisory Committee Member

Patrick Martone, Professor, Department of Botany, UBC

Supervisory Committee Member

Lacey Samuels, Professor, Department of Botany, UBC

Additional Examiner

Abstract

Parabasalid protists include may obligate termite gut symbionts that play important ecological roles, permitting their hosts to digest cellulose. One of the earliest-identified groups of parabasalids is the hypermastigote Trichonymphids, of which *Trichonympha* is the type species. The purpose of my thesis was to illustrate deep phylogenetic relationships between selected species of Trichonymphids and the other parabasalid groups Spirotrichonymphids and Cristamonads, and to compare the results of using multiple sequences against a single sequence for phylogenetic analyses in these groups. To this end, I sequenced the transcriptomes of *Trichonympha* cells isolated from a Pacific Coast population of *Reticulitermes hesperus*, and I used multi-sequence phylogenetic techniques to resolve the phylogeny of *Trichonymphida* and other Parabasalian groups beyond what has previously been done with the use of SSU-based phylogenetic analyses.

I successfully isolated two *Trichonympha* cells via single-cell picking techniques and sequenced their transcriptomes. I annotated these transcriptomes for protein-coding genes and used the subsequent translated protein sequences, along with translated transcriptomes of other Trichonymphids, to construct a 14-sequence dataset for a multi-gene phylogenetic analysis. I also constructed a SSU sequence dataset for comparison. I used maximum-likelihood algorithms to construct phylogenetic trees from each dataset to illustrate the deep phylogeny of trichonymphids. The phylogeny thus generated agreed with previous phylogenies of Parabasalia, recovering Trichonymphids, Spirotrichonymphida, and Cristamonads as monophyletic.

Additionally, I used SSU sequence data to confirm that the single *Trichonympha* species present in *R. hesperus* is phylogenetically distinct from *Trichonympha agilis*, despite previous literature claiming that *T. agilis* was present both in *R. flavipes* and *R. hesperus* and was the only

iii

Trichonympha in the latter. Therefore, the *R. hesperus Trichonympha* requires a full taxonomic description as a novel species, and the epithet *T. agilis* should be restricted to the species of *Trichonympha* in *R. flavipes*. I also confirmed, using COX-2 sequence data, a 2005 report from Austin et. al. synonymizing *R. flavipes* and *R. santonensis*.

Lay Summary

Parabasalids are small, single-celled organisms that live inside termite intestines and help the termites digest wood. The relationships of parabasalid species to each other have been analyzed previously, but usually based on one gene or protein sequence at a time. Using multiple sequences for inference can help clarify relationships; additionally, using multiple potential sequences lets future researchers know which sequences are most useful for determining relationships.

This research involved extracting RNA from two parabasalids from the termite *Reticulitermes hesperus*, translating that RNA into a protein sequence, and then using 14 of these sequences to compare determine the relationship of those two parabasalids to other parabasalids. The relationships mainly agreed with what previous researchers had determined using fewer sequences. Additionally, I found that the species I obtained my RNA from has been mistakenly considered the same as the species *Trichonympha agilis*; however, these two are in fact different species.

Preface

The work presented in this thesis is unpublished work performed by me, Caryn Cooper, in the Keeling lab at UBC. None of the text of this thesis is taken from previous publications.

Transcriptomes for Trichonympha 1 and Trichonympha 2 were generated by me. Transcriptomes other than those for *Trichonympha* 1 and *Trichonympha* 2 were generated by Nishimura et al., Filip Husnik in the Keeling Lab, and Martin Kolisko in the Keeling lab, and were obtained with the assistance of Vojtech Zarsky. Proteomes for *Trichomonas vaginalis* and *Tritrichomonas* foetus were obtained from the publicly-available UniProt online protein database as detailed in Section 2.4.3.1.

The Linux and Python code that was used in transcriptome processing and analysis was written by Vojtech Zarsky and me; previously-created Python scripts by Elizabeth Cooney and Nicholas Irwin were used with permission as detailed in Section 2.4.3.2.

Table of Contents

Abstract iii
Lay Summaryv
Preface vi
Table of Contents vii
List of Tablesx
List of Figures xi
List of Abbreviations xii
Acknowledgements xiv
1. Introduction 1
1.1. Taxonomy and Ecology of Relevant Parabasalia1
1.1.1. Phylogeny and Characteristics of Parabasalia
1.1.2. Phylogeny and Characteristics of Trichonymphida
1.1.3. Overview of <i>Trichonympha</i>
1.1.3.1 Phylogeny and Characteristics of <i>Trichonympha</i>
1.1.3.2. The host-specificity and distribution of <i>Trichonympha</i> species 6
1.2. Parabasalia in Termite Gut Symbiosis
1.3. Phylogenetic Methods in Parabasalian Research 10
1.3.1. SSU rDNA 11
1.3.2. Protein-Sequence Analysis
1.4. Purpose of Thesis 16
2. Materials and Methods 19

	2.1. Termite Collection and Maintenance	19
	2.2. Single-Cell Picking of Trichonympha Specimens	19
	2.3. RNA Isolation, Generation of cDNA Library, and Sequencing	20
	2.4. Bioinformatic Processing and Analysis of Sequenced Transcriptomes	21
	2.4.1. Trimming and Assembly of Reads and Transcriptome Annotation	21
	2.4.2. Identification of putative SSU, EF-1a, alpha-tubulin, beta-tubulin,	
	GAPDH, and actin sequences within <i>Trichonympha</i> 1 and 2	24
	2.4.3. Single- and Multi-Gene Phylogenetic Analysis	25
	2.4.3.1. Acquisition of Additional Transcriptomic Datasets	25
	2.4.3.2. Single Gene Phylogenetic Analysis	25
	2.4.3.3. SSU Sequence Phylogenetic Analysis	28
	2.4.3.4. Multi-Gene Phylogenetic Analysis	29
	2.5. Phylogenetic confirmation of <i>T. agilis</i> host range within <i>Reticulitermes</i>	30
3. R	esults	32
	3.1. Morphology of <i>Trichonympha</i> specimens 1 and 2	32
	3.2. Transcriptome Size and Completeness	32
	3.3. BUSCO Results From Transcriptomes	33
	3.4. BlobPlots	34
	3.5. SSU rDNA Sequences from Trichonympha 1 and 2	37
	3.6. Multi-Sequence Phylogenetic Analysis	39
	3.7. Distribution of <i>T. agilis</i> across Reported Termite Hosts	40
4. D	iscussion	46
	4.1. mRNA Isolation and Transcriptome Generation	46

References
5. Conclusions
4.5. "Trichonympha agilis" from a variety of termite species
4.4. Multi-Gene Phylogeny of Trichonymphida based upon selected genes
Phylogeny
4.3. Isolation of SSU Gene Sequences from Trichonympha 1 and 2 and SSU-Based
4.2. Transcriptome Quality 46

List of Tables

Table 1 . Transcriptomic details of <i>Trichonympha</i> specimens 1 and 2	34
Table 2. BUSCO scores for preliminary and deep sequencing reads of Trichonympha	
specimens	34

List of Figures

Figure 1. Light microscopy of <i>Trichonypha</i> cells illustrating morphology
Figure 2. Schematic phylogeny of investigated parabasalids reproduced from data in Cepicka et
al. 2016
Figure 3. Light microscopy image of Trichonympha specimen
Figure 4. Blobplot generated with BlobTools software for <i>Trichonympha</i> specimen 1
Figure 5. Blobplot generated with BlobTools software for <i>Trichonympha</i> specimen 2
Figure 6. Maximum-likelihood phylogenetic tree of trichonymphid, spirotrichonymphid, and
calonymphid specimens based upon SSU rDNA 39
Figure 7. Maximum-likelihood multi-sequence phylogenetic tree of 18 trichonymphid,
spirotrichonymphid, and calonymphid specimens40
Figure 8. Maximum-likelihood phylogenetic tree of SSU rDNA sequences for selected
Trichonympha species
Figure 9. Maximum-likelihood phylogenetic tree of COX-2 sequences for selected termite hosts
of Trichonympha species

List of Abbreviations

- ATP adenosine triphosphare
- BLAST Basic Local Alignment Search Tool
- bp base pair
- **BUSCO Benchmarking Universal Single-Copy Orthologs**
- cDNA complementary DNA
- EF-1a elongation factor 1 alpha
- GAPDH glyceraldehyde 3-phosphate dehydrogenase
- Gb giga-base pairs (billion base pairs)
- GHF glycoside hydrolase
- ML maximum-likelihood
- mRNA messenger RNA
- mtDNA mitochondrial DNA
- MYA million years ago
- NCBI National Center for Biotechnology Information
- PCR polymerase chain reaction
- rDNA ribosomal DNA
- RNA ribonucleic acid
- Rpb1 RNA polymerase II
- SNP single nucleotide polymorphism
- SSU small ribosomal subunit
- tRNA transfer RNA

UBC - University of British Columbia

US - United States of America

Acknowledgements

I would first like to acknowledge my thesis advisor, Dr. Patrick Keeling, for accepting me into his lab full of protists and assisting me through the construction of this thesis. I would also like to thank my supervisory committee (Dr. Naomi Fast and Dr. Patrick Martone) for their flexibility in scheduling my requisite proposal and thesis defenses and for giving me such useful advice after the proposal defence.

I would like to thank my labmates as well. My unending appreciation goes out to Vojtech Zarsky for teaching me the bioinformatics techniques that were required for the completion of this project. I must admit that, when I came to the Keeling Lab, I was one of those people who think Microsoft Excel is a perfectly fine graphing technology and scuttle away suspiciously, like a crab, should the computer they're running it on make any funny noises. You led me into the brave new world of directories and commands, scripts and novel languages, and visualization softwares. I would not have been able to complete this project, or even get through the very first steps of it, without your saintly-patient and always-available guidance. For these efforts, you deserve the whole world and possibly Mars as well - just as soon as I figure out how to wrap it neatly and leave it on your desk. I would also like to thank Elizabeth Cooney for creating many of the scripts that I used or modified in the process of my analysis, and for her guidance in showing me the overall shape of an MSc. project with Dr. Keeling.

Finally, I would like to acknowledge my family and friends for their love and support. Thanks to my mother and father, for pushing me when I needed it and being prodigal with the offspring provisioning; to my brother, for setting my days with his supportive text messages and accepting my awkward questions about computer science techniques as and when they suddenly became relevant to me; and to my friend Chanel, for never failing to believe that I could make both of us proud.

1. Introduction

1.1. Taxonomy and Ecology of Relevant Parabasalia

1.1.1. Phylogeny and Characteristics of Parabasalia

The protistan phylum Metamonada was first proposed by Grassé in 1952 and was redefined by Cavalier-Smith in 2003 as the group containing Parabasalia, *Carpediemonas*, Eopharyngia, and Anaeromonada. All groups within the Metamonada are anaerobic, and therefore do not have mitochondria; they may have hydrogenosomes instead or may have neither (Cavalier-Smith 2003). Metamonads are mainly commensals inside animal guts; however, the phylum as a whole is likely ancestrally free-living, with internal lineages having adapted to the symbiotic lifestyle (Cepicka et al. 2016)

Parabasalia is a group of flagellated anaerobic protists in the lineage Metamonada. Parabasalia was created in 1973 by Honigberg as a superorder, but was raised to phylum in 1981 by Cavalier-Smith, later reduced to class (Cavalier-Smith 2003), and currently is intermediate between class and phylum. The group is characterized by the presence of hydrogenosomes, membrane-bound organelles that share an evolutionary origin with mitochondria and similarly synthesize adenosine triphosphate (ATP) (Müer 1993, Bui et al. 1996, Shiflett and Johnson 2010), but are metabolically unlike mitochondria in that they perform anaerobic metabolism to generate ATP and hydrogen gas (Müer 1993, Leger et al. 2017). An additional characteristic feature of parabasalia is the presence of an eponymous parabasal apparatus, which constitutes a dense Golgi apparatus connected to the basal bodies of the cellular flagella (Kirby 1931, Cepicka et al. 2016). The homology between the Golgi apparatus and the parabasal apparatus was established early in the 20th century via chemical and microscopic techniques (King 1927). The basal bodies of the flagella, when grouped together in a particular arrangement, are called a mastigont (Cepicka et al. 2016).

Parabasalia were historically divided into two orders on morphological grounds: the large Hypermastigida, which can possess up to thousands of flagella and have high morphological complexity, and the much smaller Trichomonadida, which possess up to six flagella per mastigont and usually have simpler morphology (Cepicka et al. 2016). However, phylogenetic analyses indicate that hypermastigote morphology is a derived trait that has arisen in multiple parabasalian lineages, while trichomonad morphology is plesiomorphic (Cepicka et al. 2016). Hypermastigotes are therefore polyphyletic, their complex morphology having evolved multiple times independently within the termite/roach hindgut environment (Noda et al. 2012, Cepicka et al. 2016). The hypermastigote/trichomonad distinction remains in common use nonetheless.

The most recent full taxonomic revision of Parabasalia is that which was undertaken by Cepicka et al., which divided the group into 8 orders on the basis of information synthesized from multiple molecular phylogenetic analyses performed by various other research groups (2016), most of which used the short ribosomal subunit (SSU) rDNA sequence. The researchers endeavoured to include as many described genera as they could find molecular data on, creating a schematic phylogenetic tree and describing the characteristics of each order (Cepicka et al. 2016). Their schematic phylogeny divided the 8 orders into two broad clades: one containing a monophyletic Cristamonadida (including *Snyderella, Calonympha*, and *Coronympha*), a paraphyletic Tritrichomonadida (including *Tritrichomonas*), a monophyletic Spirotrichonymphida (including *Spirotrichonympha*, *Holomastigotoides*, and *Holomastigotes*), and a monophyletic Hypotrichomonadida, which was positioned basally to the three preceding

orders. The second clade contained the other four orders (Trichomonadida, including *Trichomonas*; Honigbergiellida; Lophomonadida, and Trichonymphida) which were each monophyletic, although these four could not be located basally or distally relative to one another and were shown as a 4-way polytomy.

1.1.2. Phylogeny and Characteristics of Trichonymphida

The order Trichonymphida contains the genera *Trichonympha, Staurojoenina, Hoplonympha, Euconomympha, Teranympha, Pseudotrichonympha, Leptospironympha, Urinympha,* and *Barbulanympha* (Cepicka et al. 2016). The deepest-branching groupwithin the order is the clade containing the sister groups *Trichonympha* and *Staurojoenina; Urinympha* and *Barbulanympha* also form a clade which is sister to the one containing *Leptospironympha, Pseudotrichonympha, Teranympha,* and *Eucomonympha* (Cepicka et al. 2016). Previous research from Carpenter *et al.* has also recovered the Cepicka topology, placing *Pseudotrichonympha* basal to *Barbulanympha* and *Urinympha* (Carpenter et al. 2011).

Trichonymphids are characterized by the presence of many flagella arranged along and around the anterior rostrum, and are distinguished from the morphologically-similar Spirotrichonymphida by the presence of the rostrum and the lack of a spiral flagellar arrangement (Cepicka et al. 2016). Sometimes, but not always, in Trichonymphida, the postrostral area bears flagella of its own (Cepicka et al. 2016). In trichonymphids, the parabasal complex can either compose multiple branches that surround the nucleus or multiple separate bodies (Cepicka et al. 2016).

1.1.3. Overview of Trichonympha

1.1.3.1 Phylogeny and Characteristics of Trichonympha

The type genus of the class Trichonymphea is *Trichonympha*, the first species of which to be described was *T. agilis* in 1877 by Joseph Leidy (Leidy 1881, Cepicka et al. 2010, 2016, James et al. 2013). Across the genus, *Trichonympha* species share highly similar morphological characteristics (Guichard and Gönczy 2016) and are often difficult to distinguish to the species level based upon morphology alone. Members of the genus *Trichonympha* are hypermastigotes and are therefore large and morphologically complex (James et al. 2013), characterized by a teardrop shape, a lack of mitochondria, basal bodies up to 5 µm long, and flagella that cover the entire anterior portion of the cell (Gibbons and Grimstone 1960, Biagini et al. 2006, Guichard and Gönczy 2016). The flagella are extremely numerous; in *T. campanula*, up to 14000 flagella have been estimated to be present on a single cell (Gibbons and Grimstone 1960). Cells are approximately 30-110 µm long and 21-90 µm wide (Carpenter et al. 2009). Flagella are arranged in rows in deep invaginations of the plasma membrane (Carpenter et al. 2010). *T. acuta* (and potentially other species) display elongate projections around the operculum (Carpenter et al. 2009).



Figure 1. *Trichonympha* cells visualized under inverted light microscopy to show morphology, including teardrop shape, numerous flagella, and rostrum. Image A shows three cells swimming in Trager's medium U after isolation from the termite gut; motion occurs in the direction the rostrum points.Images B and C show single cells.

Trichonympha is notable for its wide geographic and host distribution, occurring across

the globe in multiple species of termites and in Cryptocercus (Carpenter et al. 2009).

Trichonympha species are widespread obligate gut symbionts; and they also possess their own

bacterial symbionts (Yamin 1981, Carpenter et al. 2009). The most common bacterial symbionts are intracellular, but others have been observed to cover the posterior, non-flagellated region of some *Trichonympha* cells (Carpenter et al. 2009).

Trichonympha species constituted three well-supported groups in phylogenetic trees based upon the ribosomal small subunit (SSU) ribosomal DNA (rDNA) or RNA (rRNA) sequence: one group of taxa from *Cryptocercus* host cockroaches, one group from *Incisitermes* and *Porotermes* host termites, and one group from *Reticulitermes*, *Hodotermopsis*, and *Zootermopsis* host termites (Ikeda-Ohtsubo and Brune 2009, Boscaro et al. 2017). The species isolated from the termite sister group the wood-eating cockroach *Cryptocercus* form a stronglysupported clade while the species isolated from termites form two subgroups.

1.1.3.2. The host-specificity and distribution of *Trichonympha* species

Prior to the advent of nucleotide- and protein-sequencing technology, species in *Trichonympha* were described on morphological grounds. Molecular phylogenetic approaches, however, allow a better representation of the true diversity of these symbionts, as they do not depend on researcher interpretation of morphological features but on molecular sequence data.

The host termite species from which *Trichonympha agilis* was described is *Reticulitermes flavipes* (James et al. 2013). Subsequent studies have made identifications of *Trichonympha agilis* from a wide range of termite host species: Yamin's review and catalogue of the trichomonad, oxymonad, and hypermastigote flagellates of termites and *Cryptocercus* cites *Reticulitermes. flavipes, R. fukienensis, R. hesperus, R. lucifugus, R. lucifugus var. santonensis, R. speratus, R. tibialis,* and *R. virginicus* as hosts of *T. agilis* (1979). However, not all of these species epithets remain in use; *R. lucifugus var. santonensis* was promoted to species by Feytaud (1966), but subsequently synonymized with *R. flavipes* on the basis of mitochondrial DNA haplotype analysis (Feytaud 1966, Austin et al. 2005). James et al. propose that this cosmopolitan reported distribution for *T. agilis* does not reflect reality, and that the apparent wide host range of *T. agilis* is a result of researchers assuming that *Trichonympha* species with similar morphologies to *T. agilis* are *T. agilis*, regardless of host (2013). Subsequent research confirmed that specimens attributed as *T. agilis* are not monophyletic according to SSU rRNA gene sequence data when the attributed specimens come from different host species of *Reticulitermes* termites, and that the *Trichonympha* species in *T. virginicus* is sufficiently distinct from *T. agilis* to warrant the new epithet *Trichonympha burlesquei* (James et al. 2017).

R. flavipes an indigenous termite in Eastern United States, and has been introduced to other countries in South America and Europe, including France, Germany, Italy, Chile, and Uruguay (Austin et al. 2005, Ghesini et al. 2011, Baudouin et al. 2018). In British Columbia, two species of *Reticulitermes* termites are present: *R. hesperus*, the western subterranean termite, whose distribution is limited to the Pacific Coast and Western US, and *R. okanaganensis* in British Columbia, Idaho, Washington, Oregon, Nevada, and California (Smith and Rust 1994, Ye et al. 2004, Szalanski et al. 2006, McKern et al. 2007). As parabasalids with a termite species or with *Cryptocercus* as hosts show heavy host-specificity (Ohkuma and Brune 2011, Cepicka et al. 2016, Soviš n.d.), the likelihood of *Trichonympha agilis* being present in either *R. hesperus* or *R. okanaganensis* is quite low. Yamin, however, cites only *T. agilis* as a symbiont of *R. hesperus* (1979). If *Trichonympha* species are as host-specific as reported in the more recent literature, then it is probable that the *Trichonympha* species present in *R. hesperus* is in fact not *T. agilis* but a morphologically similar species that has been historically misidentified, and it would be in need of a new epithet.

1.2. Parabasalia in Termite Gut Symbiosis

Many parabasalians (and oxymonads) form obligate symbioses with lower termites (infraorder Isoptera, families Archotermopsidae, Hodotermitidae, Kalotermitidae, Mastotermitidae, Rhinotermitidae, Serritermitidae, Stolotermitidae, and Stylotermitidae) or with Cryptocercus (Inward et al. 2007, Beccaloni and Eggleton 2013, Cepicka et al. 2016). The ancestral parabasalian was likely an animal gut commensal, and existing free-living parabasalian species are most likely secondarily adapted from this lifestyle (although this ancestral parabasalian would likely have predated the origin of termites ~150 million years ago (MYA)) (Cepicka et al. 2016). The hindgut protist community in *Cryptocercus* was first described by Cleveland in 1934. This symbiotic relationship is believed to have originated from coprophagy and is implicated in the evolution of sociality within termites; as the microbial community became internalized in the gut rather than externalized in the feces, subsocial, and later social, behaviour increased the ease of inoculation of juveniles with the symbionts they needed (Nalepa et al. 2001). Nalepa also suggests that the increased social behaviour permitted the development of subsequent, more complex forms of symbiont transfer, such as proctodeal trophyllaxis (2001). The termite hindgut is morphologically adapted to containing these symbionts, with a dilated portion that accommodates a dense population (10⁷-10¹¹ cells/mL) of bacterial and protist symbionts (Ohkuma 2008, Ohkuma and Brune 2011). Hosts depend on their symbionts for the near-complete digestion of cellulose (Inoue et al. 2000, Ohkuma 2008). The system proposed for cellulose degradation in the termite digestive tract involves two parts: partial degradation via endogenous endoglucanase secreted from the salivary glands or midgut of the termite, and subsequent depolymerization of crystalline regions by gut protists in the hindgut (Ohkuma

2008). This symbiosis is ecologically significant in the decomposition of lignocellulose in terrestrial ecosystems (Abe et al. 2000, Ohkuma 2003, 2008).

Microbial communities are characteristic of their particular host species and consist mainly of lineages not found elsewhere (Ohkuma and Brune 2011). Analysis based on SSU and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) sequences shows that parabasalian symbionts of Cryptocercus have a sister-group relationship with the corresponding symbionts in Trichonymphida, indicating that the common ancestor of termites and *Cryptocercus* likely contained a set of trichonymphid flagellates that was vertically transmitted down both insect lineages (Ohkuma et al. 2009). Cospeciation is present between termite species and their parabasalian symbionts (Noda et al. 2007). There is a clear phylogenetic differentiation between Trichonympha species present in termites and those present in Cryptocercus, with those in *Cryptocercus* constituting a monophyletic group excluding those in termites (Carpenter et al. 2009); this is indicative of host-symbiont co-speciation due to vertical symbiont transfer (Ohkuma et al. 2009). However, there is no support for co-speciation within the lower termites between Trichonympha species and termite species, which would be supported by observing correlating tree topology between the phylogeny of Trichonympha species and the phylogeny of their respective hosts (Boscaro et al. 2017).

Many termite gut flagellates themselves harbour distinct bacterial ecto- or endosymbionts called Endomicrobia (Ikeda-Ohtsubo and Brune 2009), including spirochaetes in *Treponema*, *Bacteroidales*, *Synergistes*, methanogens in *Methanobrevibacter*, and the proposed phylum Termite Group 1 (TG1) (Ohkuma 2008). These endosymbionts cospeciate with their host flagellates and are likely inherited vertically in *Trichonympha* species; *Bacteroidales* and *Treponema* species occur in mutually-specific relationships with hosts (Noda et al. 2006, 2009a,

2018). *Hoplonympha* and *Strebomastix* have independently evolved morphological features including deep longitudinal furrows and vane-like extensions facilitating the attachment of their bacterial ectosymbionts (Noda et al. 2009a). The flagellate/bacterial relationship is estimated to have been established 40-70 MYA, significantly postdating the establishment of the termite/flagellate relationship (Ikeda-Ohtsubo and Brune 2009). The proposed advantages to the flagellates of hosting these symbiotic bacteria vary: some benefit their hosts through nitrogen fixation or provision of nitrogenous compounds not available in cellulose (Carpenter et al. 2009), while the endosymbiont of *Pseudotrichonympha grassii* demonstrates the ability to fix dinitrogen and recycle nitrogen wastes coupled with cellulolysis (Hongoh et al. 2008). Rare movement symbioses also occur: *Mixotricha paradoxa* is propelled by the undulation of its adherent spirochaetes (Cleveland and Grimstone 1964), while *Caduceia* is propelled by the flagella of ectosymbiotic *Synergistes* located in specialized pockets of the host membrane (Tamm 1982, Hongoh et al. 2007)

1.3. Phylogenetic Methods in Parabasalian Research

Parabasalian research and classification originated in 1836 with Donné's description of *Trichomonas vaginalis*; however, subsequent innovations in microscopy, and molecular phylogenetic approaches making use of DNA, RNA, and amino acid sequence data, have modified our understanding of phylogenetic relationships within the Parabasalia (Vandamme 2009). The most recent molecular phylogenetic research confirms the monophyly of Parabasalia as a whole within Metamonada, but does not always support the monophyly of its subgroups (Cepicka et al. 2016). Relationships are not all resolved to the same level (Cepicka et al. 2016). Parabasalids have often been placed basally within eukaryotes based upon the SSU rDNA

sequence (Dacks and Doolittle 2001); however, complicating analysis of the overall position of Parabasalids within Eukaryotes using other sequences is the fact that certain parabasalid genes appear to have been acquired through lateral gene transfer from bacteria and archaea, including two tRNA synthetases (Andersson et al. 2005).

1.3.1. SSU rDNA

The SSU rDNA sequence is frequently used as a marker gene for phylogenetic analysis of protists. The SSU gene encodes the small subunit ribosomal RNA molecule in the ribosome. The benefits of using this sequence for phylogenetic analysis are its ubiquity, due to the fact that all biological systems require the ribosome to translate mRNA into protein; its ease of isolation; and its rate of change with time (Byrne et al. 2018).

However, as with any selected marker sequences, the SSU has its own biases as well. The rDNA copy number in eukaryotes is highly variable, and extrachromosomal copies can also be generated; protists often contain multiple rDNA sequence copies (Tai et al. 2013, Wang et al. 2017). Increased copy number is associated with higher intraspecific and intragenomic SSU rDNA sequence variability (Wang et al. 2017), as any cell possessing more than one copy of the SSU gene could gain SNPs and/or indels individually per gene (Bobbett 2020). Intraspecific and intragenomic sequence variability would reduce the utility of this sequence for phylogenetic analyses, as increased variability will indicate greater divergence between sequences than is actually present in the full genome or is representative of their evolutionary history. Parfrey et al. claim that parabasalids do not appear to demonstrate any significant intraspecific genome variation that would influence phylogenetic studies using the SSU sequence (2008); however, others note significant intraspecific and intragenomic SSU sequence variability in

Trichonympha, ranging up to 1% intragenomic variation within certain individuals (Tai et al. 2013, Taerum et al. 2018, Bobbett 2020). Saldarriaga et al also cite up to 4% intraspecific variation in *Pseudotrichonympha* and high intragenomic divergence in *Kofoidia* (2011). Furthermore, sequence variability is itself variable, with some individuals having high intragenomic divergence and some having no divergence (Bobbett 2020). This creates a potential bias in the SSU sequence in that sequence variability - and therefore estimates of the rate of evolution - would be more pronounced in groups with high SSU copy numbers (Taerum et al. 2018). Depending on which members of a group with a more variable SSU sequence were sampled for an analysis, an assumption of rapid evolution could be drawn that may not be a true impression of the evolution rate of the group as a whole (Taerum et al. 2018): a critical consideration in terms of the application of the SSU as a universal marker region. The SSU sequence in Parabasalids also contains only approximately 1500 positions, which means that it is not highly informative for resolution of deeper relationships between parabasalid groups.

Additionally, as with any marker sequence, the conclusions that can be drawn are heavily dependent on the quality of the reference databases used (Dueholm et al. 2017), which is based almost entirely on depositions that skew towards certain taxonomic groups while underrepresenting others. Nonetheless, the vast majority of research into trichonymphid phylogeny has and continues to use SSU rDNA as a phylogenetic marker (Gerbod et al. 2004, Cepicka et al. 2016).

Although there exists a single report of Trichonymphida being polyphyletic on the basis of SSU sequence data (Ohkuma et al. 2005), more recent phylogenetic analyses do not bear out this claim and indicate that Trichonympha is probably truly monophyletic. Carpenter et al. found that Bayesian phylogenetic methods applied within the genus *Trichonympha* recover a

monophyletic *Trichonympha*, while maximum-likelihood trees place some *Trichonympha* specimens within Eucomonymphidae, a family with which they share few morphological features. However, *Trichonympha* and Eucomonympidae species are highly morphologically distinct in flagellar location and arrangement, suggesting that *Trichonympha* are most probably monophyletic and that the observed placement of some specimens of *Trichonympha* within Eucomonymphidae may be due to divergence within *Trichonympha* between *Trichonympha* species in termites and those in *Cryptocercus*, and subsequent long-branch attraction (Carpenter et al. 2009).

1.3.2. Protein-Sequence Analysis

Protein sampling can be used to overcome some pitfalls of SSU rRNA-based phylogeny (Noda et al. 2009b); however, investigations of Parabasalia based upon protein-sequence data have been relatively few and far between. The first study on the topic dates to 1998, investigating the sequences of glyceraldehyde-3-phosphate dehydrogenase (GADPH) in *Trichomonas vaginalis, Tritrichomonas foetus, Tetratrichomonas gallinarum, Trichomitus batrachorum,* and *Monocercomonas* (Viscogliosi and Müller 1998). These five parabasalids clearly clustered together and were separated into a distinct parabasalid group based upon a unique S-loop sequence more similar to the eubacteria than the eukaryotes (Viscogliosi and Müller 1998). Inferred enolase amino-acid sequences also placed *Trichomitus, Monocercomonas,* and *Trichomonas* in a clade separate from non-parabasalians, based mainly upon the lack of two close single-amino-acid deletions common to other eukaryotic enolase sequences (Keeling and Palmer 2000). These results support the monophyly of Parabasalia within the eukaryotes, but include an insufficient number of taxa to illustrate the group's internal phylogeny.

Gerbod et al. reported that a phylogenetic analysis of GADPH in 6 parabasalid species (Hypotrichomonas acosta, Tritrichomonas foetus, Monocercomonas sp., Trichomitus batrachorum, Trichomonas vaginalis, and Tetratrichomonas gallinarum) produced a wellresolved tree identical in phylogeny to the SSU rRNA tree (2004). However, analyses based upon enolase, a-tubulin, and b-tubulin resulted in poorly-resolved trees (Gerbod et al. 2004). This study was expanded upon in Ohkuma et al. (2007), which sequenced five parabasalids (Pseudotrichonympha grassii, Holomastigotoides mirabile, Spirotrichonympha leidyi, Devescovina sp., and Stephanonympha sp.) and also used multi-sequence methods to combine the results from each of the four sequences. They also found that GADPH was the optimal sequence with high resolution and statistical support, while enolase and both tubulins gave poorly-resolved trees (Ohkuma et al. 2007). All four of their single-protein analyses, as well as their multiprotein analysis, stably grouped *Pseudotrichonympha* and *Trichonympha* into a clade corresponding to Trichonymphida and placed Spirotrichonymphida as its sister group (Ohkuma et al. 2007). Analysis based upon concatenation of SSU rDNA and these four protein sequences also reproduced this clade (Cepicka et al. 2010). The agreement of GADPH, enolase, a-tubulin, and b-tubulin parabasalian phylogenies with the SSU rRNA-based tree was also corroborated by Hauck and Hafez (2010).

Additional single-protein analyses based on GADPH, actin, and elongation factor 1 alpha (EF-1a) support the placement of *Staurojoenina, Hoplonympha, Trichonympha, Eucomonympha, Pseudotrichonympha,* and *Teranympha* into a clade corresponding to Trichonymphida (Noda et al. 2012). GADPH also fully supports the monophyly of Cristamonadida (Noda et al. 2009b). The use of the *Rpb1* gene resolves a monophyletic Trichomonadea, Tritrichomonadea, and Hypotrichomonadea within the Parabasalia, and the use of the *Pms1* gene, which encodes a

homolog of a yeast protein that increases postmeitoic segregation, can resolve intraspecies- to genus-level relationships and supports a monophyletic Trichomonadea (Malik et al. 2011).

More recently, Nishimura et al. found that a maximum-likelihood phylogenetic analysis of chitinase sequences derived from single-cell transcriptomic data was useful in illustrating the evolutionary origins of this gene in Parabasalia and showed that *Cononympha leidyi* formed a monophyletic clade (2020); however, taxon selection in this study did not include species from Trichonymphida, and it therefore does not clarify any phylogenetic relationships within this order. Trichonymphid and spirotrichonymphid species do possess glycoside hydrolases (GHFs), which can be used in phylogenetic analyses to help identify certain termite gut symbionts and shed light on the presence or absence of these genes in ancestral protists (Sanderlin 2019). Within protists, a clear association appears in which GHF sequences are present in termiteassociated protists, indicating an acquisition of GHF genes within the termite gut (Sanderlin 2019). This study determined that the shared ancestor of Pseudotrichmonas, Lophomonas, and Trichonympha likely possessed GHF43, which was then secondarily lost as a synapomorphy of Trichonympha (Sanderlin 2019). However, the GHF phylogenies had poor resolution (Sanderlin 2019), indicating that GHF sequences are not appropriate phylogenetic markers in terms of resolving overall evolutionary relationships between parabasalians.

However, there are still some limitations to the above single-protein analyses using GADPH, EF-1a, actin, enolase, PMS1, and tubulins; taxon selection in Malik et al. did not include any trichonymphids, and therefore their results do not indicate anything about the use of PMS1 in Trichonymphida or the internal phylogeny of this group (2011). EF-1a genes display recent paralogy in some lineages and therefore lose some phylogenetic utility (Malik et al. 2011); paralogy is also present in GADPH (Viscogliosi and Müller 1998, Gerbod et al. 2004, Ohkuma

et al. 2007), although paralogs within a genome are closely related and therefore less likely to confound phylogenetic reconstruction at deeper levels.

1.4. Purpose of Thesis

Sequence-based identification and phylogenetic analysis of the trichonymphids remains limited, as analysis depends almost entirely upon SSU rDNA and reference sequences are often not identified to species level. There are no published genomes and very few published transcriptomes of termite parabasalian symbionts, and SSU-based rDNA phylogenies lose reliability for resolving relationships at deeper taxonomic levels.

The purpose of this thesis was to confirm previous phylogenetic analyses of the relationships between the parabasalid groups of Trichonymphida, Spirotrichonymphida, and Cristamonadida as well as some of the genera within each group. I hypothesized that both SSU-based and multi-sequence phylogenetic analyses would recover tree topologies similar to the tree topology obtained by Cepicka et al. (2016) (Figure 2), including a clear distinction between the Trichonymphids, Spirotrichonymphids, and Cristamonads. I hypothesized that *Trichonympha* would be the most basal branch in Trichonymphida and that that a monophyletic clade of *Urinympha* and *Barbulanympha* would be recovered.



Figure 2. Schematic cladogram of anticipated relationships between relevant parabasalid genera investigated in this thesis, drawn in accordance with relationships illustrated in Cepicka et al. (2016) but excluding taxa that were not included from this study. Length of branches does not illustrate sequence divergence or evolutionary time.

In this thesis, I use single-cell isolation and sequencing methods to generate transcriptomes for two *Trichonympha* cells from a population of *Reticulitermes hesperus* termites from Galiano Island, British Columbia. I use BLAST search methods to isolate rDNA sequences for each *Trichonympha* specimen from its transcriptome, and I use TransDecoder software to annotate those transcriptomes for protein-coding sequences. These transcriptomes are used, along with additional datasets from the Keeling Lab and publicly available databases, to generate a multi-gene phylogenetic tree to illustrate evolutionary relationships within trichonymphids. This phylogenetic tree is compared to previous phylogenies generated for the trichonymphids and used to assess the informativeness of the gene sequences used for such phylogenetic analyses.

Additionally, while I was doing the above research, it became clear that throughout the course of previous parabasalian research, *Trichonympha* specimens have frequently been misidentified during isolation and deposition of data into public repositories, as they cannot be distinguished by microscopy alone. This confounds the accuracy of phylogenetic analyses of this genus that rely on publicly-available data. In particular, *Trichonympha agilis* has been described as occurring in a wide variety of termite hosts worldwide, despite recent analyses indicating that *Trichonympha* species tend to be specific to their host termite species. In order to resolve this dilemma and to more accurately identify whether the *Trichonympha* species, I use rDNA sequence data to confirm the phylogenetic relationships between selected *Trichonympha* species, I use rDNA specimens (including the two cells above) with sequences deposited in NCBI GenBank and from Boscaro et al. (2017), many of which were identified as *T. agilis* from a wide range of *Reticulitermes* termite hosts.

I hypothesized that the sequences I had obtained from GenBank with the name *T. agilis* were not all representative of this single species, and that many had been misnamed; therefore, I hypothesized that these sequences would not form a monophyletic clade in the phylogenetic analysis. The resulting data support this hypothesis, illustrate that there is frequent misidentification of *Trichonympha* species as *T. agilis* even in non-type hosts, and that the *Trichonympha* species within *R. hesperus* is not *Trichonympha agilis* but is in fact a morphologically similar species that has been misidentified as such and therefore warrants redescription.

2. Materials and Methods

2.1. Termite collection and maintenance

Reticulitermes hesperus termites were collected from Galiano Island, B.C. Specimens were transported to and maintained in lab in sealed plastic containers to maintain humidity and provided with wood and leaf litter to mimic their natural habitat and for food.

Individual termites were dissected in March 2021 to isolate *Trichonympha* symbionts. Termites were removed from their containers with tweezers and placed into a Petri dish. Termites were immobilized with tweezers and euthanized via crushing of the head. The head and epiproct were removed using tweezers to detach the gut from the body wall. The gut was removed from the body using tweezers, visualized with the naked eye to ensure the full hindgut had been removed, and transferred to a 5 μ L microcentrifuge tube containing Tragers medium U. The microcentrifuge tube was manually agitated to facilitate mixing and suspension of the gut contents in Tragers medium U.

2.2. Single-Cell Picking of Trichonympha Specimens

Trichonympha cells were isolated using single-cell picking techniques under inverted light microscopy. Single-use, single-cell pipettes were created by flaming the center of a glass capillary tube over an alcohol burner until soft and applying traction to both ends to draw and narrow the central lumen. The capillary tube was then broken at the thinnest point, creating two pipette tips. Pipette tips were connected to manual suction via an adapter constructed using commercial micropipette tips and a rubber stopper.

 $5 \ \mu L$ of suspended termite gut contents were transferred to a slide in a single droplet, and the contents were observed under inverted light microscopy to visualize protists. Using the

pipette, a single live cell was transferred through 3-5 droplets of clean Tragers medium U to isolate it from the remainder of the gut contents. In each droplet, the cell was gently agitated to remove adhered protists or bacteria. Upon final isolation, each cleaned cell was transferred to a 0.2 mL thin-walled PCR strip tube for subsequent RNA isolation.

Two live *Trichonympha* cells were identified to genus based upon morphology and selected for RNA extraction.

2.3. RNA Isolation, Generation of cDNA Library, and Sequencing

Cell lysis, RNA isolation, reverse transcription, and PCR amplification were performed according to a published protocol (Picelli et al. 2014). cDNA library generation took place in a designated biosafety cabinet which was sterilized with UV light and spray 70% ethanol prior to and following each use. Contamination was minimized through the use of disposable protective sleeves and vinyl gloves. At indicated stopping points during the Picelli protocol, samples were frozen at -70°C for preservation. Samples were maintained on ice during transport to and from the freezer to maintain RNA stability.

The cleaning and tagmentation (a reaction that combines cDNA fragmentation and fragment tagging) steps from the Picelli protocol were not performed in the laboratory environment, but were performed as part of subsequent sequencing runs with the UBC Sequencing and Bioinformatics Consortium; therefore, the protocol was halted at step 26 and the isolated cDNA frozen.

The resulting cDNA libraries were sent for preliminary and deep Illumina sequencing with the UBC Sequencing and Bioinformatics Consortium.

2.4. Bioinformatic Processing and Analysis of Sequenced Transcriptomes

2.4.1. Trimming and Assembly of Reads and Transcriptome Annotation

Sequencing files were returned in fastq.gz format and uploaded to the Keeling Lab's Jezero bioinformatics server. Both specimens were assigned their numerical designations (Specimen 1 and Specimen 2) at this time. Reads were saved in the original fastq.gz format and in subsequent formats. Generated data were periodically downloaded to a backup server to protect against data loss in case of breakdown in the main server.

Reads were trimmed using Trimmomatic software (Bolger et al. 2014) to remove low quality bases from the beginnings and ends of reads and to remove remaining adapter sequences from the returned data. The parameters used were as follows: Illumina adapters were provided in a .fa file, and Trimmomatic was set to look for seed matches with these adapter sequences with a maximum of 2 mismatches. The seeds were extended and clipped if paired-end reads reached a score of 30 and single-end reads reached a score of 10. Leading and trailing bases were removed if they had a quality score below 3. The read was scanned with a 4-base-wide sliding window, and the selected bases were cut when the average quality per base dropped below 15. The influence of the base quality threshold during trimming on the completeness of the subsequent assembly was assessed, but it was determined that the number of reads retained remained relatively constant regardless of base quality threshold. The reads were subsequently scanned for length, and reads less than 100 bp long were dropped as they were unlikely to be sufficiently informative for subsequent transcriptome assembly to outweigh the computational power required to process. Read quality was assessed for both trimmed and original (untrimmed) reads using the FastQC quality control tool (n.d.).
Trimmomatic software finds matching pairs of forward and reverse reads and pairs them together. Reads without a matching forward/reverse read are designated as "unpaired". Unpaired reads for each specimen were merged with the Linux zcat command to reduce the number of files necessary for SPAdes processing.

Reads were assembled into full transcriptomes using the SPAdes assembly algorithm (Bankevich et al. 2012). The SPAdes algorithm is an A-Bruijn assembler which adjusts k-values according to coverage to appropriately balance the tendencies to collapse repeats into a single contig at smaller k-values and the tendency to miss overlaps at larger k-values (Bankevich et al. 2012, Nikolić 2021). The RNA flag was used to indicate that an RNA-Seq dataset was being assembled. A memory limit of 200 Gb and a thread number of 24 were set.

To indicate how the SPAdes assembly corresponded to the original provided read library, read mapping was performed using the BWA software package (Li and Durbin 2009). The resulting files were subsequently sorted and indexed using the Samtools program suite (Li et al. 2009).

Blobplots were generated using BlobTools software (Laetsch and Blaxter 2017) to visualize the proportion of the dataset successfully mapped and to visualize the taxonomic partitioning (compared to the NCBI Taxonomic Database), GC proportion, and coverage of the dataset. The contiguous sequences ("contigs") were compared to the NCBI non-redundant protein database using the DIAMOND blastx alignment algorithm (Buchfink et al. 2015). The contigs were plotted based upon the order of the protein sequence to which they had the highest sequence similarity. Matches were displayed at the taxonomic level of order, as the NCBI non-redundant protein database is lacking in reference sequences from *Trichonympha* species.

To remove contamination from the dataset, the DIAMOND search results were parsed using a Python script, and a subset of contiguous sequences that had aligned to sequences designated in the taxon Parabasalia was isolated. This subset of transcripts was then used to calculate BUSCO scores.

Completeness score data was generated for the shallow sequencing reads and the deep sequencing reads using Benchmarking Universal Single-Copy Orthologs software (BUSCO) (Waterhouse et al. 2018). This tool compares the genes within the dataset to an established lineage-specific dataset, generating a BUSCO score that indicates how many of the genes from said dataset were found in the transcriptome in question (Waterhouse et al. 2018). *Trichonympha* specimens 1 and 2 were compared using the BUSCO tool to the BUSCO.V4 Eukaryota Odb10 dataset, as this is the most specific dataset that contains Parabasalia (n.d.); this dataset contains 255 total BUSCO groups. The transcriptome assessment mode was selected.

Protein-coding regions were identified using blastx with the DIAMOND algorithm against the UniProt database reference proteomes (Buchfink et al. 2015, n.d.). The UniProt reference proteomes are a series of proteomes that have been selected as a representative and non-redundant cross-section of the taxonomic diversity present in the UniProt database (n.d.).

Transcriptomes were annotated for open reading frames using TransDecoder software filtered via blastp with the DIAMOND algorithm against the Uniprot database reference proteomes.

2.4.2. Identification of putative SSU, EF-1a, alpha-tubulin, beta-tubulin, GAPDH, and actin sequences within *Trichonympha* 1 and 2

Small ribosomal RNA sequences were isolated from the assembled transcriptomes by cmscan and then assigned taxonomic identifications by blastn search against the SILVA rRNA database (n.d.). The output was formatted in a tab-separated text file with high-scoring pairs reported only for the first 1000 target sequences. The SILVA blastn search was parsed using a Python in order to assign taxonomic identities to the isolated SSU sequences.

At this point, identification of both specimens to genus level was confirmed by performing a BLASTn search of the isolated ribosomal rRNA sequences against the NCBI GenBank nucleotide database; this procedure was used to account for phenotypic plasticity among *Trichonympha* species (Boscaro et al. 2017). Small ribosomal RNA sequences were saved on Geneious Prime genome annotation software for subsequent processing.

Ef-1a, alpha-tubulin 1, beta-tubulin, actin, and GAPDH sequences from *Trichonympha* specimens 1 and 2 were isolated via BLASTn search against the generated transcriptomes using as query sequences vouchers downloaded from NCBI GenBank for these genes in *Trichomonas vaginalis*. Sequence U63122.1 was selected as the query sequence for actin, AF327848.1 for alpha-tubulin 1, HM217352.1 for Ef-1a, L05468.1 for beta-tubulin 1, and AF022414.1 for GAPDH. BLAST nucleotide databases were generated from the assembled transcriptomes of *Trichonympha* specimens 1 and 2, and these five sequences were compared to the databases via a megablast search set to return up to 50 of the closest hits.

2.4.3. Single- and Multi-Gene Phylogenetic Analysis

2.4.3.1. Acquisition of Additional Transcriptomic Datasets

The two transcriptomes acquired through the above techniques were both determined to be derived from the sole *Trichonympha* species symbiotic to *R. hesperus* via morphology and SSU BLAST comparison against the NCBI Genome Database. Transcriptomes for other representative genera in Trichonymphida were obtained from previous datasets. Eight transcriptomes of *Pseudotrichonympha grassii* were obtained from Ninshimura et al.; of these, three were hiseq runs and 5 were miseq runs. All eight transcriptomes were combined into a single dataset for *Pseudotrichonympha grassii*. Transcriptomes for *Coronympha* and *Snyderella* were obtained from Filip Husnik in the Keeling Lab, and transcriptomes for two specimens of *Calonympha*, two specimens of *Holomastigotoides*, *Spirotrichonympha*, two specimens of *Barbulonympha*, *Pseudotrichonympha*, and two specimens of *Uronympha* were obtained from Martin Kolisko in the Keeling Lab. Reads from these datasets were assembled into full transcriptomes using the SPAdes assembly algorithm, and ORF prediction was performed using TransDecoder software.

Trichomonads were represented by *Trichomonas vaginalis* and *Tritrichomonas foetus*. Proteomes were downloaded for these two species from the UniProt database (n.d.). The proteome ID for the *T. vaginalis* proteome was UP000001542 and for the *T. foetus* proteome was UP000179807.

2.4.3.2. Single Gene Phylogenetic Analysis

The Keeling lab is in possession of a dataset of 263 eukaryotic protein-coding genes. BLAST search techniques were used to identify these genes in the generated transcriptomes in order to identify appropriate genes for multi-gene tree building. A Python script was developed for this purpose.

Each protein-annotated transcriptome generated with TransDecoder, whether from *Trichonympha* specimens 1 and 2 or obtained from collaborators, was renamed with its appropriate specimen ID, and a BUSCO analysis was run on each of the transcriptomes. A BLAST database was generated from each of the new annotated transcriptomes. The Keeling Lab's 263-gene dataset was compared to the transcriptome databases using BLASTp analysis, and the outputs from this analysis were parsed using a Perl script previously generated by Elizabeth Cooney from the Keeling lab for ease of interpretation.

The original transcriptome datasets were then concatenated, and each of the 263 genes was searched within this concatenated dataset using BLAST search to identify homologs, if present of each of these 263 genes within the transcriptomes. The genes that were identified were added to new files containing the pre-existing sequences of each of these 263 genes from a wide range of eukaryotes.

A BLAST database was generated for the UniProt SwissProt database, and the identified gene homologs from each transcriptome were compared against this database to identify sequences which required trimming prior to tree construction. A previously-generated Perl script from Elizabeth Cooney was used to trim these sequences to remove extensions that might have been retained from the sequencing process. These extensions would not be present in the UniProt SwissProt database.

The trimmed homologs were included with the sequences of the 263 eukaryotic genes to generate single-gene phylogenetic trees showing the relative position of the parabasalian species represented by the novel transcriptomes within the wider phylogeny of eukaryotes. As homologs

of the 263 eukaryotic genes used were previously identified from the novel transcriptomes by BLAST search, these homologs were added to the previous dataset to generate 263 .fasta files, each one containing all the translated protein sequences for the eukaryotic homologs of a specific eukaryotic gene. These sequences were aligned using the MAFFT sequence alignment program using the L-INS-i iterative refinement method (Katoh 2005) and overhanging flanking sequences were trimmed using trimAL software with a gap threshold set at 0.8 (Capella-Gutiérrez et al. 2009). As occasionally the entirety of a sequence may be trimmed if it does not align at any point along its length with the majority of the other sequences in the analysis, a Python script from Elizabeth Cooney was used to remove any sequences that had been entirely trimmed.

263 maximum-likelihood phylogenetic trees were generated from these single-gene alignment files using IQTree software (Nguyen et al. 2015) using the LG+G substitution model (Tamura et al. 2013). The generated trees were viewed in FigTree v1.4.4 graphical software (n.d.) and the search feature was used to identify paralogous sequences from each of the novel taxa that had been included in the tree. Paralogs were filtered manually, with paralogs that clustered with highly unrelated sequences (indicating significant sequence divergence) and had short (and therefore less informative) sequences colour-coded for subsequent removal from the dataset. This filtering resulted in a maximum of 1 retained paralog per eukaryotic gene per novel taxon. The colour-coded paralogs were removed from the dataset using a pre-existing Python script provided by Elizabeth Cooney and created by Nicholas Irwin.

Visualization of the single-gene phylogenetic trees was also used to inform selection of the gene sequences that were used for subsequent multi-gene phylogenetic analysis. Genes that displayed significant paralogy in multiple eukaryotes were selected against, whereas genes that generated phylogenetic trees with low paralogy and high concordance with pre-existing analyses

of parabasalian phylogeny were preferred. Additionally, genes that were absent in a large proportion of the novel transcriptomes were excluded.

After single-gene analysis, sixteen gene sequences were selected for inclusion in the multi-gene phylogenetic analysis due to high conservation and low presence of paralogues in the dataset: alpha-tubulin, actin, GAPDH, EF-1a, beta-tubulin, rps2, rpl30, VPS26B, RPS24, RPS21, RPF1, POLR2F, HSP75mito, EIF2a, COPG2, and ABCE. Hsp90 and ARP2 were suggested by Elizabeth Cooney as potential sequences and searched for, but were not found in the dataset and were therefore not included.

2.4.4.3. SSU Sequence Phylogenetic Analysis

From each of the annotated transcriptomes, SSU rRNA sequences were isolated by cmscan and then assigned taxonomic identifications by blastn search against the SILVA rRNA database (n.d.). The output was formatted in a tab-separated text file with high-scoring pairs reported only for the first 1000 target sequences. The SILVA blastn search was parsed using a Python in order to assign taxonomic identities to the isolated SSU sequences. These isolated SSU rRNA sequences were concatenated in Geneious Prime, aligned using MAFFT with L-INS-i and trimmed using TrimAL, and a maximum-likelihood phylogenetic tree was constructed using IQTree software. This tree was rooted on the branch between the clade containing the Trichonymphids and the clade containing the Cristamonads and Spirotrichonymphids, as this is the division illustrated in the most recent overall phylogenetic and taxonomic revision of Parabasalia (Cepicka et al. 2016).

2.4.3.3. Multi-Gene Phylogenetic Analysis

The sequences of the retained paralogs of the sixteen selected genes from the novel transcriptomes were extracted from the overall eukaryotic dataset into sixteen new .fasta files. These sequences were re-aligned using MAFFT with L-INS-i (Katoh 2005) and overhanging sequences were re-trimmed using trimAL software with a gap threshold set at 0.8 (Capella-Gutiérrez et al. 2009). The trimmed alignments were uploaded to the SCaFoS (Selection, Concatenation, and Fusion of Sequences) phylogenetic inference software (Roure et al. 2007) which was run on the Keeling lab's Jezero server but visualized in graphical mode through the XQuartz windowing system for macOS (n.d.). Sequence names were edited manually within each file to reflect the required format for interpretation by SCaFoS.

A list of included taxa (OTUs) was generated by running SCaFoS's automatic "Species Presence" feature. Eighteen different OTUs, corresponding to the eighteen novel transcriptomes obtained and annotated, were identified, ensuring that in subsequent steps, each sequence would be appropriately linked to the transcriptome from which it has been isolated. New alignment files for each of the sixteen selected gene sequences were generated by running the "File Selection" feature, which isolated only sequences with sequence names corresponding to the previous 18 OTUs from the new .fasta files above. These new alignment files were written out in .ali format.

Finally, a concatenated dataset was created by running SCaFoS's "Dataset Assembling" feature. As previous steps had filtered the dataset to contain a single sequence per gene per OTU, the "longer sequence" option was selected to determine the best sequence within each OUT, as this option would return results faster but with less processing power than the "minimal evolutionary distance" option, which would have been optimal had this pre-filtering not been performed and there had been multiple sequences per gene per OTU. The sequences of all

sixteen genes were concatenated, and the concatenated aligned multi-gene file was output in .fasta format. A phylogenetic tree was generated from this alignment using IQTree software set to automatically test the best-fitting model for the data and 1000 bootstrap replications.

A statistical file summarizing the concatenated alignment was also generated to illustrate the percentage of missing positions per sequence, the percentage of missing OTUs per gene, and the sequence length of each sequence. It was determined that RPS24 contained a high proportion (>40%) of missing OTUs, and a decision was made to remove this gene from the analysis. The updated concatenated alignment contained sequences from 15 genes, and an updated phylogenetic tree was generated using the parameters above. In accordance with the previous SSU-based tree, this tree was rooted on the branch between the clade containing the Trichonymphids and the clade containing the Cristamonads and Spirotrichonymphids.

2.5. Phylogenetic confirmation of T. agilis host range within Reticulitermes

Twenty-eight full rDNA sequences with a specified identity as a *Trichonympha* species were downloaded from the NCBI GenBank database using the search query "ORGN=Trichonympha" followed by manual selection. Sequence lengths ranged from 409 bp - 1593 bp. This dataset was expanded with the 158 sequences utilized by Boscaro et al. (Boscaro et al. 2017), the full records of which were downloaded from the NCBI GenBank database using the Batch Entrez software after compilation of a list of sequence IDs from the trimmed dataset.

The total rDNA dataset was aligned using MAFFT multiple sequence alignment software using the L-INS-i algorithm (Katoh 2005) and was trimmed using Block Mapping and Gathering with Entropy (BMGE) software (Criscuolo and Gribaldo 2010). One duplicate sequence was removed, and a maximum-likelihood phylogenetic tree was constructed using IQTree phylogenetic inference software (Nguyen et al. 2015). One thousand bootstrap replications were used. The resulting tree was visualized in FigTree software version 1.4.4.

The NCBI GenBank database allows researchers to include information about the host species of a symbiont from which a particular sequence was extracted. The reported hosts of each of the sequences used in this analysis were recorded in conjunction with the accession number, and each sequence was colour-coded in accordance with the reported host species for easy analysis of the illustrated relationships. As sequences AB434787.1, KJ778600.1, KJ778601.1, KJ778602.1, KJ778603.1, and KJ778604.1 had been deposited with *R. santonensis* as the reported host despite the 2005 synonymization of *R. santonensis* with *R. flavipes (Austin et al. 2005);* this name was retained in the list of host species and the tree annotated to reflect the more recent classification.

In order to illustrate the phylogenetic relationships of the host species, COX2 sequences for each of the reported termite hosts were obtained from the GenBank nucleotide database using the Advanced Search function with the species name of each host as the query. Although COX1 is more favoured as an animal barcode, not all of the reported termite hosts had an available COX1 sequence in the NCBI GenBank database, while all possessed a deposited COX2 sequence, which has also proven useful as a potential barcode (Ahmed 2022). The obtained COX2 sequences were concatenated into a single .fasta file and aligned using MAFFT and trimmed with BMGE. A maximum-likelihood phylogenetic tree was constructed using IQTree phylogenetic inference software with 1000 bootstrap replications and visualized in FigTree software version 1.4.4 for concordance with the previous tree of *Trichonympha* species.

3. Results

3.1. Morphology of *Trichonympha* specimens 1 and 2

When viewed under inverted light microscopy, specimens *Trichonympha* 1 and *Trichonympha* 2 both possessed the morphology typical of *Trichonympha*. Each cell was teardrop-shaped, with a distinct anterior rostrum and a large number of trailing flagella. Cells appeared colourless under light microscopy.



Figure 3. Light microscopy image of *Trichonympha* specimen 2 viewed during single-cell picking. The anterior rostrum is visible in the upper left-hand corner, and numerous flagella can be seen pointing posteriorly along the cell.

3.2. Transcriptome Size and Completeness

Transcriptomes were generated for both Trichonympha 1 and Trichonympha 2. Details of the transcriptome assemblies are provided in Table 1. The transcriptome from Trichonympha

specimen 2 was 184% larger than the transcriptome from Trichonympha 1, and had 7.40% greater GC content, as well as having a larger median contig size (Table 1).

Parameter	Data for ID				
	Trichonympha specimen 1	Trichonympha specimen 2			
No. of raw sequencing	34515	51751			
reads					
Largest contig size	12301	9002			
Transcriptome size (bp)	20 251 471	37 449 287			
N50 (bp)	681	939			
GC content (%)	40.22956632	47.6285236			
% N	0.014961876	0.007263156			

Table 1. Transcriptomic details of *Trichonympha* specimens 1 and 2.

3.3. BUSCO Results

BUSCO scores were increased in the deep sequencing reads compared to the shallow

sequencing reads, indicating greater transcriptome completeness. The results of the BUSCO

analysis are provided in Table 2.

Table 2. BUSCO scores for preliminary and deep sequencing reads of *Trichonympha* specimens 1 and 2. There was an increase in complete BUSCO proportion for the deep sequencing reads compared to the preliminary reads.

Specimen	Read Depth	Complete #	Fragmented #	Missing #	Total #
		(%)	(%)	(%)	
1	Preliminary	24 (9.41%)	10 (3.92 %)	211 (82.75)	255
	Deep	27 (10.59%)	13 (5.10 %)	215 (84.31 %)	255
2	Preliminary	44 (17.25 %)	13 (5.10%)	198 (77.65 %)	255
	Deep	56 (21.96%)	12 (4.71 %)	187 (73.33 %)	255

3.4. BlobPlots

Blobplots were generated using BlobTools software, alloweing for visualization of the taxonomic partitioning of the transcriptomes. In both *Trichonympha* isolates, and in both shallow and deep sequencing reads, strong signals were obtained from parabasalids (Tritrichomonadida and Trichomonadida). Signals were also obtained from Corynebacteriales, Clostridiales, Hymenoptera, and Blattodea.

Both specimens displayed two clusters in GC proportion, one between 0.2 and 0.4 GC content and one around 0.6 GC content. Coverage in these sequences ranged between 1x and 1000x. It is probable that a large proportion of the "no-hit" and "other" sequences are also derived from *Trichonympha* specimens 1 and 2 and are unlabeled due to the reference dataset used for taxonomic partitioning not possessing a closer category to *Trichonympha* than *Tritrichomonadida* and *Trichomonadida*. The overall proportion of reads mapped was 97.36% for specimen 1 and 93.62% for specimen 2.

trichonympha10.blobDB.json.bestsum.order.p8.span.100.blobplot.bam0



Figure 4. Blobplot generated with BlobTools software for *Trichonympha* specimen 1. Parabasalian sequences are colour-coded in red and orange. The upper portion of each plot indicates the taxonomic partitioning, coverage, and GC proportion of each read, while the lower portion indicates the proportion of paired and unpaired reads within the assembly and the proportion of reads that were assigned to each taxonomic partition via BLAST.

trichonympha12.blobDB.json.bestsum.order.p8.span.100.blobplot.bam0



Figure 5. Blobplot generated with BlobTools software for *Trichonympha* specimen 2. Parabasalian sequences are colour-coded in red and orange. The upper portion of each plot indicates the taxonomic partitioning, coverage, and GC proportion of each read, while the lower portion indicates the proportion of paired and unpaired reads within the assembly and the proportion of reads that were assigned to each taxonomic partition via BLAST.

3.5. SSU rDNA Sequences from Trichonympha 1 and 2

Small subunit rDNA sequences were successfully obtained from both *Trichonympha* specimen 1 and *Trichonympha* specimen 2. When these sequences were compared to the NCBI Genome database using BLASTn search, 99 out of the 100 results with the lowest e-value for both *Trichonympha* 1 and *Trichonympha* 2 were sequences that had also been identified as *Trichonympha* species, confirming the placement of these two specimens in the genus *Trichonympha*.

A successful phylogenetic tree rooted on the trichonymphids in accordance with previous analyses (Cepicka et al. 2016) was generated from the dataset of the SSU rDNA sequences obtained from each of the transcriptomes with the addition of SSU rDNA sequences obtained from GenBank for each genus (Figure 6). Distinct clades representing the spirotrichonymphids, and trichonymphids, with 100% bootstrap support, were observed. The calonymphids were paraphyletic, containing the spirotrichonymphids.

Monophyly was observed for every genus included within this analysis (Figure 6). This analysis obtained a monophyletic *Trichonympha* located as a sister group to the other trichonyphids with 100% bootstrap support (Figure 6). *Pseudotrichonympha* (100% bootstrap support) was located sister to a clade containing *Urinympha* and *Barbulanympha* (Figure 6)

Within the Spirotrichonymphida, *Holomastigotoides* was located basally to a clade consisting of *Spirotrichonymphida* and *Holomastigotes* (Figure 6). Within the calonymphids, *Calonympha* and *Snyderella* formed a clade with 89% bootstrap support. This clade and the Spirotrichonymphids formed a single larger clade with 62% bootstrap support, to which the genus *Coronympha* was the sister group (Figure 6).



Figure 6. Maximum-likelihood phylogenetic tree of trichonymphid, spirotrichonymphid, and calonymphid specimens generated using GTR+F+I+G4 model based upon SSU rDNA sequences. Nodal values represent percentage bootstrap support in 1000 replications. Sequences with accession numbers were obtained from the NCBI GenBank public database, while sequences without accession numbers were isolated from sequenced and assembled

transcriptomes. The root was placed between the Trichonymphids and the clade containing the Spirotrichonymphids and Cristamonads.

3.6. Multi-Sequence Phylogenetic Analysis

A multi-sequence phylogenetic tree was generated from the novel transcriptomic data (Figure 7). Three clades are present, representing the calonymphids, spirotrichonymphids, and trichonymphids.



Figure 7. Maximum-likelihood unrooted phylogenetic tree of 18 trichonymphid, spirotrichonymphid, and calonymphid specimens based upon translated alpha-tubulin, actin, GAPDH, EF-1a, beta-tubulin, rps2, rpl30, VPS26B, RPS21, POLR2F, HSP75mito, EIF2a,

COPG2, and ABCE protein sequences from transcriptome sequencing, including sequences from specimens with transcriptomes in this study (*Trichonympha* specimen 1 and *Trichonympha* specimen 2). Nodal values represent percentage of bootstrap support in 1000 bootstrap replications. The root was placed between the Trichonymphids and the clade containing the Spirotrichonymphids and Cristamonads.

Holomastigotoides and *Holomastigotes* formed a clade with high bootstrap support (86%), and these two species also formed a clade with *Spirotrichonympha* representing *Spirotrichonymphida* (Figure 7). Both specimens of Calonympha did not form a monophyletic clade; one specimen grouped with *Snyderella* with high bootstrap support (100%), whereas the other *Calonympha* specimen was localized within *Pseudotrichonympha* (Figure 7), grouping with *P. grassii* with 100% bootstrap support.

Pseudotrichonympha clustered most closely with *Barbulanympha*. *Trichonympha* was not monophyletic in this tree but was paraphyletic (Figure 7). The monophyletic clade that contained both *Trichonympha* specimens also contained *Pseudotrichonympha*, *Urinympha*, and *Barbulanympha* (Figure 7).

All nodes had high bootstrap support, with the lowest bootstrap value across the entire maximum-likelihood tree being 86%.

3.7. Distribution of T. agilis across reported termite hosts

A successful phylogenetic tree was generated from NCBI-sourced SSU rDNA sequences for *Trichonympha* species combined with sequences provided by Boscaro et al. to illustrate the phylogeny of *Trichonympha* as a genus (Figure 8). Two major clades were observed; one containing the taxa with *Incisitermes* termites as hosts, and one containing the taxa with *Reticulitermes* termites as hosts. Taxa that had been assigned the name of *Trichonympha agilis* in NCBI GenBank did not form a clade; however, a clade containing all the taxa whose host termites were *R. hesperus* was observed (Figure 8, coloured teal).



Figure 8. Maximum-likelihood phylogenetic tree of SSU rDNA sequences for selected *Trichonympha* species using the GTR+F+G4 model, illustrating phylogenetic structure of the genus. Sequences have been colour-coded according to host termite species from which the specimen was isolated, according to the same colour code used in Figure 9 below. Taxon names and accession numbers have been retained from the NCBI GenBank deposition metadata, and names may not represent the most accurate taxonomic assignment for each taxon.

A phylogenetic tree was generated from COX-2 sequences isolated from GenBank to illustrate the phylogeny of the termite hosts for the above *Trichonympha* species (Figure 9)



Figure 9. Maximum-likelihood phylogenetic tree of COX-2 sequences for selected termite hosts of *Trichonympha* species, illustrating host phylogeny. Sequences have been colour-coded according to termite species for ease of viewing, and to aid in the interpretation of Figure 8 above.

Two distinct clades were apparent: one which contained all *Reticulitermes* host species, one containing *Hodotermopsis* and *Incisitermes* (Figure 9).

This phylogenetic tree contained a monophyletic *Incisitermes* with 100% bootstrap support, as well as a monophyletic *Reticulitermes virginicus*, *R. lucifugus*, and *R. hesperus*, each with 100% bootstrap support. *R. aculabialis* was also monophyletic in this tree with 68% bootstrap support. This tree contained a paraphyletic *R. speratus* and *R. chinensis* (Figure 9). All specimens of *R. flavipes* and *R. santonensis* grouped into a single clade with 100% bootstrap support with no clear phylogenetic distinction between the specimens identified with each epithet.

There was no nodal correlation between the phylogenetic tree of *Trichonympha* species and the phylogenetic tree of their termite hosts, indicating that co-speciation has not occurred between *Trichonympha* species and their hosts.

4. Discussion

4.1. mRNA Isolation and Transcriptome Generation

The successful generation of a transcriptome from both *Trichonympha* specimen 1 and *Trichonympha* specimen 2 indicates that the Picelli et al. protocol is effective for isolation and purification of RNA libraries from parabasalian cells (Picelli et al. 2014). Picelli et al. note that for optimal results cells should be processed rapidly at near-physiological conditions (Picelli et al. 2014); this was a particular concern during this study as parabasalian protists are highly specialized for the anaerobic environments within their host termites and wood-eating cockroaches, and during the picking process non-target cells could be observed losing motility and obtaining a hypertonic, swelled shape despite the use of Trager's medium U to approximate their anaerobic environment. As RNA quality degrades following cell death or under cell stress, cells were frozen rapidly after isolation. For future research in this vein using single-cell picking for transcriptomic analyses - particularly in research generating larger numbers of transcriptomes - the optimal procedure would likely be to pick and freeze cells in small batches to prevent degradation of RNA in each cell.

4.2. Transcriptome Quality

The transcriptomes used in this study varied in quality. In particular, the transcriptome for *Pseudotrichonympha grassii* was obtained in raw form as three different Illumina HiSeq runs and five different Illumina MiSeq runs, which were concatenated into a single raw transcriptome prior to assembly with the intention of obtaining greater coverage. However, this transcriptome still had only 5 out of the 16 genes used to construct the multi-gene phylogenetic tree present after BLAST search, indicating that the concatenation was not advantageous in terms of

increasing transcriptome coverage in this situation. The most complete transcriptome was *Trichonympha* specimen 2, which contained a sequence for each of the genes used in the multigene tree, which illustrates that the Picelli et al. protocol can successfully generate high-coverage transcriptomes for phylogenetic analysis (Picelli et al. 2014).

4.3. Isolation of SSU Gene Sequences from *Trichonympha* 1 and 2 and SSU-Based Phylogeny

The protocols used were successfully able to isolate SSU gene sequences from both *Trichonympha* specimens. Both sequences were similar in length and GC content to the other SSU gene sequences from *Trichonympha* available on NCBI GenBank.

The SSU-based phylogenetic tree constructed using the SSU sequences isolated from each transcriptome in this analysis has some alignments and some disagreements with previously-generated phylogenies of Parabasalia. Cepicka et al.'s 2016 tree, which is a synthesis of results from previous phylogenetic analysis that mainly used SSU sequences, positioned the clades Spirotrichonympha and Cristamonadida closer to one another than to Trichonymphida (Cepicka et al. 2016); this overall topology was maintained in the SSU-based phylogenetic tree.

With regards to Trichonymphids, a monophyletic *Trichonympha* was observed, which was in accordance previous research. Additionally, this tree contained a monophyletic clade consisting of *Barbulanympha* and *Urinympha* that was sister to a clade containing *Pseudotrichonympha* (Figure 6), which is in accordance with the Parabasalian trees generated by Carpenter et al. (Carpenter et al. 2011) and Cepicka (Cepicka et al. 2016). The analysis performed by Carpenter et al. also placed *Teranympha* and *Eucomonympha* with *Pseudotrichonympha* (2011); however, for this analysis, transcriptomes for these two genera

could not be obtained due to the limited amount of transcriptomic work that has been done on parabasalids, and so they were not able to be included in the analysis. Follow-up research would ideally include transcriptome generation from these two species as well, so that they could be included in subsequent multi-sequence phylogenetic analyses, where they would be expected to be monophyletic and present within Trichonymphida.

This SSU-based phylogenetic tree also placed Calonympha and Snyderella in a monophyletic clade; this result agrees with Gile et al. 2011, whose SSU-based analysis of calonymphids and other cristamonads obtained a clade consisting of Snyderella, Calonympha, and Stephanonympha (the latter of which did not have a transcriptome useable in the current study) (Gile et al. 2011). However, the tree generated by Gile et al. failed to show monophyly for Calonympha (2011). while the analysis in this thesis does indicates monophyly for both Snyderella and Calonympha (Figure 6). Additionally, the Carpenter et al. analysis placed Spirotrichonympha basally to the clade containing Coronympha, Calonympha, and Snyderella, whereas this SSU-based tree positioned Spirotrichonympha, Holomastigotes, and Holomastigotoides in a clade sister to Calonympha and Snyderella, and placed Coronympha as a sister group to that clade. This spirotrichonymphid clade is observed in the Cepicka 2016 tree; however, Cepicka et al.'s analysis nonetheless does not place Coronympha as a sister group of that clade (Cepicka et al. 2016). All three *Coronympha* sequences formed a monophyletic clade with 100% bootstrap support; however, as Coronympha loses its position from this tree in the multi-sequence phylogenetic tree (Figure 7), this suggests that the true phylogeny likely has the Spirotrichonymphid clade positioned basally to a clade containing the cristamonads (and including Coronympha), and the location of Coronympha in this SSU-based tree (Figure 6) does not necessarily indicate the true position.

Carpenter et al. obtained sequences for their phylogenetic analysis via DNA extraction from manually isolated cells, followed by PCR with SSU-specific primers and sequencing of PCR products (Carpenter et al. 2011). They also used PCR amplification of whole termite gut contents with the same primers (Carpenter et al. 2011). In contrast, this analysis used RNA isolation, and obtained SSU sequences directly from the assembled transcriptome via BLAST search. The high degree of conformity between the SSU-based phylogenetic tree constructed in this analysis and the phylogenetic tree generated by Carpenter et al. indicates that both methods can be suitable to obtain the rDNA sequence from a specimen for subsequent phylogenetic analyses.

4.4. Multi-Gene Phylogeny of Trichonymphida based upon selected genes

The higher number of positions used in the multi-gene phylogeny of Trichonymphida taxa compared to single-gene phylogenies resulted in significantly higher bootstrap support for each node. The lowest bootstrap support for any node in the multi-gene phylogeny was 86% support for the split between *Holomastigotes* and *Holomastigotoides*.

The multi-gene phylogenetic tree generated in this analysis (Figure 7) disagreed with previous phylogenies of Parabasalia in multiple ways. While the Cepicka tree also placed *Coronympha* within Cristamonadida in a clade shared with other cristamonads (including *Foaina, Macrotrichomonoides, Macrotrichomonas, Metadevescovina, Kofoidia,* and *Devescovina,* which were not included in this analysis as no transcriptomes could be obtained for them) (Cepicka et al. 2016), this tree placed *Coronympha* with 100% bootstrap support in a clade with *Tritrichomonas foetus*, which Cepicka et al.'s analysis did not place within the Cristamonadida.

Focussing on the Trichonymphida, the Cepicka tree places the clade containing *Staurojoenina* and *Trichonympha* as the deepest-branching group in the order (Cepicka et al. 2016). The Cepicka tree shows *Trichonympha* as monophyletic, which is corroborated by Ikeda-Ohtsubo and Brune's SSU rRNA tree and Ohkuma et al.'s analysis (Ohkuma et al. 2000, Ikeda-Ohtsubo and Brune 2009). In contrast, the multi-gene tree generated in this study showed a paraphyletic *Trichonympha* rather than a deep-branching monophyletic group. The smallest clade that contained both *Trichonympha* specimens also contained *Urinympha*,

Pseudotrichonympha, *Barbulanympha*, and *Calonympha* specimen 2 (Figure 7). *Trichonympha* specimen 2, which was more complete, was placed as a sister group to *Trichonympha* specimen 1 and the additional clade. This was speculated to be an influence of the fact that *Trichonympha* specimen 2 contained a sequence for every one of the 14 protein sequences used in the multi-sequence analysis, while *Trichonympha* specimen 1, along with *Barbulanympha* sp. 1 and both *Pseudotrichonympha* specimens, lacked identified sequences for the POLR2F sequence. This shared lack of presence could have led to SCaFoS clustering *Trichonympha* specimen 1 more closely with the clade containing *Barbulanympha* sp. 1 and both *Pseudotrichonympha*

As a follow-up analysis to determine which of the gene sequences was most significant in this positioning of *Trichonympha* specimen 2 as a sister group to the clade of *Trichonympha* specimen 1 and *Barbulanympha*, *Pseudotrichonympha*, and *Urinympha*, single-gene phylogenetic trees for each of the involved genes were constructed using FastTree analysis to allow for easy visualization of each gene's contribution to the overall topology. EF1a and beta-tubulin had the greatest contribution to the lack of monophyly observed for *Trichonympha* within the overall multi-sequence tree. Beta-tubulin sequences placed *Trichonympha* sp. 2 very close to

Holomastigotes, and EF1a sequences placed *Trichonympha* specimen in a clade with *Holomastigotes* and *Spirotrichonympha*. These similarities between the beta-tubulin and EF-1a sequences of *Trichonympha* specimen 2 and *Holomastigotes* likely "pulled" *Trichonympha* specimen 2 away from *Trichonympha* specimen 1 and into a position as a sister group to the rest of the Trichonymphida.

The reason for the higher sequence similarity of beta-tubulin and EF-1a sequences of Trichonympha specimen 2 to Holomastigotes and Spirotrichonympha specimens than to Trichonympha specimen 1 was not clarified by this analysis. It is possible that during single-cell picking and RNA isolation, the manual washing of the cell was insufficient to prevent contamination from introduced to the final transcriptome assigned to *Trichonympha* specimen 2. Because the termite hindgut contains multiple different species of parabasalid within a single host, environmental DNA derived from *Holomastigotes* or a species more closely related to it may have been amplified and sequenced and assigned to the transcriptome of Trichonympha specimen 1 under the assumption that it was an RNA sequence from *Trichonympha* specimen 2. When these contaminating sequences were then isolated and used in the phylogenetic analysis, it could have resulted in the above "pulling" of Trichonympha specimen 2 away from its congeneric. This hypothesis could be tested via using BLAST techniques to directly assess the sequence similarity of the beta-tubulin and EF-1a sequences of Trichonympha specimen 2 to those of Trichonympha specimen 1 and to those of the two Holomastigotes specimens; if the beta-tubulin and EF-1a sequences of *Trichonympha* specimen 2 were, in truth, contamination from Holomastigotes or a closely related species, a high sequence similarity would be observed between the sequences of Trichonympha specimen 2 and the Holomastigotes sequences, and a lower sequence similarity would be observed between Trichonympha specimen 2 and

Trichonympha specimen 1. If, however, the beta-tubulin and EF-1a sequences of *Trichonympha* specimen 2 were derived from that cell, they should have a high sequence similarity to those in the congeneric *Trichonympha* specimen 1. Furthermore, a BLAST analysis could be performed against sequences in a public database with the EF-1a and beta-tubulin sequences of *Trichonympha* specimen 2 as the queries, and the taxonomic partitioning of the results could indicate the source of these two sequences.

If the similarity between *Trichonympha* specimen 2 beta-tubulin and EF-1a sequences and those of the *Holomatigotes* sequences is not the result of *Holomastigotes* contamination attributed to *Trichonympha* specimen 2 and was rather the result of *Trichonympha* specimen 2 having significant intraspecific sequence divergence from *Trichonympha* specimen 1 in these genes, the inclusion of additional *Trichonympha* sequences within the tree could potentially have resolved the lack of monophyly displayed by the genus in this analysis. However, the simplest potential treatment of beta-tubulin and EF-1a that could have generated a tree with a monophyletic *Trichonympha* would have been the removal of those sequences from the analysis.

Within the clade containing *Barbulanympha*, *Pseudotrichonympha*, and *Urinympha*, the multi-gene tree generated in this study placed *Barbulanympha* and *Pseudotrichonympha* in a clade with 89% bootstrap support, with *Urinympha* sister to this clade with 100% bootstrap support (Figure 7). This disagrees with the topology of the Cepicka tree, which placed *Barbulanympha* and *Urinympha* as forming a monophyletic clade with *Pseudotrichonympha* branching as the sister group to this clade (Cepicka et al. 2016). The presence of a *Barbulanympha-Urinympha* clade is likely the correct topology, as research from Carpenter *et al.* has also aligned with the Cepicka topology, placing *Pseudotrichonympha* as a sister group to a clade containing *Barbulanympha* and *Urinympha* (Carpenter et al. 2011).

Calonympha sp. 2 and Pseudotrichonympha grassii formed a clade with 99% bootstrap support within *Pseudotrichonympha* during the first analysis, where it was expected that this analysis would recover a monophyletic *Calonympha* (Figure 7). Although it is possible that this result was influenced by the fact that Calonympha sp. 2 and Pseudotrichonympha grassii both had a relatively high number of missing sites and missing genes within the multi-gene analysis, it is more likely that this was the result of contamination within the transcriptomes labelled as Calonympha sp. 2 or Pseudotrichonympha grassii or a misidentification of one of these two taxa during the generation of their transcriptomes, as both transcriptomes were obtained from other researchers and were not generated in the course of this thesis work. An ideal follow-up analysis would remove the *Calonympha* sp. 2 and *Pseudotrichonympha grassii* transcriptomes from the analysis, or would include a greater number of Calonympha and Pseudotrichonympha transcriptomes to provide more points of comparison; however, the limited availability of transcriptomic data from parabasalids means it would likely require generating these transcriptomes as part of the follow-up study. Additional transcriptomes of either Calonympha or *Pseudotrichonympha* would be useful in determining whether this result was caused by contamination or misidentification of the original taxon, as a comparison of additional transcriptomes of good quality with either of the existing transcriptomes would show lower sequence similarity between congenerics if the existing transcriptome contained contamination or had been misidentified before being used in this analysis.

This analysis clearly indicates that sequence selection in phylogenetic analyses of Parabasalia has a critical influence on subsequent tree topology. The most commonly used sequence is the SSU rDNA sequence; however, when multiple gene sequences are used, the tree topology is altered. A phylogenetic analysis using the SSU sequence along with multiple

translated protein sequences for each taxon would provide the most accurate view of relationships within *Trichonymphida*. An analysis of the specific dataset would be required to select the sequences used, as the completeness and coverage of those transcriptomes would influence which sequences were present.

The use of more sequence positions resulted in higher bootstrap support for all the nodes in the tree than is commonly found when SSU rDNA sequences (which are approximately 1500 bp for Trichonymphids) are used alone (Figure 7). Additionally, the higher conservation in protein-coding sequences due to the redundancy of the genetic code, and the possibility for 20 different amino acids per position rather than 4 nucleotides in protein sequences, also likely contributed to the higher bootstrap support. However, this higher bootstrap value did not result in an increased overall tree strength and confidence in the topology, due to the aforementioned concerns about potential contaminant sequences being attributed to *Trichonympha* specimen 1 and the proportionally-lower number of points of comparison between *P. grassii* and *Calonympha* sp. 2 relative to the other specimens included in the analysis. Therefore, this tree should be considered to have relatively low strength despite the high bootstrap values as the causes of its atypical topology have not been determined and accounted for through subsequent analyses.

4.5. "Trichonympha agilis" from a variety of termite species

The *Trichonympha* type species, *T. agilis*, was described in the host termite *Reticulitermes flavipes* (James et al. 2013), and subsequent studies identified *T. agilis* from many different termite hosts on morphological grounds. However, subsequent research using molecular phylogenetic approaches indicate that *T. agilis* has a much less cosmopolitan distribution than

expected, and that *Reticulitermes* termite species other than *R. flavipes* should not be assumed to contain *T. agilis* solely based on symbiont morphology (James et al. 2013, Boscaro et al. 2017). Had the flagellates isolated in this study, *Trichonympha* specimens 1 and 2, been identified to species level solely based on the historical reported distribution of the species, they would have been designated as *T. agilis*, as *T. agilis* is the only *Trichonympha* species yet described from *Reticulitermes hesperus* (Yamin 1979). However, as has been shown by James et al., the historical reported distribution cannot be depended upon for identification of *Trichonympha* species (2013), and the isolated flagellates were isolated from *R. hesperus* rather than from *R. flavipes* (the original type host for *T. agilis*). This reduced the likelihood of the *T. agilis* designation being accurate, since *Trichonympha* species are highly host-specific and are not found in many different termite species. Therefore, both specimens have been designated only as *Trichonympha* sp.

In the phylogenetic analysis, a clear distinction was observed between *Trichonympha* species identified from *R. hesperus* and those identified from *R. flavipes* (Figure 8). This agrees with the hypothesis proposed by James et al. (2013), and indicates that the *Trichonympha* species in *R. hesperus* is not *T. agilis* but is separate enough to merit its own species epithet. These findings confirm the conclusions of James et. al that *Trichonympha* identified on morphological grounds have historically been presumed to be the same species without sufficient evidence (James et al. 2013). *Trichonympha agilis*, as the type species of *Trichonympha*, has had multiple distinct species placed under its name, leading to an overestimation in the host range of and diversity of *T. agilis* and an underestimation of the total diversity of *Trichonympha*. Future research should use phylogenetic methods to re-examine the *Trichonympha* populations in *Reticulitermes* termite species other than *R. flavipes*; previous studies have identified novel

Trichonympha species in already-described termite hindgut communities through these methods (James et al. 2013, Tai et al. 2013). At least one *Trichonympha* species from *R. hesperus* (the species to which the specimens in this study belong) has not been fully described due to a longstanding presumption that it was T. agilis; this species will need a full taxonomic description and deposition of a type specimen, preferentially performed in a similar manner to those in (Boscaro et al. 2017). The two transcriptomes generated in this study may be useful reference sequences in this endeavour, and illustrate that, beyond the SSU rRNA gene sequence, generation of a type transcriptome may be possible. Neither Trichonympha specimen used in this study was suitable for designation as a type specimen due to a small degree of uncertainty in the identity of the host termite species; while the coastal location and habitat of collection, as well as the gross morphology, suggest a *R. hesperus* host, a minor possibility still exists that their hosts were range-edge *R. okanagenensis*, as the range of the latter overlaps with the former (Szalanski et al. 2006). These host termites would require DNA barcoding to distinguish with absolute confidence, due to the extreme morphological similarity between the two species and the overlap between the ranges of *R. okanagenensis* and *R. hesperus*; this was not possible in the current study due to time constraints but is a prerequisite for description of the *R*. hesperus Trichonympha as a novel species.

Furthermore, the COX-2 phylogenetic tree created for the host termite species for these *Trichonympha* species agreed with Austin et al.'s SSU-based haplotype analysis illustrating that *R. santonensis* is not phylogenetically distinct from *R. flavipes* and that both species should be synonymized under the older epithet of *R. flavipes* (2005). This phylogeny adds more support to the synonymy of the populations previously designated *R. santonensis* and *R. flavipes*; in fact, Austin et al. proposed that data from the COX-2 mitochondrial gene would likely corroborate

their conclusions in reference to a then-unpublished study (Austin et al. 2005). However, during the course of constructing these trees, it became apparent that many sequences from both termite and endosymbionts have been deposited into GenBank with *R. santonensis* given as the organism or as the host descriptor even post-2005. This shows that the epithet *R. santonensis* remains in common use despite the synonymization, potentially confounding estimates of the diversity and distribution of *Trichonympha* species within *Reticulitermes* termite hosts. As the benefits of biological sequence database information in genomic, phylogenetic, and other types of bioinformatic research are very significant, unreliability of results deposited in public databases represents a major loss in possibilities for future research and can cause significant problems if the information connected with deposited sequence data is erroneous (Korning et al. 1996, Hofstetter et al. 2019), including if the associated error has been compounded from previous depositions. Optimally, publically-available data associated with the epithet *R. santonensis* should be edited to refer to *R. flavipes* instead, which would prevent future research from misleadingly using the outdated epithet.
5. Conclusions

Phylogenetic analyses in Parabasalia are heavily dependent on the selection of the gene or proteins sequences being used. Although the SSU rDNA sequence is the most commonly-used due to its reliable rate of evolution and ease of isolation, which contributes to its frequency in existing databases, it is less useful for deep phylogenetic analyses due to the limited number of positions available. The use of multiple sequences increases the number of usable positions for each taxon, and therefore the bootstrap support in the final tree for each node and clade. However, not all sequences are beneficial for ultimate tree construction, or produce trees that agree with the overall consensus. The phylogeny generated in this study broadly agrees with previous phylogenies of parabasalia, placing *Trichonympha* basal within Trichonymphida compared to *Pseudotrichonympha*, *Barbulanympha*, and *Urinympha*, and obtaining clades representing Cristamonadida and Spirotrichonympha; however, it failed to obtain a monophyletic *Trichonympha*, which is most likely the true state.

Additionally, this study provides phylogenetic support for the conclusion that species of *Trichonympha* are highly host-specific rather than being present in multiple termite host species, and that multiple phylogenetic species of *Trichonympha* have been misidentified as *T. agilis* despite not being present in the *T. agilis* type host and being phylogenetically distinct. This includes the *Trichonympha* species present in *R. hesperus*, which was sequenced successfully using Picelli et al.'s 2014 protocol for single-cell transcriptome sequencing. This study also confirms Austin et al.'s 2005 conclusion that there is no phylogenetic distinction between the termite populations designated as *Reticulitermes flavipes* and *R. santonensis*, despite the latter's persistent use as an epithet.

Based on these results, the most immediate topic I would suggest for future research is a full description and taxonomic designation of the *Trichonympha* species present in *R. hesperus*, including both morphology-based and sequence-based identification of the type host. The transcriptomes generated in this study may be useful points of comparison for this project, although they are not suitable to be type sequences themselves.

References

- Abe, T., Bignell, D.E. & Higashi, M. (Eds.) 2000. Termites: Evolution, Sociality, Symbioses, Ecology. Springer Netherlands, Dordrecht.
- Ahmed, S. 2022. DNA Barcoding in Plants and Animals: A Critical Review.
- Andersson, J.O., Sarchfield, S.W. & Roger, A.J. 2005. Gene Transfers from Nanoarchaeota to an Ancestor of Diplomonads and Parabasalids. *Molecular Biology and Evolution*. 22:85–90.
- Austin, J.W., Szalanski, A.L., Scheffrahn, R.H., Messenger, M.T., Dronnet, S. & Bagnères, A.-G. 2005. Genetic Evidence for the Synonymy of Two Reticulitermes Species: Reticulitermes flavipes and Reticulitermes santonensis. *Annals of the Entomological Society of America*. 98:395–401.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M. et al. 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol.* 19:455–77.
- Baudouin, G., Bech, N., Bagnères, A.-G. & Dedeine, F. 2018. Spatial and genetic distribution of a north American termite, Reticulitermes flavipes, across the landscape of Paris. *Urban Ecosyst.* 21:751–64.
- Beccaloni, G. & Eggleton, P. 2013. Order Blattodea. In: Zhang, Z.-Q. (Ed.) Animal Biodiversity: An Outline of Higher-level Classification and Survey of Taxonomic Richness (Addenda 2013). Zootaxa. 3703:46.
- Biagini, G.A., Finlay, B.J. & Lloyd, D. 2006. Evolution of the hydrogenosome. FEMS Microbiology Letters. 155:133–40.
- Bobbett, B. 2020. Quantifying Intragenomic Variability in the 18S Gene of Trichonympha from Zootermopsis.
- Bolger, A.M., Lohse, M. & Usadel, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114–20.
- Boscaro, V., James, E.R., Fiorito, R., Hehenberger, E., Karnkowska, A., del Campo, J., Kolisko, M. et al. 2017. Molecular characterization and phylogeny of four new species of the genus Trichonympha (Parabasalia, Trichonymphea) from lower termite hindguts. *International Journal of Systematic and Evolutionary Microbiology*. 67:3570–5.
- Buchfink, B., Xie, C. & Huson, D.H. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods*. 12:59–60.
- Bui, E.T., Bradley, P.J. & Johnson, P.J. 1996. A common evolutionary origin for mitochondria and hydrogenosomes. *PNAS*. 93:9651–6.

- Byrne, S.J., Butler, C.A., Reynolds, E.C. & Dashper, S.G. 2018. Chapter 7 Taxonomy of Oral Bacteria. *In* Gurtler, V. & Trevors, J. T. [Eds.] *Methods in Microbiology*. Academic Press, pp. 171–201.
- Capella-Gutiérrez, S., Silla-Martínez, J.M. & Gabaldón, T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 25:1972–3.
- Carpenter, K.J., Chow, L. & Keeling, P.J. 2009. Morphology, Phylogeny, and Diversity of Trichonympha (Parabasalia: Hypermastigida) of the Wood-Feeding Cockroach Cryptocercus punctulatus. *Journal of Eukaryotic Microbiology*. 56:305–13.
- Carpenter, K.J., Horak, A., Chow, L. & Keeling, P.J. 2011. Symbiosis, morphology, and phylogeny of Hoplonymphidae (Parabasalia) of the wood-feeding roach Cryptocercus punctulatus. *J Eukaryot Microbiol*. 58:426–36.
- Carpenter, K.J., Horak, A. & Keeling, P.J. 2010. Phylogenetic Position and Morphology of Spirotrichosomidae (Parabasalia): New Evidence from Leptospironympha of Cryptocercus punctulatus. *Protist.* 161:122–32.
- Cavalier-Smith, T.Y. 2003 2003. The excavate protozoan phyla Metamonada Grassé emend. (Anaeromonadea, Parabasalia, Carpediemonas, Eopharyngia) and Loukozoa emend. (Jakobea, Malawimonas): their evolutionary affinities and new higher taxa. *International Journal of Systematic and Evolutionary Microbiology*. 53:1741–58.
- Cepicka, I., Dolan, M. & Gile, G. 2016. Parabasalia. *In Handbook of the Protists*. Springer International Publishing, pp. 1–44.
- Cepicka, I., Hampl, V. & Kulda, J. 2010. Critical Taxonomic Revision of Parabasalids with Description of one New Genus and three New Species. *Protist.* 161:400–33.
- Cleveland, L.R. & Grimstone, A.V. 1964. The Fine Structure of the Flagellate Mixotricha paradoxa and Its Associated Micro-Organisms. *Proceedings of the Royal Society of London. Series B, Biological Sciences*. 159:668–86.
- Criscuolo, A. & Gribaldo, S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*. 10:210.
- Dacks, J.B. & Doolittle, W.F. 2001. Reconstructing/Deconstructing the Earliest Eukaryotes: How Comparative Genomics Can Help. *Cell*. 107:419–25.
- Dueholm, M.S., Karst, S.M., McIlroy, S.J., Kirkegaard, R.H., Nielsen, P.H. & Albertsen, M. 2017. High throughput sequencing of full-length SSU rRNA sequences from complex microbial communities without primer bias and how it affects our ability to study microbial ecology.

Feytaud, J. de 1966. Le peuple des termites.

- Gerbod, D., Sanders, E., Moriya, S., Noël, C., Takasu, H., Fast, N.M., Delgado-Viscogliosi, P. et al. 2004. Molecular phylogenies of Parabasalia inferred from four protein genes and comparison with rRNA trees. *Mol Phylogenet Evol*. 31:572–80.
- Ghesini, S., Pilon, N. & Marini, M. 2011. A new finding of Reticulitermes flavipes in northern Italy. *Bulletin of Insectology*. 64.
- Gibbons, I.R. & Grimstone, A.V. 1960. On Flagellar Structure in Certain Flagellates. *The Journal of Biophysical and Biochemical Cytology*. 7:697.
- Gile, G.H., James, E.R., Scheffrahn, R.H., Carpenter, K.J., Harper, J.T. & Keeling, P.J. 2011. Molecular and morphological analysis of the family Calonymphidae with a description of Calonympha chia sp. nov., Snyderella kirbyi sp. nov., Snyderella swezyae sp. nov. and Snyderella yamini sp. nov. *International Journal of Systematic and Evolutionary Microbiology*. 61:2547–58.
- Guichard, P. & Gönczy, P. 2016. Basal body structure in Trichonympha. Cilia. 5:9.
- Hauck, R. & Hafez, H.M. 2010. Systematic Position of Histomonas meleagridis Based on Four Protein Genes. *Journal of Parasitology*. 96:396–400.
- Hofstetter, V., Buyck, B., Eyssartier, G., Schnee, S. & Gindro, K. 2019. The unbearable lightness of sequenced-based identification. *Fungal Diversity*.
- Hongoh, Y., Sato, T., Dolan, M.F., Noda, S., Ui, S., Kudo, T. & Ohkuma, M. 2007. The Motility Symbiont of the Termite Gut Flagellate Caduceia versatilis Is a Member of the "Synergistes" Group. Applied and Environmental Microbiology. 73:6270–6.
- Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Toh, H., Taylor, T.D., Kudo, T. et al. 2008. Genome of an endosymbiont coupling N2 fixation to cellulolysis within protist cells in termite gut. *Science*. 322:1108–9.
- Ikeda-Ohtsubo, W. & Brune, A. 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: Trichonympha species and 'Candidatus Endomicrobium trichonymphae.' *Molecular Ecology*. 18:332–42.
- Inoue, T., Kitade, O., Yoshimura, T. & Yamaoka, I. 2000. Symbiotic Associations with Protists. *Termites: Evolution, Sociality, Symbioses, Ecology*. 275–88.
- Inward, D., Beccaloni, G. & Eggleton, P. 2007. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol Lett.* 3:331–5.
- James, E.R., Tai, V., Scheffrahn, R.H. & Keeling, P.J. 2013. Trichonympha burlesquei n. sp. from Reticulitermes virginicus and evidence against a cosmopolitan distribution of Trichonympha agilis in many termite hosts. *International Journal of Systematic and Evolutionary Microbiology*. 63:3873–6.

- Katoh, K. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*. 33:511–8.
- Keeling, P.J. & Palmer, J.D. 2000. Parabasalian flagellates are ancient eukaryotes. *Nature*. 405:635–7.
- King, S.D. 1927. The Golgi Apparatus of Protozoa. *Journal of the Royal Microscopical Society*. 47:342–55.
- Kirby, H. 1931. The Structure and Reproduction of the Parabasal Body in Trichomonad Flagellates. *Transactions of the American Microscopical Society*. 50:189–95.
- Korning, P.G., Hebsgaard, S.M., Rouzé, P. & Brunak, S. 1996. Cleaning the GenBank Arabidopsis Thaliana Data Set. *Nucleic Acids Research*. 24:316–20.
- Laetsch, D.R. & Blaxter, M.L. 2017. BlobTools: Interrogation of genome assemblies. F1000Research.
- Leger, M.M., Kolisko, M., Kamikawa, R., Stairs, C.W., Kume, K., Čepička, I., Silberman, J.D. et al. 2017. Organelles that illuminate the origins of Trichomonas hydrogenosomes and Giardia mitosomes. *Nat Ecol Evol.* 1:0092.
- Leidy, J. 1881. Parasites of the Termites. Collins, Printer. 112 pp.
- Li, H. & Durbin, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25:1754–60.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25:2078–9.
- Malik, B., Brochu, C., Bilic, I., Yuan, J., Hess, M., Logsdon, J. & Carlton, J. 2011. Phylogeny of Parasitic Parabasalia and Free-Living Relatives Inferred from Conventional Markers vs. Rpb1, a Single-Copy Gene. *PloS one*. 6:e20774.
- McKern, J., Szalanski, A., Austin, J., Messenger, M., Mahn, J. & Gold, R. 2007. Phylogeography of Termites (Isoptera) from Oregon and Washington. *Sociobiology*. 50:607.
- Müer, M. 1993 1993. Review Article: The hydrogenosome. *Microbiology*. 139:2879-89.
- Nalepa, C.A., Bignell, D.E. & Bandi, C. 2001. Detritivory, coprophagy, and the evolution of digestive mutualisms in Dictyoptera: *Insectes soc.* 48:194–201.
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A. & Minh, B.Q. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*. 32:268–74.

- Nikolić, V. 2021. Scalable methods for improving genome assemblies. University of British Columbia.
- Nishimura, Y., Otagiri, M., Yuki, M., Shimizu, M., Inoue, J., Moriya, S. & Ohkuma, M. 2020. Division of functional roles for termite gut protists revealed by single-cell transcriptomes. *ISME J.* 14:2449–60.
- Noda, S., Hongoh, Y., Sato, T. & Ohkuma, M. 2009a. Complex coevolutionary history of symbiotic Bacteroidales bacteria of various protists in the gut of termites. *BMC Evol Biol*. 9:158.
- Noda, S., Inoue, T., Hongoh, Y., Kawai, M., Nalepa, C.A., Vongkaluang, C., Kudo, T. et al. 2006. Identification and characterization of ectosymbionts of distinct lineages in Bacteroidales attached to flagellated protists in the gut of termites and a wood-feeding cockroach. *Environ Microbiol.* 8:11–20.
- Noda, S., Kitade, O., Inoue, T., Kawai, M., Kanuka, M., Hiroshima, K., Hongoh, Y. et al. 2007. Cospeciation in the triplex symbiosis of termite gut protists (Pseudotrichonympha spp.), their hosts, and their bacterial endosymbionts. *Mol Ecol.* 16:1257–66.
- Noda, S., Mantini, C., Bordereau, C., Kitade, O., Dolan, M.F., Viscogliosi, E. & Ohkuma, M. 2009b. Molecular phylogeny of parabasalids with emphasis on the order Cristamonadida and its complex morphological evolution. *Molecular Phylogenetics and Evolution*. 52:217–24.
- Noda, S., Mantini, C., Meloni, D., Inoue, J.-I., Kitade, O., Viscogliosi, E. & Ohkuma, M. 2012. Molecular Phylogeny and Evolution of Parabasalia with Improved Taxon Sampling and New Protein Markers of Actin and Elongation Factor-1α. *PLoS One*. 7.
- Noda, S., Shimizu, D., Yuki, M., Kitade, O. & Ohkuma, M. 2018. Host-Symbiont Cospeciation of Termite-Gut Cellulolytic Protists of the Genera Teranympha and Eucomonympha and their Treponema Endosymbionts. *Microbes Environ*. 33:26–33.
- Ohkuma, M. 2003. Termite symbiotic systems: efficient bio-recycling of lignocellulose. *Appl Microbiol Biotechnol*. 61:1–9.
- Ohkuma, M. 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends in Microbiology*. 16:345–52.
- Ohkuma, M. & Brune, A. 2011. Diversity, Structure, and Evolution of the Termite Gut Microbial Community. In Bignell, D. E., Roisin, Y. & Lo, N. [Eds.] Biology of Termites: A Modern Synthesis. Springer Netherlands, Dordrecht, pp. 413–38.
- Ohkuma, M., Iida, T., Ohtoko, K., Yuzawa, H., Noda, S., Viscogliosi, E. & Kudo, T. 2005. Molecular phylogeny of parabasalids inferred from small subunit rRNA sequences, with emphasis on the Hypermastigea. *Molecular Phylogenetics and Evolution*. 35:646–55.

- Ohkuma, M., Noda, S., Hongoh, Y., Nalepa, C.A. & Inoue, T. 2009. Inheritance and diversification of symbiotic trichonymphid flagellates from a common ancestor of termites and the cockroach Cryptocercus. *Proc Biol Sci.* 276:239–45.
- Ohkuma, M., Ohtoko, K., Iida, T., Tokura, M., Moriya, S., Usami, R., Horikoshi, K. et al. 2000. Phylogenetic Identification of Hypermastigotes, Pseudotrichonympha, Spirotrichonympha, Holomastigotoides, and Parabasalian Symbionts in the Hindgut of Termites. *Journal of Eukaryotic Microbiology*. 47:249–59.
- Ohkuma, M., Saita, K., Inoue, T. & Kudo, T. 2007. Comparison of four protein phylogeny of parabasalian symbionts in termite guts. *Mol Phylogenet Evol.* 42:847–53.
- Parfrey, L.W., Lahr, D.J.G. & Katz, L.A. 2008. The Dynamic Nature of Eukaryotic Genomes. *Molecular Biology and Evolution*. 25:787–94.
- Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S. & Sandberg, R. 2014. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc*. 9:171–81.
- Roure, B., Rodriguez-Ezpeleta, N. & Philippe, H. 2007. SCaFoS: a tool for Selection, Concatenation and Fusion of Sequences for phylogenomics. *BMC Evolutionary Biology*. 7:S2.
- Saldarriaga, J.F., Gile, G.H., James, E.R., Horák, A., Scheffrahn, R.H. & Keeling, P.J. 2011. Morphology and molecular phylogeny of Pseudotrichonympha hertwigi and Pseudotrichonympha paulistana (Trichonymphea, Parabasalia) from Neotropical Rhinotermitids. *Journal of Protozoology*. 58:487–96.
- Sanderlin, V. 2019. Glycoside Hydrolase Gene Families Of Termite Hindgut Protists.
- Shiflett, A.M. & Johnson, P.J. 2010. Mitochondrion-Related Organelles in Eukaryotic Protists. *Annual Review of Microbiology*. 64:409–29.
- Smith, J.L. & Rust, M.K. 1994. Temperature preferences of the western subterranean termite, Reticulitermes hesperus Banks. *Journal of Arid Environments*. 28:313–23.
- Soviš, M. n.d. Charles University Faculty of Science. 45.
- Szalanski, A.L., Austin, J.W., Mckern, J. & Messenger, M.T. 2006. Genetic Evidence for a New Subterranean Termite Species (Isoptera: Rhinotermitidae) from Western United States and Canada. *The Florida Entomologist*. 89:299–304.
- Taerum, S.J., De Martini, F., Liebig, J. & Gile, G.H. 2018. Incomplete Co-cladogenesis Between Zootermopsis Termites and Their Associated Protists. *Environmental Entomology*. 47:184–95.
- Tai, V., James, E.R., Perlman, S.J. & Keeling, P.J. 2013. Single-Cell DNA Barcoding Using Sequences from the Small Subunit rRNA and Internal Transcribed Spacer Region

Identifies New Species of Trichonympha and Trichomitopsis from the Hindgut of the Termite Zootermopsis angusticollis. *PLOS ONE*. 8:e58728.

- Tamm, S.L. 1982. Flagellated ectosymbiotic bacteria propel a eucaryotic cell. *J Cell Biol*. 94:697–709.
- Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution*. 30:2725– 9.
- Vandamme, A.-M. 2009. Basic concepts of molecular evolution. *In* Lemey, P., Salemi, M. & Vandamme, A.-M. [Eds.] *The Phylogenetic Handbook*. 2nd ed. Cambridge University Press, Cambridge, pp. 3–30.
- Viscogliosi, E. & Müller, M. 1998. Phylogenetic Relationships of the Glycolytic Enzyme, Glyceraldehyde-3-Phosphate Dehydrogenase, from Parabasalid Flagellates. *J Mol Evol*. 47:190–9.
- Wang, C., Zhang, T., Wang, Y., Katz, L.A., Gao, F. & Song, W. 2017. Disentangling sources of variation in SSU rDNA sequences from single cell analyses of ciliates: impact of copy number variation and experimental error. *Proc. R. Soc. B.* 284:20170425.
- Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V. et al. 2018. BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol Biol Evol.* 35:543–8.
- YAMIN, M. 1979. Flagellates Of The Orders Trichomonadida Kirby, Oxymonadida Grasse, And Hypermastigida Grassi And Foa Reported From Lower Termites (Isoptera Families Mastotermitidae, Kalotermitidae, Hodotermitidae, Termopsidae, Rhinotermitidae, And Serritermitidae) And From The Wood-Feeding Roach Cryptocercus (Dictyoptera: Cryptocercidae). Flagellates Of The Orders Trichomonadida Kirby, Oxymonadida Grasse, And Hypermastigida Grassi And Foa Reported From Lower Termites (Isoptera Families Mastotermitidae, Kalotermitidae, Hodotermitidae, Termopsidae, Rhinotermitidae, And Serritermitidae) And From The Wood-Feeding Roach Cryptocercus (Dictyoptera: Cryptocercidae).
- Ye, W., Lee, C.-Y., Scheffrahn, R.H., Aleong, J.M., Su, N.-Y., Bennett, G.W. & Scharf, M.E. 2004. Phylogenetic relationships of nearctic Reticulitermes species (Isoptera: Rhinotermitidae) with particular reference to Reticulitermes arenincola Goellner. *Molecular Phylogenetics and Evolution*. 30:815–22.
- N.d. Babraham Bioinformatics FastQC A Quality Control tool for High Throughput Sequence Data. Available At: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (last accessed October 21, 2021a).
- N.d. BUSCO.v4 datasets. Available At: https://busco.ezlab.org/busco_v4_data.html (last accessed October 21, 2021b).

- N.d. UniProt. Available At: https://www.uniprot.org/ (last accessed September 24, 2021c).
- N.d. Silva. Available At: https://www.arb-silva.de/ (last accessed September 24, 2021d).
- N.d. FigTree. Available At: http://tree.bio.ed.ac.uk/software/figtree/ (last accessed March 16, 2022e).
- N.d. XQuartz. Available At: https://www.xquartz.org/ (last accessed March 16, 2022f).