AEM Accepts, published online ahead of print on 20 March 2009
Appl. Environ. Microbiol. doi:10.1128/AEM.00178-09

High genetic diversity but no population structure of the peritrichous ciliate *Carchesium polypinum* in the Grand River basin (North America) inferred from nuclear and mitochondrial markers

Running title: Intraspecific genetic variation of *C. polypinum*

Section: Environmental Microbiology

E. Gentekaki* and D. H. Lynn

Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada

Correspondence:

Eleni Gentekaki

Department of Integrative Biology, University of Guelph

Guelph, ON, CANADA N1G 2W1

Phone: 519-824-4120 x58381

Fax: 519-767-1656

E-mail: egenteka@uoguelph.ca

ABSTRACT

Studies that assess intraspecific genetic variation in ciliates are few and quite recent. Consequently, knowledge on the subject and understanding of the processes that underlie it are limited. We sought to assess the degree of intraspecific genetic variation in *Carchesium polypinum* (Ciliophora: Peritrichia), a cosmopolitan, freshwater, ciliate. We isolated colonies of *C. polypinum* from locations in the Grand River basin in Southwestern Ontario, Canada. We then used the nuclear markers - ITS1, ITS2 and the hypervariable regions of the large subunit (LSU) rRNA, and an 819-bp fragment of the mitochondrial cytochrome *c* oxidase I (*cox-1*) gene, to investigate the intraspecific genetic variation of *C. polypinum* and the degree of resolution of the above-mentioned markers at the population level. We also sought to determine whether the organism demonstrated any population structure that mapped onto the geography of the region. Our study shows that there is a high degree of genetic diversity at the isolate level revealed by the mitochondrial but not the nuclear markers. Furthermore, our results indicate that *C. polypinum* is likely not a single morphospecies as previously thought.

INTRODUCTION

Studies of intraspecific genetic variation in macro-organisms have provided insights into micro-evolutionary processes and speciation in these organisms. However, our understanding of intraspecific genetic diversity of microbial eukaryotes and its temporal and spatial distribution is neither deep nor broad due to the limited number of studies that deal with these topics (2, 5, 7, 33).

Many of the studies assessing within-species genetic variation and its distribution have used random amplified polymorphic DNA (RAPD) fingerprinting. Typically, the amount of genetic variation is very low within a species (37, 38, 62) with some exceptions (21, 63). However, there are several problems arising from the use of RAPDs as the sole source of data, including reproducibility and scoring concerns (59).

Sequences of the internal transcribed spacers, namely ITS1 and ITS2, of the rRNA region have also been used for studying the genetic diversity and population structure of protists (2, 6, 46). In several instances the ITS sequences of populations of ciliates from across the globe have been nearly identical, indicating very low genetic variation (4, 13, 65, 69). On the other hand, the hypervariable regions of the large subunit (LSU) rRNA have revealed moderate genetic diversity in ciliates at the population level (16, 46, 55).

Mitochondrial DNA (mtDNA) has been employed only sparingly in population studies of protists, even though it is routinely used in similar studies in metazoans (3, 7, 8). Of all the mitochondrial genes, the apocytochrome *b* (*cob*) and the cytochrome *c* oxidase I (*cox-1*) have been used for population level analyses of ciliates. Estimation of

intraspecific nucleotide diversity of the *cob* gene from three species of the prostome ciliate *Coleps* showed only minor differences (5). On the other hand, the *cox-1* gene revealed genetic diversity that nuclear markers failed to detect in members of the genera *Paramecium* and *Tetrahymena*. However, the implications and the extent of this diversity were not examined in detail, since the datasets used in these studies were small (4, 44). Finally, *cox-1* uncovered a modest degree of genetic diversity in members of the *Paramecium aurelia* sibling species complex (11).

To broaden our understanding of population-level processes in ciliates, we sought to assess the level of intraspecific genetic diversity within the morphologically cosmopolitan, freshwater peritrich ciliate, *Carchesium polypinum*. A Canadian river system - the Grand River drainage basin, was intensively sampled. We developed primers for the cox-1 gene of *C. polypinum* to assess the degree of genetic variation of this organism. Furthermore, we compared the resolving power of this gene at the population level to that of the ITS regions and the LSU rRNA hypervariable region. Our study demonstrates the efficacy of *cox-1* in identifying genetic diversity in *C. polypinum* and implies that it can be applied in other ciliate species as well.

MATERIALS AND METHIDS

Sample collection, culturing, and DNA extraction. Colonies of *C. polypinum* were collected from localities throughout the Grand River basin in Southwestern Ontario. This river basin consists of the main river (Grand River) and four tributaries, the Conestogo and Nith Rivers on the west and Eramosa and Speed Rivers on the east. A table indicating exact collection localities is available in the supplemental material

section (Supplemental Table S1). Plastic Petri dishes were left in the river for one week and were subsequently retrieved and transported to the laboratory. Individual peritrich colonies were scraped from the bottom of the dish, washed several times in Canadian Springs (Grenville, QC) mineral water, and placed in a clean glass Petri dish with mineral water, barley grains, and 20 µl of Cerophyl™ (Cerophyl Laboratories Inc., Kansas city, MI), which had been inoculated with *Enterobacter aerogenes* at least 24 h previously. The colonies were confirmed as *C. polypinum* with live microscopy and silver staining.

After a few days of growth and replication, colonies were picked and DNA was isolated from a minimum of 30 cells using the MasterPure™ DNA Purification Kit (Epicentre, Madison, WI, USA).

Amplification and sequencing. *The ITS and LSU RRNA regions***.** The 3'-end of the small subunit (SSU) rRNA, the full ITS region containing the 5.8S rRNA gene, and part of the large subunit (LSU) rRNA containing the hypervariable region were amplified using the 300 forward (5'-AGGGTTCGATTCCGGAG-3') and reverse C primers (5'-TGGTCCGTGTTTCAAGACG-3'; 32). Each PCR reaction contained 4 µl of DNA, 1 µM of each primer, 2 mM of $MgCl_2$, 1 µM of each dNTP, 1x PCR buffer, and 2.5U of Diamond Taq DNA polymerase (Medicorp, Montreal, QC, Canada) for a total volume of 25 µl. The PCR amplification was performed in a Perkin-Elmer GeneAmp 2400 Thermal Cycler (PE Applied Biosystems, Mississauga, ON, Canada) with the amplification conditions for this fragment as follows: initial denaturation at 94 °C for 4 min, 35 amplification cycles (94 °C for 60 s, 55 °C for 120 s, 72 °C for 150 s), and final extension step at 72 °C for 10 min.

*The cox-1 gene region***.** Initially, the *cox-1* sequences of *Tetrahymena thermophila* and *Paramecium tetraurelia* were aligned and degenerate primers were designed based on highly conserved protein domains. The resulting PCR product was sequenced and a nested forward primer (coxFper; 5'-GTTGGAAGTAAAGATGTTGC-3') was designed, while the reverse primer (339R; 5'- ATAGGATCACCTCCGTAAGC-3') was maintained. The resulting fragment of the cox-1 gene was 819 bp long. The PCR reaction mix was identical to the one used for amplification of the ITS region. The PCR amplification was performed in a Perkin-Elmer GeneAmp 2400 Thermal Cycler (PE Applied Biosystems, Mississauga, ON, Canada) with the amplification conditions for this fragment as follows: initial denaturation at 94 °C for 4 min, 35 amplification cycles (5 cycles at 94 °C for 30 s, 45 °C for 60 s, 72 °C for 105 s and 30 cycles at 94 °C for 30 s, 55 °C for 60 s, 72 °C for 105 s), and final extension step at 72 °C for 10 min.

Sequencing. The resulting amplicons for the rRNA region and *cox-1* were purified with the Qiagen MinElute gel extraction kit (Qiagen, Mississauga, ON, Canada). The amplification primer reverse C and the internal forward primer 1055F (5'-GGTGGTGCATGGCCG-3') were used for sequencing the rRNA region in both directions with a 3730 DNA analyzer (Applied Biosystems Inc., Foster City, CA, USA), an ABI Prism BigDye Terminator (ver 3.1), and a Cycle Sequencing Ready Reaction Kit. The coxFper and 339R were used for sequencing the cox-1 gene fragment.

Sequence accession numbers. All sequences are available through GenBank. For the *cox-1* dataset the accession numbers are FJ810309-FJ810353. For the nuclear marker dataset the corresponding numbers are FJ810354-FJ810408.

Phylogenetic analyses. The sequence fragments were imported into Sequencher, ver 4.0.5 (Gene Codes Corp) and assembled into contigs. Subsequently, the sequences of the ITS regions were uploaded to our DCSE (Dedicated Comparative Sequence Editor; 14) database and aligned against existing peritrich sequences. The alignment was further refined by eye.

The sequences of *cox-1* were aligned with MEGA, ver 4.0, which uses the CLUSTALW algorithm (36).

The 3'-end of the SSU rRNA, 5.8S rRNA, the two ITS regions, and the hypervariable region of the LSU rRNA of 55 colonies of *C. polypinum* were sequenced for this study, but only the latter three regions were used for analysis. Three separate files were constructed for phylogenetic analysis of the nuclear sequences: 1) the ITS1 region; 2) the ITS2 region; and 3) the hypervariable region of the LSU rRNA. The files consisted of 128, 183, and 407 nucleotides, respectively. The 5.8S gene was omitted from the analysis due to its high degree of conservation. An 819-bp fragment of *cox-1* of 42 colonies of *C. polypinum* was also sequenced along with one of *Epistylis plicatilis* (organism generously provided by Dr. Chris Lobban, University of Guam, 41) and one of *Epistylis* sp. as well as a *Vorticella* sp.

The DNADIST program in PHYLIP ver 3.65b (15) was used to calculate genetic distances with the Kimura 2-parameter model (15). Subsequently, a neighbour-joining (NJ) tree was inferred. Using SEQBOOT, the data were bootstrap resampled 1,000 times and CONSENSE was subsequently used to construct a consensus tree. For the Bayesian inference (BI) and maximum likelihood (ML) analyses, missing nucleotides were treated as missing and the gaps as a fifth character state. To determine the model of DNA

substitution that best fit the data for the ML and BI calculations, Modeltest (54) was used. The most suitable models for the ITS1, ITS2, and LSU rRNA data sets were F81, Jukes Cantor, and F81+I+Γ respectively, while for the *cox-1* dataset, the GTR+Γ was used. The models were applied in MrBayes ver 3.1.1 (30, 57) and the corresponding phylogenetic trees were inferred. In all cases, the HLRT criterion of Modeltest was used. However, the datasets were also analyzed employing the AIC criterion proposed models (F81+I, GTR+Γ, and TrN+I+ Γ, respectively) and the tree topologies were identical. The models were also applied in PHYML for the ML analysis (25, 26), and the data were bootstrap resampled 500 times.

Genetic structure analysis. The genetic structure of *C. polypinum* was investigated with the software ARLEQUIN 3.0 (58). An analysis of molecular variance (AMOVA) was employed to investigate the genetic structure within and between populations using the cox-1 gene sequences. The sequences were first grouped by tributary and then by clade as defined in the phylogenetic analysis of the *cox-1* dataset.

RESULTS

Sequence analysis of the nuclear datasets. The beginning and the end of ITS1, 5.8S rRNA, and ITS2 were determined by taking into consideration the master alignment available in our laboratory, which contains data from species across the different classes and subclasses of ciliates. The lengths of the ITS1, 5.8S rRNA, and ITS2 for *C. polypinum* were 113, 142, and 172 nucleotides, respectively. A detailed analysis of the dataset is available from the author upon request.

Sequence analysis of the cox-1 dataset. Five distinct amino acid sequences were

identified for the mitochondrial *cox-1*. These correspond to the clades described below (see section on **Phylogenetic Analysis of *Cox-1* Dataset**). The intraclade nucleotide sequence divergences within all the clades that contained more than one individual were <1.0%. The interclade genetic distances as calculated from the nucleotide sequences ranged from 11--18% (Table 1).

There were a total of 243 polymorphisms noted: 67% were at the third nucleotide position, 10% at the second, and 23% at the first position. Of all the substitutions, 18% were non-synonymous resulting in an amino acid change while the rest were synonymous.

Phylogenetic analyses of the nuclear datasets. The tree topologies for ITS1, ITS2, and the 5'-end of the LSU rRNA were similar with some notable differences (see below and Figures 1, 2, 3). In addition, for the individual markers the topologies of NJ, ML, and BI were nearly identical. Based on these trees, no pattern of geographic distribution is apparent: individual genotypes collected from the five rivers were distributed throughout the tree rather than forming distinct monophyletic clades.

In the ITS1 tree (Figure 1), four clades are immediately obvious. All clades are weakly to moderately supported. Clade 2 and Clade 3 were not completely recovered when employing the BI analysis.

In the ITS2 tree (Figure 2), there are three clades. All three clades were recovered with all the methods used. The support values for the clades of the ITS2 tree are higher than those of the ITS1 tree, but still weak to moderate. The exception is Clade 1, which is now strongly supported.

For the LSU rRNA tree (Figure 3), five clades were recovered. Clades 2 and 3 were not completely recovered when employing BI analysis, and were very weakly supported when employing NJ and ML (bootstrap values below 60%, not shown). The other three clades, however, were strongly supported.

Phylogenetic analyses of the cox-1 dataset. For the cox-1 tree, the tree topologies were identical for NJ, ML, and BI analyses (Figure 4). For this dataset, five clades were identified. Clades 3 and 4 were identified with very strong support in all three analyses as were three putative clades (Clades 1, 2, and 5). All clades are deeply diverging. Clades 1 and 2 consist of individuals that belonged to Clade 1 of the nuclear datasets. The rest of the individuals in Clade 1 of the nuclear datasets and all the individuals from Clade 2 form the strongly supported *cox-1* Clade 3. Clades 4 and 5 of the *cox-1* dataset are identical to Clades 3 and 4 of both the ITS1 and LSU datasets.

Similarly to the nuclear trees, the clades did not correspond to the collection localities, rather individual genotypes collected from various localities were distributed throughout the clades.

Population structure analysis. Analysis of the *cox-1* dataset revealed 26 haplotypes. The sequences of *C. polypinum* were partitioned in two ways; initially a population was defined and partitioned as the individuals collected from each tributary. The AMOVA analysis based on the *cox-1* dataset indicated that 71.2% of the genetic variation observed was within populations and 28.9% between populations (Table 2). The pairwise F_{ST} values were for the most part non significant (data not shown). Similarly, when the sequences were grouped by clade, as defined in the phylogenetic analysis of the *cox-1* dataset 71.15% of the variation was within populations and 28.85% among

populations. However, the pairwise F_{ST} values were high and statistically significant indicating a deep divergence in clades 2, 3, and 4 (Table 3). When the analysis was performed separately for the densely sampled clades (clades 3 and 4), the corresponding F_{ST} values were non-significant (data not shown).

DISCUSSION

Marker resolution. When assessing intraspecific genetic variation, it is important to establish the appropriate markers for the organisms under investigation. In metazoan populations, such investigations involve the use of mtDNA markers with the cox-1 gene being one of the most popular due to its faster rate of evolution (27, 31). Alternatively, the ITS2 of the rRNA region is commonly used in plants (45). Traditionally, in ciliates, studies of intraspecific genetic variation are based on RAPDs (21, 37). Over the past few years, the focus has shifted from RAPDs to the use of nuclear and mitochondrial sequences (4, 5, 16, 46). Despite this, there are very few studies on intraspecific variation of ciliates, and so it is still unclear which marker is the most suitable for studying their genetic diversity. In ciliates ITS1, ITS2, and the hypervariable regions of the LSU rRNA have been used most commonly. However, there has never been an attempt to compare all three nuclear markers on one ciliate species at the sequence level. In this study we used all three nuclear markers in addition to one mitochondrial marker in order to compare and contrast the resolution and suitability of these molecules at the intraspecific level. We picked the cox-1 gene as the mitochondrial marker of choice for *C. polypinum*. The use of *cox-1* in intraspecific genetic variation in ciliates is extremely limited (4). This is likely due to the fact that the only available

sequences of *cox-1* in ciliates are those from species of the genera *Paramecium* and *Tetrahymena*. To date, only six complete ciliate mtDNA genomes are available - *Tetrahymena thermophila*, *Tetrahymena pyriformis*, *Tetrahymena paravorax*, *Tetrahymena malaccensis*, *Tetrahymena pigmentosa*, and *P. aurelia* (9, 10, 48, 56).

Tetrahymena and *Paramecium* belong to the class Oligohymenophorea, but different subclasses – the subclasses Peniculia and Hymenostomatia, respectively (43). Despite this, the percent divergence of *cox-1* is so high that it was not possible to align the sequences with the default settings of the Sequencher alignment program (personal observation, E. Gentekaki). Thus, it has been extremely difficult to design primers for the *cox-1* gene of other groups of ciliates. After considerable experimentation, we developed a set of primers that amplifies an 819-bp fragment of *cox-1* in *C. polypinum.*

In our study, the best resolution was obtained with the cox-1 gene as assessed by the resulting clades being deeply divergent and robustly and strongly supported. The nuclear gene that matched the cox-1 the closest in terms of separation of clades, evolutionary distances, robustness, and support was that of the 5'-end of the LSU rRNA. This is likely because the regions chosen for the analysis are among the most highly variable in the LSU rRNA and therefore suitable for analysis below the morphospecies level (1, 46). The analysis of the ITS1 and ITS2 regions gave somewhat similar topologies but the resulting clades were not strongly supported. Our study is in agreement with Barth *et al*. (4) who found that the cox-1 gene provided better separation of *Paramecium* species than the corresponding ITS markers. Since ITS regions contain signals used for processing of the rRNA transcript (29, 40), these spacers are constrained to evolve more slowly. Consequently, they did not have adequate time to accumulate

polymorphisms that would be useful for an intraspecific level analysis. Alternatively, the high degree of variation and faster evolutionary rates in the mtDNA genome versus the nuclear genome are well documented in metazoans, and it seems also to be the case in protists (23). Unlike metazoans, the mutation rate of mtDNA in protists is unknown. It has been speculated that it is 10% per million years (4), which would make mtDNA genes good candidates for assessing intraspecific variation.

Intraspecific genetic variation within *Carchesium polypinum*. Studies on intraspecific genetic diversity of ciliates have yielded variable results. On one hand, low genetic diversity was revealed in *T. thermophila* (34), *Stentor coeruleus* (37), and *Euplotes daidaleos* (38). On the other hand, high genetic diversity was evident in others, including *Paramecium multimicronucleatum* and *Paramecium caudatum* (5), *Halteria grandinella* (33), *Cyclidium glaucoma* (16), *Coleps spetai* (4), and *Coleps hirtus hirtus* (4). It is worth noting, however, that all the studies that used mitochondrial markers revealed a high intraspecific genetic diversity (4, 5). In agreement with these studies, ours also indicated a high genetic diversity of *C. polypinum* populations particularly in the case of *cox-1*. Moreover, the degree of variation revealed when analyzing the *cox-1* data set indicated that the majority of the variation was at the isolate level rather than among populations. A similar degree of genetic variation within populations of *C. polypinum* was also found when inter-simple sequence analysis (ISSR) was employed (70). These authors suggested that the high genetic diversity of *C. polypinum* at the isolate level was due to high gene flow and the ability of telotrochs, the dispersal stage, to range over large distances, resulting in a large panmictic population.

Phylogeography. It has been a long standing debate whether or not protists form uniform global populations or if, at least in the case of some, they form distinct endemic populations (17, 18, 19, 20, 47, 66). So far the debate remains inconclusive due to the very small number of studies that deal with the spatial and temporal distribution of protists (7). For instance, no geographic isolation of populations was observed in the heterotrich *Stentor* or the hypotrich *Euplotes* (37, 39). Alternatively, when RAPDs were used along with morphological characters in the hypotrich *Gonostomum affine*, a rudimentary population structure was revealed (21). Use of the ITS1 region, a nuclear marker, has yielded conflicting results. On one hand Miao *et al.* (46) demonstrated distinct population structure in populations of *C. polypinum* in China represented by two distinct clades that mapped onto the geography of the region. Similarly, *H. grandinella* isolates showed variation among geographic isolates (33). Conversely, there was no geographic structure for populations of *Meseres corlissi*, *Isotricha prostoma*, and *P. caudatum,* despite the fact that some individuals from all of these species were from across the globe (4, 65, 69). In agreement with the latter studies, we did not detect any pattern that maps to the geography of the Grand River basin when we used any of the four markers. In the case of the mitochondrial marker we obtained deeply divergent clades but those were not geographically partitioned. The absence of population structure could be due to a high effective population size and high gene flow or a combination of both (3). It has been argued that effective population size is inversely proportional to body size: as body size decreases the effective population size increases (42). Effective population size estimates of ciliates are extremely limited and restricted to members of the genera *Paramecium* and *Tetrahymena* (34, 60). These analyses indicated that the

effective population of *T. thermophila* is small but that of *Paramecium* is large. However, recent studies on the protein evolution of ciliates suggest an elevated rate of substitutions that might make analysis of their effective population sizes difficult to determine (71). At present, we do not know the effective population size of *C. polypinum.*

Morphospecies and cryptic species. Over the past couple of decades the advent of molecular techniques has uncovered a large degree of genetic diversity. Several species that were once considered cosmopolitan are n fact, discovery of cryptic species has grown exponentially (53). The most common current method of identifying cryptic species is based on the amount of genetic divergence as defined by genetic distances. Analyses of our mitochondrial dataset show that at least three genetically distinct and deeply divergent groups of *C. polypinum* occur in sympatry in the Grand River basin. The minimum genetic distance between the groups was 11% while the maximum was 18%. Given this amount of genetic divergence, *C. polypinum* is very likely a cryptic species complex when compared to genetic divergences of this gene for known biological species of *Tetrahymena* and *Paramecium*.

Currently, it is speculated that the concept of morphospecies might be too conservative for assessing protist diversity (64), or that their morphological and molecular evolution might be decoupled (20). Previous research has shown that the amount of genetic divergence differs between groups. For instance, a 10% divergence in the cox-1 gene of *Tetrahymena* would be enough to designate a new species while this value is higher for *Paramecium* (4, 12). Since there are very few studies of such nature

in ciliates we cannot be sure what percentage of variation constitutes a new species or a cryptic species of *C. polypinum*.

If the uncovered diversity is indeed an indication of cryptic species of *C. polypinum*, then how did seemingly sympatric species occur in a continuous system like a river where there are no barriers to disrupt gene flow long enough for isolation to occur? Lately there have been speculations that sympatric speciation might be more common than initially thought (51). However studies that make a strong case for this are very few (50). Alternatively, the observed genetic diversity could be due to transfer of a few founder cells from other bodies of water. While we currently have no way to determine which of the two processes account for the observed variation, it seems that there is a lot more underlying genetic diversity in protists than initially thought.

ACKNOWLEDGEMENTS

This research was partly funded by the National Research Council of Canada (NSERC). The research was also supported through funding to the Canadian Barcode of Life Network from Genome Canada through the Genomics Institute, awarded to D. H. L. and other sponsors (listed at http://www.BOLNET.ca).

We thank Dr. Chris Lobban for providing us with the *E. plicatilis* isolate. We also thank Dr. Strüder-Kypke for critical review of this manuscript. We are grateful to Guelph wastewater treatment plant personnel for their help in obtaining the sludge *C. polypinum* isolate. We also thank Angela Holliss and Elizabeth Holmes for their invaluable help in sequencing all of our samples. Finally we thank two anonymous reviewers and their constructive comments.

REFERENCES

1. **Ali, A. B., J. Wuyts, R. De Wachter, A. Meyer, and Y. Van de Peer.** 1999. Construction of a variability map for eukaryotic large subunit ribosomal RNA. *Nucleic Acids Res*. **27**: 2825-2831.

2. **Amato, A., W. H. C. F. Kooistra, J. H. L. Ghiron, D. G. Mann, T. Pröschold, and M. Montresor.** 2007. Reproductive isolation among sympatric cryptic species in marine diatoms. *Protist*. **158**: 193-207.

3. **Avise, C. J.** 2000. Phylogeography: the history and formation of species, 1st edn, Harvard University Press.

4. **Barth, D., S. Krenek, S. I. Fokin, and T. U. Berendonk. 2006.** Intraspecific Genetic Variation in *Paramecium* Revealed by Mitochondrial Cytochrome c Oxidase I Sequences. *J. Eukaryot. Microbiol*. **53**: 20-25.

5. **Barth, D., K. Tischer, H. Berger, M. Schlegel, and T. U. Berendonk.** 2008. High mitochondrial haplotype diversity of *Coleps* sp. (Ciliophora: Prostomatida). *Environ Microbiol*. **10**: 626-634.

6. **Bass, D., T. A. Richards, L. Matthai, V. Marsh V, and T. Cavalier-Smith. 2007.** DNA evidence for global dispersal and probable endemicity of protozoa. *BMC Evol. Biol*. 7: 162-175.

7. **Beheregaray, L. B.** 2008. Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Mol. Ecol*. **17**: 3754-3774.

8. **Blair, D., T. Agatsuma, T. Watanobe, M. Okamoto, and A. Ito.** 1997. Geographical genetic structure within the human lung fluke, *Paragonimus westermani*, detected from DNA sequences. *Parasitology*. **115:**411-417.

9. **Brunk, F.C., L. C. Lee, A. B. Tran, and J. Li.** 2003. Complete sequence of the

373 mitochondrial genome of *Tetrahymena thermophila* and comparative methods for
374 identifying highly divergent genes. *Nucleic Acids Res*. **31:**1673-1682.
375 10. **Burger, G., Y. Zhu, T. G. Littlejohn, S. J. Greenwood, M. N. Schnare, B.F. Lang,**
376 **and M. W. Gray.** 2000. Complete sequence of the mitochondrial genome of
377 *Tetrahymena pyriform*is and comparison with *Paramecium aurelia* mitochondrial DNA.
378 *J. Mol. Biol*. **297**:365-380.
379 11. **Catania, F., F. Wurmser, A. A. Potekhin, E. Przybos´, and M. Lynch**. 2009.
380 Genetic diversity in the *Paramecium aurelia* species complex. *Mol. Biol. Evol.* **26**:421-
381 431.
382 12. **Chantangsi, C., D. H. Lynn, M. T. Brandl, J. C. Cole, N. Hetrick, and P.**
383 **Ikonomi.** 2007. Barcoding ciliates: a comprehensive study of 75 isolates of the genus
384 *Tetrahymena*. *Int. J. Syst. Evol. Microbiol* **57**:2412-2425.
385 13. **Coleman, A. W.** 2005. *Paramecium aurelia* revisited. *J. Eukaryot. Microbiol*. **52**:68–
386 77.
387 14. **De Rijk, P., R. De Wachter.** 1993. DCSE, an interactive tool for sequence alignment
388 and secondary structure research. *CABIOS*. **9**:735–740.
389 15. **Felsenstein, J.** 2004. PHYLIP (Phylogeny Interference Package) version 3.6a2.
390 Distributed by the author, Dept. of Genetics, University of Washington, Seattle,
391 Washington.
392 16. **Finlay, J. B., G. F. Esteban, S. Brown, T. Fenchel, and K. Hoef-Emden.** 2006.
393 Multiple cosmopolitan ecotypes within a microbial eukatyote morphospecies. *Protist*.
394 **157**:377-390.
395 17. **Finlay, J. B., G. F. Esteban, and T. Fenchel.** 1998. Protozoan diversity, converging

396 estimates of the global number of free-living ciliate species. *Protist*. **149**:29-37.

397 18. **Finlay, J. B., and T. Fenchel.** 1999. Divergent perspectives on protist species
398 richness. *Protist News*. **150**:229-233.

399 19. **Foissner, W.** 2006. Biogeography and dispersal of micro-organisms, a review
400 emphasizing protists. *Acta Protozool*. **45**:111-136.

401 20. **Foissner, W., A. Chao, L. A. Katz.** 2008. Diversity and geographic distribution of
402 ciliates (Protista: Ciliophora). *Biodivers. Conserv*. **17**:345-363.

403 21. **Foissner, W., T. Stoeck, H. Schmidt, and H. Berger.** 2001. Biogeographical
404 differences in a common soil ciliate, *Gonostomum affine* (Stein), as revealed by
405 morphological and RAPD-fingerprint analysis. *Acta Protozool*. **40**:83-97.

406 22. **Gomez, A., M. Serra, G. R. Carvalho, and D. H. Lunt.** 2002. Speciation in ancient
407 cryptic species complexes: Evidence from the molecular phylogeny of *Brachionus*
408 *plicatilis* (Rotifera). *Evolution*. **56:**1431-1444.

409 23. **Gray, M. W., B. F. Lang, and G. Burger.** 2004. Mitochondria of protists. *Annu.*
410 *Rev. Genet.*. **38**:477-524.

411 24. **Gruchy, D. F.** 1955. The breeding system and distribution of Tetrahymena
412 pyriformis. *J. Protozool*. **2**:178.

413 25. **Guindon, S., and O. Gascuel.** 2003. PhyML , A simple, fast, and accurate algorithm
414 to estimate large phylogenies by maximum likelihood. *Syst. Biol*. **52**:696-704.

415 26. **Guindon, S., F. Lethiec, P. Duroux, and O. Gascuel.** 2005. PHYML Online--a web
416 server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.*
417 **33**(Web Server issue):W557-9.

418 27. **Hajibabaei, M., G. A. C. Singer, P.D.N. Hebert, and D. A. Hickey.** 2007. DNA
419 barcoding: how it complements taxonomy, molecular phylogenetics and population
420 genetics, *Trends Genet*. **23**:167-172.
421 28. **Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs.**
422 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical
423 skipper butterfly *Astraptes fulgerator*. *Proc. Natl. Acad. Sci. U.S.A*. **101**:14812-14817.
424 29. **Hillis, D.M., and M. T. Dixon.** 1991. Ribosomal DNA: molecular evolution and
425 phylogenetic inference. *Q Rev Biol*. **66**:411-453.
426 30. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES, Bayesian inference of
427 phylogenetic trees. *Bioinformatics*. **17:**754–755.
428 31. **Hwang, U.W., and W. Kim.** 1999. General properties and phylogenetic utilities of
429 nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular
430 systematics. *Korean J. Parasitol*. **37**:215-228.
431 32. **Jerome, C.A., and D. H. Lynn.** 1996. Identifying and distinguishing sibling species
432 in the *Tetrahymena pyriformis* complex (Ciliophora, Oligohymenophorea) using
433 PCR/RFLP analysis of nuclear ribosomal DNA. *J. Eukaryot. Microbiol*. **43**:492–497.
434 33. **Katz, L. A., G. B. McManus, L. O. Snoeyenbos-West, A. Griffin, K. Pirog, B.**
435 **Costas, and W. Foissner.** 2005. Reframing the "everything is everywhere" debate:
436 evidence for high gene flow and diversity in ciliate morphospecies. *Aquat. Microb. Ecol.*
437 **41**:55-65.
438 34. **Katz, L.A., L. O. Snoeyenbos-West, and F. P. Doerder.** 2006. Unusual patterns of
439 molecular evolution at the SerH surface antigen locus in Tetrahymena thermophila:
440 implication for estimates of effective population size. *Mol. Biol. Evol*. **23:**608-614.

441 35. **King, J. L., and R. Hanner.** 1998. Cryptic species in a "Living Fossil" lineage:
442 Taxonomic and phylogenetic relationships within the genus Lepidurus (Crustacea :
443 Notostraca) in North America. *Mol. Phylogenet. Evol*. **10**:23-36.
444 36. **Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei.** 2001. MEGA2, molecular
445 evolutionary genetics analysis software. *Bioinformatics*. **17**:1244–1245.
446 37. **Kusch, J.** 1998. Local and temporal distribution of different genotypes of pond-
447 dwelling *Stentor coeruleus*. *Protist*. **149**:147-154.
448 38. **Kusch, J., and K. Heckmann.** 1996. Population structure of Euplotes ciliates
449 revealed by RAPD fingerprinting. *Ecoscience*. **3**:378-384.
450 39. **Kusch, J., H. Welter, M. Stremmel, and H. J. Schmidt.** 2000. Genetic diversity in
451 populations of a freshwater ciliate. *Hydrobiologia*. **431**:185-192.
452 40. **Leary, D. J., and J. Huang.** 2001. Regulation of ribosome biogenesis within the
453 nucleolus, *FEBS Letters*. **509**:145-150.
454 41. **Lobban, C.S., and M. Schefter.** 2008. Freshwater biodiversity in Guam, 1.
455 Introduction with new records of ciliates and a heliozoan. *Micronesica*. **40**:253-273.
456 42. **Lynch, M.** 2005. The origins of eukaryotic gene structure. *Mol. Biol. Evol*. **23**:450-
457 468.
458 43. **Lynn, D. H.** 2008. The ciliated protozoa. Characterization, classification and guide
459 to the literature, 3rd edn. Springer, New York.
460 44. **Lynn, D. H., and M. C. Strüder-Kypke.** 2006. Species of *Tetrahymena* identical by
461 small subunit rRNA gene sequences are discriminated by mitochondrial cytochrome c
462 oxidase I gene sequences. *J. Eukaryot. Microbiol*. **53**:385-387.
463 45. **Mai, J.C., and A. W. Coleman.** 1997. The internal transcribed spacer 2 exhibits a

464 common secondary structure in green algae and flowering plants. *J. Mol. Evol*. **44**:258-
465 271.
466 46. **Miao, W., Y. Yu, Y. H. Shen, and X. Zhang.** 2004. Intraspecific phylogeography of
467 *Carchesium polypinum* (Peritrichia, Ciliophora) from China, inferred from 18S-ITS1-
468 5.8S ribosomal DNA. *Science in China*. **47**:11-17.
469 47. **Mitchell, E. A. D., and R. Meisterfeld.** 2005. Taxonomic confusion blurs the
470 debate on cosmopolitanism versus local endemism of free-living protists. *Protist*.
471 **156**:263-267.
472 48. **Moradian, M.M., D. Beglaryan, J. M. Skozylas, and V. Kerikorian.** 2007.
473 Complete Mitochondrial Genome Sequence of Three *Tetrahymena* Species Reveals
474 Mutation Hot Spots and Accelerated Nonsynonymous Substitutions in *Ymf* Genes. *PLoS*
475 *ONE*. **2**:e650. doi:10.1371/journal.pone.0000650
476 49. **Nanney, L. D., and J. W. McCoy.** 1976. Characterization of the species of the
477 *Tetrahymena pyriformis* complex. *Trans. Am. Microsc. Soc.* **95:**664-682.
478 50. **Niemiller, M. L., B. M. Fitzpatrick, and B. T. Miller.** 2008. Recent divergence
479 with gene flow in Tennessee cave salamanders (Plethodontidae: Gyrinophilus) inferred
480 from gene genealogies. *Mol. Ecol.* **17**:2258-2275.
481 51. **Nosil, P.** 2008. Speciation with gene flow could be common. *Mol. Ecol.* **17**:2103-
482 2106.
483 52. **Page, T. J., and J. M. Hughes.** 2007. Radically different scales of phylogeographic
484 structuring within cryptic species of freshwater shrimp (Atyidae: Caridina). *Limnol.*
485 *Oceanogr.* **52**:1055-1066.
486 53. **Pfenninger, M., and K. Schwenk.** 2007. Cryptic animal species are homogeneously

487 distributed among taxa and biogeographical regions. *BMC Evol. Biol*, **7**, on line.

488

489 54. **Posada, D., and K. A. Crandall.** 1998. MODELTEST, testing the model of DNA
490 substitution. *Bioinformatics*. **14**:817–818.

491 55. **Preparata, R.M., E. B. Meyer, F. P. Preparata, E. M. Simon, C. R. Vossbrinck,**
492 **and D. L. Nanney.** 1989. Ciliate Evolution: The Ribosomal Phylogenies of the
493 Tetrahymenine Ciliates. *J. Mol. Evol*. **28**:427-441.

494 56. **Pritchard, A. E., J. J. Seilhamer, R. Mahalingam, L. C. Sable, E. S. Venuti, and**
495 **J. D. Cummings.** 1990. Nucleotide sequence of the mitochondrial genome of
496 *Paramecium*. *Nucleic Acids Res*. **18**:173-180.

497 57. **Ronquist, F., and J. P. Huelsenbeck.** 2003. MrBayes3, Bayesian inference under
498 mixed models. *Bioinformatics*. **19**:1572–1574.

499 58. **Schneider, S., D. Roessli, and L. Excoffier.** 2000. Arlequin ver 2.000: a software for
500 population genetics data analysis. Genetics and Biometry Laboratory, University of
501 Geneva, Geneva, Switzerland.

502 59. **Skroch, P., and J. Nienhuis.** 1995. Impact of scoring error and reproducibility of
503 RAPD data on RABD based estimates of genetic distance. *Theor. Appl. Genet.* **91**:1086-
504 1091.

505 60. **Snoke, M.S., T. U. Berendonk, and D. Barth.** 2006. Large global effective
506 population sizes in *Paramecium*. *Mol. Biol. Evol*. **23**:2474-2479.

507 61. **Sonneborn, T. M.** 1975. The *Paramecium aurelia* complex of 14 sibling species.
508 *Trans. Am. Microsc. Soc*. **94**:155-178.

509 62. **Stoeck, T., and H. J. Schmidt.** 1998. Fast and accurate identification of European

510 species of the *Paramecium aurelia* complex by RAPD-fingerprinting. *Microb. Ecol.*
511 **35**:311-317.
512 63. **Stoeck, T., E. Przybos, J. Kusch, and H. J. Schmidt.** 2000. Intra-specific
513 differentiation and level of inbreeding of different sibling species of the *Paramecium*
514 *aurelia* complex.
515 *Acta Protozool*. **39**:15-22.
516 64. **Weisse, T.** 2008. Distribution and diversity of aquatic protists: an evolutionary and
517 ecological perspective. *Biodivers. Conserv*. **17**:243-259.
518 65. **Weisse, T., M. C. Strüder-kypke, H. Berger, and W. Foissner.** 2008. Genetic,
519 Morphological, and Ecological Diversity of Spatially Separated Clones of *Meseres*
520 *corlissi* Petz & Foissner, 1992 (Ciliophora, Spirotrichea). *J. Eukaryot. Microbiol*. **55**:257-
521 270.
522 66. **Wilkinson, D. M.** 2001. What is the upper size limit for cosmopolitan distribution in
523 free-living microorganisms? *J. Biogeogr*. **28**:285-291.
524 67. **Wirnsberger, E., W. Foissner, and H. Adam.** 1986. Biometric and morphogenetic
525 comparison of the sibling species *Stylonychia mytilus* and *Stylonychia lemnae* including a
526 phylogenetic system of the Oxytrichids (Ciliophora, Hypotrichia). *Arch. Protistenkunde*.
527 **132**:167–185.
528 68. **Witt, S. D. J. and P. D. N. Hebert.** 2000. Cryptic species diversity and evolution in
529 the amphipod genus *Hyalella* within central glaciated North America: a molecular
530 phylogenetic approach. *Can. J. Fish. Aquat. Sci*. **57**:687-698.
531

532 69. **Wright, A. D. G.** (1999) Analysis of intraspecific sequence variation among eight
533 isolates of the rumen symbiont, *Isotricha prostoma* (Ciliophora) from two continents. *J.*
534 *Eukaryot. Microbiol*. **46**:445–446.
535 70. **Zhang, W. J., J. Yang, Y. H. Yu, S. W. Shu, and Y. F. Shen.** 2006. Population
536 Genetic Structure of *Carchesium polypinum* (Ciliophora: Peritrichia) in Four Chinese
537 Lakes Inferred from ISSR Fingerprinting: High Diversity but Low Differentiation. *J.*
538 *Eukaryot. Microbiol*. **53**:358-363.
539 71. **Zufall, R. A., C. McGrath, S. V. Muse, and L. A. Katz.** 2006. Genome architecture
540 drives protein evolution in ciliates. *Mol. Biol. Evol*. **23**:1681-1687.
541

542 **Figure Legends**

543 Figure 1. Maximum likelihood tree of the ITS1 region of environmental isolates of the
544 peritrich ciliate *Carchesium polypinum* computed with PHYML, based on the Felsenstein
545 81 (F81) model, determined by Modeltest. The first and second values at the nodes
546 represent bootstrap values for neighbour joining and maximum likelihood analyses,
547 respectively, while the third value represents the posterior probability of the Bayesian
548 inference analysis. The scale bar represents 5 changes per 100 positions. Asterisks
549 denote nodes that were not recovered in Bayesian inference analysis. The sequences of
550 *Meseres corlissi* and *Tetrahymena mimbres* were used for rooting the tree. Only
551 bootstrap values above 40 are shown.
552

553 Figure 2. Maximum likelihood tree of the ITS2 region of environmental isolates of the
554 peritrich ciliate *Carchesium polypinum* computed with PHYML, based on the Jukes

555 Cantor (JC) model, determined by Modeltest. The first and second values at the nodes
556 represent bootstrap values for neighbour joining and maximum likelihood analyses,
557 respectively, while the third value represents the posterior probability of the Bayesian
558 inference analysis. The scale bar represents 5 changes per 100 positions. The sequences
559 of *Meseres corlissi* and *Tetrahymena mimbres* were used for rooting the tree. Only
560 bootstrap values above 40 are shown.
561
562 Figure 3. Maximum likelihood tree of the 5'-end of the hypervariable region of the large
563 subunit rRNA of environmental isolates of the peritrich ciliate *Carchesium polypinum*
564 computed with PHYML, based on Felsenstein 81 (F81) model with gamma distribution
565 and an estimate of invariable sites, determined by Modeltest. The first and second values
566 at the nodes represent bootstrap values for neighbour joining and maximum likelihood
567 analyses, respectively, while the third value represents the posterior probability of the
568 Bayesian inference analysis. The scale bar represents 5 changes per 100 positions. Only
569 nodes with strong support are shown. The sequence of Meseres corlissi was used for
570 rooting the tree. Only bootstrap values above 80 are shown.
571
572 Figure 4. Maximum likelihood tree of an 819-bp fragment of *cytochrome c oxidase 1*
573 gene of environmental isolates of the peritrich ciliate *Carchesium polypinum* computed
574 with PHYML, based on the General Time Reversible (GTR) model with gamma
575 distribution, determined by Modeltest. The first and second values at the nodes represent
576 bootstrap values for neighbour joining and maximum likelihood analyses, respectively,
577 while the third value represents the posterior probability of the Bayesian inference

analysis. The scale bar represents 5 changes per 100 positions. The sequences of *Tetrahymena thermophila*, *Colpidium campylum* and *Glaucoma chattoni* were used for rooting the tree.

Table 1. Mean intraspecific nucleotide sequence divergence (%) of an 819-bp fragment of the *cytochrome c oxidase 1* gene between the 5 clades of the peritrich ciliate *Carchesium polypinum* (see Figure 4).

	Clade1	Clade2	Clade3	Clade4	Clade5
CLADE 1	**-**	**12**	**16**	**18**	**18**
CLADE 2	-	**-**	**15**	**18**	**16**
CLADE 3	-	-	**-**	**18**	**11**
CLADE 4	-	-	-	**-**	**17**
CLADE 5	-	-	-	**-**	**-**

Table 2. AMOVA analysis of an 819-bp fragment of the *cytochrome c oxidase 1* gene of the peritrich ciliate *Carchesium polypinum,* showing percent of genetic variation within and among populations collected from different localities as defined by the tributaries in the Grand River basin, southwestern Ontario, Canada (see Supplemental Table S1).

Source of variation	Degrees of freedom	Sum of squares	% of variation
Among populations	4	452.3	28.9
Within populations	36	947.2	71.2

Table 3. Pairwise F_{ST} values based on AMOVA analysis and uncorrected P values using an 819-bp fragment of the *cytochrome c oxidase 1* gene of five clades of the peritrich ciliate *Carchesium polypinum* (see Figure 4).

	Clade 1	Clade 2	Clade 3	Clade 4	Clade 5
Clade 1		0.377	0.046	0.085	0.999
Clade 2	0.988		0.005	0.008	0.999
Clade 3	0.980 *	0.978 *		0.000	0.999
Clade 4	0.930	0.934 *	0.961 *		0.999
Clade 5	1.000	0.991	0.981	0.879	

Numbers in the lower diagonal are pairwise F_{ST} values, while those in the upper diagonal are the uncorrected P values. Asterisks denote values of statistical significance ($p \leq 0.05$)