Barcodes Reveal 48 New Species of Tetrahymena, Dexiostoma, and Glaucoma: Phylogeny, Ecology, and Biogeography of New and Established Species

F. Paul Doerder
Department of Biological, Geological and Environmental Sciences, Cleveland State University, 2121 Euclid Avenue, Cleveland, Ohio 44115

Keywords
Cox1; cryptic species; endemism; group I intron; SSUrRNA.

Correspondence
F.P. Doerder, Department of Biological, Geological and Environmental Sciences, Cleveland State University, 2121 Euclid Avenue, Cleveland, OH 44115
Telephone number: 216-687-2440; FAX number: 216-687-6972; e-mail: f.doerder@csuohio.edu

Received: 21 March 2018; revised 30 April 2018; accepted May 21, 2018.
Early View publication June 26, 2018

doi:10.1111/jeu.12642

ABSTRACT
Tetrahymena mitochondrial cox1 barcodes and nuclear SSUrRNA sequences are particularly effective at distinguishing among its many cryptic species. In a project to learn more about Tetrahymena natural history, the majority of >1,000 Tetrahymena-like fresh water isolates were assigned to established Tetrahymena species with the remaining assigned to 37 new species of Tetrahymena, nine new species of Dexiostoma and 12 new species of Glaucoma. Phylogenetically, all but three Tetrahymena species belong to the well-established “australis” or “borealis” clades; the minority forms a divergent “paravorax” clade. Most Tetrahymena species are micronucleate, but others are exclusively amicronucleate. The self-splicing intron of the LSUrRNA precursor is absent in Dexiostoma and Glaucoma and was likely acquired subsequent to the “australis/borealis” split; in some instances, its sequence is diagnostic of species. Tetrahymena americana, T. elliotti, T. gruchyi n. sp., and T. borealis, together accounted for >50% of isolates, consistent with previous findings for established species. The biogeographic range of species found previously in Austria, China, and Pakistan was extended to the Nearctic; some species show evidence of population structure consistent with endemism. Most species were most frequently collected from ponds or lakes, while others, particularly Dexiostoma species, were collected most often from streams or rivers. The results suggest that perhaps hundreds of species remain to be discovered, particularly if collecting is global and includes hosts of parasitic forms.

“We suspect that the named species are a small fraction of the total existing in nature, and that a substantial fraction of the whole are facultative parasites in a wide array of invertebrate hosts.” (Preparata et al. 1989)

WHEN the above was written there were approximately 30 named species of Tetrahymena, a large number compared to other ciliate genera. The optimism was based on the observation that nearly every field collection, including the writers’ own, found new species, though many were unnamed. Most were biological species separated by mating incompatibility, while others displayed particular morphological or life history traits. Some species were described based solely on genetic distance. Presently, there are 44 recognized species (Lynn and Doerder 2012; Pitsch et al. 2016; Quintela-Alonso et al. 2013; Zahid et al. 2014), some of which are parasitic, including the most recently described species (Lynn et al. 2017). This paper describes the collection of another 37 species of Tetrahymena as well as many new species in related genera Dexiostoma and Glaucoma. The results imply that there are dozens, perhaps hundreds more, sustaining the 1989 prediction.

Since 1996 neither morphology nor mating specificity that traditionally defined Tetrahymena species have played decisive roles in describing new species. Rather, the decision to declare new species has been based on differences in SSUrRNA and/or cox1 sequences. Examples include T. empidokyrea (Jerome et al. 1996), T. farleyi (Lynn et al. 2000), (amicronucleate) T. aquasubterranea
Tetrahymena species are cryptic, indistinguishable by microscopy. Failure of morphology to distinguish among most Tetrahymena species was recognized long ago (Corliss 1953) and is accepted as requiring alternatives (e.g., DNA sequences) for species description (Warren et al. 2017). Second, mating tests not only require investigators to perfectly maintain a panel of now dozens of mating type testers but such tests fail to identify the numerous isolates that are immature, selfing or asexual. Third, species are readily distinguished by molecular criteria, first by isozyme profiles (Nanney and McCoy 1976), and now primarily by cox1 barcodes and secondarily by SSUrRNA sequences (Chantangsi et al. 2007; Kher et al. 2011). Molecular description of species is increasingly common (Renner 2016).

The cox1 barcodes successfully identify Tetrahymena species, including those with identical SSUrRNA sequences, because the typical 0–2% intraspecies variation is considerably less than the average 10.5% interspecific variation (Chantangsi et al. 2007; Kher et al. 2011). The barcodes essentially confirmed past species assignments, including questionable ones. Both papers designated type strains and provided diagnostic cox1 barcodes and SSUrRNA sequences for most species. In addition to their pivotal role in decisions to name the species listed above, the sequences were used by Doerder (2014) to associate numerous wild amicronucleates with known (mostly micronucleate) species and to assign others to new species as further described here. The barcodes also were used to identify wild isolates as T. australis in a re-description of this species (Liu et al. 2016). The few exceptions to barcoding success appear to be due to mislabeling or misidentification of certain stock strains, a situation partially remedied here and discussed in more detail below. There are, however, six named Tetrahymena species for which living strains appear not to exist and therefore have little prospect of being barcoded.

Despite their morphological similarity, tetrahymenans are biologically diverse. Many possess a micronucleus and are sexual, but others, averaging 25% of wild isolates (Doerder 2014), are amicronucleate and asexual. Most species were isolated as free-living bacterivores, but some were found as parasites, primarily of aquatic invertebrates. Among the bacterivores are several species capable of transforming into large-mouthed predators which eat other tetrahymenas. Some species are globally distributed, appearing in two or more biogeographic zones, while others give evidence of endemism. All of these features cut across the “australis”, “borealis” and “paravorax” clades revealed by molecular data (Brunk et al. 1990; Chantangsi et al. 2007; Lynn and Struder-Kypke 2006; Nanney et al. 1998; Preparata et al. 1989; Williams et al. 1984). The “borealis” clade is larger and more diverse while the “australis” clade is somewhat smaller and contains a cluster of closely related species; only one traditional species belongs to the “paravorax” clade.

The new species of Tetrahymena, Dexiostoma and Glaucoma formally named here are easily diagnosed by their unique cox1 barcodes and SSUrRNA sequences. Species descriptions also include information on the micronucleus, the self-splicing intron (SSI) of the large ribosomal subunit rRNA (LSUrRNA) and habitat data. The micronucleus by itself has no diagnostic value, but it is an important life cycle trait because only cells with a micronucleus are sexual, usually in the form of conjugation but also autogamy. The loss of the micronucleus happens frequently across a wide range of species; some amicronucleates, for which collecting has not identified their micronuclear counterparts, may be ancient and evolutionarily stable as species (Doerder 2014).

The SSI presence too is not diagnostic as species may be SSI+, SSI– or mixed; however, in some cases its sequence is diagnostic. The SSI is a Group I intron that removes itself from the LSUrRNA precursor and joins the two exons (Zaug and Cech 1980). Its presence is easily detected by PCR. From the lack of congruence of SSI and SSUrRNA sequence differences, Sogin et al. (1986) concluded that SSIs were independently acquired subsequent to splitting of the Tetrahymena major clades. In this survey, the presence/absence of the SSI for most isolates was determined and new information on SSI population biology and evolution is presented.

An advantage of cox1 and other sequences is that they can be used to make inferences regarding population biology, including population structure. This is relevant to the often cited “everything is everywhere” debate regarding the biogeography of microorganisms (reviewed in Caron 2009). Some investigators (Fenchel et al. 1997; Finlay 2002) have argued that species of microorganisms, including ciliates, are globally distributed with large populations that exchange genes through frequent migration. Others (including Foissner 1999; Foissner et al. 2008) have argued for endemism. For Tetrahymena, some species are indeed global, found in 2–3 biogeographic zones (though the Ethiopian zone has rarely been sampled) (Simon et al. 2008), but they have not been studied regarding population structure. Other species, however, have more restricted distributions, and one, T. thermophila, gives evidence of population structure consistent with “moderate” endemism (Zufall et al. 2013). In this report, species with sufficient sample sizes are examined for evidence of population structure and endemism.

MATERIALS AND METHODS

Water samples were collected from freshwater sources primarily in the northeast quadrant of the United States; most were collected since 2006 with the goal of finding species in addition to T. thermophila. Samples were collected and processed as described by Doerder and Brunk (2012). Following isolation as clonal populations, cells...
were cultured in Cerophyll inoculated with *Klebsiella pneumoniae* or in axenic PPY (1% proteose peptone, 0.15% yeast extract, 0.001M FeCl$_3$; autoclaved). Some isolates failing to grow in PPY grew successfully in axenic LP (per liter: 15 g proteose peptone; 1 g yeast extract; 2.5 g bacterotryptone; 2.5 g liver fraction L; 5 g glucose; 1 g KH$_2$PO$_4$; 1 g Na$_2$HPO$_4$; autoclaved). Of isolates that could be grown only in Cerophyll, many could not be so maintained for long periods. A complete list of isolates, their collecting sites, Tetrahymena Stock Center accession numbers and associated GenBank accession numbers, is contained in Table S1; this table includes previously barcoded amicronucleates (Doerder 2014). Initials FPD and KD (Kristen Dimond) are used to identify primary collectors. GenBank accession numbers for the previously sequenced wild isolates and ATCC (American Type Culture Collection) strains (Chantangsi et al. 2007; Kher et al. 2011) are contained in those publications and are not repeated here.

Prior to any molecular work, most FPD isolates were challenged with a panel of all seven *T. thermophila* mating type testers to determine if they were mature *T. thermophila* and if so, their mating type. If conjugants were observed in the unmated control, the isolate was marked as a “sselfer”. Starved cells of the control well were vitally stained with acridine orange and examined for the presence/absence of the micronucleus by fluorescence microscopy (250x, 400x).

DNA from Cerophyll (typically 15 ml) or PPY (or LP) (8–12 ml) grown cells was purified with a modified microwave procedure (Goodwin and Lee 1993) as previously described (Zufall et al. 2013). Standard PCR with primers listed in Table S2 was used to amplify DNA for the cox1 barcode region, nuclear SSUrRNA, nuclear D2LSUrRNA, mtSSUrRNA, and a portion of the LSUrRNA containing the SSI. For the latter, the primers bind to highly conserved flanking sequences resulting in products that are SSI+ (~900 nt) or SSI− (~500 nt) as revealed by agarose gel electrophoresis. For sequencing, all PCR products were prepared with shrimp alkaline phosphatase and exonuclease III, and, in most instances, were sequenced in both directions with the same or additional primers. Sequences were assembled, edited and aligned with Geneious Versions 5 and 6 created by Biomatters at http://www.geneious.com/. Trees were drawn and edited with Mega 6.0 (Tamura et al. 2013). Networks were drawn by Network and Network Publisher (fluxus-engineering.com) (Bandelt et al. 1999; Forster et al. 2001). Population parameters were calculated with DnaSP5.10 (Librado and Rozas 2009) and Arlequin 3.5.1.2 (Excoffier and Lischer 2010). The Automatic Barcode Gap Detection (ABGD) program (http://wwwabi.snv.jussieu.fr/public/abgd/) was used to bin barcoded isolates into hypothetical species. The cox1 alignment for this purpose contained 778 sequences of 588 nucleotides (85% of barcode length, to allow the inclusion of more isolates with shorter sequences). The alignment included sequences of wild isolates, unnamed ATCC strains, and type strains of named species; it has been deposited at the Tetrahymena Stock Center. Upper value of P was set to 3%, and the distance value was set to “Distance K80 Kimura” with the transition/transversion ratio set to 1.65 as calculated by Mega.

Most isolates were identified as to species by 98–100% identity of their cox1 sequence to the cox1 barcode of one of the type strains (Table 1) as designated by Chantangsi et al. (2007) and Kher et al. (2011); type strain GenBank accession numbers are found in Lynn and Doerder (2012). Identification was usually corroborated by 100% identity to type SSUrRNA sequence, though in some instances species have identical SSUrRNA sequences. Barcodes are not available for the following recognized species: *T. chironomi*, *T. dimorpha*, *T. edaphoni*, *T. rotunda* (amicronucleate), *T. sialis*, and *T. stegomyiae*. It is possible, but likely unverifiable, that one or more of the new species reported here is one of these species. For six named *Tetrahymena* species, identification by cox1 barcode was verified by comparison of D2LSUrRNA sequences (Nanney et al. 1998) (Table 1). For *T. elliotii*, initial identification was by D2LSUrRNA. This is because, as explained in Text S1, the strain originally chosen for barcoding (Chantangsi et al. 2007) is not *T. elliotii*. The D2LSUrRNA sequences of 12 isolates (collected at seven different sites and including micronucleates, amicronucleates and selfers) were identical to those of 13/15 *T. elliotii* strains sequenced by Nanney et al. (1998). Following identification by D2LSUrRNA, *T. elliotii* isolates were identified using the cox1 barcode of the designated type. As discussed in Text S1 and Text S2, four “australis” species with identical SSUrRNA sequences, *T. cosmopolitanus*, *T. nipissingi*, *T. nanneyi*, and *T. sonneborni*, do not have reliable cox1 barcodes, and thus isolates suspected as belonging to them were grouped as “cosnipnanson”. Identification of *T. americana* required the designation of a type strain because as mentioned in Kher et al. (2011) and explained in Text S1, strain FL71 (ATCC 205052) chosen to represent *T. americana* (Chantangsi et al. 2007) is not that species. Therefore, two of the original *T. americana* (“T. pyriformis” syngen 2) as collected by Gruchy (1955) were obtained from ATCC, cox1 barcodes were determined and a type strain was designated. New *Tetrahymena*, *Dexiostoma* and *Glaucoma* species were designated based on cox1 sequence differences of >4%, as described in Results.

This paper makes use of a database (the NS database) of strain and biogeographical information collected by David Nanney and Ellen Simon and described in Text S1. This unpublished database formed the basis for assigning species to biogeographical regions in Simon et al. (2008) and is the basis for much of the biogeographical information at the website http://www.life.illinois.edu/nanney/tetrahymena/biogeography.html. The NS database was reconstructed from printouts gifted by David Nanney and is here made available in Excel spreadsheet form as Table S3 with caveats discussed in Text S1. Problematic strains highlighted in the spreadsheet were not used either for species identification or in population
| Species                    | Number observed | Wild isolates | Mic | Amic | SSIa | SSU clade | Dominant habitat | Biogeographical zonec | Evidence of population structured | Designated standard strain | N (cox1 length) | # haplo- types | Hd | π | Collector | TSC | TSC (cells;DNA)f |
|---------------------------|-----------------|---------------|-----|------|------|-----------|------------------|-----------------------|-------------------------------|------------------------|----------------|----------------|----|---|----------|-----|-----------------|
| **Tetrahymena**           |                 |               |     |      |      |           |                  |                       |                               |                        |                |                |    |   |          |     |                 |
| T. americanis             | 107             | 107           | 0   |      |      |           | Ponds, lakes    | NA, PA, NT          | Yes                           | ATCC30316               | 105 (674)       | 24             | 0.868 | 0.0088 | 21.3     |     |                 |
| T. australis              | 1               | 1             | 0   |      |      |           | Ponds, lakes    | NA, NT, PA, AU      | No                            | ATCC30171              | 2 (695)         | 2              | 1.000 | 0.0088 | 0.1      |     |                 |
| T. borealis               | 62              | 25            | 37  |      |      |           | Ponds, lakes    | NA                     | No                            | ATCC30317              | 62 (675)        | 17             | 0.762 | 0.0033 | 22.2     |     |                 |
| T. canadensis             | 22              | 16            | 6   |      |      |           | Streams         | NA, OR               | No                            | ATCC30368              | 23 (641)        | 19             | 0.987 | 0.0142 | 5.1      |     |                 |
| T. cosmopolitanis T. nippisingi T. nanneyi T. sonneborni | 12   | 12           | 0   |      |      |           | –               | –                      | –                            | –                      | 20 (635)        | –              | –     | –       | 0.4      |     |                 |
| T. elliotti               | 87              | 49            | 35  |      |      |           | Ponds, lakes    | NA, PA, NT, OR      | Yes                           | 21057-3               | 67 (643)        | 23             | 0.912 | 0.0084 | 7.2      |     |                 |
| T. farahensis             | 2               | nd            | nd  | nd   |      |           | Pond, wastewater| NA                   | –                            | Isolate 1.7            | 3 (286)         | 3              | NA    | NA     | 0.0      |     |                 |
| T. fusigasani             | 1               | 1             | 0   |      |      |           | Ponds, lakes    | NA                   | NA                            | ATCC30026              | 2 (843)         | 2              | 1.000 | 0.0089 | 0.1      |     |                 |
| T. hegewischii            | 25              | 23            | 0   |      |      |           | Ponds, lakes    | NA                   | Yes                          | ATCC30822              | 30 (638)        | 10             | 0.887 | 0.0063 | 0.1      |     |                 |
| T. mobilis                | 6               | 0             | 6   |      |      |           | River           | NA, PA               | No                            | CCAP1630/22           | 7 (627)         | 4              | 0.810 | 0.0039 | 0.1      |     |                 |
| T. pigmentosa             | 19              | 19            | 0   |      |      |           | Ponds, lakes    | NA, PA, NT          | No                            | ATCC30278              | 20 (677)        | 11             | 0.900 | 0.0046 | 0.1      |     |                 |
| T. pyriformis             | 13              | 0             | 13  |      |      |           | Ponds, lakes    | NA, PA               | No                            | ATCC30237              | 11 (524)        | 4              | 0.600 | 0.0017 | 4.0      |     |                 |
| T. shanghaiensis          | 1               | 0             | 1   |      |      |           | Stream          | NA, PA               | Yes                          | ATCC205039             | 2 (790)         | 2              | 1.000 | 0.0076 | 0.0      |     |                 |
| T. thermophila            | 1070            | 147           | 8   |      |      |           | Ponds, lakes    | NA                   | Yes                          | Strain B               | 111 (654)       | 28             | 0.867 | 0.0131 | 22.4     |     |                 |
| T. tropicalis             | 9               | 1             | 8   |      |      |           | Streams         | NA, NT              | –                            | ATCC30276             | 18 (487)        | 14             | 0.967 | 0.0331 | 0.1      |     |                 |
| T. vorax                  | 28              | 0             | 28  |      |      |           | Streams, ponds  | NA                   | No                            | ATCC30421             | 24 (685)        | 2              | 0.083 | 0.0001 | 12.2     |     |                 |
| bSSU/NIRA/CO              | 4               | 0             | 4   |      |      |           | Stream, soil, parasite | NA, PA, OR | –                            | See note              | 6 (655)         | –              | –     | –       | –        |     |                 |
| **Glaucoma**              |                 |               |     |      |      |           |                  |                       |                               |                        |                |                |    |   |          |     |                 |
| G. scintillans            | 8               | 8             | 0   |      |      |           | Stream, river   | NA, OR              | –                            | sAJ611861              | 8 (758)         | 4              | 0.9210 | 0.00297 | 0.3      |     |                 |
| G. reniformis= chattoni   | 22              | 13            | 2   |      |      |           | See text        | Pond, lake          | NA, AU                        | cEF070328              | 22 (712)        | 15             | 0.9610 | 0.00826 | 0.3      |     |                 |
| Colpidium striatum        | 3               | nd            | nd  |      |      |           | River           | NA, PA              | –                            | sx65533               | 2 (689)         | 2              | 1.0000 | 0.00871 | 0.2      |     |                 |
| Deixiotrema campylum      | 1               | 1             | –   |      |      |           | Lake            | NA, PA              | –                            | cEF070326              | –               | –              | –     | –       | 0.1      |     |                 |

Mic, micronucleate; Amic, amicronucleate; Hd, haplotype diversity, includes type strain and other ATCC isolates; π, nucleotide diversity per site, includes type strain and other ATCC strains.
aSSI, self-splicing intron of LSUrRNA.
bD2 LSU, 190 nucleotide variable region of large ribosomal subunit RNA.
cBiogeographical zone: NA, Nearctic; PA, Palearctic; NT, Neotropical; AU, Australian; OR, Oriental.
dPopulation structure indicated by significant $F_{ST} (p < 0.05)$, where appropriate (see text).
eType designation follows Chantangsi et al. (2007) and Kher et al. (2011); for cox1 and SSU accession numbers, see Table 1 of Lynn and Doerder (2012); for stains without ATCC or CCAP designation, GenBank accession numbers for cox1(c) or SSU(rRNA) (s) are given; ATCC, American Type Culture Collection; CCAP, Cambridge Collection of Algae and Protozoa.
fNumber of cell isolates and DNA samples deposited with the Tetrahymena Stock Center, Cornell University.
gAmbiguity among ATCC strains precludes identification of T. cosmopolitanis, T. nippisingi, T. nanneyi, and T. sonneborni (see Text S1 and Text S2).
Calculations. The NS database has also been deposited at the Tetrahymena Stock Center.

RESULTS

Species identification

Water samples were collected at a total of 939 sites (Fig. 1) mostly in May-September. Since 2006 when most collections were made with the goal of finding species other than *T. thermophila*, 49.8% of FPD sites (407/812) and 34% of FPD samples (857/2521) yielded *Tetrahymena*-like cells. In all, 1036 wild isolates (954 isolates collected by FPD and 82 collected by KD in 2011) were examined for *cox*1 barcodes and/or nuclear SSUrRNA sequence and assigned to existing or new species by matches to type strains (see Materials and Methods).

*Cox*1 barcodes identified 651 isolates as belonging to 15 (possibly 17) *Tetrahymena* species named prior to 2017 and to one previously barcoded unnamed strain as listed in Table 1 and discussed species-by-species in Text S2. Of these, 252 isolates belonged to *T. thermophila*, many of which were confirmed by mating tests. This number should not be interpreted as indicating that *T. thermophila* is a common species; it is in fact a relatively rare species, and the above number is inflated for reasons associated with collecting (see below). Another 275 isolates were assigned to 32 new species of *Tetrahymena* (Table 2, Text S2), using criteria discussed below; after it was independently discovered during the course of this project, one of these new species, *T. glochidiophila*, was named in a separate publication (Lynn et al. 2017). Among *Colpidium/Dexiostoma*, three isolates were assigned to *C. striatum*, one to *D. campyllum* and 29 to nine new *Dexiostoma* species (Table 3). Among *Glaucoma*, eight belonged to *G. scintillans*, 22 belonged to *G. chattoni* and 47 belonged to 12 new species (Table 3, Text S2).

**New species of Tetrahymena, Dexiostoma, and Glaucoma**

Isolates whose *cox*1 and SSUrRNA sequences found no match to those of type strains were candidates for new species (Tables 2 and 3). A *cox*1 sequence difference of >4% was used to assign isolates to new species, the same as previously used to declare amicronucleates as putative new species (Doerder 2014). This difference was chosen based on three considerations. The first, as cited in the Introduction, was that interspecific differences average 10.5% whereas intraspecific differences typically are 0–2%. The second consideration was that a plot of 222,778 pairwise differences among individual isolates and type strains (Figure S1A) shows a break at 4%; with exceptions described below, pairwise differences <4% are intraspecific. The third consideration is that in most instances species with >4% *cox*1 barcode differences also have divergent SSUrRNA sequences, which in the near absence of SSUrRNA intraspecific variation, is indicative of species level differences. There are, however, in the “australis” clade two groups of biological species that not only have *cox*1 differences of <4% but also have identical SSUrRNA sequences (see below). These are nevertheless distinguished by their unique *cox*1 haplotypes. In any event, it is recognized that the value of 4% is arbitrary and a compromise between false positives and false negatives. As with past species criteria, future studies may alter the conclusions presented here.

The validity of diagnosis of *Dexiostoma/Glaucoma/Tetrahymena* species by *cox*1 barcodes was further investigated with the distance program ABDG (see Materials and Methods). ABDG uses the barcode “gap” that occurs when intraspecific differences are smaller than interspecific differences. By recursively inferring intraspecific limits and detecting barcode gaps, the program bins barcode sequences into hypothetical species until the number of groups is stable. For the present analysis, the alignment

© 2018 International Society of Protistologists

Journal of Eukaryotic Microbiology 2019, 66, 182-208
| New species | Proposed name                  | Number observed | Wild isolates | Mic | Amic | SSI | nuSSU clade | Dominant habitat | Evidence of population structure | cox1 N (length) | # haplo types | Hd   | p    | Collector |
|-------------|--------------------------------|----------------|---------------|-----|------|-----|-------------|-------------------|----------------------------------|----------------|---------------|------|------|-----------|
| nsp01       | Tetrahymena eriensis n. sp.   | 4              | 4             | 0   |      |     |             | australis Ponds, lakes |                                 | 3 (689)        | 3             | 1.000 | 0.0048 | fpd       |
| nsp07       | Tetrahymena gruchyi n. sp.    | 85             | 4             | 27  | +/−  |     |             | australis Ponds, lakes |                                 | 59 (516)       | 12            | 0.866 | 0.0100 | fpd, kd   |
| nsp09       | Tetrahymena kentuckyensis n. sp. | 40            | 36            | 2   | +/−  |     |             | australis Ponds, lakes |                                 | 38 (649)       | 5             | 0.529 | 0.0009 | fpd       |
| nsp10       | Tetrahymena glochidiophila    | 29             | 27            | 0   |      |     |             | paravorax Ponds, lakes | No                              | 24 (658)       | 23            | 0.996 | 0.0163 | fpd       |
| nsp14       | Tetrahymena holzi n. sp.      | 2              | 2             | 0   | +    |     |             | australis Streams      |                                 | 2 (689)        | 2             | 1.000 | 0.0029 | fpd       |
| nsp15       | Tetrahymena alphathermophila n. sp. | 17            | 0             | 17  |      |     |             | borealis Ponds, lakes | Yes                             | 11 (655)       | 2             | 0.436 | 0.0087 | fpd       |
| nsp16       | Tetrahymena welliae n. sp.    | 4              | 1             | 3   |      |     |             | borealis Streams      |                                 | 4 (685)        | 3             | 0.833 | 0.0110 | fpd       |
| nsp18       | Tetrahymena newhampshirensis n. sp. | 13            | 4             | 9   | +    |     |             | borealis Ponds, lakes |                                 | 12 (678)       | 7             | 0.864 | 0.0029 | fpd       |
| nsp19       | Tetrahymena kidderi n. sp.    | 6              | 0             | 6   |      |     |             | borealis Ponds, lakes |                                 | 6 (689)        | 2             | 0.333 | 0.0005 | fpd       |
| nsp20       | Tetrahymena deweyae n. sp.    | 6              | 0             | 6   |      |     |             | borealis Ponds, river |                                 | 6 (652)        | 2             | 0.533 | 0.0033 | fpd       |
| nsp21       | Tetrahymena williamsi n. sp.  | 8              | 3             | 5   |      |     |             | borealis Ponds, river |                                 | 7 (675)        | 3             | 0.667 | 0.0157 | fpd       |
| nsp22       | Tetrahymena aqufluente n. sp. | 4              | 0             | 4   | +    |     |             | borealis Streams      |                                 | 4 (689)        | 1             | NA    | NA    | fpd       |
| nsp23       | Tetrahymena colorunensis n. sp. | 3             | 0             | 3   |      |     |             | borealis Ponds        |                                 | 3 (682)        | 1             | NA    | NA    | fpd       |
| nsp25       | Tetrahymena betathermophila n. sp. | 2             | 0             | 2   | +    |     |             | borealis Ponds        |                                 | 2 (632)        | 2             | 1.000 | 0.0079 | fpd       |
| nsp30       | Tetrahymena mcdonaldae n. sp. | 8              | 8             | 0   |      |     |             | borealis Ponds, lakes |                                 | 8 (683)        | 3             | 0.464 | 0.0020 | fpd       |
| nsp31       | Tetrahymena rayi n. sp.       | 5              | 2             | 2   | +    |     |             | borealis Ponds, river |                                 | 5 (685)        | 3             | 0.800 | 0.0038 | fpd       |
| nsp35       | Tetrahymena engbergii n. sp.  | 2              | Self          | 0   |      |     |             | australis Ponds       |                                 | 2 (689)        | 1             | NA    | NA    | fpd       |
| nsp40       | Tetrahymena allevheiniensis n. sp. | 2             | 2             | 0   |      |     |             | australis Streams     |                                 | 2 (689)        | 2             | 1.000 | 0.0015 | fpd       |
| nsp41       | Tetrahymena conneri n. sp.    | 2              | 2             | 0   |      |     |             | borealis Lakes        |                                 | 3 (661)        | 3             | 1.000 | 0.0272 | fpd       |
| nsp44       | Tetrahymena sellolouisa n. sp. | 7              | nd            | nd  |      |     |             | borealis Ponds        |                                 | 3 (653)        | 1             | NA    | NA    | kd        |
| nsp45       | Tetrahymena daisymae n. sp.   | 12             | 12            | nd  |      |     |             | borealis Ponds        |                                 | 12 (615)       | 7             | 0.894 | 0.0077 | kd        |
| nsp46       | Tetrahymena alphapyriformis n. sp. | 4             | nd            | nd  |      |     |             | borealis Ponds        |                                 | 5 (666)        | 2             | 0.600 | 0.0009 | kd        |
| nsp51       | Tetrahymena scolopax n. sp.   | 1              | 1             | 1   |      |     | borealis    | Stream              |                                 | 1 (886)        | 1             | NA    | NA    | fpd       |
| nsp52       | Tetrahymena pennsylvaniaensis n. sp. | 1             | 1             | 1   |      |     | paravorax   | stream              |                                 | 1 (892)        | 1             | NA    | NA    | fpd       |
| nsp53       | Tetrahymena zeutheni n. sp.   | 1              | 1             | 1   |      |     | australis   | Stream              |                                 | 1 (864)        | 1             | NA    | NA    | fpd       |
| nsp55       | Tetrahymena cleffmanni n. sp. | 1              | 1             | 1   |      |     | borealis    | Stream              |                                 | 1 (832)        | 1             | NA    | NA    | fpd       |
| nsp56       | Tetrahymena frigidoflumine n. sp. | 1             | 1             | 1   |      |     | australis   | Stream              |                                 | 1 (759)        | 1             | NA    | NA    | fpd       |
| nsp57       | Tetrahymena prescottii n. sp. | 1              | 1             | 1   |      |     | borealis    | Stream              |                                 | 1 (777)        | 1             | NA    | NA    | fpd       |
| nsp58       | Tetrahymena aemila n. sp.     | 1              | 1             | 1   |      |     | australis   | Lake                |                                 | 1 (844)        | 1             | NA    | NA    | fpd       |
| nsp59       | Tetrahymena eicheli n. sp.    | 1              | 1             | 1   |      |     | borealis    | Lake                |                                 | 1 (849)        | 1             | NA    | NA    | fpd       |
| nsp60       | Tetrahymena pocoroeensis n. sp. | 1             | 1             | 1   |      |     | borealis    | Lake, guppy         |                                 | 1 (844)        | 1             | NA    | NA    | fpd       |
| SIN         | Tetrahymena alphapoecilia n. sp. | 1             | 1             | 0   |      |     | australis   | Lake, guppy         |                                 | 2 (762)        | 2             | 1     | 0.0116 | fpd       |

For explanation, see legend to Table 1.
Table 3. Properties of wild isolates of new species of Dexiostoma and Glaucoma

| New species number | Proposed name | Number observed | Dominant habitat | SSI | N (cox1 length) | # haplo types | Hd | π |
|--------------------|---------------|----------------|-----------------|-----|----------------|--------------|----|---|
| Dexiostoma         |               |                |                 |     |                |              |    |   |
| Dnsp12             | D. maupasi, n. sp. | 5 4 0          | Steam           | 4   | 677            | 3            | 0.833 | 0.0032 |
| Dnsp29             | D. saccharum, n. sp. | 5 4 1          | Lake, stream    | 5   | 864            | 1            | NA | NA |
| Dnsp36             | D. venango, n. sp. | 2 2 0          | Lake            | 2   | 686            | 1            | NA | NA |
| Dnsp37             | D. laceyvillensis, n. sp. | 4 4 0          | River, lake     | 4   | 560            | 2            | 0.833 | 0.0060 |
| Dnsp38             | D. sabulum, n. sp. | 6 5 0          | River           | 6   | 675            | 3            | 0.733 | 0.0013 |
| Dnsp39             | D. pemaquidensis, n. sp. | 4 3 0          | Stream, pond    | 4   | 676            | 2            | 0.500 | 0.0007 |
| Dnsp40             | D. pinus, n. sp. | 1 1 0          | Stream          | 1   | 909            | 1            | NA | NA |
| Dnsp41             | D. novaeangliae, n. sp. | 1 1 0          | Lake            | 1   | 777            | 1            | NA | NA |
| Dnsp42             | D. tobeco, n. sp. | 1 1 0          | Stream          | 1   | 842            | 1            | NA | NA |
| Glaucoma           |               |                |                 |     |                |              |    |   |
| Gnsp01             | G. michiganensis, n. sp. | 5 5 0          | Lake            | 5   | 897            | 2            | 0.400 | 0.0005 |
| Gnsp02             | G. gramina, n. sp. | 3 3 0          | River, stream   | 5   | 880            | 3            | 0.800 | 0.0130 |
| Gnsp03             |               | 9 9 0          | Stream, pond    | 9   | 716            | 9            | 1.000 | 0.0071 |
| Gnsp04             |               | 7 7 0          | Stream, river   | 7   | 773            | 3            | 0.667 | 0.0081 |
| Gnsp05             |               | 7 5 2          | Stream          | 6   | 688            | 2            | 0.400 | 0.0017 |
| Gnsp06             |               | 4 4 0          | Stream          | 4   | 764            | 4            | 1.000 | 0.0081 |
| Gnsp07             |               | 2 0 2          | Stream          | 2   | 726            | 1            | NA | NA |
| Gnsp08             |               | 2 2 0          | Stream, pond    | 2   | 758            | 2            | 1.000 | 0.0040 |
| Gnsp09             |               | 1 1 0          | Stream          | 1   | 866            | 1            | NA | NA |
| Gnsp10             |               | 1 1 0          | Stream          | 1   | 785            | 1            | NA | NA |
| Gnsp11             |               | 1 1 0          | Lake            | 1   | 837            | 1            | NA | NA |

For explanation, see legend for Table 1.

(Materials and Methods) of 788 sequences included wild isolates assigned by distance to both known and new species, designated type strains of named species, and miscellaneous strains for which cox1 sequences are available. The inclusion of named biological species and the wild isolates assigned to them provides a test of the ability of barcodes to delineate species. In particular, should ABGD fuse species, the use of cox1 barcodes to diagnose new species would be compromised.

ABGD runs typically produced ~114 hypothetical species. Based on the number of named species and new species using the 4% rule, 104 species were expected. ABGD correctly grouped named species in all genera and species using the 4% rule, 104 species were expected. Based on the number of named species and new species, there is a barcode “gap” between clusters of closely related haplotypes. Tetrahymena americanis was split into three hypotheticals, one consisting of 89 strains including type strain UM616, one consisting of 16 strains including strain UM351, and one consisting solely of isolate 20866-1. Strains UM616 and UM351 differ by 2.4% and are bone fide mating strains of T. americanis (Text S1). Notably, ABGD did not fuse any T. americanis group with closely related T. hegewischii; both species have identical SSUrRNA sequences. Tetrahymena borealis was split into two hypothetical species differing by 1.5%, one containing 52 isolates (<0.3% intragroup difference) and the other containing five isolates (<0.5% intragroup difference). Tetrahymena thermophila was split into two hypotheticals, with 93 and 17 isolates, respectively. These divergent haplotype clusters have been described before (see figure 2 of Zufall et al. 2013). The most extreme haplotypes differ by 3.8%, but the average pairwise sequence difference is 1.3% (mode 0.9%). Last, T. tropicalis was split into four hypothetical species. For this and reasons discussed in Text S2, this problematic species is in need of reinvestigation.

ABGD did not fuse any new species defined by the 4% rule but did split four into two hypothetical species each (see also Text S2). Again, the splits occurred because distinct haplotype clusters created a barcode gap. Amicronucleate T. alphathermophila n. sp. (nsp15) contains two sets of identical haplotypes collected at two separate locations and differing by 2%. For T. williamsi n. sp. (nsp21) one hypothetical species consisted of five isolates (two nucleotide differences) and the other consisted of two identical isolates, the two hypotheticals differing by 3.1%. SSUrRNA sequences of both hypotheticals are identical. The species T. conneri n. sp. (nsp41) was formed from two wild micronucleate isolates and amicronucleate strain ATCC50413. The ATCC50413 cox1 sequence is 99.1% identical to one wild isolate, but is missing the 5’ region,
where many of the differences between the two wild isolates occur (96.2% identical). The SSUrRNA sequences of the two wild isolates are identical, but there is no SSUrRNA sequence for the ATCC strain. Of the new species split into two groups by ABGD, _T. conneri_ is the most likely to consist of two species. Mating tests were not performed. ABGD results indicate that cox1 barcodes readily distinguish among species and suggest that the 4% threshold is a conservative approach, less likely to split species.

For _Dexiostoma_ and _Glaucoma_ average intrageneric pairwise differences were 7.8% and 11.6%, respectively. Among the new _Tetrahymena_ species the average pairwise cox1 difference was 11.8% (Figure S1A; for phylogenetic tree see Fig. 2). For all _Tetrahymena_ species, the average intragroup pairwise differences in the “australis”, “borealis”, and “paravorax” clades were 8.4%, 10.7%, and 13.5%, respectively, with the “australis” and “borealis” clades showing considerable overlap (Figure S1B). The relatively few differences (<4%) were mostly among previously named species, none of which were fused by ABGD. For the “australis” clade, these include _T. americanis_ and _T. hegewischii_ (2.1–2.5% difference), and these two species and ATCC strain FL71 (4.4% difference) (see Text S1). Also, <4% are differences among members of the “cosnipanason” group (differences ranging from 0.3% to 3.4%) and that group compared to _T. pigmentosa_ (3.1–3.6% difference). Significantly, the only new species at <4% is _T. mcdonaldae_ n. sp. (nsp30) which differed from _T. pigmentosa_ by 3.4% (differences between _T. mcdonaldae_ and “cosnipanason” are 4.3–5.8%). In this case, _T. mcdonaldae_ was declared a new species based on cox1 phylogeny as supported by its unique SSUrRNA sequence. In the “borealis” clade, _T. borealis_ and _T. canadensis_ differ by 3.9% and are the only two historical “borealis” species below the 4% threshold. The other “borealis” instance involves strains CO/NI/R9/bSSU and their relationship with _T. furgasoni_, one of the originally named amicronucleate species (Nanney and McCoy 1976). As explained in Text S2, though cox1 difference is 2.7% (NI vs. _T. furgasoni_), uncertainties with respect to SSUrRNA sequences (Fig. 2) and SS1 divergence led to restraint in declaring these four strains and similar wild isolates as belonging to _T. furgasoni_.

The separation of species by differences in cox1 barcodes is consistent with low per site nucleotide diversity (π) of most species despite their often high haplotype diversity (Tables 1–3). The high π values of _T. tropicalis_ and _T. conneri_ n. sp. (nsp41) may indicate subspecies or even cryptic species (see ABGD results above; see also Text S2). The separation of species by cox1 barcodes is also consistent, in most cases, with differences in SSUrRNA sequences (see Phylogeny below). In some cases, the converse is true, as some species have identical SSUrRNA sequences but are distinguished by cox1 differences. For named species, these include members of the “australis” clade as mentioned above and in the “borealis” clade species _T. borealis_ and _T. canadensis_. Other species pairs with identical SSUrRNA include _T. alphapyriformis_ n. sp. (nsp46) and _T. pyriformis_; _T. newhampshirensis_ n. sp. (nsp18) and _T. colorunensis_ n. sp. (nsp23); _T. holzi_ n. sp. (nsp14) and _T. zeutheni_ n. sp. (nsp53); and several new species of _Dexiostoma_ and _Glaucoma_.

**Phylogeny**

The SSUrRNA NJ phylogenetic tree shown in Fig. 2 places _Dexiostoma, Glaucoma_ and _Tetrahymena_ in separate clades and supports the “australis” and “borealis” clades of _Tetrahymena_. It also places _T. paravorax_ and two relatives in a separate “paravorax” clade. Except for the inclusion of many more species, this tree strongly resembles the SSUrRNA NJ of Chantangsi and Lynn (their Supplementary figure 3, 2008). It is also similar to Chantangsi and Lynn’s (their figure 2, 2008) SSUrRNA Bayesian tree in which _T. paravorax_ is in its own clade but within monophyletic _Tetrahymena_. Trees with full length cox1 sequences (Chantangsi and Lynn 2008), regardless of method, while generally agreeing with SSUrRNA trees often differ in detail. For instance, such cox1 trees often include _Dexiostoma_ and _Colpidium_ within _Tetrahymena_ and variably place _T. paravorax_ in either the “australis” or “borealis” clade. NJ trees (Figure S2, also ML and MP trees, not shown) with the cox1 barcode (representing less than half of the gene) generally lack deep branch bootstrap support and therefore meaningful phylogeny. For instance, some “australis” members are grouped with “borealis” members, and _T. malaccensis_ and _T. thermophila_, close relatives in SSUrRNA and full-length cox1 trees, are not closely related in this tree. The cox1 barcode tree of Chantangsi et al. (2007) also groups species in unexpected combinations with low bootstrap values.

To explore the possibility that cox1 sequences are indicative of different trajectories between mitochondrial and nuclear genomes, and to explore their possible use in species identification, two nonoverlapping regions of the mitochondrial small subunit rRNA (mtSSUrRNA) of selected isolates were sequenced (Fig. 3). MtSSUrRNA sequences also were extracted from the complete mitochondrial genomes of _T. malaccensis_ (Strain MP75, GenBank DQ927303.1), _T. paravorax_ (Strain RP, GenBank DQ927304.1), _T. pigmentosa_ (Strain UM1060, DQ927305.1), and _T. pyriformis_ (Strain GL, GenBank L28677). Although the R (right) region is represented by more sequences, the L (left) region yielded similar results. Both R and L regions were highly variable, with R region average pairwise difference of ~11% for the “borealis” clade and ~3% for the “australis” clade. For the R region, unlike for nuclear SSUrRNA sequences, intraspecific variability was found among 5/16 species for which two or more isolates were examined; these include _T. americanis_ (two transition sites), _T. borealis_ (two transition sites), _T. canadensis_ (one transition site), and _T. pyriformis_ (one transition site). The more complex situation with _T. tropicalis_ (three transitions, one transversion) contributes to the need to reinvestigate this species as possibly consisting of subspecies or more anciently diverged asexuals.
Figure 2  SSUrRNA NJ phylogenetic tree of previously named species/strains and new species. This tree was based on 1376 nucleotides of nuclear SSUrRNA. Evolutionary distances were computed using the Kimura 2-parameter method and are in the units of the number of base substitutions per site. Bootstrap (500 replicas) values of <51% are not shown. All new species named here are shown in bold font. Branch labeled “trop” refers to wild isolates related to *T. tropicalis* (see Supplementary Text S2). Strains without species names are discussed in Text S2. Brackets indicate major *Tetrahymena* clades (in quotes), and the genera *Dexiostoma, Colpidium*, and *Glaucoma*. Other genus abbreviations: Gl, Glaucomides; B, Bromelio-phyra. Open circle: SSI−; Filled circle: SSI+; absence of circle, species/strain not examined. SSI data from present study and Sogin et al. (1986).

© 2018 International Society of Protistologists
Journal of Eukaryotic Microbiology 2019, 66, 182–208
Figure 3 MtSSUrRNA NJ phylogenetic tree. A. Map of the mitochondrial SSUrRNA gene (msb) showing left and right portions of the mtSSUrRNA. B. NJ phylogenetic tree of the right fragment of mtSSUrRNA. The tree was derived from 491 positions. Evolutionary distances were computed using the Kimura two-parameter method and are shown as the number of base substitutions per site. Bootstrap (500 replicas) of <51% are not displayed. With the exception of species for which GenBank accession numbers and strain designations are shown, species were identified here by cox1 and/or nuclear SSUrRNA sequences. Newly named species are shown in bold. Brackets indicate species belonging to the major SSUrRNA clades as shown in Fig. 2.
The relationship between *T. americanis* and *T. hegewischi* also is complex. Among six *T. americanis* isolates were three variants (n = 3, n = 2, n = 1); the sole *T. hegewischi* was identical to the n = 2 variant of *T. americanis*. For *T. thermophila*, the two identical wild isolates differed from the GenBank sequence of the Strain B type strain at one transversion site, a difference not only likely due to the older sequencing method used to produce the type strain sequence, but also possibly due to mutation in separately maintained laboratory strains. Among polymorphic *T. pyriformis*, one wild isolate was identical to its type strain. At the L region, with fewer sequences, 5/6 species with multiple isolates were polymorphic, including *T. glochidiophila* (nsp10) and *G. chattoni* which were not polymorphic at the R end. Full length sequences likely will reveal more polymorphisms and polymorphic species.

The NJ phylogenetic tree of the R region of mtSSUrRNA (Fig. 3) is in substantial agreement with SSUrRNA trees (e.g., Fig. 2). The genera *Glaucoma*, *Colpidium*/*Dexiostoma*, and *Tetrahymena* form separate clades, and *T. paravorax* and *T. glochidiophila* (nsp10) together also form a separate clade. The “australis” clade is separated with strong bootstrap support, though it is included in a moderately supported clade with two “borealis” species. For the remaining members of “borealis” clade, their positions are similar to those SSUrRNA trees (Fig. 2). Unlike in the cox1 barcode tree (Figure S2), *T. thermophila*, *T. malaccensis*, and *T. elliotti* form a well-supported clade that also contains *T. gruchyi* n. sp. (nsp7) and *T. alphathermophila* (nsp15) as in nuclear SSUrRNA trees. Although sequences from additional species are desirable, the present mtSSUrRNA data do not indicate substantially different nuclear and mitochondrial histories as suggested by some cox1 phylogenies. The variability in cox1 that makes it useful as a barcode simultaneously renders it of limited use in phylogeny except for recently separated species (Chantangsi and Lynn 2008).

Among species which have identical nuclear SSUrRNA sequences, mtSSUrRNA sequences resolved *T. borealis* and *T. canadensis* (2.3% difference) and *T. thermophila* and *T. alphathermophila* (3% difference). They did not, however, resolve *T. "cosnipnanson"* where all five isolates were identical and are expected by cox1 sequences to contain three species.

**The self-splicing intron (SSI)**

Most FPD isolates were examined for their SSI phenotype, and SSIs of SSI+ species were sequenced. As shown in Tables 1 and 2 and Fig. 2 most *Tetrahymena* species lack the SSI, as do all members of *Dexiostoma* and *Glaucoma*. Among the SSI+ species, four contained both SSI+ and SSI− isolates. SSIs were found in *Tetrahymena* species throughout the collecting area (Figure S3).

Among the newly sequenced SSIs, those of 11 new species (Table 9) and *T. malaccensis* are unique and therefore potentially diagnostic of their species.

As suggested by mapping onto the SSUrRNA phylogenetic tree (Fig. 2) SSIs likely were acquired subsequent to the “australis/borealis” split as previously concluded (Sogin et al. 1986). This is further supported by SSI phylogeny that does not cleanly break along “australis” and “borealis” clades (Figure S4B) and by the related observation that SSIs of close relatives such as *T. thermophila* and *T. malaccensis* are among the most divergent, differing by 13%. Conversely, among certain lineages SSIs appear to be vertically inherited from a common acquisition, for example, “cosnipnanson” and *T. mcdonaldae*, *T. betathermophila*, and *T. thermophila*. A fuller discussion of SSI acquisition and phylogeny is provided in supplemental Text S3.

**Population biology, ecology and biogeography**

Four species, *T. americanis*, *T. elliotti*, *T. gruchyi*, and *T. borealis* together account for over half of all *Tetrahymena* isolates (Table 4). Among the 15 most abundant species, six are new species, and nine belong to previously named species, including the unresolved “cosnipnanson”. *T. thermophila* and *T. vorax* are omitted from Table 4 because the present collection was not random with respect to them. *T. thermophila* was sometimes purposely collected by revisiting ponds where it is resident. For *T. vorax*, all instances in which an isolate ate the *T. thermophila* mating type testers (see Materials & Methods) were further examined as there are several tetrahymenas that are predatory macrostomes; with one exception, all such isolates had the typical *T. vorax* shape, and cox1 sequences showed that they were *T. vorax*. Estimates place *T. thermophila* just below *T. glochidiophila* and *T. vorax* just below *T. pyriformis* in relative abundance. The relative abundance of previously named *Tetrahymena* species in the present collection is compared to North American entries in the NS database in Table 5, again with the exception of *T. thermophila* and

| Species             | Relative abundance (%) |
|---------------------|------------------------|
| *T. americanis*     | 16.5                   |
| *T. elliotti*       | 13.4                   |
| *T. gruchyi* (nsp7) | 13.1                   |
| *T. borealis*       | 9.6                    |
| *T. kentuckyensis*  | 6.2                    |
| *T. glochidiophila* (nsp10) | 4.5              |
| *T. hegewischi*     | 3.9                    |
| *T. canadensis*     | 3.4                    |
| *T. pigmentosa*     | 2.9                    |
| *T. alphathermophila* (nsp15) | 2.6               |
| *T. pyriformis*     | 2.0                    |
| *T. newhampshirens* (nsp18) | 2.0            |
| *T. daisyae* (nsp45) | 1.9                    |
| "cosnipnanson"      | 1.4                    |
| *T. tropicalis*     | 1.4                    |
| Others              | 15.3                   |

*T. thermophila* and *T. vorax* omitted, see text. Species named in this paper are shown in bold.
**Table 5.** Rank order abundance of species found in this survey compared to North American isolates in the NS database

| Species                  | This survey | NS NA<sup>a</sup> |
|--------------------------|-------------|-------------------|
| *T. americanis*          | 28.8%       | 20.7%             |
| *T. elliotti*            | 22.6%       | 14.4%             |
| *T. borealis*            | 16.7%       | 13.2%             |
| *T. pigmentosa*          | 7.3%        | 8.3%              |
| *T. helegischi*          | 6.2%        | 7.2%              |
| *T. canadensis*          | 5.9%        | 3.4%              |
| *T. pyriformis*          | 3.5%        | 5.2%              |
| *T. cos/nip/nan/son*     | 3.2%        | 5.5%              |
| *T. tropicalis*          | 2.4%        | 5.5%              |
| *T. mobilis*             | 1.6%        | Absent<sup>b</sup>|
| SSU/NI/RA/CO             | 1.1%        | Absent            |
| *T. australis*           | 0.3%        | 11.2%             |
| *T. furgasoni*           | 0.3%        | 0.6%              |
| *T. shanghaiensis*       | 0.3%        | 0.0%              |
| n                        | 372         | 348               |

*T. thermophila* and *T. vorax* omitted, see text.

<sup>a</sup>Nanney-Simon database, North America.

<sup>b</sup>Species/strain not recorded in database.

T. vorax. In both collections, the most frequent species was *T. americanis* followed by *T. elliotti* and *T. borealis*, collectively accounting for 68% of previously named species in the present collection and 48% of previously named species in the NS database; these three species account for ~40% of all isolates (Table 4). Rank order of other species is similar in both databases; the major exception is *T. australis* which was more frequent in the NS database (found frequently in FL and New England) than in the present collection (found once in NH). The third place rank of *T. gruchyi* (Table 4) is skewed by its abundance in KD’s southern collection in which it was the most frequent isolate. In FPD’s collection, *T. gruchyi* ranks between *T. kentuckynsis* n. sp. (nsp9) and *T. glochidia* in relative abundance. These results hint at geographical variation in species occurrence.

Of the pre-2017 named species found in this collection all except three were previously reported as present in North America (Simon et al. 2008). The new Nearctic species are: *T. shanghaiensis* previously known from Shanghai, China (Feng et al. 1988), here represented by a single amicronucleate isolate collected in OH; *T. mobilis* originally from sludge in Salzburg, Austria (Schiftner and Foissner 1998) here collected in KY, OH, and VA; *T. farahensis* previously collected from industrial wastewater in Pakistan (Zahid et al. 2014) here represented by two GA isolates. Representatives of strains designated bSSU/NI/RA/CO (n = 4, “borealis”) and SIN (n = 1, “australis”) were also found. Guppy parasites, NI and SIN were isolated in Singapore, and RA and CO were isolated in Israel (Leibowitz and Zilberg 2009). The new species *T. alphapoeilia* includes strain SIN. More detail is presented in Text S2; see also below for ecological preferences.

The richness of *Tetrahymena* species was observed both regionally (Table 6) and locally (Table 7). The heavily sampled ANF (209 hectares) yielded as many species as the lesser sampled (with more sites) broadly defined Midwest states. As few as 15 eastern PA sites yielded 10 species. Among 351 collecting sites (FPD), yielding one or more *Tetrahymena*-like isolates, ~42% yielded two or more species, and one site, the most sampled CRWP, yielded eight species (Table 7), including three new species. One or more of the new species in Tables 2 and 3 was found in 17/18 collecting sites yielding four or more isolates; among the 33 sites yielding three species, seven yielded no new species and one consisted entirely of them. Table 7 indicates that relatively few samples per site are needed to potentially yield both named and new species.

Water bodies were not sampled so as to allow reliable quantitative estimates of proportions of species in them. However, finding two or more species in bodies with relatively few samples suggests that species abundance in these instances may be similar. Some ponds, however, appear to have dominant species. For example, Lake Gaston on the NC/VA border was sampled nine times at four locations, and all six *Tetrahymena*-positive samples yielded only *T. gruchyi*. This species also was the sole species at many of KD’s collecting sites. For some ANF, western PA, eastern PA and New England ponds *T. thermophila* is the dominant species, present in the majority of *Tetrahymena*-positive water samples. This species has been found without fail since 1987 in CRWP, even though the pond has been incompletely drained for repairs at least twice; a more or less permanent beaver pond just downstream of the drainpipe likely is a refuge. The frequency of minority species in CRWP is small. Of 44 isolates which were either refractory to *T. thermophila* mating type testers or consumed the testers, nine (~20%) belonged to other species as determined molecularly; the others were (immature) *T. thermophila*. Since approximately 20% of all CRWP isolates were refractory to the mating testers, this sample suggests that for CRWP, about 4% of isolates belong to a species other than *T. thermophila*. *T. elliotti* was found twice, *T. vorax*, an easily noticed species because of its carnivory and shape,
Table 7. Collection sites yielding five or more species

| Collection site | Geo area | Type      | Samples | Named species | Tetrahymena new species | Total species |
|-----------------|----------|-----------|---------|---------------|-------------------------|--------------|
| CRWP            | ANF      | Pond      | 5,958   | 5             | 3                       | 8            |
| SG69-1          | WPA      | Pond      | 66      | 4             | 3                       | 7            |
| Houghton Lake   | MidW     | Lake      | 35      | 3             | 2                       | 6            |
| FS136 Eculvert  | ANF      | Stream    | 20      | 3             | 1                       | 5            |
| Little Brokenstraw | WPA     | Stream    | 21      | 2             | 2                       | 5            |
| SG69-6          | WPA      | Pond      | 65      | 1             | 3                       | 5            |

Samples from 1987 to 2013. GPS coordinates: CRWP 41.418N 78.865W; SG69-1 41.620N 79.918W; Houghton Lake 44.300N 84.725W; FS136Eculvert 41.457N 78.882W; Little Brokenstraw 41.843N 79.390W; SG69-6 41.640N 79.912W.

*Includes Dexiostoma and Glaucoma.

The best studied tetrahymena, *T. thermophila*, has restricted distribution (Zufall et al., 2013), low effective population size ($N_e$) (Katz et al., 2006; Zufall et al., 2013), and significant population structure ($F_{ST}$) (Zufall et al., 2013), all of which are consistent with what has been called “moderate endemicity”. Species in the present survey with sufficient cox1 sequences were examined for both $N_e$ and $F_{ST}$. $N_{st}$ values ($N_{st}$ proxy) are of the same order of magnitude as for *T. thermophila* or smaller (Table S4). There is a strong tendency for amicronucleate species to have the smaller values; the exceptional *T. alphathermophila* is an artifact of the fusion of two homogeneous populations (see Text S2). $F_{ST}$ values (Table 8) indicate population structure in 6/9 species. The exceptions are *T. borealis*, *T. canadensis*, and *T. glochidiophila*. The latter has a large number of haplotypes distributed throughout the collecting region (see Text S2). Surprisingly, globally distributed species such as *T. americanis* and *T. elliotti* also gave evidence of population structure. For more discussion of population structure along with appropriate maps and haplotype networks for specific species, see Text S2.

In addition to evidence of population structure in some species, there also appear to be ecological preferences. For instance, *T. thermophila* was typically found in smaller ponds and lakes, rarely in streams, whereas *T. canadensis* was collected more often from streams. *T. aquafluente* (nsp22) was found in streams, and of the seven *Tetrahymena* singletons, six were found in streams. *Dexiostoma* was isolated primarily from flowing water, as were many of the new species of *Glaucoma*. *T. alphapoeicila* previously reported as guppy parasite Sin (Leibowitz and Zilberg 2009) was here found in a stream, as were representatives of unnamed strains guppy parasites NI/RA/CO; strain bSSU previously found in soil (Brandl et al. 2005) was also found in a stream. These species may have diverse ecological preferences. Aquatic invertebrates, larval and adult, that have in the past yielded species of *Tetrahymena* (see list in Lynn and Doerder 2012) were not sampled here. Such invertebrates have been sampled even less than streams, which, in general, are under-sampled (see Discussion).

Table 8. $F_{ST}$ (AMOVA) analysis of population structure

| Species             | Among populations | Within populations |
|---------------------|-------------------|--------------------|
| *G. chattoni*       | 52.5**            | 47.5               |
| *T. americanis*     | 14.5**            | 85.5               |
| *T. borealis*       | 2.4               | 97.6               |
| *T. canadensis*     | 6.4               | 93.6               |
| *T. elliotti*       | 26.3**            | 73.7               |
| *T. hegewischi*     | 60.7**            | 39.3               |
| *T. gruchyi* (nsp7) | 12.3**            | 87.7               |
| *T. kentuckyensis*  | 14.2*             | 85.8               |
| *T. alphathermohila*| (nsp15)           | 100                |

$F_{ST}$ values were calculated by Arlequin V3.5. *p < 0.05, **p < 0.001.

**DISCUSSION**

**The number of species and biogeography**

The biogeography of *Tetrahymena* reviewed in 2008 (Simon et al. 2008) integrated the results of collecting by Nanney’s group with the earlier collecting by Elliott and his students. Though both groups collected mostly in North America, between them they also sampled parts of South and Central America, Europe, Australia, Asia, and several Pacific islands. However, vast portions of these areas and Africa were not sampled, and in North America, most collecting was in the northeast quadrant of the United States, as in this study. The Simon et al. review added to the list of named species present in two or more biogeographic zones, and listed 12 as found only in North America. There are now over twice as many described species, more species now have been found in two or more biogeographic zones, and several species, including ones with global distribution, give evidence of population structure predicted by endemicity.

The finding of 32 new *Tetrahymena* species in the FPD/ KD collection is in accord with the prediction (see Introduction) that there are many more species and is consistent with the apparent ease which past investigators (e.g.,...
Batson 1983, 1985; Cho 1971b; Elliott 1973; Gruchy 1955; Jerome et al. 1996; Lynn et al. 2000; Nyberg 1981b; Pitsch et al. 2016; Quintela-Alonso et al. 2013; Simon et al. 1985; Zahid et al. 2014) discovered new species. Not only does the finding of additional new species in each successive collecting season of this study suggest that the sampled area has more species yet to be discovered, it follows that collecting on a larger, global, scale would identify dozens, more likely hundreds, of additional species, especially if collecting includes hosts for parasitic forms and moist soils (see below). In both the present collection and NS database, *T. americanus*, *T. elliotti*, and *T. borealis* were among the most abundant species, suggesting rather stable populations. By contrast, given its abundance (third rank) and range, it is perhaps surprising that new species *T. gruchyi* was not discovered in previous collections, particularly because it gives evidence of endemism.

The large number of *Tetrahymena* species and the implication of the same for *Dexistoma* and *Glaucoma* are consistent with observations on other ciliates. The morphospecies Paramecium “aurelia” contains 15 cryptic species, and the genus itself contains many more named species, some of which also consist of cryptic species, for example, *P. putrinum* (Tarcz et al. 2014) and *P. bursaria* (Zagata et al. 2015). There are similar examples among the numerous less well studied ciliates (Bastos Gomes et al. 2017; Chen et al. 2013; Chi et al. 2017; Fernandes and Neto 2013; Gao et al. 2012; Gentekaki and Lynn 2010; Katz et al. 2011; McManus et al. 2010; Przyboś and Tarcz 2016; Santoferrara et al. 2015; Simon et al. 2008; Tarcz et al. 2012, 2014). These occurrences suggest that most (all?) ciliate morphospecies consists of several to many cryptic species. Curiously, despite extensive collecting (e.g., and references therein: Przybos and Surmacz 2010; Przybos and Rautian 2012; Przybos et al. 2013, 2014; Przybos and Prajar 2015; Rautian et al. 2015) no new species have recently been added to the *P. “aurelia“* complex. This study suggests that *Tetrahymena* may be exceptionally species rich, even if limited to micronucleate (sexual) forms.

The amicronucleate condition is rare in ciliates but is particularly common in *Tetrahymena*, constituting about 25% of isolates (Doerder 2014). In most instances, cox1 haplotypes of amicronucleates are either identical to or sufficiently similar to those of micronucleate isolates to indicate recent origin (see Text S2). In other instances, however, there are no close haplotypes, for example, *T. thermophila* and its relatives *T. alphathermophila* and *T. betathermophila*, or a species is represented solely by amicronucleate isolates. For reasons discussed by Doerder (2014), some of these could be ancient, their micronucleate progenitors having become extinct. *T. pyriformis*, for example, is on an SSuRNA branch (Fig. 2) that includes many amicronucleates, and neither the NS database nor this study reports its micronucleate counterpart. Similarly, amicronucleate *T. aquasubterranea* (Quintela-Alonso et al. 2013) from South Africa, though most closely related to micronucleate *T. ehrithsii* n. sp. (6.3% cox1 difference), places in a clade that also contains many amicronucleates. Further collecting, especially globally, could possibly locate close micronucleate relatives of these and other amicronucleates.

This survey extended the biogeographic range of three previously named species into the Nearctic: *T. shanghaiensis* first reported from China (Feng et al. 1988), *T. mobilis* first reported from Austria (Schiftner and Foissner 1998), and *T. farahensis*, first reported from Pakistan (Zahid et al. 2014). It also extended the habitat and range of guppy parasites found in Singapore (NI, SIN) and Israel (RA, CO) (Leibowitz and Zilberg 2009). All of the above were rare, isolated just 1–4 times. As this paper was being prepared, the new species *T. gruchyi* with evidence of endemism was found on the Pacific island of Guam (R. Zufall, personal communication). Whether other new species described here have ranges that extend beyond North America can only be determined by more extensive collecting.

The distribution of *Tetrahymena* species is relevant to the question of whether microorganism species are cosmopolitan or endemic (see Introduction). The former are found in multiple biogeographic regions, whereas the latter have limited, local distribution with little or no migration. Endemism predicts significant population structure whereas global distribution maintained by migration among large populations does not. While many *Tetrahymena* species are found in 2–4 regions and those above were added by this study, the matter is more nuanced. For instance, global species *T. americanus* and *T. elliotti* gave evidence of (North American) population structure, while *T. canadensis* did not (see below for haplotype distribution). Similar conflicting results were obtained with species found (so far) only in North America. *T. thermophila*, a species presently confined to eastern United States (Zufall et al. 2013), gives evidence of moderate endemism having limited migration from New England and significant population structure. The present *F*ST and *N*e results suggest similar endemism for *G. chattoni*, *T. hegewischii*, *T. gruchyi*, and *T. kentuckyensis*. Other new species with insufficient sample sizes also may be endemic. However, *T. borealis*, confined to North America in the NS database does not have significant population structure. In this context, the issue of population size is relevant to the question of geographical distribution as it affects the ability to disperse. There are many variables to consider, such as seasonal fluctuation in population density and, in most species, the absence of cysts. Based on collection volumes, the average density of *T. thermophila* is 6–7 cells/liter in summer months when sampling efficiency is highest, a seemingly low density. Density is highest in the benthos, lowest in open water free of vegetation (Doerder and Brunk 2012). Effective population sizes (Table S4) of *Tetrahymena* species are similar to that of *T. thermophila*, and to the extent that *N*e is a proxy for census size, results suggest that population density is unlikely to contribute to dispersal. Both broadly distributed *T. americanus* and *T. elliotti* have *N*e proxies similar to endemism candidates listed above. Those *Tetrahymena*
species have small population sizes that mitigate against ready dispersal should, however, be considered provisional. The present analysis does not include isolates from throughout the range of globally dispersed species. *Paramecium* species with global distributions are reported to have large effective population sizes (Smoke et al. 2006). In sum, what the present results indicate, more than any resolution of the question of cosmopolitanism versus endemism, is the need for extensive, global sampling. In this respect, *Tetrahymena* and related species are behind *Paramecium* (Przybos and Surmacz 2010).

Dispersal of cystless species has always raised questions as to mechanism, especially when long distances are concerned. For instance, what accounts for the occurrence of the same haplotype of *T. canadensis* in CO and Malaysia, a distance of over 10,000 km? Or, what is the explanation for the observation that *T. thermophila* and its closest relative, *T. malaccensis*, have so far only been found in separate life zones on opposite sides of the globe? Over shorter distances and perhaps timescales, there is the question as to how *T. thermophila* present in natural ponds in New England and eastern PA came to be present in artificial ponds further west in the ANF and western PA or further south in FL. These questions likely have different answers, but a facile answer is under sampling; despite new data presented here, the biogeography of *Tetrahymena* is woefully incomplete. Evidence of population structure suggests dispersal is rare, and but population intrafertility, though rarely studied (Nyberg 1981a), suggests it is sufficiently common to maintain species integrity, as also observed for widely dispersed and intrafertile species of *Paramecium* (Przybos and Surmacz 2010). As for dispersal itself, if some species are indeed tens of millions of years old (Doerder 2014; Wright and Lynn 1997), continental drift may be important. On a more recent time scale, dispersal may have been facilitated by an abundance of wetlands following glaciation. Species such as *T. glochidiophila* might be dispersed when infected glochidia are attached to fish gills; the ranges of *T. glochidiophila* and its host are roughly congruent (Lynn et al. 2017). Other natural means of dispersal might include storm-borne water droplets or facultative parasitism in which infected hosts are carried long distances by storms. In the case of *T. thermophila*, however, the apparent migration west and south from New England is against prevailing weather patterns. In this case and perhaps others, human activity should be considered. People have transported water across long distances in ships either as drinking water or as ballast, a possible explanation for *T. gruchyi* appearing in Guam. On land, ciliates may have occupied water barrels of wagon trains or been carried in water tanks of the early steam railroads. The close connection between New England and ANF populations of both *T. thermophila* and *T. vorax* might be due to the spread of steam driven logging railroads in the late 19th and early 20th centuries. Fish stocking might also have contributed to the spread of some species. It could be interesting to look at *Tetrahymena* distribution in more anciently settled areas of Africa, Europe, and Asia.

**Phylogeny**

*Tetrahymena* phylogeny as historically based on morphological, life history, and ecological traits (reviewed by Lynn and Doerder 2012) has long been abandoned in favor of phylogeny based molecules (Brunk et al. 1990; Chantangsi and Lynn 2008; Preparata et al. 1989; Sadler and Brunk 1992). Rather than various species complexes (“pyriformis” for bacterivores, “rostrata” for parasites, “patula” for macrostomes), molecules clearly separate *Tetrahymena* into “australis” and “borealis” clades of mostly cryptic species with each clade including bacterivores, parasites, and macrostomes. The SSUrRNA phylogeny (Fig. 2) with double the number of species of previous studies also supports these clades. Similarly, the mtSSUrRNA tree (Fig. 3), though consisting of fewer species clearly define an “australis” clade and suggest that *T. eicheli* and *T. williamsi* are its closest “borealis” relatives; low mtSSUrRNA interspecific variability in the “australis” clade suggests more recent radiation. The SSUrRNA and mtSSUrRNA trees also identify a third, smaller “paravorax” clade of more distantly related species. *Tetrahymena paravorax* was an outlier in studies cited above and has the largest cox1 difference compared to other *Tetrahymena* species. Although this suggests that the “paravorax” clade should be reexamined for possible elevation to the generic level, morphologically both *T. paravorax* and *T. glochidiophila* resemble other *Tetrahymena* species (Lynn et al. 2017). The mtSSUrRNA sequences obtained here also corroborate nuclear SSUrRNA phylogeny that clearly separates *Tetrahymena, Dexiostoma, Colpidium* and *Glaucoma* into separate clades, indicating that nuclear and mitochondrial genomes are not on separate trajectories as suggested by some cox1 trees that place *Dexiostoma/Colpidium* within *Tetrahymena*.

The present results indicate that mtSSUrRNA sequences, even with their polymorphisms, may be useful in diagnosing most *Tetrahymena, Dexiostoma*, and *Glaucoma* species, including some with identical nuclear SSUrRNA. However, a more extensive survey with the complete mtSSUrRNA sequence is needed to properly assess its utility. The intraspecific polymorphisms may be of use in analyzing migration and population structure.

The SSUrRNA phylogeny (Fig. 2) is not congruent with phylogeny of the SSI (Figure S4B). Absent in *Dexiostoma, Glaucoma* and the “paravorax” clade, as concluded by Sogin et al. (1986) SSIs appear to be acquired subsequent to the split between the “borealis” and “australis” clades, with multiple instances of independent acquisition. Though likely acquired from bacterial food sources, no similar bacterial Group I intron is detected in BLAST searches of GenBank; the closest related intron is that of a slime mold. No intraspecific sequence variation was observed, and no SSI gave evidence of retaining the homing endonuclease gene, suggesting that SSIs sweep through a species upon acquisition. Mixed SSI+/SSI− species are likely in the process of losing, rather than acquiring, the intron (Text S3). Sequences of SSIs from globally collected
Tetrahymena may help resolve their origin. SSI sequences may in some instances be diagnostic of species. Another reason for further SSI study is their possible relevance to the question of Tetrahymena speciation. Two main drivers of speciation are partitioning of food sources and avoidance of predators. As envisioned by NaNney (1999), the evolution of the basic tetrahymena form was followed by species radiation facilitated by an abundance of bacterial food. Unfortunately, nothing is known about food preferences in wild populations; in the lab, most species can be successfully maintained on Klebsiella pneumoniae, though the exceptions suggest that specialization exists. Another hint that there may be differences in Tetrahymena food preferences is if the different SSIs were acquired from different, albeit related, bacterial food sources. A test of this hypothesis requires the rather unlikely finding, particularly if millions of years are involved, of similar bacterial SSIs. In this context, it would be useful to know whether Tetrahymena species indeed have food preferences, as such preferences might also help explain species distribution.

Parasitism and the edaphic environment

Insect hosts of Tetrahymena species include larvae and sometimes adults of black flies (Simulidae), mosquitoes (Culicidae), midges (Chironomidae), and alder flies (Sialidae); other hosts include slugs (Limacidae), glochidia (Unionidae), guppies (Poeciliidae), and (!) dogs (Canidae) (for references see Lynn and Doerder 2012). Additional hosts infected with tetrahymena-like ciliates are reported in the older literature (for summary see Corliss 1979). More recent host examples include planaria (Planariidae) (Wright 1981) and amphibians (Leiopelmatida) (Shaw et al. 2011), and doubtless there are many others. Infection rates typically are low, usually just a few percent or less. An unknown species of Tetrahymena was found in dead mosquito larvae (Takahashi et al. 2004), and I have found Tetrahymena in casts of dragonfly larvae and in remains of dead invertebrates and fish, perhaps attracted by bacteria. From the edaphic environment, Tetrahymena species have been isolated from mosses (Kahl 1926; Kozloff 1948, 1957; Roque et al. 1970) and soil (Brandt et al. 2005; Foissner 1987). Were this survey to have included aquatic invertebrates, mosses, and soils more species likely would have been discovered, as predicted (see Introduction). There is a long history of sampling from these sources, but relatively little study in most cases beyond initial species descriptions. For parasites, available evidence suggests that there is host species specificity in some instances (Batson 1983, 1985; Jerome et al. 1996; Lynn et al. 1981), but other species are known to infect multiple host species (Van As and Basson 2004). Conversely, there is evidence that some hosts may be infected by multiple Tetrahymena species, for example, guppies infected by T. alphapoeiica (SIN), strain NI and possibly T. corlissi (based on morphology Hoffman et al. 1975).

There are currently 12 Tetrahymena species listed as parasites (aka histophages) (Lynn and Doerder 2012; Lynn et al. 2017) and characterized as either facultative or obligate, the latter meaning that free-living cultures could not be established. Though some may disperse and be acquired as cysts, most others (because cysts were not observed) likely disperse and are acquired by their hosts as free-living forms. For instance, T. rostrata and T. lilimus, better known as gastropod parasites, have been found in moss where they could be easily acquired by feeding snails. Transmission of other parasitic tetrahymenas is less obvious, but free-living forms of T. farleyi (dog parasite) have been reported (Quintela-Alonso et al. 2013), and this study found free-living forms of the guppy parasites NI/SIN and mussel parasite T. glochidiophila. Similarly, this study found aquatic forms of the soil strain bSSU.

The parasitic species of Tetrahymena do not cluster on SSUrRNA phylogenetic trees, and the placement of histophag T. glochidiophila in the “paravorax” clade (see below) strengthens the suggestion that histophagy/parasitism is an ancestral trait (Lynn et al. 2017; Struder-Kypke et al. 2001). As with many parasites and pointed out by Struder-Kypke et al. (2001), tetrahymenas have complex nutritional requirements, requiring 10 amino acids, sterols, and two nitrogenous bases, among others (Hill 1972), consistent with their having originated as parasites in which nutrient rich environments such as hemocoels or renal glands made metabolic losses tolerable. Tetrahymenas commonly regarded as free-living bacterivores are capable of histophagy, evidenced by their successful long-term storage with axenic rat gut (Williams et al. 1980), and laboratory infection of invertebrate hosts using standard laboratory strains has been known for some time (e.g., Grassmick and Rowley 1973; Thompson 1958). The presence of variant surface antigens, with structure similar to variant surface antigens of other protist parasites (Simon and Schmidt 2007), might also indicate an ancient disposition toward parasitism. In short, using barcodes to unambiguously identify tetrahymenas, many ecological and life-cycle hypotheses can now be tested.

Finally, a deliberate search for parasites likely would find many more parasitic species but also might find parasitic forms of species currently known only as free-living. With barcodes, isolates can now be unambiguously identified even if the parasitic and free-living forms are morphologically different as was found, for example, in T. dimorpha (Batson 1983). It is unlikely, however, that named parasitic species without barcodes can be unambiguously rediscovered. Several of these are reported only once, decades ago, and may not be recoverable from original collecting sites, and even then species identity is not certain.

Criteria for diagnosing species

Species are artificial constructs with no single all-encompassing definition. Species diagnosis by DNA sequences is increasingly common and is especially useful for cryptic species in which, by definition, morphology fails to distinguish among them (Jöger and Schrödl 2013; Renner
Sonneborn (1957) long ago proposed that obligate inbreeders (e.g., selfers) and asexuals, neither of which satisfies biological species criteria, could be declared species based on some minimum threshold of (genetic) divergence. Modern experts have supported this suggestion (Schlegel and Meisterfeld 2003), and others, noting the difficulties associated with protist alpha-taxonomy, recommend that species definitions be made “on a case-by-case basis” (Boenigk et al. 2012). Abandoning morphology in favor of genetics is not, however, so straightforward. For instance, in a recent paper on best practices, Warren et al. (2017) place considerable emphasis on morphology as part of a species description. While acknowledging that cryptic species present special challenges, they nevertheless encourage (though explicitly do not require) investigators to search for new morphological traits suitable for multivariate analysis as a way of distinguishing among species. While this might reveal interesting evolutionary detail, the analysis involved is beyond the expertise (and perhaps resources) of most investigators and is clearly impractical for field studies such as this one. A more detailed response to Warren et al. and references to ICZN rules with respect to molecular description appear in the paper formally naming *T. glochidiophila* (Lynn et al. 2017). It is, however, perhaps worthwhile in the present context to cite two examples in which expert morphological description has failed. The first concerns the naming of the micronucleate *T. aquasubterranea* (Quintela-Alonso et al. 2013). Despite elegant, detailed (light, electron) microscopic description these authors did not find distinguishing morphological characters, though to be fair, few such detailed descriptions exist for *Tetrahymena* species. Other traits such as its “biphasic life cycle” consisting of feeding trophonts and starvation-induced theronts are found in *T. mobilis* (Quintela-Alonso et al. 2013), and forms similar to theronts have been seen in other species (Nelsen 1978; Nelsen and DeBault 1978). Moreover, the description of highly starved theronts as “very small, broadly lenticular or ovate lying motionless on the bottom” (Quintela-Alonso et al. 2013) applies to many wild isolates observed in this collection; indeed, the phenomenon was sufficiently common that instances were not recorded. Ultimately, the naming of *T. aquasubterranea* was, like other recent examples, “Based on the considerable genetic distance…” of its SSUrRNA and cox1 sequences (Quintela-Alonso et al. 2013). The second example concerns a discrepancy regarding *G. chattoni* discovered in this study. In this case, the original strain and the independently identified reference strain appear to be different species. As detailed in Text S2, the D2LSUrRNA sequence (Nanney et al. 1998; Preparata et al. 1989) of the original strain described by Corliss (Corliss 1959) differs from that of present isolates identified as *G. chattoni* by cox1 similarity to the designated type strain (Chantangsi et al. 2007). The latter was identified as *G. chattoni* by an expert, and assuming no error or strain mix-up it likely is a separate, cryptic species, possibly (and unknowably) one of the syngens of *G. chattoni* morphospecies studied by Cho (1971a). Since the original Corliss isolate exists in ATCC, its D2LSUrRNA, cox1, and SSUrRNA sequences should be determined.

In lieu of morphology, this paper names new species of *Tetrahymena, Dexiostoma, and Glaucoma* using as species diagnostics cox1 barcodes and SSUrRNA sequences as supplemented by SS1 and mtSSUrRNA sequences. As the equivalent of unique morphological traits (Lynn et al. 2017), these sequences permit the unambiguous identification (with the few exceptions that have not been adequately barcoded) of isolates independent of actual morphology or sexual status. The sequences can be obtained from isolates that are difficult to culture in sufficient quantities for morphological and breeding studies, and they are currently the only practical way to identify isolates in field studies where identification restricted to morphospecies severely underestimates biodiversity. Although obtaining barcodes requires some molecular expertise, they are relatively easily obtained by standard PCR and sequencing techniques; indeed, with support, the procedures are suitable for high school students ([https://tetrahymenaasset.vet.cornell.edu/science-modules/](https://tetrahymenaasset.vet.cornell.edu/science-modules/)). Sequence data have the additional advantage of providing insight into evolutionary relationships and population biology. Perhaps most importantly, sequences are archivable in public databases, useful in the absence of living or preserved material. Based on these arguments, it is unfortunate that in applying barcoding to *Paramaecium* (Krenek et al. 2015) the authors refrained from formally naming three cryptic species, although both SSUrRNA and cox1 sequences differences clearly indicated their status as separate species. Rather, Krenek et al. assigned provisional names with the prefix “Eucandidatus” to indicate their tentative status as species, a solution that compromises their scientific utility. Przyboś and Tarcz (2016) recently named three morphologically indistinguishable species of “P. jenningsi” based both on molecular divergence and mating incompatibility.

### New species of *Tetrahymena, Dexiostoma, and Glaucoma*

Forty-seven new species are described in Table 9 (*Tetrahymena*) and Table 10 (*Dexiostoma* and *Glaucoma*). The same information in conventional paragraph form is presented in Supplemental Information Text S4. The 48th new species found in this survey is *T. glochidiophila* reported by Doerder (2014) as unnamed species nsp10; it was formally named in a separate publication (Lynn et al. 2017).

Type strains (holotypes) were designated based on combination of one or more of the following (i) successful accession of live cells and/or DNA at the Tetrahymena Stock Center at Cornell University (Ithaca, NY); (ii) successful cox1 sequence; (iii) availability of SSUrRNA sequence; (iv) presence of a micronucleus; (v) presence of the SS1. For some type specimens, SSUrRNA sequence was not available in which case Tables 9 and 10 indicate the GenBank accession number of a secondary type so indicated by parentheses. Unfortunately, some isolates,
Table 9. Description and diagnostic characters of new *Tetrahymena* species

| NSP No. | Proposed name and ZooBank ID | GenBank | mSSU/URRNA | Designated type strain properties |
|---------|-----------------------------|---------|-------------|----------------------------------|
| nsp01   | *Tetrahymena* eriensis n. sp. | E2E8FBB920C1419F-EDE6B9F4A4D7FB1B | + – | The name refers to Lake Erie, acknowledging that this is the first *Tetrahymena* species isolated from that water body |
| nsp07   | *Tetrahymena* gruchyi n. sp. | CDA2D75F-6F2C-439F-8AC2-C45A421EBF8 | +/- +/- | The name honors Daryl Gruchy, early collector of breeding *Tetrahymena* species |
| nsp09   | *Tetrahymena* kentuckyensis n. sp. | 7059907A-D6D3-4F91-9627-63E649D0D8B8 | +/- +/- | The name refers to the frequent collection of this species at sites in Kentucky |
| nsp14   | *Tetrahymena* newhampshirensis n. sp. | 5FE17DA1-BF55-4A8B-8A7A-11008A4A-D44B-4F91-8710-7ED2E5A8-F036 | +/- | The name emphasizes a close relationship to *T. thermophila* |
| nsp16   | *Tetrahymena* williamsi n. sp. | 3797EB3E-720B-4E83-854C-84A68B7C2671 | +/- +/- | The name honors Norman Williams for numerous contributions to *Tetrahymena* biology and who was the first to note the "austrials" and "borealis" clades |
| nsp22   | *Tetrahymena* aquatilis n. sp. | DDA7AC0D-6D0C-4EAD-8D0A-3E7F98E52A4F | +/- | The name derives from Latin for flowing water. All isolates were isolated from larger streams |
| nsp23   | *Tetrahymena* caloereri n. sp. | 535272C2-9D0A-4D01-9A4A-8112446B81B6 | +/- | The type strain was isolated from Cole Run Pond, a pond containing several *Tetrahymena* species |
| nsp25   | *Tetrahymena* taylorii n. sp. | 18B72255-723B-4488-8881-A8E65648B420 | +/- | The name emphasizes a close relationship to *T. thermophila* |
| nsp30   | *Tetrahymena* mcdonaldae n. sp. | 69964204-A41B-4C27-AE30-656A615D80A | +/- | The name honors Barbara McDonald who showed that *Tetrahymena* exchange cytoplasm at conjugation |

(continued)
### Table 9. (continued)

| NSP No. | Proposed name and zoonBank ID | mc | SSi | SSi GenBank | msSUrRNA | GenBank | Etyymology | Designated type strain properties |
|---------|-------------------------------|----|-----|-------------|----------|---------|------------|-----------------------------------|
|         |                               |    |     |             |          |         |            | Isolate number   | Cox1 GenBank | SSU GenBank | TSC Accession | TSC DNA | Water Source | State | Latitude  | Longitude  |
| nsp31   | Tetrahymena syv. n. sp.        |    |     |             |          |         |            | K Y217814# | JKO28642 | JKO28659 | SDO0315 (1) | (1)     | GrandRiverMI | MI     | 43.040    | -86.065    |
| nsp35   | Tetrahymena engbergi n. sp.   |    |     |             |          |         |            | K Y217820 | JKO28519 | (0)     | CRWP     | PA          | 41.418   | -78.880    |
| nsp40   | Tetrahymena afghanensis n. sp.|    |     |             |          |         |            | 21121-1   | JKO28599 | JKO28482 | SDO03055 (0) | (0)     | RedLiRn     | PA      | 41.485    | -78.880    |
| nsp41   | Tetrahymena connorii n. sp.   |    |     |             |          |         |            | FLO51-15  | JS319264 | JKO28506 | ATCC30143 | (-2)    | South Haven | MI      | 43.637    | -85.166    |
| nsp44   | Tetrahymena salt/ourius n. sp.|    |     |             |          |         |            | SC141-3   | JKO28322 | (0)     | CampArrowhead pond | FL | 30.637 | -85.166    |
| nsp45   | Tetrahymena desyrmare n. sp.  |    |     |             |          |         |            | MLD4u     | EF070302 | ATCC30562 | (0)     | Lake Shelbyville | IL | 34.968 | -81.951    |
| nsp51   | Tetrahymena scolopax n. sp.   |    |     |             |          |         |            | 19903-1   | JKO28669 | JKO28504 | SDO01709 (0) | (0) | WoodcockCreek | PA | 41.709 | -80.145    |
| nsp52   | Tetrahymena pennan/kariensis n.sp. |    |     |             |          |         |            | 19518-2   | JKO28745 | JKO28516 | (0)     | SpringCo/Jriving | PA | 41.516 | -78.994    |
| nsp53   | Tetrahymena zeutheni n. sp.   |    |     |             |          |         |            | 18808-1   | JKO28681 | JKO28508 | (0)     | SpringCo/Jriving | PA | 41.516 | -78.994    |
| nsp55   | Tetrahymena defmanni n. sp.   |    |     |             |          |         |            | 19604-3   | JKO28701 | JKO28514 | SDO01689 (0) | (0) | ItalianShantyRun | PA | 41.454 | -78.848    |
| nsp56   | Tetrahymena frigidaformis n. sp.|    |     |             |          |         |            | 20149-1   | JKO28647 | JKO28501 | (0)     | BobbsCr | PA | 41.593 | -79.217    |
| nsp57   | Tetrahymena passcottii n. sp. |    |     |             |          |         |            | 20152-2   | JKO28645 | JKO28500 | SDO03013 (0) | (0) | BeaverRun2 | PA | 41.589 | -79.361    |
| nsp58   | Tetrahymena aemilia n. sp.    |    |     |             |          |         |            | 20760-1   | JKO28696 | JKO28487 | (0)     | LakeEmily | MI | 46.112 | -88.495    |
| nsp59   | Tetrahymena eicheli n. sp.    |    |     |             |          |         |            | 20811-1   | JKO28603 | JKO28486 | SDO03049 (0) | (0) | ManiqureLake | MI | 46.215 | -85.807    |

(continued)
| NSP No. | Proposed name and ZooBank ID | GenBank | mSSU rRNA GenBank | Etyymology | Combined Isolate number | Cox1 GenBank | SSU GenBank | TSC Accession | TSC DNA Source | Water Source | State | Latitude | Longitude |
|---------|-----------------------------|---------|-------------------|------------|----------------------|-------------|-------------|---------------|---------------|--------------|------|----------|----------|
| nsp60  | Tetrahymena poconoensis n. sp. | 72F04B28-8E48-4D78-826B-70FB79D2CE1F | KY217826# | The name refers to the Poconos, the region in which the type strain was collected | 20984-1 | KJ029572 | KJ028484 | SD030332 (0) | (0) | FairviewLake | PA | 41.395 | -75.193 |
| SIN    | Tetrahymena alphacoelica n. sp. | 7A837B00-F23C-4F1C-83EC-78BD9DA25BDD | KY217828# | KY218597(1) | 21101-1 | KY273141 | (0) | (0) | TamarackLake3 | PA | 41.417 | -80.130 |
|        | Unnamed strains deposited in ATCC by David Nanney and Ellen Simon |         |       |                  |            |               |         |               |               |              |        |         |          |
| nsp61  | Tetrahymena saltuflumine n. sp. | 8D136250-B097-4422-9B48-2491C968281B |         | The Latin words saltu and flumine refer to the jungle stream in which this species was first found | MP51 | GU439287 | ATCC50402 | Jungle stream, Malaysia | Unknown | Unknown |      |          |
| nsp62  | Tetrahymena allanae n. sp. | 28F9C11E-F7DA-46B6-A1CE-2DC14074C4F | +        | The name honors Sally Allen for her many contributions to Tetrahymena genetics; this name was proposed by David Nanney | WK5 | GU439251 | ATCC50403 | West of Muskegon, MI | MI | Unknown |      |          |
| nsp63  | Tetrahymena simonae n. sp. | C80ADD48-1F60-4EB5-BEEA-F045BD81C03B | +        | The name honors Ellen Simon for developing methods for cryopreservation and collecting of Tetrahymena; the name was proposed by David Nanney | T1 | GU439285 | ATCC205048 | River Falls, Inthanout, Thailand | Unknown | Unknown |      |          |
| nsp64  | Tetrahymena colurium n. sp. | F3F98A8D-2B92-4832-8527-92A03A0D40C | +        | The Latin word colurium refers to ancient Roman fort near the collecting site. | UK 20i | GU439291 | ATCC205049 | Roman Fort, Chester, England | Unknown | Unknown |      |          |
| nsp65  | Tetrahymena terraevirentia n. sp. | 899170D1-3418-4C0E-AAC0-272303B4396 | +        | The name means "green earth" and refers to Greenland from which this species was isolated | GR 90e | GU439274 | ATCC205079 | South of Godhaven, Greenland | Unknown | Unknown |      |          |

SSI GenBank, # indicates unique SSI sequence; mtSSU rRNA GenBank, values in parentheses () indicate number of sequences deposited; SSUrRNA GenBank, () indicate accession number of a secondary type; TSC Accession, () indicate number of cell lines deposited at Tetrahymena Stock Center; TSC DNA, + indicates DNA of type strain deposited; () indicates number of additional samples; Water Source, official or assigned name, spelling as in GenBank.
### Table 10. Description and diagnostic characters of new species of Dexiostoma and Glaucoma

| NSP No. | Proposed name and Zoobank ID | mtSSUrRNA GenBank | Isolate number | Cox1 GenBank | SSU GenBank | TSC DNA | Water Source | Location |
|---------|-----------------------------|-------------------|----------------|--------------|-------------|---------|--------------|----------|
| Dexiostoma maupasi n. sp. | + The name honors Emile Maupas who described many ciliates | KY218545(1) | KY218028 | + (2) BobbsCr | PA (41.593, –79.217) |
| Dexiostoma saccharum n. sp. | + The name refers to Sugar Lake from which the type strain was isolated | KY218546(2) | KY218032 | KY218610 | + (1) SugarLake | PA (41.564, –79.946) |
| Dexiostoma venango n. sp. | + The name refers to a river near the collecting site. | KY218548(2) | KY218033 | + (0) LakePleasant | PA (42.003, –79.895) |
| Dexiostoma laceyvillensis n. sp. | + The name refers to a town that was submerged when the dam forming Tappan Lake was built | KY218549(2) | KY218034 | KY218611 | + (1) TappanLake3 | OH (40.359, –81.216) |
| Dexiostoma sabulum n. sp. | + The Latin word sabulum (gravel) shares the root name of the AuSable River which in the French means “at the sand” | KY218550(2) | KY218035 | KY218612 | + (1) AuSableREBr | MI (44.789, –84.596) |
| Dexiostoma novaeangliae n. sp. | + The name refers to the New England states, which includes New Hampshire where this species was collected | KY218551(2) | KY218036 | KY218613 | + (1) MudPond | NH (42.899, –72.017) |
| Dexiostoma tobeco n. sp. | + The name is derived from the Native American word “Tobeco” for Clarion River from which the type strain was collected | KY218552(2) | KY218037 | KY218614 | + (1) ClarionRiverWBr | PA (41.575, –78.693) |

(continued)
| Species properties | Designated type strain properties |
|-------------------|----------------------------------|
| Proposed name and | Isolate number | Cox1 GenBank | SSU GenBank | TSC DNA | Water Source | Location |
| ZooBank ID | Mic | SSI | mtSSUrRNA GenBank | Etymology |
|----------------|-------|------|-----------------|-----------|
| **Glaucoma** | | | | |
| Gns01 | Glaucoma michiganensis n. sp. | + | − | KY218101 | | KY218625 | ThunderBayAlpena | MI | (45.068, −83.449) |
| | 54D08B9A-50B2-4D67-AF10-CFA325F18B3C | | | | | | |
| | Gns02 | Glaucoma gramina n. sp. | + | − | KJ026553 | | KY218623 | BarrenRiver1 | KY | (37.031, −86.445) |
| | 09A66B70-54D3-4AC7-B34D-7FDBD0D91FA4 | | | | | | |
| | Gns03 | | + | − | KY218563(2) | | | ColoradoRiver1 | UT | (39.087, −109.102) |
| | Gns04 | | + | − | KJ026833 | | KY218618 | CraigPond | ME | (44.588, −68.678) |
| | Gns05 | | + | − | KJ026741 | | KY218612 | FrenchCreek | PA | (41.903, −79.854) |
| | Gns06 | | + | − | KJ026895 | | KY218613 | TionestaCr | PA | (41.617, −79.119) |
| | Gns07 | | + | − | KJ026822 | | KY218620 | ThompsonBrook | NH | (43.069, −72.334) |
| | Gns08 | | + | − | KJ026825 | | KY218619 | WarrenBrook | NH | (43.122, −72.297) |
| | Gns09 | | + | − | KJ026834 | | KY218617 | CardMillStream1 | ME | (44.597, −68.191) |
| | Gns10 | | + | − | KJ026889 | | KY28512 | GilfoyleRun | PA | (41.530, −79.016) |
| | Gns11 | | + | − | KJ026892 | | KY28488 | HavenCreek | MI | (47.382, −88.028) |
| | Gns12 | | + | − | KJ026842 | | KY28542 | RoundLake | MN | (46.330, −93.817) |

mtSSUrRNA GenBank, () indicate number of sequences deposited; SSUrRNA GenBank, values in parentheses () indicate accession number of a secondary type; TSC Accession, () indicate number of cell lines deposited at Tetrahymena Stock Center; TSC DNA, + indicates DNA of type strain deposited; () indicates number of additional samples; Water Source, official or assigned name, spelling as in GenBank.
particularly not only *Dexiostoma* and *Glaucoma*, but also *Tetrahymena*, were lost because they were difficult or impossible to maintain in the laboratory. While they reproduced for a time in bacterized media, they resisted transfer to axenic medium (e.g., PPY) or grew slowly. Some fastidious isolates were successfully grown axenically in LP medium. Other investigators have also reported difficulty growing certain isolates under laboratory conditions (e.g., Corliss 1960; Lynn et al. 1981, 2000). A few isolates could not be successfully frozen in liquid nitrogen. The cox1 barcodes and SSUrRNA and other sequences provided here will permit investigators to unambiguously assign new isolates to these species. Though GPS coordinates of type strain collecting sites are provided, there is no guarantee that the species can be collected again from that water source. In addition, it should not be assumed that the type locality (e.g., stream or pond) is typical.

Of the *Tetrahymena* species, 32 are the result of the collecting reported here, and five are strains deposited at ATCC by David Nanney and Ellen Simon (see Acknowledgements). The cox1 and SSUrRNA sequences of the latter were reported by Kher et al. (2011). Only two species of *Glaucoma* (Table 10) are named here. The other 10 are not named for two related reasons. First, though all isolates belong to a well-supported clade (Fig. 2), this clade contains not only *Glaucoma* but also the genera *Glaucomides* and *Bromeliophrya*. Only Gns1p1 and Gns2p2 are considered congeneric with *G. chattoni* and *G. scintillans*. Based on genetic distance, the remaining 10 species likely constitute two new genera. Second, the Catalog of Life (catalogueoflife.org) lists 33 species of *Glaucoma*, many described decades ago. Although several names are clearly synonymous (e.g., variant spellings) and others are likely invalid, five of these species bear some resemblance to *G. bromelicola* (Foissner 2013). Until barcodes are provided for these species (if indeed any still exist in culture or are unambiguously identified among new isolates), naming of new *Glaucoma* species here risks misclassification and duplication. Unfortunately, isolates reported here did not survive to be archived as cryopreserved specimens; DNA samples, however, were deposited. The barcodes should allow any new isolate to be assigned to one of Gns1p1–Gns1p2, or, as may be more likely, yet another new species.

**ACKNOWLEDGMENTS**

I am especially grateful to Kristen Dimond (University of Houston) for permission to include the results of her southern collections in this paper. She is responsible for naming nsp44 and nsp45. Also contributing to this paper are former CSU students Shannon Valentine who determined SSI presence/absence for most of FPD’s isolates, Ryan Phillips who sequenced portions of the mitochondrial SSUrRNA, and Scott Fulton who sorted out the relationship between *T. americanus* and *T. hegewischi* and designated the type strain for the former. Renee Heberle meticulously transcribed old computer printouts to create the NS database. I also acknowledge the continued encouragement from Denis Lynn and Rebecca Zufall in pursuing this work. Christopher McGinnis provided advice on Latin names. The following individuals assisted in the field and/or laboratory or generously provided water samples: Hasan Amir, Ferdinand Apollonio, Muhttin Arslanyolu, James Bartram, Laurie Booth, Cal Borden, Chris Caprette, Ioan Chelbezan, Michael Churns, Daniel Coulton, Joe Deak, Nora Doerder, Lowell Doerder, Mary Doerder, Daryl Doerder, Cory Doerder, Evan Dvorin, Frank Eberhardt, Barb Evans, Scot Fulton, Michael Gates, Cari Gerber, Michelle Gresham, Ray Harvey, Tom Jones, Michelle Kaczmarec, John Kandray, Elizabeth Katoa, Kyle Koloini, Kevin Kunzelman, Greg LaCrosse, Vinnie LaSalvia, Daniel Laszlo, Peter Laszlo, James Little, Alex Lopez, Brandi Martin, Justin McCray, Manuel Mendoza, Christina Merwin, Barry Mita, Jennie Murphy, Ryan Phillips, Beth Popadak, Ari Daniel Shapiro, Steve Shook, Susan Slotnick, Josh Svedlow, Paul Sweeney, Jack Thatcher, Shannon Valentine, Ben Wolfe, Jason Young, Andrew Young, and Jennifer Zimmer. In particular, FPD remembers James Little with whom he first collected *Tetrahymena thermophila* in 1987; Michael Gates, Nora Doerder and Ben Wolfe assisted on several major collecting trips. Apologies are due to those who have been inadvertently omitted. This work was supported over the years by the CSU Graduate College, the BGES department and personal funds. This paper is dedicated to the memory of Dr. David L. Nanney who had an abiding interest in *Tetrahymena* natural history and evolution. It was Nanney’s intention (personal communication) to name two species from his collection in honor of long-time collaborators Dr. Sally Allen and Dr. Ellen Simon. Sadly, though his wish is fulfilled here, all three have passed before it could be realized.

**ZOOBANK**

This project was registered at ZooBank with the identification number urn:lsid:zoobank.org:pub:381EFA71-76F2-4D34-B85A-C03D8F701B0B. Species IDs are given in Table 9 and Text S4.

**LITERATURE CITED**

Bandelt, H.-J., Forster, P. & Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.*, 16:37–48.

Bastos Gomes, G., Jerry, D. R., Miller, T. L. & Hutson, K. S. 2017. Current status of parasitic ciliates *Chilodonella* spp. (Phyllopharyngea: *Chilodonellidae*) in freshwater fish aquaculture. *J. Fish Dis.*, 40:703–715.

Batson, B. S. 1983. *Tetrahymena dimorpha* sp nov. (Hymenostomatida: Tetrahymenidae), a new ciliate parasite of *Simulidae* (Diptera) with potential as a model for the study of ciliate morphygenesis. *Phil. Trans. R. Soc. Lond. B*, 301:345–363.

Batson, B. S. 1985. A paradigm for the study of insect-ciliate relationships: *Tetrahymena sialidosa* sp. nov. (Hymenostomatida: Tetrahymenidae), parasite of larval *Sialis lutaria* (Linn.) (Megalopectera: Sialidae). *Phil. Trans. R. Soc. Lond.*, 310:123–144.

Boenigk, J., Ereshefsky, M., Hoef-Emeden, K., Mallet, J. & Bass, D. 2012. Concepts in protistology: species definitions and boundaries. *Eur. J. Protistol.*, 48:96–102.
Brandl, M. T., Rosenthal, B. M., Haxo, A. F. & Berk, S. G. 2005. Enhanced survival of Salmonella enterica in vesicles released by a soilborne Tetrahymena species. Appl. Environ. Microbiol., 71:1562–1569.

Brunk, C. F., Kahn, R. W. & Sadler, L. A. 1990. Phylogenetic relationships among Tetrahymena species determined using the polymerase chain reaction. J. Mol. Evol., 30:290–297.

Caron, D. A. 2009. Past president’s address: protistan biogeography: why all the fuss? J. Eukaryot. Microbiol., 56:105–112.

Chantangsi, C., Lynn, D. H., Racicot, A. C., Gray, H. M., Guzetta, K. E., Lin, H. Y., Song, Y. L., Tung, C. H., Zenke, K., Yoshinaga, T., Cheng, C. Y., Chang, W. J. & Gong, H. 2017. High genetic diversities between isolates of the fish parasite Cryptocaryon irritans (Ciliophora) suggest multiple cryptic species. Mol. Phylogenet. Evol., 112:47–52.

Cho, P. L. 1971a. Cortical patterns in two syngens of Glaucoma. J. Protozool., 18:180–183.

Cho, P. L. 1971b. The genetics of mating type in a syngen of Glaucoma. Genetics, 67:377–390.

Corliss, J. O. 1953. Comparative studies on holotrichous ciliates in the Colpidium-Glaucoma-Leucophrys-Tetrahymena group. II. Morphology, life cycles and systematic status of strains in pure culture. Parasitology, 43:49–87.

Corliss, J. O. 1959. Glaucoma chattoni n. sp., an experimental organism which has been confused with the type species of the genus, G. scintillans Ehrlbg., 1830. J. Protozool., 6:24.

Corliss, J. O. 1960. Tetrahymena chironomi sp. nov., a ciliate from midge larvae, and the current status of facultative parasitism in the genus Tetrahymena. Parasitology, 50:111–153.

Corliss, J. O. 1979. The ciliated protozoa: characterization, classification, and guide to the literature. Trans. Amer. Microsc. Soc., 98:465.

Doerder, F. P. 2014. Abandoning sex: multiple origins of asexuality in the ciliate Tetrahymena. BMC Evol. Biol., 14:112.

Doerder, F. P. & Brunk, C. 2012. Natural populations and inbred strains of Tetrahymena. In: Collins, K. (ed.), Methods in Cell Biology. Elsevier Academic Press, Cambridge, MA: 109:277–300.

Elliott, A. M. 1973. Life cycle and distribution of Tetrahymena. In: Elliott, A. M. (ed.), Biology of Tetrahymena. Dowden, Hutchinson & Ross, Inc, Stroudsburg, PA. p. 259–286.

Excolfier, L. & Lischer, H. E. L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour., 10:564–567.

Fenchel, T., Esteban, G. F. & Finlay, B. J. 1997. Local versus global diversity of microorganisms: cryptic diversity of ciliated protozoa. Oikos, 80:220–225.

Feng, X. L., Sun, Q., Cao, T. G., Li, L. Y. & Chen, Y. Z. 1988. The S1 strain of Tetrahymena from Shanghai - Tetrahymena shanghaiensis Sp-nov. Acta Zool. Sin., 34:42–51.

Fernandes, N. M. & Neto, I. D. D. 2013. Morphology and 18S rDNA gene sequence of Spirostrongylus minus and Spirostrongylus teres (Ciliophora: Heterotrichida) from Rio de Janeiro, Brazil. Zoologia, 30:72–79.

Finlay, B. J. 2002. Global dispersal of free-living microbial eukaryote species. Science, 296:1061–1063.

Foissner, W. 1987. Neue terrestri sche und limnische Ciliaten (Protozoa, Ciliophora) aus Österreich und Deutschland. Sber. Akad. Wiss. Wien, 195:217–268.

Foissner, W. 1999. Prost diversity: estimates of the near-imponderable. Protist, 150:363–368.

Foissner, W. 2013. Description of Glaucomides bromelica n. gen., n. sp. (Ciliophora, Tetrahymenida), a macrostome forming inhabitant of bromeliads (Bromeliaceae), including redescriptions of Glaucoma scintillans and G. reniformis. J. Eukaryot. Microbiol., 60:137–157.

Foissner, W., Chao, A. & Katz, L. A. 2008. Diversity and geographic distribution of ciliates (Protista: Ciliophora). Biodivers. Conserv., 17:345–363.

Forster, P., Torroni, A., Renfrew, C. & Rohl, A. 2001. Phylogenetic star contraction applied to Asian and Papuan mtDNA evolution. Mol. Biol. Evol., 18:1864–1881.

Gao, F., Katz, L. A. & Song, W. 2012. Insights into the phylogenetic and taxonomy of philasterid ciliates (Protozoa, Ciliophora, Scuticociliatia) based on analyses of multiple molecular markers. Mol. Phylogenet. Evol., 64:308–317.

Gentekaki, E. & Lynn, D. H. 2010. Evidence for cryptic speciation in Carchesium polygonum Linnaeus, 1758 (Ciliophora: Peritrichida) inferred from mitochondrial, nuclear, and morphological markers. J. Eukaryot. Microbiol., 57:508–519.

Goodwin, D. C. & Lee, S. B. 1993. Microwave miniprep of total genomic DNA from fungi, plants, protists and animals for PCR. Biotechniques, 15:438–444.

Grassmick, R. A. & Rowley, W. A. 1973. Larval mortality of Culex tarsalis and Aedes aegypti when reared with different concentrations of Tetrahymena pyriformis. J. Invertebr. Pathol., 22:86–93.

Gruchy, D. F. 1955. The breeding system and distribution of Tetrahymena pyriformis. J. Protozool., 2:178–185.

Hill, D. L. 1972. The Biochemistry and Physiology of Tetrahymena. Academic Press, New York. p. 230.

Hoffman, G. L., Landolt, M., Camper, J. E., Coats, D. W., Stookey, J. L. & Burek, J. D. 1978. A disease of freshwater fishes caused by Tetrahymena corlissi Thompson, 1955, and a key for identification of holotrich ciliates of freshwater fishes. J. Parasitol., 64:217–223.

Jerome, C. A., Simon, E. M. & Lynn, D. H. 1996. Description of Tetrahymena empidodyrea n.sp., a new species in the Tetrahymena pyriformis sibling species complex (Ciliophora, Oligohymenophorea), and an assessment of its phylogenetic position using small-subunit rRNA sequences. Can. J. Zool., 74:1898–1906.

Jörgen, K. & Schrödl, M. 2013. How to describe a cryptic species? Practical challenges of molecular taxonomy. Front. Zool., 10:59.

Kahl, A. 1926. Neue und wenig bekannte Formen der holotrichen und heterotrichen Ciliaten. Arch. Protistenk., 55:197–438.

Katz, L., DeBerardinis, J., Hall, M., Kovner, A., Dunthorn, M. & Muse, S. 2011. Heterogeneous rates of molecular evolution among cryptic species of the ciliate morphospecies Chilodonella uncincta. J. Mol. Evol., 73:266–272.

Katz, L. A., Snoeyenbos-West, O. & Doerder, F. P. 2006. Patterns of protein evolution in Tetrahymena thermophila: implications for estimation of effective population size. Mol. Biol. Evol., 23:608–614.

Kher, C. P., Doerder, F. P., Cooper, J., Ikonomi, P., Achilles-Day, U., Kupper, F. C. & Lynn, D. H. 2011. Barcoding Tetrahymena.
discriminating species and identifying unknowns using the cytochrome c oxidase subunit 1 (cox-1) barcode. *Protist*, 162:2–13.

Kozloff, E. N. 1946. The morphology and systematic position of a holotrichous ciliate parasitizing *Deroceras agreste* (L.). *J. Morphol.*, 79:445–465.

Kozloff, E. N. 1957. A species of *Tetrahymena* parasitic in the renal organ of the slug *Deroceras reticulatum*. *J. Protozool.*, 4:75–79.

Krenek, S., Berendonk, T. & Fokin, S. 2015. New *Tetrahymena* (Ciliophora, Oligohymenophorea) congeners shapes our view on its biodiversity. *Org. Divers. Evol.*, 15:215–233.

Leibowitz, M. P. & Zilberg, D. 2009. *Tetrahymena* sp. infection in guppies, *Poecilia reticulata* Peters; parasite characterization and pathology of infected fish. *J. Fish Dis.*, 32:845–855.

Librado, P. & Rozas, J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25:1451–1452.

Liu, M., Fan, X., Gao, F., Gao, S., Yu, Y., Warren, A. & Huang, J. 2016. *Tetrahymena australis* (Protozoa, Ciliophora): a well-known but “non-existing” taxon – consideration of its identification, definition and systematic position. *J. Eukaryot. Microbiol.*, 63:760–770.

Lynn, D. H. & Doerder, F. P. 2012. The life and times of *Tetrahymena*. In: Collins, K. (ed.), Methods in Cell Biology. Elsevier Academic Press, Cambridge, MA. p. 109:11–27.

Lynn, D. H., Doerder, F. P., Prosser, R. & Gillis, P. 2017. *Tetrahymena* glochidiophila n. sp., a new species of *Tetrahymena* (Ciliophora) that causes mortality to glochidia larvae of freshwater mussels (*Bivalvia*). *Dis. Aquat. Org.*, 127:125–136.

Lynn, D. H., Gransden, S. G., Wright, A.-D. G. & Josephson, G. 2000. Characterization of a new species of the ciliate *Tetrahymena* (Ciliophora: Oligohymenophorea) isolated from the urine of a dog: first report of *Tetrahymena* from a mammal. *Acta Protozool.*, 39:289–294.

Lynn, D. H., Molloy, D. & LeBrun, R. 1981. *Tetrahymena rotundata* n. sp. (Hymenostomatida: Tetrahymenidae), a ciliate parasite of the homolymph of *Simulium* (Diptera: Simuliidae). *Trans. Amer. Microsc. Soc.*, 100:134–141.

Lynn, D. H. & Struder-Kypke, M. C. 2006. Species of *Tetrahymena* identical by small subunit rRNA gene sequences are discriminated by mitochondrial cytochrome c oxidase I gene sequences. *J. Eukaryot. Microbiol.*, 53:385–387.

McManus, G. B., Xu, D. P., Costas, B. A. & Katz, L. A. 2010. Genetic identities of cryptic species in the *Strombidium* stylifer/ apolatum/ocularatum cluster, including a description of *Strombidium rassoulzadegani* n. sp. *J. Eukaryot. Microbiol.*, 57:369–378.

Nannen, D. L. 1999. When is a rose?: the kinds of *Tetrahymena*. *In: Wilson, R. A. (ed.), Species, New Interdisciplinary Essays. The MIT Press, Cambridge, MA. p. 93–118.

Nannen, D. L. & McCoy, J. W. 1976. Characterization of the species of the *Tetrahymena pyriformis* complex. *Trans. Amer. Microsc. Soc.*, 95:664–682.

Nannen, D. L., Park, C., Preparata, R. & Simon, E. M. 1998. Comparison of sequence differences in a variable 235 rRNA domain among sets of cryptic species of ciliated protozoa. *J. Eukaryot. Microbiol.*, 45:91–100.

Nelsen, E. M. 1978. Transformation in *Tetrahymena thermophila*: Development of an inducible phenotype. *Develop. Biol.*, 66:17–31.

Nelsen, E. M. & DeBault, L. E. 1978. Transformation in *Tetrahymena pyriformis*: description of an inducible phenotype. *J. Protozool.*, 25:113–119.

Nyberg, D. 1981a. Fertility is not a function of geographic distance in *Tetrahymena*. *J. Hered.*, 72:94–96.

Nyberg, D. 1981b. Three new “biological” species of *Tetrahymena* (*T. hegewischii* n. sp., *T. sonnebornii* n. sp., *T. nipissingi* n. sp.) and temperature tolerance of members of the “pyriformis” complex. *J. Protozool.*, 28:65–69.

Pitsch, G., Adamc, L., Dirren, S., Nitsche, F., Šimek, K., Sirová, D. & Posch, T. 2016. The Green *Tetrahymena utriculariae* n. sp. (Ciliophora, Oligohymenophorea) with its endosymbiotic algae (*Microactinum* sp.), living in traps of a carnivorous aquatic plant. *J. Eukaryot. Microbiol*. https://doi.org/10.1111/jeu.12369

Preparata, R. M., Meyer, E. B., Preparata, F. P., Simon, E. M., Vossbrinck, C. R. & Nannen, D. L. 1989. Ciliate evolution: the ribosomal phylogenies of the *tetrahymenine* ciliates. *J. Mol. Evol.*, 28:427–441.

Przybos, E. & Prajer, M. 2015. New stands of species of the *Paramecium aurelia* complex; Is the occurrence of particular species limited by temperature barriers? *Folia Biol. (Krakow)*, 63:215–220.

Przybos, E. & Rautian, M. 2012. New stands of species of the *Paramecium aurelia* complex (Ciliophora, Protozoa) in the Mediterranean region (Italy, Greece, Morocco). *Folia Biol. (Krakow)*, 60:147–149.

Przybos, E., Rautian, M. & Sursmacz, M. 2014. New stands of species of the *Paramecium aurelia* complex (Ciliophora, Protozoa) in Europe and South America (Ecuador). *Folia Biol. (Krakow)*, 62:13–16.

Przybos, E., Rautian, M., Sursmacz, M. & Bielavskaya, A. 2013. New stands of species of the *Paramecium aurelia* complex (Ciliophora, Protozista) in Russia (Siberia, Kamchatka). *Folia Biol. (Krakow)*, 61:41–45.

Przybos, E. & Sursmacz, M. 2010. New, worldwide data on the distribution of species of the *Paramecium aurelia* complex (Ciliophora, Protozoa). *Folia Biol. (Krakow)*, 58:185–188.

Przyboś, E. & Tarcz, S. 2016. *Paramecium jenningsi* complex: existence of three cryptic species confirmed by multi-locus analysis and strain crosses. *Syst. Biodivers.*, 14:140–154.

Quinta-Alonso, P., Nitsche, F., Wylezich, C., Arndt, H. & Foissner, W. 2013. A New *Tetrahymena* (Ciliophora, Oligohymenophorea) from groundwater of Cape Town, South Africa. *J. Eukaryot. Microbiol.*, 60:235–246.

Rautian, M., Przybos, E. & Beliaevskia, A. 2015. New stands of species of the *Paramecium aurelia* complex in Yakutia, Russia. *Folia Biol. (Krakow)*, 63:289–293.

Renner, S. S. 2016. A return to Linnaeus’s focus on diagnosis, not description: the use of DNA characters in the formal namin of species. *Syst. Biol.*, 65:1085–1086.

Roque, M., De Puytorac, P. & Savoe, A. 1970. Caractéristiques morphologiques et biologiques de *Tetrahymena bergeri* sp. nov., cilié hyménostome tetrahyménien. *Protistologica*, 6:343–351.

Sadler, L. A. & Brunk, C. F. 1992. Phylogenetic relationships and unusual diversity in histone H4 proteins within the *Tetrahymena-pyriformis* complex. *Mol. Biol. Evol.*, 9:70–84.

Santoferreira, L. F., Tian, M., Alder, V. A. & McManus, G. B. 2015. Discrimination of closely related species in tintinnid ciliates: new insights on crypticity and polymorphism in the genus *Helicostomella*. *Protist*, 166:78–92.

Schiftner, U. & Foissner, W. 1998. Revision of *Tetrahymena* species with caudal cillum. *J. Protozool.*, 45:15A.

Schigel, M. & Meisterfeld, R. 2003. The species problem in protozoa revisited. *Eur. J. Protistol.*, 39:349–355.

Shaw, S., Speare, R., Lynn, D. H., Yeates, G., Zhao, Z., Berger, L. & Jakob-Hoff, R. 2011. Nematode and ciliate nasal infection in captive Archey’s frogs (*Leiopelma archeyi*). *J. Zoo Wildl. Med.*, 42:473–479.
**Figure S7.** Geographic distribution of new *Glaucoma* species.

**Figure S8.** *Glaucoma chattoni* distribution and cox1 haplotype network.

**Figure S9.** Map of distribution of *T. americanis*.

**Figure S10.** Cox1 haplotype network for *T. americanis*.

**Figure S11.** NJ cox1 phylogenetic tree showing separation of *T. americanis* and *T. hegewischi*.

**Figure S12.** Map of *T. americanis* haplotypes by geographical region.

**Figure S13.** *T. hegewischi* map and cox1 haplotype network.

**Figure S14.** Distribution of *T. borealis* cox1 haplotypes and amicronucleates.

**Figure S15.** Distribution of *T. canadensis* cox1 haplotypes and amicronucleates.

**Figure S16.** Distribution of *T. elliotti* isolates, cox1 haplotypes and amicronucleates.

**Figure S17.** Distribution and cox1 haplotypes of *T. pigmentosa* isolates.

**Figure S18.** Distribution of *T. thermophila* mating types in six heavily sampled ANF ponds.

**Figure S19.** Distribution and cox1 haplotypes of *T. tropicalis* isolates.

**Figure S20.** Distribution and cox1 haplotypes of *T. gruchyi* (nsp7) isolates.

**Figure S21.** Distribution and cox1 haplotypes of *T. kentuckyensis* (nsp9) isolates.

**Figure S22.** Distribution and cox1 haplotypes of *T. glochidiophila* (nsp10) and *T. ennsylvaniensis* (19518-2) isolates.

**Figure S23.** Distribution and cox1 haplotypes of *T. newhampshirensis* (nsp18) and *T. colorunensis* (nsp23).

**Table S1.** Wild isolates (Excel spreadsheet).

**Table S3.** Nanney-Simon (NS) database (Excel spreadsheet).