

Intraspecies variation in *Tetrahymena rostrata*

Watt, A.E.¹, Young, N.D.², Haites, R.E.¹, Dunse, K.², Russell, D.A.¹ and Billman-Jacobe, H.¹

¹ Asia-Pacific Centre for Animal Health, Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, University of Melbourne

² Melbourne Veterinary School, Faculty of Veterinary and Agricultural Sciences, University of Melbourne

* Correspondence: H. B-J hbj@unimelb.edu.au

Abstract

Two distinct isolates of the facultative parasite, *Tetrahymena rostrata* were compared, identifying and utilising markers that are useful for studying clonal variation within the species. The mitochondrial genomes and several nuclear genes were determined using Illumina short read sequencing. The two *T. rostrata* isolates had similar morphology and were pathogenic for *Deroceras reticulatum* slugs. The linear mitogenomes had the gene content and organisation typical of the *Tetrahymena* genus, comprising 8 tRNA genes, 6 ribosomal RNA genes and 45 protein coding sequences (CDS), twenty-two of which had known function. The two isolates had nucleotide identity within common nuclear markers encoded within the histone H3 and H4 and small subunit ribosomal RNA genes and differed by only 2-4 nucleotides in a region of the characterised actin genes. Variation was observed in several mitochondrial genes and was used to determine intraspecies variation and may reflect the natural history of *T. rostrata* from different hosts or the geographic origins of the isolates.

Keywords: Ciliophora; mtSSUrRNA; Histone; cytochrome oxidase 1, origin of replication

1. Introduction

Tetrahymena rostrata (Kahl 1926, Corliss) are protists which can be free-living in edaphic environments [1-4] and can occur as facultative parasites of terrestrial molluscs [2,5-7]. Most reports of natural infections of slugs and snails have shown that *T. rostrata* favours the renal tissues where it can multiply up to large numbers [7,8]. Histological examination of wild slugs has shown that the albumen gland and genital tract organs can become heavily infected with *T. rostrata* which can result in trans-ovarial transmission of the parasite [8]. There are few isolates of *T. rostrata* available for genetic comparison. Characterisation of isolates from *Helix aspera* snails and *Deroceras reticulatum* slugs in Spain included morphometric measurements and molecular data derived from the mitochondrial cytochrome oxidase 1 (*cox1*) gene and the nuclear small subunit ribosomal RNA gene (SSUrRNA) [9]. A further four isolates from *Zonitoides nitidus* and *Cochlicopa lubrica* snails from Poland were compared to the Spanish isolates using partial *cox1* and SSUrRNA sequences. These comparisons showed the Spanish and Polish isolates were identical in the SSUrRNA but clustered as two subgroups based on *cox1* [5]. The first group, consisting of isolates TR01, TR02, TR03, TR1034 and TR1035, had 0.9-1.3% sequence divergence. The second group, consisting of TR1015 and TR1016 had 0.6% divergence between them but a 4.4-5.2% sequence divergence from TR01, the type strain, which is greater than expected for intraspecies variation. The guideline for declaring that *Tetrahymena* strains belong to different species is 4-5% divergence in the *cox1* gene sequences [10,11]. Kaczanowski *et al* suggested that the natural environment of *T. rostrata* is variable and that there may be a corresponding clonal diversity within the species [5].

The Grey field slug, *D. reticulatum* is an invasive pest in Australia and has been present since at least the 1930s, probably longer [12]. *T. rostrata* was recently identified in

Australia when it was isolated from the egg of a *D. reticulatum* slug and has been established in axenic culture [13]. The isolate, designated TRAUS, had the same morphological and phenotypic characteristics of the Spanish and Polish isolates. We recently reported the mitochondrial genome of TRAUS from Illumina libraries of whole cell DNA extracts [13]. While the TRAUS isolate grouped with Polish isolates TR1015 and TR1016 on the basis of the *cox1* sequence data, the lack of whole mitochondrial genome sequence data for other *T. rostrata* isolates made detailed comparisons of taxonomic relationships challenging.

Phylogenetic studies of ciliates are increasingly being performed using additional nuclear and mitochondrial molecular markers, such as the mitochondrial SSUrRNA (mtSSUrRNA) and nuclear ITS1-5.8S-ITS2 [11,14]. In this paper, we report the mitochondrial genome and selected nuclear genes of the type strain TR01, an isolate from Spain, and make a comparison with our Australian isolate (TRAUS) to gain insights into variation within the *T. rostrata* species. We chose these isolates because their collection sites, host species and sites of infection in the host were different.

2. Materials and Methods

2.1 Strains and culture

T. rostrata TR01 was obtained from the American Type Culture Collection (ATCC®PRA326™). *T. rostrata* TRAUS was isolated from the egg of a laboratory-reared *D. reticulatum* in 2015. The parents of the slug had been previously collected from Ringwood North, SE of Melbourne, Australia. The isolates were routinely cultured at 20°C in sterile PPYE medium consisting of 0.5% w/v proteose peptone (Oxoid LP0085), 0.5% w/v yeast extract (Oxoid LP0021), and 0.125% w/v glucose. The source of isolates and relevant sequences are listed in Table 1.

Table 1. *Tetrahymena rostrata* isolates and sequences

Strain ID	Host	Tissue	Accession numbers for each isolate	Ref
TR01	<i>Helix aspersa aspersa</i>	renal organ	GU439231 (<i>cox1</i>), JQ045342 (18S rRNA), MT375014 (mtDNA), MT506240 (histone H3 H4), MT420428 (18S-5.8S-28S rRNA), SRR12315381(short read archive)	[9] This study
TR02	<i>Helix aspersa maxima</i>	renal organ	GU439232 (<i>cox1</i>)	[9]
TR03	<i>Deroceras reticulatum</i>	renal organ	GU439233 (<i>cox1</i>)	[9]
TR1015	<i>Zonitoides nitidus</i>	renal organ	KR778771 (<i>cox1</i>); KR778775 (18S rRNA)	[5]
TR1016	<i>Zonitoides nitidus</i>	renal organ	KR778772 (<i>cox1</i>), KR778776 (18S rRNA)	[5]
TR1034	<i>Cochlicopa lubrica</i>	renal organ	KR778773 (<i>cox1</i>), KR778777 (18S rRNA)	[5]
TR1035	<i>Cochlicopa lubrica</i>	renal organ	KR778774 (<i>cox1</i>), KR778778 (18S rRNA)	[5]
TRAUS	<i>Deroceras reticulatum</i>	egg	MN025427 (mtDNA), MN167836 (histone H3 H4), MN158348 (18S, 5.8S, 28S rRNA), SRR12315411 (short read archive)	[12] and this study

2.2 Morphological and phenotypic characterisation

T. rostrata TRAUS cells from a densely growing PPYE culture (cell density: $0.7-1 \times 10^5$ cell/ml), were subcultured in PP (1% w/v proteose peptone (Oxoid LP0085) and 0.125% w/v glucose) for 7 days and were harvested (800 x g 5-10 min) and washed twice in 10mM HEPES pH7 NaOH. Cells were resuspended in an aqueous infusion of composted pine bark/ 10mM HEPES pH7 NaOH and then dispensed into 6 well tissue culture plates. Plates were incubated at 26°C for 24 hours to allow cysts to form. Light microscopy was used to confirm that cells with rounded morphologies were cysts. Excystment was stimulated by transferring the trays to 20°C and the released theronts were examined. Wet mounts and fixed Giemsa stained cells were imaged with a Leica DMLS light microscope. Cell measurements were made using ImageJ using a calibrated scale [15].

For scanning electron microscopy, trophonts were collected by centrifugation, washed and then fixed with 2.5% glutaraldehyde and applied to polyethyleneimine coated glass coverslips. Cells were dehydrated in increasing concentrations of ethanol and the coverslips were dried in a Balzers CPD030 critical point dryer (Balzers, Liechtenstein, Germany) and mounted onto 25 mm aluminium stubs with double-sided carbon tabs. The coverslips were coated with gold using a Xenosput sputter coater (Dynavac, Wantirna South, Australia). The cells were imaged with the Philips XL30 field-emission scanning electron microscope (Philips, Eindhoven, Netherlands) at a voltage of 2.0 kV and a spot size of 2. Lineart was made in Adobe Photoshop.

T. rostrata TR01 and TRAUS trophonts were compared in a bioassay to determine their pathogenicity against *D. reticulatum*. Both strains were maintained at 18°C in ATCC 357 medium without antibiotics and were subcultured every 3 to 5 weeks. Cultures used for infection experiments were subcultured and grown at 25°C for 24 hr. Slugs for experiments were offspring of laboratory reared slugs [16].

Adult slugs used for experiments were removed from the rearing boxes and starved in clean boxes lined with moist paper and then placed in groups of 3 in Petri dishes coated with water agar. Each treatment comprised of 15 groups of three slugs and the treatments were *T. rostrata* TR01 or TRAUS (5×10^3 cells in 1 ml water) or water only. The slugs were moved to new Petri dishes lined with moist filter paper after a week and the dishes were cleaned, food added and mortality recorded weekly for the duration of experiment (42 days).

In a separate experiment, *D. reticulatum* neonates were obtained from eggs laid on clean, moist paper and Stage V eggs [16,17] were randomly assigned to groups of 10 eggs and incubated for another 12 hours to hatch. Hatchlings were transferred to fresh dishes. *T. rostrata* TR01 or TRAUS (5×10^3 cells in 1 ml water) were added to six groups each and four groups were treated with water only. All groups were fed with lettuce. Mortality was recorded every 2-3 days for 15 days.

2.3 Genome sequencing

Genomic DNA was extracted from axenic cultures of TR01 (ATCC PRA326) and TRAUS using a Promega genomic DNA kit. The DNA was fragmented by sonication and fragments in the suitable size range were purified and end-repaired and A-tailed using the polymerase activity of Klenow fragment. Indexed adapters were ligated to the DNA fragments by DNA ligase followed by performing a PCR reaction of 15 cycles to enrich the adapter-modified DNA fragments using KAPA HiFi HotStart ReadyMix (KK2602 Kapa Biosystems). After validating the libraries by TapeStation, each library was sequenced using Illumina MiSeq. The reads were mapped to the mitochondrial genome of the closest *cox1* relative for which there was sequence in Genbank, *T. pigmentosa* (DQ927305) and then the mapped reads were self-assembled and used as scaffolds for rounds of gap filling and extension using the total read libraries until the whole mitochondrial genome was assembled as a single contig. TRAUS and TR01 mitochondrial DNA sequences are available in GenBank (MN025427 and MT375014 respectively). Illumina read data is available from the GenBank Sequence Read Archive for TR01 SRR12315381

and TRAU SRR12315411. The contigs were annotated with reference to the mitogenomes of *T. thermophila* (AF396436) [18], *T. pyriformis* (AF160864) [19], *T. paravorax* (DQ927304), *T. malaccensis* (DQ927303) and *T. pigmentosa* (DQ927305) [20].

2.4 Phylogenetic analysis

The *cox1* sequences for *T. rostrata* strains were obtained from Genbank and a 1796 nt *cox1* barcode region was selected based on previous studies [21] [10]. The mitochondrial SSUrRNA barcode region selected was the 541 bp identified by Doerder (2019) [11]. Alignments of *cox1*, and mtSSUrRNA were performed using MAFFT version 7.388 (22) and Bayesian phylogenetic inference was performed using a Markov chain Monte Carlo (MCMC) analysis in MrBayes version 3.2.6 (<https://github.com/NBISweden/MrBayes>) using a 1,100,000 MCMC generation chain length with consensus trees generated using the 50% majority rule criterion and the final 90% of trees generated by Bayesian inference after a burn-in of 100,000 generations. Estimates of evolutionary divergence between the 1796 nt *cox1* and 689 nt *cox1* barcode [21] were conducted using the Kimura 2-parameter model [22]. All ambiguous positions were removed for each sequence pair (pairwise deletion option). Evolutionary analyses were conducted in MEGA X [23] using the settings for protozoal mitochondria codon usage. Each functional pair of coding DNA sequences was extracted from the assembled TR01 and TRAU mtDNA contigs and aligned in Geneious using the Translate align function with the mold-protzoan mitochondrial genetic code table and Blosum 62 cost matrix [24]. Single stranded DNA topology was determined using Mfold on the UNAFold server using default parameters except the temperature, which was set at 25°C [25].

3. Results and Discussion

3.1 Morphological and phenotypic examination of isolate TRAU

The *Tetrahymena* isolate, designated TRAU, showed the typical morphology of *T. rostrata* and was capable for forming cysts and releasing theronts.

Trophonts were ovoid and may have a rostrate anterior end (Fig. 1A). Fresh, unfixed trophonts grown in PPYE for 7 days were an average size of $56.45 \pm 7.91 \times 42.08 \pm 6.69 \mu\text{m}$ (n=78). (Supp Material 1). The oval buccal opening was situated in the top quarter of the anterior end and was lined with ciliary membranelles (Fig 1C). Trophonts contained a macronucleus and associated micronucleus (Fig 1D). Trophonts developed into cysts under nutrient deprivation at 26°C and the cysts released theronts with characteristic lobed macronuclear analgen either side of a micronucleus (Figure 1 D-F). The trophonts had ~28 kineties and the oral opening was ~11 x 9 microns. These data show that TRAU has the same morphological and life history characteristics as *T. rostrata* described by other authors [8,9].

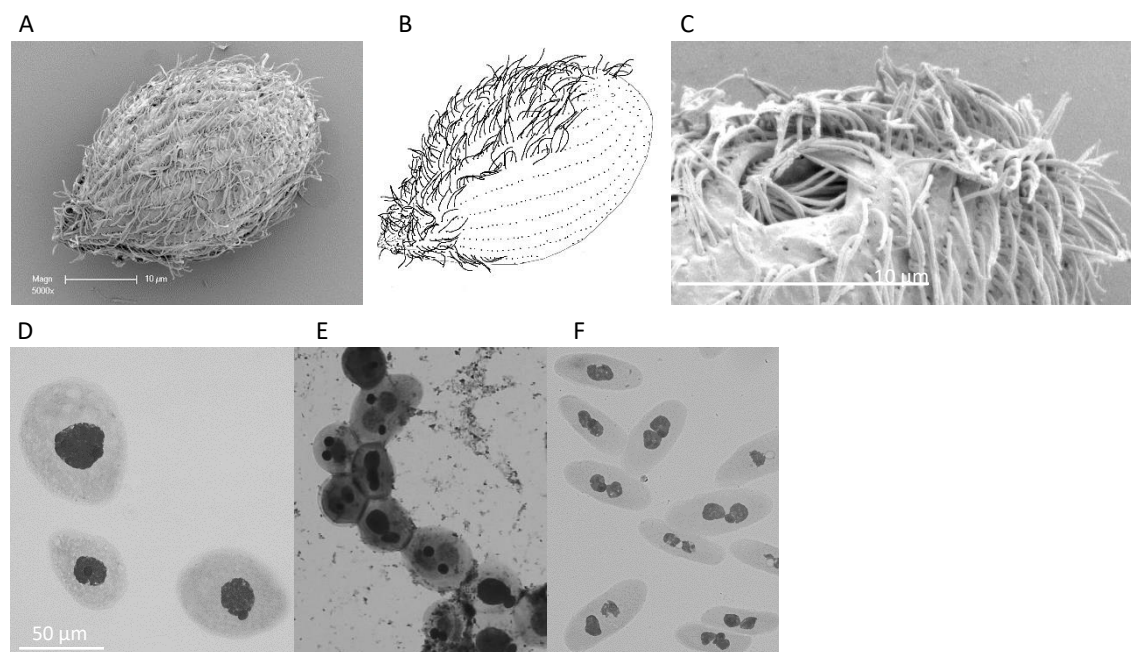


Figure 1. A) Scanning electron micrograph and B) line drawing of *T. rostrata* TRAUS showing a trophont with 28 rows of cilia and a tapered rostrum; C) scanning electromicrograph of the buccal opening and anterior rostrum; Giemsa stained D) trophonts, E) cysts and F) theronts.

3.2 Pathogenicity

Exposure of adult *D. reticulatum* resulted in high mortality of slugs exposed to TRAUS (day 42, 89%) and TR01 (day 42, 60%) compared to the controls (24%). The result showed that both *T. rostrata* TR01 and TRAUS were pathogenic for *D. reticulatum*. Eggs of *D. reticulatum*, laid in the first week of the experiment, when no adults had died, were incubated until hatching occurred and the survival of neonates was assessed. Few neonates of the TRAUS and TR01 exposed parents survived after 2 weeks (8% and 20% respectively) compared to 82% of neonates of the non-exposed control parents. In a second experiment, direct exposure of neonates of non-exposed parents resulted in high mortality within 15 days in the TRAUS and TR01 exposed groups (6.7% and 1.7% survival respectively) compared to non-exposed controls (80% survival). These results established that both isolates were pathogenic for *D. reticulatum*.

3.3 Mitogenomes of TR01 and TRAUS

The mitogenomes of TRAUS and TR01 were characterised using 17,797,914 and 33,260,090 Illumina DNA short-read sequences, respectively [13]. Assembly of the mitochondrial DNA sequences of TRAUS resulted in a 47,235nt contig (1,498,614 reads, average sequence depth of 4806 reads) and for TR01 a 47,310nt contig (1,204,576 reads, average sequence depth of 3808 reads). The mitogenomes were linear and the assemblies reached the telomeric repeats which was taken as an indication of assembly of the complete mitochondrial genomes.

The TRAUS mitogenome had a percent nucleotide composition of T (40.6) C (10.5), A (37.6) and G (11.2) which was almost identical to the TR01 mitogenome composition which was T (40.6), C (10.6), A (37.7) and G (11.1). Both had a low G+C content (21.7-21.8%). Each had 45 protein coding sequences (CDS), 8 tRNA genes and 56 ribosomal RNA genes. There are 2 ORFs for the SSUrRNA (*rns a* and *rns b*) and two ORFs for the

LSU rRNA rnl a and b. The rnl a and b genes are duplicated at each end of the linear mitochondrial DNA. Twenty-two of the coding DNA sequences (CDS) encoded proteins of known function and the remaining CDS were open reading frames designated *yfmf*, which are conserved in *Tetrahymena* but have no known function [18-20]. The difference in the length of the contigs of TR01 and TRAUS, excluding the telomer repeats, consist of 24 InDels in intergenic regions and InDels in rnlb1 (4 SNPs), rnsb (2 SNPs) and rnlb2 (1 SNP).

The organisation and gene arrangements are syntenic with the other *Tetrahymena* mitogenomes; *T. thermophila* [18], *T. pyriformis* [19], *T. paravorax*, *T. malaccensis* and *T. pigmentosa* [20] except that there was no duplication of *nad9*, as has occurred in *T. thermophila* and *T. malaccensis*. The genes are arranged divergently from a central region which is thought to contain elements that control transcription and replication (Figure 2). Phylogenetic placement of the concatenated amino acid coding regions of the mitogenomes placed TRAUS and TR01 closely together (Figure 3).

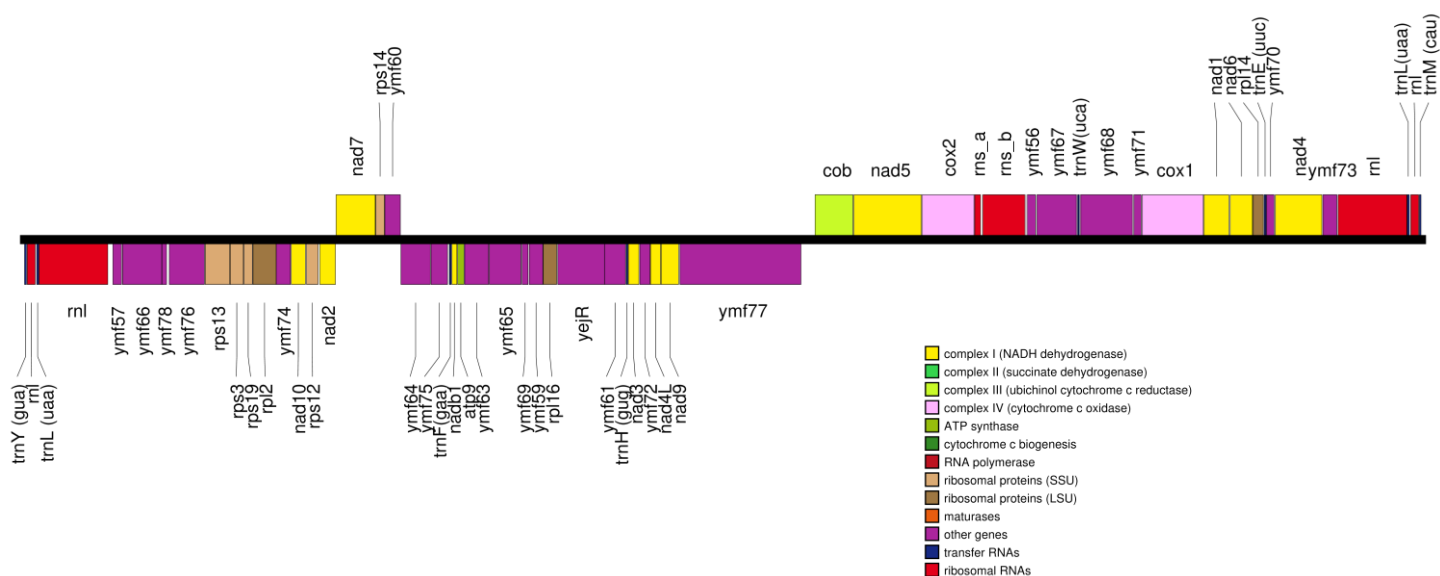


Figure 2. Arrangement of ORFs in the *T. rostrata* mitogenome

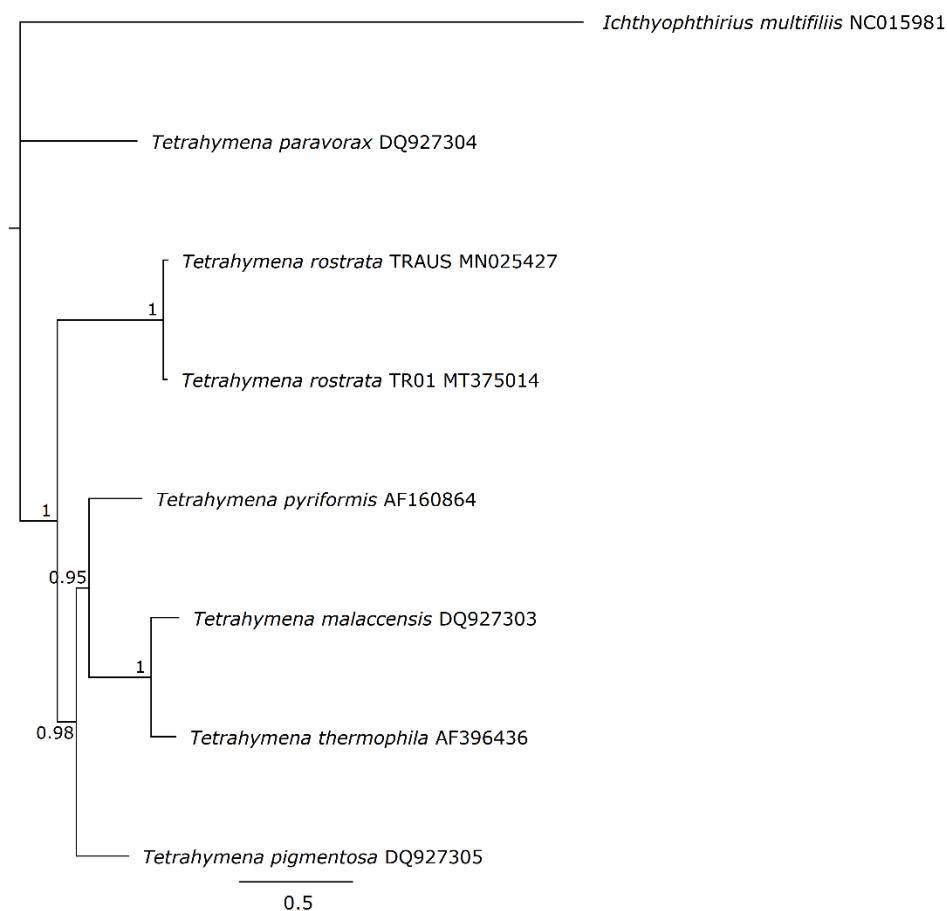


Figure 3. MrBayes tree of the translated, concatenated amino acid coding regions of the mt genomes of *T. rostrata* TRAUS, TR01 and other *Tetrahymena* species

The sequence conservation of the *yfm* ORFs is in the same order as the genes encoding proteins of known function, except for the large >4kb ORF *yfm77* (Table 2). Although the *yfm77* ORF is maintained, the low complexity of the encoded polypeptide can support many synonymous substitutions without much impact on the structure. Indeed, the hydrophobic domains of the predicted Ymf77 protein are very well conserved between TRAUS and TR01 (data not shown). Ymf ORFs may be particularly useful as target genes for examining intraspecies and clonal diversity because they are generally highly variable [18-20]. Comparisons of the ORFs between the two isolates identified many conserved regions that would be amenable as PCR primer binding sites which would permit the amplification of the variable regions.

Of particular interests is the *nad9* gene. There is only a single *nad9* CDS in TRAUS and TR01 whereas duplication of *nad9* has occurred in *T. thermophila* and *T. malaccensis*. This gene encodes part of the oxidative phosphorylation NADH Fo complex. Gershgorin suggests the *nad9* duplications in *T. thermophila* and *T. malaccensis* are due to independent duplication events rather than a single duplication event in a recent common ancestor [26]. Comparison of the *nad9* alleles show that the differences within *T. malaccensis* and *T. thermophila* respectively are small (*T. malaccensis* 7 SNPS, 1 amino acid difference, 100% similarity; *T. thermophila*, 10 SNPS, 1 amino acid difference, 100% similarity) and less than the inter-species variation of the two pairs of duplicated genes (76-81 SNPS, 17-18 amino acid differences, 97.49% similarity). The differences between *nad9* of TRAUS and TR01 are less than the differences between the *T. thermophila* and *T. malaccensis* but more than the differences between the *nad9* duplicates (TRAUS vs TR01 *nad9*, 30 SNPS, 7 amino acid differences, 98.99% similarity (Figure 4). The *T. thermophila* and *T. malaccensis* *nad9* display concerted evolution, where paralogs in the same species show higher sequence similarity to each other than to orthologs in other species. This trend is heightened those cases because the paralogs are evolving in the same DNA molecules.

Table 2. Comparisons of protein coding sequences of TRAUS and TR01

CDS	bp	DNA seq identity (%)	Amino acid seq similarity (%)
Atp9	228	99.12	100
Rps14	306	99.02	100
Ymf57	303	98.68	100
Ymf78	168	98.21	100
Ymf56	294	97.96	100
Ymf72	357	97.76	99.16
Nad4L	351	97.44	100
RpS12	402	97.26	100
Ymf69	216	97.22	100
Ymf61	720	97.08	99.58
Ymf65	2652	97.05	100
Nad2	537	97.02	100
Ymf64	999	97	99.7
Nad10	489	96.93	100
Ymf76	1194	96.82	99.75
RpS3	456	96.77	100
RpL16	465	96.77	100
Ymf74	492	96.75	100
Cob	1287	96.74	99.77
Nad1b	180	96.67	100
Ymf70	270	96.67	98.89
Nad7	1329	96.61	99.77
RpS13	834	96.52	99.28
Nad3	366	96.45	100
Ymf75	567	96.3	98.41
Ymf63	816	96.2	100
Cox1	2067	96.18	99.71
RpL2	789	96.07	100
RpS19	297	95.96	98.99
Ymf66	1329	95.71	100
Nad1a	855	95.67	100
Ymf68	1761	95.57	99.32
RpL14	360	95.56	100
Nad4	1578	95.56	97.72
YejR	1581	95.51	98.67
Nad6	762	95.41	98.92
Ymf60	534	95.13	99.44
Ymf67	1344	95.09	98.44
Ymf73	483	95.03	100
Nad9	597	94.97	98.99
Nad5	2280	94.91	99.21
Cox2	588	94.9	98.64
Ymf59	471	94.69	98.73
Ymf71	264	94.32	97.73
Ymf77	4077	92.62	96.98

¹In order of DNA identity

Consensus	MQNKYMQPIQLFIVFEKLNLSKIWIWSKKLNSNHSIXLIPSNWFYSINLFLKKELLFSNITL	247
TPARA DQ927304 - nad9	. .KS.H. .L.T.L.M.N.A.I.M. TS. .	248
TRAUS MN025427 - nad9	. I.I.T.VI.	249
TR01 MT375014 - nad9	. I.I.VI.N.	250
TPYR AF160864 - nad9L.I.M.A.I.L.	251
TPIG DQ927305 - nad9L. V. N.L. I.L.LS.	252
TMAL DQ927303 - nad9 1A.I.S.	253
TMAL DQ927303 - nad9 2T.I.S.	254
TTHERM AF396436 - nad9 1M.A.R.	255
TTHERM AF396436 - nad9 2M.A.R.	256
Consensus	IENSAIDTXXYNXETNQXLDXNXIENYFFKNKLLVFNYYNYFLKSKLTXFIIILNXLXK	257
TPARA DQ927304 - nad9CQN. . ID. D. . S. . Q. FL. . F. . N. . I. Y. F. L. I. K. S. .	258
TRAUS MN025427 - nad9TH. . Q. . KF. SISV. S. I. N. I. L. I. SFY. .	259
TR01 MT375014 - nad9TH. . Q. . F. SISV. S. I. N. I. L. I. NFY. .	260
TPYR AF160864 - nad9TN. . L. N. . D. NK. S. . F. R. . I. L. . . N. S. .	261
TPIG DQ927305 - nad9SY. . L. . . . N. . K. S. N. . . . F. . . . S. S. .	262
TMAL DQ927303 - nad9 1IN. . Q. . . . N. . FSC. I. F. . . . Q. N. .	263
TMAL DQ927303 - nad9 2IN. . Q. . . . N. . FSC. I. F. . . . Q. N. .	264
TTHERM AF396436 - nad9 1IS. . L. . . . D. . KTN. M. L. . . . N. N. .	265
TTHERM AF396436 - nad9 2IS. . L. . . . D. . KTN. M. L. . . . N. N. .	266
Consensus	NIDSIDRIFSNANWLERETSEMYGINYKWKIDTRKLLLDYSKIENPMLKEFQSEGTQDAF	267
TPARA DQ927304 - nad9	. L.Y.L.S.Y. T. . . . V.	268
TRAUS MN025427 - nad9	. . . V. . . P. R. E. V.	269
TR01 MT375014 - nad9	. . . V. . . P. T. E. V.	270
TPYR AF160864 - nad9	SM. . . . V. VH.	271
TPIG DQ927305 - nad9	. MN. T. C.	272
TMAL DQ927303 - nad9 1 A.	273
TMAL DQ927303 - nad9 2 A.	274
TTHERM AF396436 - nad9 1	S. A.	275
TTHERM AF396436 - nad9 2 A.	276
Consensus	YNIFENQVVVLKNETVEL*	277
TPARA DQ927304 - nad9 TAT.	278
TRAUS MN025427 - nad9 I.	279
TR01 MT375014 - nad9	. S. I.	280
TPYR AF160864 - nad9	281
TPIG DQ927305 - nad9	. S.	282
TMAL DQ927303 - nad9 1	283
TMAL DQ927303 - nad9 2	284
TTHERM AF396436 - nad9 1	285
TTHERM AF396436 - nad9 2	286

Figure 4 Alignment of *nad9* from *T. paravorax* (TPARA), *T. rostrata* (TRAUS, TR01), *T. pyriformis* (TPYR), *T. pigmentosa* (TPIG), *T. malaccensis* (TMAL) and *T. thermophila* (TTHERM)

3.4 Cytochrome oxidase 1 gene

Comparison of the 1796 nucleotides available for the *cox1* gene of the contemporary *T. rostrata* isolates is shown in Figure 5. The phylogenetic analysis shows that the sequences group into two clades. Sequences of TRAUS, TR1015 and TR1016 cluster together. The genetic distances between TRAUS and TR1015 and TR1016 were 1.52 and 1.07 respectively. The sequences of the other Spanish and Polish isolates clustered together as previously reported [5]. There was no discrimination between the host animal or geographic origin of the isolates. The *cox1* sequence previously deposited in Genbank for TR01 is included (GU439231). There were 13 SNP differences between the *cox1* of TR01 entered into Genbank in 2016 [10] and TR01 (ATCCPRA326) (MT375014) as examined by us. The Illumina reads across the TR01 (ATCCPRA326) *cox1* gene are homogeneous indicating that the template is clonal. The most likely reason for the discrepancy is that GU439231 was derived from DNA extracted from a *T. rostrata* culture that was not clonal as indicated in Segade *et al* 2009 [9] but the material deposited at ATCC is clonal. The *cox1* sequence of TR01 (GU439231) and TR01_PRA326 (MT375014) agree over the first 1064 bases of common sequence. Indeed, TR02 and TR03 sequences are also the same over this region which covers the barcode region (nt 220-908, MT375014). The 18S sequence did not differ between TR01 (JQ 045342), TR01 (ATCCPRA326) and TRAUS. Comparison of each protein coding DNA sequence showed that there was a high level of conservation across

the mitogenomes. There were no particular hotspots of nucleotide substitutions or indels that might indicate genetic drift.

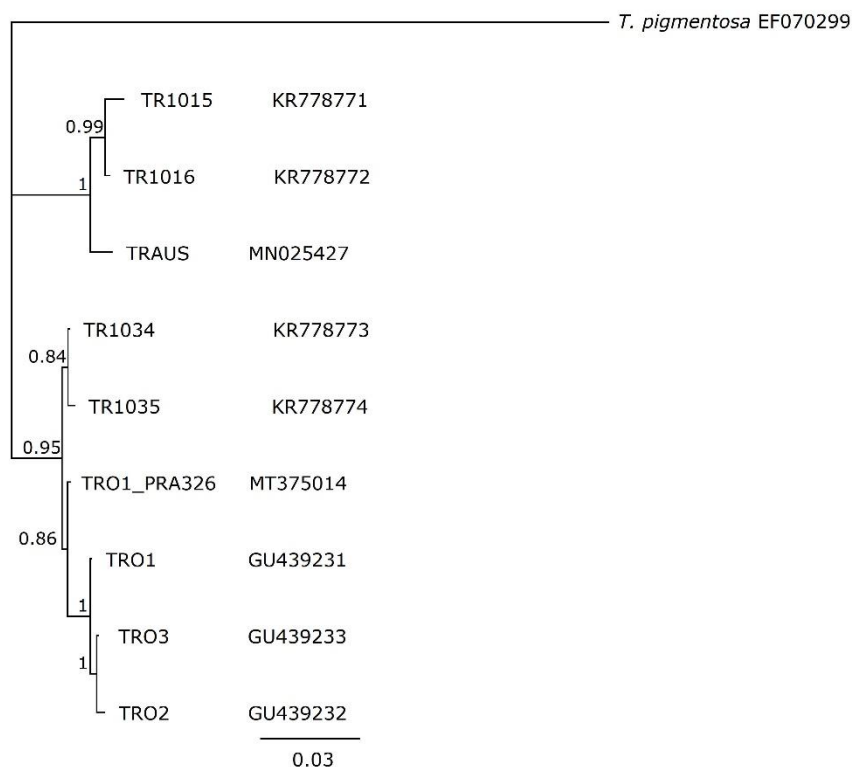


Figure 5. Phylogenetic tree of *Tetrahymena rostrata* strains based on 1796 bp of the *cox1* genes. A MrBayes tree was built in Geneious Prime v.1.2 using *Tetrahymena pigmentosa* as an outgroup

3.5 HCEs and central repeat region

Five highly conserved elements (HCE) have been identified in the mitogenomes of other *Tetrahymena* [20,26]. The HCEs occur in both TR01 and TRAUS at the same sites. HCE287 was 31nt downstream of *ymf57*. Both HCE234 and 290 are overlapping and occur within the coding region of *ymf78*. HCE299 occurs in the *nad2* CDS and has been suggested to be the promoter of *nad7* [26]. HCE315 and HCE138 occur between the *ymf77* and *cob* ORFs which are arranged divergently, on either side of the central control region. Notably HCE138, 29 bp (5'AATAGCCGCACCAAAAAGAAAAAATCAA) was shared by the other species of *Tetrahymena* (Table 3). The motif contains the very highly conserved, GC-rich motif, GCCGCACC [20]. The only pair to agree completely were TRAUS and TR01 (Table 2). The distance between HCE315 and HCE138 was 113 nt for TRAUS and 120 nt for TR01. Each isolate had 5 tandem repeats between HCE138 and HCE315 but the repeats differed. The TRAUS repeat was TAAATTTAAAATAAT and the TR01 repeat TAATTAATTTAAAATAA. Repeats were not apparent in the other species. The region between HCE138 and 315 contains the most variable spans of sequence between TR01 and TRAUS, however secondary structures were almost identical and were significantly different from the other species (Figure 6). Zhang et al [27] compared the central region of a number of linear ciliate mitogenomes and found the repeats in two species of *Euplotes* were identical. However comparison shows that the translated *nad9* genes in these two species are only 63.92% similar highlighting how multiple sequences are needed to build taxonomies.

Table 3. Comparison of HCE sequences in between *yml77* and *cob* in *Tetrahymena* species

	HCE 315	space	HCE 138
Consensus	ATAACGAATTTACAATAAAAAATAAT		AATAGCCGCACCTAAAAGAAAAAATCAA
<i>T. rostrata</i> TR01	.C.....T.....T.....	120A.....
<i>T. rostrata</i> TRAUS	.C.....T.....T.....	113A.....
<i>T. pigmentosa</i>	TC....T.....T....A	126C.
<i>T. paravorax</i>GAC..A.....T.A...	116T..T..A.....
<i>T. thermophila</i>	T....T..C.....A	113T.
<i>T. malaccensis</i>T.....	91T.
<i>T. pyriformis</i>	T.....A	65	...G.....

¹GC box in bold type

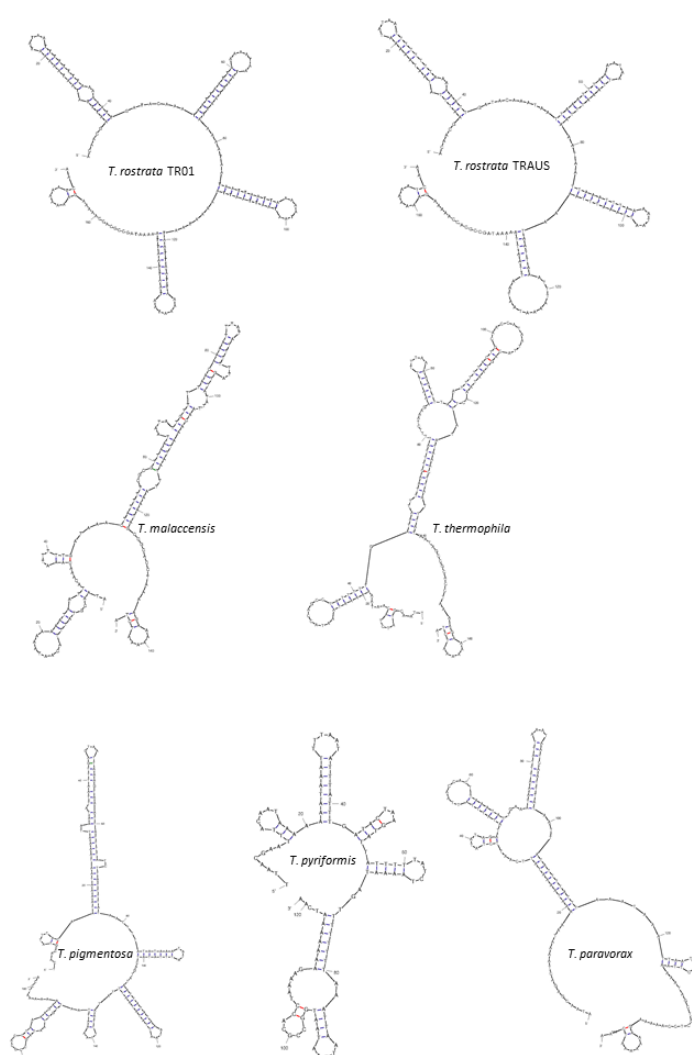


Figure 6. Secondary structures of the central repeat region of *T. rostrata* TR01 and TRAUS and other *Tetrahymena* species based on single stranded DNA

3.6 mtSSUrRNA

Doerder (8) highlighted the use of the mtSSUrRNA for species identification where there are multiple isolates available. Doerder found that the “left” and “right” region of the mtSSUrRNA were highly variable and that the right region was represented by more strains in GenBank. The right mtSSUrRNA region was analysed for its potential ability to provide clarity around complex species. The two *T. rostrata* mtSSUrRNA have 2 SNP differences (1 transition and 1 transversion) indicating a low sequence divergence (2.7%). Both isolates had identical V4 regions (TRAUS nt847-1008; TR01 nt848-1009). A phylogenetic tree of the mtSSUrRNA sequences shows TR01 and TRAUS on a separate branch from other *Tetrahymena* (Figure 7). Intra-species variability has been found in some other *Tetrahymena* species [11]. In this case, we are able to compare the whole mtSSU genes of the two isolates. Overall, there are 18 SNP differences in the 1431 nt gene and 8 of the SNPs are focused in a 22-25nt hypervariable domain. The differences are one 3nt insertion (TTT), 3 transversions and 2 transitions. This region (TR01 nt1283-1304) may be particularly useful for examining diversity in species complexes. It is situated 3’ to the region most commonly available in Genbank which was used in recent *Tetrahymena* phylogenies which revealed many new species [11]

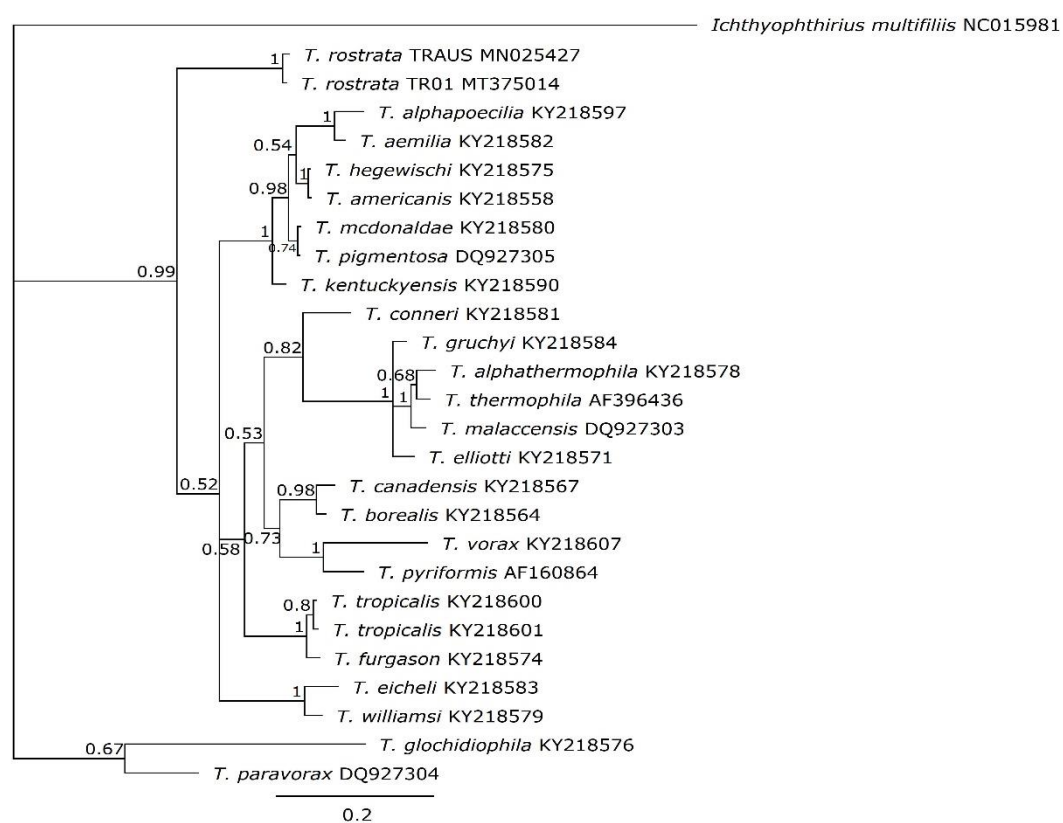


Figure 7. Phylogenetic tree of *Tetrahymena* based on the mtSSUrRNA, 541 bp right-hand region of TRAUS, TR01 and other *Tetrahymena* species. A MrBayes tree was built in Geneious Prime v.1.2 using a 1,100,000-generation chain length and a 1,000,000-generation burn-in with *Ichthyophthirius multifiliis* as an out group. Posterior probability values are shown to indicate branch support.

3.5 Nuclear genes

Nuclear genes are expected to evolve at a different rate from mitochondrial genes. However, the life cycle of *T. rostrata* is different from ciliates which have genetic exchange via conjugation. During encystment-induced autogamy, the pronuclei from the same cell fuse and the new macronuclei derived from them are expected to be homozygous [5,28]. Recessive mutations in the micronuclei will accumulate though successive rounds of

autogamy unless there is purifying selection. We mined the short read files to extract several nuclear genes. The entire small subunit, 18S rRNA-ITS1-5.8S-ITS2-28S rRNA sequences of TR01 and TRAUS were assembled as ~6kb single contigs for each strain (MT420428 and MN158348 respectively). Contigs were assembled with 168,388 reads for TR01 and 656,235 reads for TRAUS with an average coverage of 4,299 for TR01 and 16,024 for TRAUS. These were compared to Genbank entries for 18S rRNA available for a *T. rostrata* from Spain (JQ045342) (strain not specified) and the Polish strains, TR1015, TR1016 TR1034 and TR1035.. The entire set are identical across the 18S-ITS1-5.8S rRNA sequence and 28S sequence. Neither *T. rostrata* TR01 or TRAUS have a Group 1 intron in the 28S rRNA [29].

Histone H3 – intergenic region – histone H4 sequences of TR01 and TRAUS were each assembled as single contigs for each strain (MT506240 and MN167836 respectively). The entire H3 and H4 genes were covered. Contigs were assembled with 259 reads for TR01 and 345 reads for TRAUS with an average coverage of 20 reads for TR01 and 30 reads for TRAUS. The H3 and H4 histone genes of TR01 and TRAUS were identical, two SNPs were identified in the intergenic region.

Actin plays an essential role in multiple eukaryotic cellular process such as cytoskeletal structure, motility and intracellular transport. Many eukaryotes have families of actin genes however *T. thermophila* and *T. pyriformis* have just one actin gene without exons. There are low levels of genetic polymorphism within the actin gene of different populations of *T. thermophila* [30] but other ciliates have variation within and between actin gene families [31]. We attempted to extract the actin gene sequence from TRAUS and TR01 for comparison, using the actin gene of *T. thermophila* template. Surprisingly, 4 contigs were assembled from each short read library, which were labelled ACT1-4. Each encoded a 377 aa protein which had 99.73% similarity to *T. thermophila* actin. The paralogs for each pair had low sequence divergence between TRAUS and TR01 (ACT1, 0; ACT2, 0.5%; ACT3, 0; ACT4, 0.4%). However the divergence between alleles from TRAUS was between 4.2% and 7.3% and the equivalent comparison for TR01 paralogs was 4.4-7.2% divergence. These findings indicate that duplication of actin genes has occurred in *T. rostrata* and there must be multiple actin genes in the micronuclear genome which are destined for the macronucleus. The multiplication of the actin genes is most likely to have occurred in the last common ancestor of TRAUS and TR01 because the divergence between the pairs (e.g. the 2 ACT1 genes) is less than the divergence of the ACT genes within the strain. It appears that there have been three duplication events from an ancestral actin gene, first to produce two genes which in turn both were duplicated. Between the ACT gene pairs in TRAUS and TR01 the copies of ACT4 and of ACT3 were identical and there were 4 and 2 SNPS between the ACT1 and ACT2 pairs showing the between pair conservation is high and suggesting that, if all mutations were equally likely that ACT4 and ACT3 were the more recent duplicates.

4. Conclusion

We have compared *T. rostrata* TRAUS with TR01 on the basis of morphology, life cycle and pathogenicity and by comparison of 58 DNA sequences (9 nuclear genes and 3 ITS, and 45 mitochondrial genes and the central repeat region). This is the most extensive comparison of any two *Tetrahymena* isolates. Classical mating experiments to assign them to the same species are not possible; with *T. rostrata* because it does not form mating pairs and therefore we must rely on morphology, behaviour and molecular comparisons to determine the relationships between isolates. In every respect, TRAUS and TR01 are more like each other than with any other *Tetrahymena*. There was a high level of identity in nuclear genes although there is some diversity in the mitochondrial genes, which are expected to evolve at a higher rate.. The variability in the protein coding mitochondrial CDS was similar among some *yfmf* genes and genes encoding known proteins. Some mitochondrial genes, such as *nad9*, are particularly useful for comparison because we can compare

the rate on change in the duplicates in *T. thermophila* and *T. malaccensis* and *cox1* because it is represented by many sequences in the databases and is variable. The difference in the *cox1* barcode show the two isolates to be very closely related and are not sufficient to split them into different species. None of the other comparisons suggested difference beyond intraspecies variation which might be attributed to clonal variation between two isolates from very distant geographical sites and possibly due to some selection during their parasitic phase in different hosts. The whole genome sequencing was a relatively easy way to derive the mitochondrial genomes and had the added bonus that we could mine the short read archive for additional nuclear genes. A complete genome assembly using long reads will further elucidate features of this species.

Supplementary Materials:

Author Contributions: Conceptualization, H.B-J., A.E.W. and N.D.Y.; methodology, A.E.W., R.E.H and N.D.Y.; experimentation, K.D., A.E.W, R.E.H and H.B-J.; writing—original draft preparation, H.B-J. and A.E.W; writing—review and editing, H.B-J, N.D.Y, A.E.W, R.E.H, D.A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Grains Research and Development Corporation, project number UM00059, and University of Melbourne.

Data Availability Statement: All sequences have been deposited in Genbank

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Corliss, J.O. Le cycle autogamique de *Tetrahymena rostrata*. *Comptes Rendus Hebdomadaires Des Seances De L Academie Des Sciences* **1952**, *235*, 399-402.
2. Kozloff, E.N. A species of *Tetrahymena* parasitic in the renal organ of the slug, *Deroceras reticulatum* *Journal of Protozoology* **1957**, *4*, 75-79.
3. McArdle, E.W.; Bergquist, B.L.; Ehret, C.F. Structural changes in *Tetrahymena rostrata* during induced encystment. *Journal of Protozoology* **1980**, *27*, 388-397.
4. Stout, J. The ecology, life history and parasitism of *Tetrahymena [Paraglaucoma] rostrata* (Kahl) corliss. *The Journal of Protozoology* **1954**, *1*, 211-215.
5. Kaczanowski, A.; Brunk, C.F.; Kazubski, S.L. Cohesion of clonal life history, senescence and rejuvenation induced by autogamy of the histophagous ciliate *Tetrahymena rostrata*. *Protist* **2016**, *167*, 490-510.
6. Segade, P.; Garcia-Estevez, J.; Arias, C.; Iglesias, R. Parasitic infections in mixed system-based heliciculture farms: Dynamics and key epidemiological factors. *Parasitology* **2013**, *140*, 482-497.
7. Van As, J.; Basson, I. *Ciliophoran (Ciliophora) parasites of terrestrial gastropods*. In *Natural Enemies of Terrestrial Molluscs*, GM Barker (Ed), CABI Publishing: Wallingford, UK, 2004; p 559-578.
8. Brooks, W. Tetrahymenid ciliates as parasites of the gray garden slug. *Hilgardia* **1968**, *39*, 205-276.
9. Segade, P.; Kher, C.P.; Lynn, D.H.; Iglesias, R. Morphological and molecular characterization of renal ciliates infecting farmed snails in Spain. *Parasitology* **2009**, *136*, 771-782.
10. Kher, C.P.; Doerder, F.P.; Cooper, J.; Ikonomi, P.; Achilles-Day, U.; Küpper, F.C.; Lynn, D.H. Barcoding *Tetrahymena*: Discriminating species and identifying unknowns using the cytochrome c oxidase subunit i (*cox-1*) barcode. *Protist* **2011**, *162*, 2-13.
11. Doerder, F.P. Barcodes reveal 48 new species of *Tetrahymena*, *Dexiostoma*, and *Glaucoma*: Phylogeny, ecology, and biogeography of new and established species. *Journal of Eukaryotic Microbiology* **2019**, *66*, 182-208.
12. South, A. *Terrestrial slugs: Biology, Ecology and Control*. Springer Science & Business Media: 2012.

13. Watt, A.; Haites, R.; Young, N.; Billman-Jacobe, H. The mitochondrial genome of *Tetrahymena rostrata*. *Mitochondrial DNA Part B-Resources* **2020**, *5*, 53-54. 480
481
14. Gao, F.; Li, J.; Song, W.; Xu, D.; Warren, A.; Yi, Z.; Gao, S. Multi-gene-based phylogenetic analysis of oligotrich ciliates with emphasis on two dominant groups: Cyrtostrombidiids and strombidiids (Protozoa, Ciliophora). *Molecular Phylogenetics and Evolution* **2016**, *105*, 241-250. 482
483
484
15. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH image to imagej: 25 years of image analysis. *Nature Methods* **2012**, *9*, 671-675. 485
486
16. Billman-Jacobe, H.; Watt, A.; Heap, D.; Haites, R. A practical guide for managing a self-sustaining colony of *Deroceras reticulatum* (Müller) (Mollusca: Pulmonata). *Biocontrol Science and Technology* **2020**, *30*, 920-928. 487
488
17. Carrick, R. Xxi.—the life-history and development of *agriolimax agrestis* L., the gray field slug. *Earth and Environmental Science Transactions of The Royal Society of Edinburgh* **1939**, *59*, 563-597. 489
490
18. Brunk, C.F.; Lee, L.C.; Tran, A.B.; Li, J. Complete sequence of the mitochondrial genome of *Tetrahymena thermophila* and comparative methods for identifying highly divergent genes. *Nucleic Acids Research* **2003**, *31*, 1673-1682. 491
492
19. Burger, G.; Zhu, Y.; Littlejohn, T.G.; Greenwood, S.J.; Schnare, M.N.; Lang, B.F.; Gray, M.W. Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. *Journal of Molecular Biology* **2000**, *297*, 365-380. 493
494
495
20. Moradian, M.M.; Beglaryan, D.; Skozylas, J.M.; Kerikorian, V. Complete mitochondrial genome sequence of three *Tetrahymena* species reveals mutation hot spots and accelerated nonsynonymous substitutions in ymf genes. *PLoS One* **2007**, *2*, e650. 496
497
498
21. Chantangsi, C.; Struder-Kypke, M.C.; Lynn, D.H. Barcoding protists: A trailblazing study on a ciliate genus of closely related species, *Tetrahymena*. *Journal of Eukaryotic Microbiology* **2007**, *54*, 10S-10S. 499
500
22. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **1980**, *16*, 111-120. 501
502
23. Kumar, S.; Stecher, G.; Li, M.; Nnyaz, C.; Tamura, K. Mega X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **2018**, *35*, 1547-1549. 503
504
24. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647-1649. 505
506
507
25. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **2003**, *31*, 3406-3415. 508
26. Gershgorin, R.A.; Gorbunov, K.Y.; Zverkov, O.A.; Rubanov, L.I.; Seliverstov, A.V.; Lyubetsky, V.A. Highly conserved elements and chromosome structure evolution in mitochondrial genomes in ciliates. *Life* **2017**, *7*, 9. 509
510
27. Zhang, T.; Li, C.; Zhang, X.; Wang, C.; Roger, A.J.; Gao, F. Characterization and comparative analyses of mitochondrial genomes in single-celled eukaryotes to shed light on the diversity and evolution of linear molecular architecture. *International Journal of Molecular Sciences* **2021**, *22*, 2546. 511
512
513
28. Segade, P.; Garcia, N.; Garcia Estevez, J.M.; Arias, C.; Iglesias, R. Encystment/excystment response and serotypic variation in the gastropod parasite *Tetrahymena rostrata* (Ciliophora, Tetrahymenidae). *Parasitol Res* **2016**, *115*, 771-777. 514
515
29. Nielsen, H.; Engberg, J. Sequence comparison of the rDNA introns from six different species of *Tetrahymena*. *Nucleic Acids Research* **1985**, *13*, 7445-7455. 516
517
30. Katz, L.A.; Snoeyenbos-West, O.; Doerder, F.P.; Evolution. Patterns of protein evolution in *tetrahymena thermophila*: Implications for estimates of effective population size. *Molecular Biology and Evolution* **2006**, *23*, 608-614. 518
519
31. Yi, Z.; Huang, L.; Yang, R.; Lin, X.; Song, W.. Actin evolution in ciliates (Protist, Alveolata) is characterized by high diversity and three duplication events. *Molecular Phylogenetics and Evolution* **2016**, *96*, 45-54. 520
521
522

