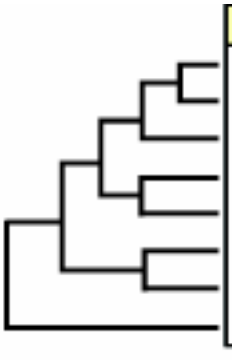
Tetrahymena Comparative Genomics Sequencing Project

A Revised White Paper Submitted to the NHGRI
Comparative Genome Evolution Working Group
March 28, 2007

Summary of Request *Tetrahymena thermophila* is by far the most well-developed model organism for basic and biomedical research among the entire Alveolate clade, a major and diverse assemblage of organisms that includes many human pathogens. The comparative genome sequence made available by this project would be invaluable to its further development and have direct benefits to the study of processes relevant to human health, such as chromatin structure, telomeres and telomerase, ciliary motility, RNAi, phagocytosis, regulated secretion, etc. In addition, the comparative *Tetrahymena* species were chosen not only for their varying degrees of phylogenetic relatedness to *T. thermophila*, but also for particular interesting traits: parasitism, inducible cellular differentiation, and asexuality. We expect this project to generate a wealth of valuable information and spur further research in ciliate biology for years to come.

- Phylogeny and requested WGS sequence coverage: This figure shows the relationships of the species in question (for more detail, see Figs. 2 and 3), along with the requested levels of sequence coverage. Numbers in parentheses indicate community priorities for each requested genome (Priority #5 = cDNA sequencing; MAC = macronucleus, MIC = micronucleus).



	MAC (~105 Mb)	MIC (~125 Mb)
T. thermophila	Done	6X (2)
T. malaccensis	3X (1)	3X (6)
T. elliotti	3X (1)	3X (6)
T. mobilis	6X (3)	-
T. rostrata	6X (3)	-
T. pyriformis	6X (4)	N.A.
T. vorax	6X (4)	N.A.
T. patula	6X (7)	-

- Major benefits of MAC comparative genomics:
  - Improved annotation of coding and non-coding features of the genome of *T. thermophila*, a valuable and popular model organism for molecular and cellular biology and free-living representative of the alveolate clade
  - Insight into adaptations within the genus such as parasitism, cyst formation, inducible morphological transformation, and asexuality

- Major benefits of MIC comparative genomics:
  - Better understanding of molecular basis and evolutionary history of genome-wide developmental DNA rearrangement
  - Improved power of forward genetics in *T. thermophila*
  - Understanding of germline chromosome structural elements, such as centromeres and telomeres
  - Sequence of MAC chromosomes lost during early vegetative proliferation, and further improvement of *T. thermophila* MAC genome assembly

Introduction:

The ciliated protozoan *Tetrahymena thermophila* is a model organism for a wide variety of basic, biomedical, and biotechnological studies. It is also the first free-living member of the alveolate clade to have had its genome sequenced (Eisen et al., 2006). The compelling arguments for the genome sequencing effort were thoroughly laid out in a white paper submitted to the NHGRI in 2002 (http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/*Tetrahymena*_Genome.pdf) and a concept paper submitted to the Trans-NIH NonMammalian Models Committee in 2001 (http://www.lifesci.ucsb.edu/~genome/*Tetrahymena*/SeqInitiative/SeqInitiative.htm). These arguments, and the unified support of the *Tetrahymena* research community, led the NHGRI to classify *T. thermophila* as one of six "high priority" projects in the first round of model organism genome priority assignments in 2002 (http://www.genome.gov/10002851). The importance of this project was also recognized by the NIGMS and the NSF/USDA Microbial Genome Sequencing Program, and it was through these sources that funding was obtained to sequence the macronucleus.

Whole Genome Shotgun (WGS) sequencing was performed at The Institute for Genomic Research (TIGR) on purified macronuclear DNA to nine-fold coverage, and the first assemblies were released in 2003 (http://www.tigr.org/tdb/e2k1/ttg/). The *Tetrahymena* Genome Database (TGD; http://www.ciliate.org/) was established to develop and disseminate genomic resources for the research community (Stover et al., 2006). After extensive efforts in closure, EST sequencing, annotation, and other analyses, the initial description of the genome sequence was published in 2006 (http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.137/journal.pbio.0040286) (Eisen et al., 2006). Other publications, including more complete analysis of EST sequence data (http://tbestdb.bcm.umontreal.ca/searches/organism.php?orgID=TT), metabolic reconstruction (http://pathema.tigr.org:1555/), and comparative genomics with the distantly related ciliate *Paramecium tetraurelia* (Aury et al., 2006), are in progress.

The purpose of this white paper is to propose sequencing of the macronuclear (MAC) genomes of five to seven additional species of *Tetrahymena* as well as the micronuclear (MIC) genomes of *T. thermophila* and its two closest known relatives, *T. malaccensis* and *T. elliotti*. These sequence data will be invaluable in fully interpreting and making use of the *T. thermophila* genome sequence and provide critical insights into the remarkable genome reorganization processes of ciliates and several diverse adaptations found in the comparison species.

Sections of the white paper:

1. *T. thermophila* as an experimental organism
2. Summary of the *T. thermophila* macronuclear genome sequence analysis
3. Rationale for micronuclear genome sequencing
4. Phylogeny of the genus *Tetrahymena*
5. Characteristics of the selected species
6. Strategy and practical considerations
7. Plans for sequence analysis
8. Priority List

Tetrahymena thermophila **as an experimental organism**

Tetrahymena species are distributed globally in freshwater habitats. The genus was first brought into laboratory culture by Nobel laureate Andre Lwoff in 1923 and has been used extensively in a variety of experimental studies ever since (Asai and Forney, 2000; Collins and Gorovsky, 2005). The development (in the 1950s and 1960s) of inbred mating strains of the species that eventually became known as *T. thermophila* established it as the principal genetic model within the genus (Nanney and Simon, 2000). The efforts of many *T. thermophila* researchers have developed an impressive toolbox of experimental methods (Turkewitz et al., 2002), including gene knockouts by homologous recombination (Cassidy-Hanley et al., 1997; Yao and Yao, 1991), gene silencing by RNAi (Howard-Till and Yao, 2006) and "antisense ribosomes" (Sweeney et al., 1996), high frequency transformation by high or low copy-number vectors (Gaertig et al., 1994), conditional gene promoters (Boldrin et al., 2006; Shang et al., 2002), epitope and affinity tagging (Yu and Gorovsky, 2000), genetic tricks to make recessive mutations homozygous in a single generation (Cole and Bruns, 1992), and genetic mapping by either deletions or meiotic recombination (Hamilton and Orias, 2000). A pending NIGMS proposal (1R24GM081560-01) by the presenters of this white paper seeks funding to develop "cloning by complementation" capability (N.B. this proposal has received a favorable priority score). In addition to its genetic strengths, *Tetrahymena* grows very rapidly in simple or even chemically defined media to high cell density and can be easily fractionated, subjected to proteomic analysis (Bowman et al., 2005; Garcia et al., 2006; Jacobs et al., 2006), labeled with radioactive tracers, patch-clamped (Hennessey and Kuruvilla, 2000), microinjected (Chalker et al., 2000), and cytologically stained to reveal its complex animal-like cellular structure (Figure 1).

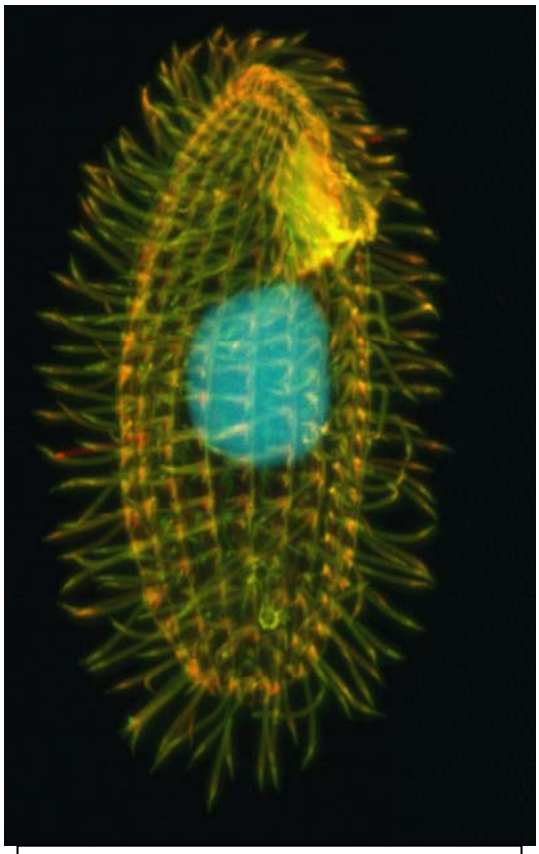


Figure 1: *Tetrahymena thermophila* stained with DAPI and anti-tubulin antibody (image courtesy of J. Gaertig).

Groundbreaking discoveries made using *T. thermophila* include catalytic RNA (Zaug and Cech, 1986), the structure of telomeres (Blackburn and Gall, 1978), telomerase (Greider and Blackburn, 1985), the function of histone acetylation (Brownell et al., 1996), the first microtubule motor (Gibbons and Rowe, 1965), and RNAi-guided genome rearrangement (Mochizuki et al., 2002). It is an active model organism for many studies with direct relevance to human health, including (only representative references given) telomerase function (Blackburn et al., 2006), regulated secretion (Turkewitz, 2004), ciliary motility (Liu et al., 2005), the post-translational modification of tubulins (Janke et al., 2005), chromatin modification (Taverna et al., 2002), phagocytosis (Jacobs et al., 2006), RNAi (Mochizuki and Gorovsky, 2004), chromosome condensation and segregation (Cervantes et al., 2006a; Cervantes et al., 2006b), signal

transduction (Christensen et al., 2003; Li et al., 2006), apoptosis (Mpoke and Wolfe, 1996), mitotic checkpoints (Yakisich et al., 2006), immunoglobulin-like DNA rearrangement (Coyne et al., 1996), and DNA replication (Morrison et al., 2005), amplification (Kapler, 1993), and copy number control (Blomberg et al., 1997). Because it can be grown on inexpensive media to very high cell density, *T. thermophila* has been developed in the biotechnology industry as a cellular factory for protein production (Weide et al., 2006). In particular, its phylogenetic proximity to other alveolates makes it potentially suitable for production of vaccines targeted against parasites such as *Ichthyophthirius* (a fish-parasitic ciliate that is a common pest in aquaculture and home aquariums) (Gaertig et al., 1999) and *Plasmodium* (a group of apicomplexan parasites that cause malaria and other diseases in humans and other mammals (Peterson et al., 2002).

The ciliate research community currently includes more than 300 active molecular and cell biologists in ~150 research groups (382 subscribe to the Ciliate Molecular Biology list-serve); the majority works with *Tetrahymena* species (mainly *thermophila*). Additional ciliate investigators work on areas of ecology and evolution. The community has in the most recent five years published over 100 papers listed in PubMed per year. This is comparable to the number published on two other unicellular model organisms, *Chlamodymonas* and *Dictyostelium*. Many of the recent papers benefited directly from the availability of the complete *T. thermophila* macronuclear genome sequence, EST sequences, and structural annotation.

Summary of the *T. thermophila* macronuclear genome sequence analysis

Only a brief summary is presented here. Full details of the analyses are freely available at: http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0040286 (Eisen et al., 2006).

Whole genome shotgun sequencing of purified MAC DNA was performed to 9X coverage at TIGR. The resulting ~104 Mb assembly is of very high quality and completeness. Nearly 50% of the genome is completely closed. Most of the physical gaps that remain are short (<1 kb). The Orias laboratory has recently received support from the NCRR to complete the HAPPY mapping of the MAC genome (Hamilton et al., 2006) (and map the MIC; see below), which is expected to order and orient nearly all of the remaining incomplete scaffolds (those not capped with telomeres at both ends) within one year. In addition, the pending NIGMS proposal to develop a "cloning by complementation" resource for *T. thermophila* involves generation of a tiled set of large-insert (12-20 kb) end-sequenced clones. If funded, this effort will further advance complete closure of the *T. thermophila* MAC genome. Because the MAC genome is derived from that of the MIC, the MIC sequencing proposed below will also greatly contribute to this goal of a 100% closed genome sequence, a rarity for genomes of this size.

Automated and manual annotation of the MAC genome assembly resulted in a prediction of over 27,000 protein-coding genes. EST evidence supports over 40% of these gene predictions (Coyne et al., ms. in preparation). This gene number is over four times that of the most popular unicellular eukaryotic model, *Saccharomyces cerevisiae*. Analysis of orthologs shared between humans, yeast, and *Tetrahymena* shows that *Tetrahymena* has retained many ancestral eukaryotic functions lost in the more streamlined yeast genome. Even after collapsing recent gene expansions into ortholog sets, there are 874 human genes with orthologs in *Tetrahymena* but not *S. cerevisiae*, 58 of which correspond to loci associated with human disease (Eisen et al., 2006). This highlights the importance and utility of genetic and genomic analysis with this organism for the complete understanding of many general eukaryotic functions, including those with direct relevance to human

health. In addition, 419 ortholog sets are shared between *Tetrahymena* and *Plasmodium falciparum* (the major causative agent of malaria) but not yeast. The power of *Tetrahymena* genetics and cell biology may prove useful in elucidating the functions of these genes and developing tools to combat this disease.

Rationale for micronuclear genome sequencing

As is typical of ciliates, *T. thermophila* contains two very different types of nucleus within its single cell. The macronucleus is the site of all known gene expression. It is polyploid (rDNA ~9,000C; all other loci ~45C), and its chromosomes segregate amitotically, with no mechanism for equal partitioning of alleles. The micronucleus is diploid, transcriptionally silent, and divides mitotically during vegetative growth. During the sexual process of conjugation, the MICs of each mating partner undergo meiosis, followed by exchange of haploid gametic nuclei and cross-fertilization to generate zygotic MICs (Karrer, 2000). New MACs develop from mitotic copies of the MICs. The parental MAC is destroyed through an apoptosis-like mechanism (Davis et al., 1992; Endoh and Kobayashi, 2006; Mpoke and Wolfe, 1996) and does not directly contribute DNA to the new generation. Thus, ciliates display the most evolutionarily basal form of separation of germ line (MIC) and somatic (MAC) genetic functions.

Although the MAC is derived from the MIC, the two genomes are dramatically different in sequence content and organization. There are five MIC chromosomes, but programmed chromosome breakage and *de novo* telomere formation at sites of the 15 bp Chromosome breakage sequence (Cbs) result in approximately 200-250 MAC chromosomes (Eisen et al., 2006; Yao et al., 1990). In addition, several thousand Internally Eliminated Sequences (IESs) are removed, ranging in length from several hundred to at least tens of thousands of base pairs (Coyne et al., 1996). Most of the repetitive sequences in the MIC genome are eliminated from the MAC, either as IESs or as (so far uncharacterized) terminal deletions, accounting for an estimated 15 to 20 additional Mb specific to the germ line (Yao and Gorovsky, 1974). Prior to the MAC genome project, only a handful of IESs had been cloned and sequenced, but minor (unavoidable) contamination of TIGR's MAC libraries with MIC DNA inserts led to a small, but significant representation of MIC-specific reads, revealing a surprising diversity of transposon-related sequences (Eisen et al., 2006). It is hypothesized that the elaborate genome rearrangement process of ciliates is a form of self-defense against invasion of the somatic genome by such disruptive mobile elements (Yao et al., 2003).

Given that the MIC-specific sequences are primarily repetitive and transcriptionally silent, what are the reasons for determining their sequence organization?

- The study of genome rearrangement is one of the most active fields within the ciliate research community. The phenomenon of genome-wide programmed rearrangement is governed by RNAi interactions and specific histone modifications that show its clear affinity to heterochromatin formation mechanisms nearly universal among eukaryotes. Despite this progress, only a few IESs have been studied in detail; the overall sequence organization of eliminated DNA remains a mystery. It is unknown whether any IESs are found within protein-coding regions of genes, as is common in some other ciliates. The basis for the apparent alternative genome rearrangements that determine the establishment of mating type (Orias, 1981) and surface antigen (Doerder and Berkowitz, 1987) mutually exclusive expression (and perhaps other gene expression systems) are also unknown.
- For the purposes of standard Mendelian genetics, the MIC is the nucleus on which recombinational maps are based. Thus, for a viable system of forward genetics in this

genomic era, the full MIC genome sequence will be a vital tool. *T. thermophila* is very well-suited to forward and reverse genetics, and many mutant stocks with phenotypes of interest to biomedicine (such as ciliary motility, phagocytosis, regulated secretion, etc.) are maintained in the research community, awaiting the proper tools, such as positional cloning and cloning by complementation, to allow molecular analysis. Availability of such tools would spur further genetic screens, as evidenced by the large number of letters in support of the pending "cloning by complementation" proposal (1R24GM081560-01).

- Functional non-coding elements are also present in the MIC-specific DNA, such as centromeres and MIC telomeres.
  - MAC chromosomes divide amitotically, and immunolocalization of the centromeric histone Cna1p during MAC differentiation indicates that centromeres are among the sequences eliminated (M.A. Gorovsky; pers. comm.). Chromatin immunoprecipitation could be used to isolate Cna1p-binding DNA, but the genomic context of the MIC centromeres can only be revealed by large scale sequencing. Knowledge of the structure of centromeres of an evolutionarily divergent and genetically tractable organism will further our understanding of this essential chromosomal component. Although assembly of centromeric DNA sequences is generally problematic, the overall organization can be iteratively refined using centromeric markers (H. S. Malik, pers. comm.) Because loss of one or more (but not all) MIC chromosome is not a lethal event, genetic dissection of centromeres is very feasible.
  - MIC telomeres are also eliminated from the MAC and have a distinctive and intriguing structure (Kirk and Blackburn, 1995). It is unknown how much telomere-adjacent DNA is eliminated or what role it may play in the function of MIC telomeres. As with centromeres, genetic dissection is highly feasible.
  - Sequence organization around eukaryotic centromeres and telomeres is typically influenced, on an evolutionary timescale, by the silencing associated with the unusual chromatin around these structures (Barry et al., 2003; Mefford and Trask, 2002). Because the micronucleus is transcriptionally silent, it should be immune to such selective forces, which may have resulted in unusual, informative organizations.
- Strong evidence exists that at least two chromosomes found in MACs immediately following conjugation are lost within at most 100 vegetative divisions (Cassidy-Hanley et al., 2005); thus, they are not represented in the current MAC genome sequence assembly, which was based on genomic DNA isolated long after 100 fissions. It is possible that these and/or other "lost" chromosomes contain functional genes required for early proliferation, in particular the period of sexual immaturity that persists for approximately 65 vegetative fissions following conjugation (Rogers and Karrer, 1985).
- For reasons not understood, the MIC of *T. thermophila* (as well as other micronucleate ciliates that have been investigated) is essential for vegetative viability. Induction of an amicronucleate state by chemical disruption of mitosis leads to gradual loss of the oral apparatus, disordering of cortical structure, and eventual death (Haremaki et al., 1995; Haremaki et al., 1996). The only case of a viable amicronucleate *T. thermophila* strain arose by an unknown mechanism in a mutagenesis experiment (Kaney, 1985; Kaney and Speare, 1983) and was shown to contain sequences that are ordinarily MIC-limited within its MAC (Karrer et al., 1984). Understanding the basis of this MIC requirement for vegetative growth, in particular cortical structure maintenance, will require knowledge of the MIC genome sequence.

- Preliminary results indicate that the arrangement of IESs between even closely related *Tetrahymena* species is highly variable (Huvos, 1995), and in press). Because at least some IESs are thought to be derived from mobile elements, these changes may reflect recent invasion events and/or continued mobilization within the MIC genome that may influence speciation. Whole MIC genome sequence comparison between *T. thermophila*, *T. malaccensis*, and *T. elliotti* will shed light on the nature of these variations in IES location and sequence content.
- Whereas the Chromosome breakage sequence (Cbs) is highly conserved among tetrahymenine ciliates, the number of copies and orientation of Cbs at each end of the MIC rDNA locus was shown to be variable (Coyne and Yao, 1996), suggesting the possibility of MIC chromosomal rearrangements associated with the endpoints of MAC chromosomes. Again, comparative MIC genome sequencing will shed light on this hypothesis. Even within the genome of *T. thermophila*, there is evidence of possibly recent Cbs duplication events; among 30 Cbs junctions characterized (Cassidy-Hanley et al., 2005), two pairs of junctions display exceptionally high sequence conservation on both sides of Cbs itself. A full survey of the 200-250 *T. thermophila* Cbs sites will reveal further details of the evolutionary history of chromosome breakage sites.

Phylogeny of the genus *Tetrahymena*

Over 40 species have been recognized within the *Tetrahymena* genus, but because a.) free-living and parasitic species are found throughout the world, b.) only limited collections have been attempted, and c.) analysis has focused on readily culturable specimens, it is doubtless that many other species exist in nature (D. Nanney, pers. comm.). Several studies of conserved phylogenetic markers (Brunk et al., 1990; Nanney et al., 1998; Preparata et al., 1989; Sadler and Brunk, 1992; Struder-Kypke et al., 2001) support the phylogenetic tree shown in Figure 2.

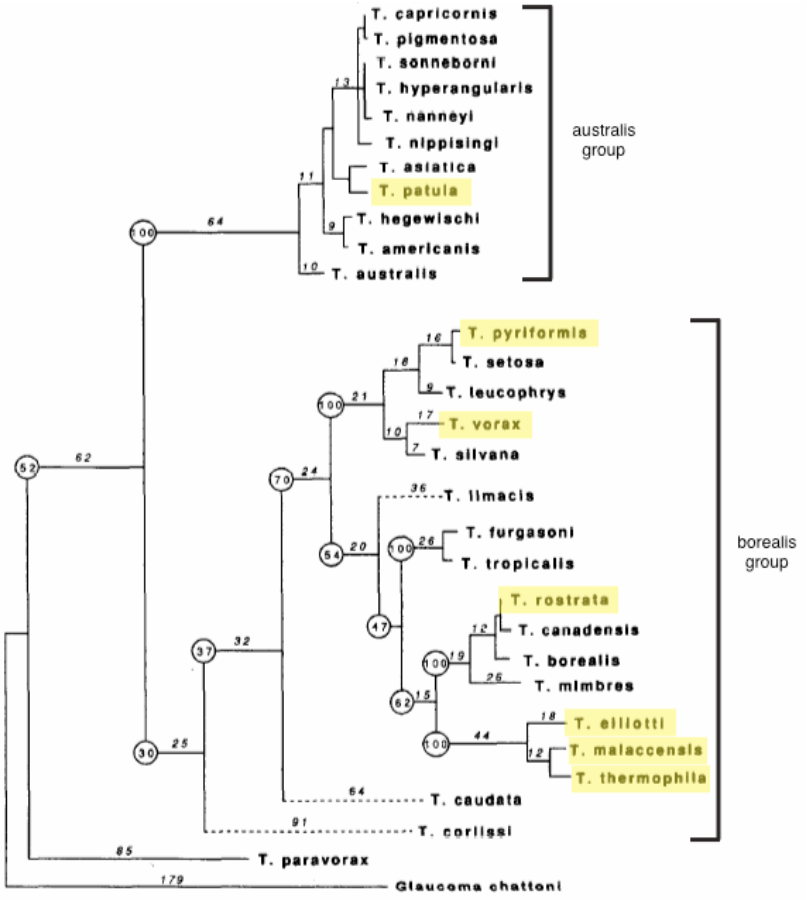


Figure 2: Maximum parsimony tree for Tetrahymena genus, based on Histone H3II/H4II intergenic region. Italic numbers are branch lengths in nucleotide substitutions. Numbers in circles are bootstrap values. Species of interest to this proposal are boxed in yellow. Reprinted from Sadler and Brunk, 1992.

Characteristics of the selected species

Numerous recent studies have demonstrated the enormous value of sequence information from closely related species to aid in the annotation of coding and non-coding features of a reference genome (a model organism, or one of medical or economic importance) (Bergman et al., 2002; Cliften et al., 2003; El-Sayed et al., 2005; Kellis et al., 2003; Stein et al., 2003). Theoretical considerations (Eddy, 2005) and empirical evidence indicate that the sequencing of several species at a range of evolutionary distances is most valuable for recognizing conserved sequence elements of various lengths, from transcription factor binding sites, which may be as short as six base pairs, to exons up to several thousand base pairs. To maximize the usefulness of the genome sequence for comparisons with the reference *T. thermophila*, and also sample the genetic, behavioral, morphological, and ecological diversity within the *Tetrahymena* genus, we have chosen six species from the borealis group (see Figure 2): *T. malaccensis*, *T. elliotti*, *T. mobilis*, *T. rostrata*, *T. pyriformis*, and *T. vorax*. For a member of the more distant australis group, we recommend *T. patula.*

Figure 3 shows a comparison of the relative genetic distances between *T. thermophila* and the borealis group species vs. the distances between the model organisms *D. melanogaster* and *S. cerevisiae* and several of their related species that have been fully or partially sequenced to provide invaluable comparative data. *T. malaccensis*, *T. elliotti*, *T. mobilis*, and *T. rostrata* are positioned at the low to moderate genetic distances from *T. thermophila* that have proven ideal for comparative annotation purposes in *Drosophila* and *Saccharomyces* (Bergman et al., 2002; Cliften et al., 2003; Kellis et al., 2003; Richards et al., 2005). The CNA1 gene sequences of *T.*

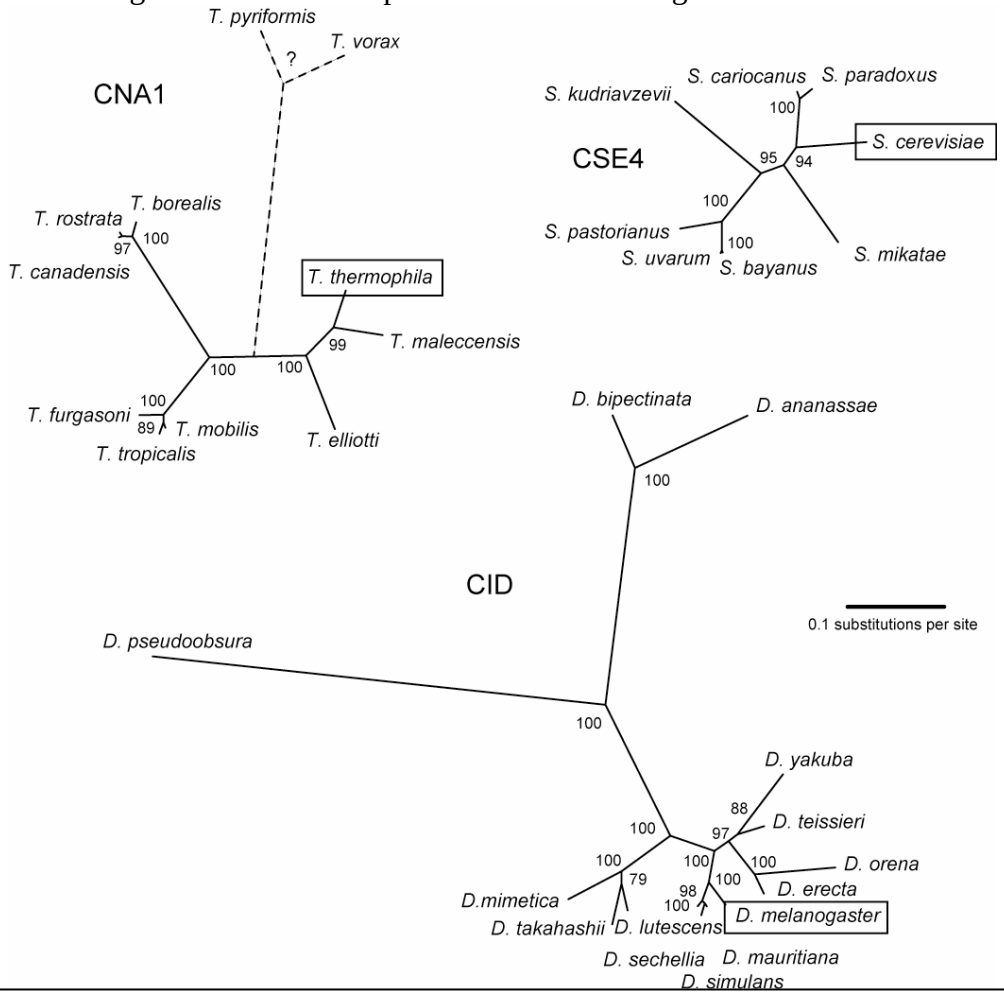


Figure 3. Unrooted neighbor-joining phylogenetic trees of the centromere-specific histone whole gene sequences for species of *Drosophila*, *Saccharomyces* and *Tetrahymena*. Bootstrapping was performed with 100 data-sets and trees were adjusted to the same scale such that branch lengths are comparable between the genera.

vorax, *T. pyriformis* and *T. patula* are not yet known, but based on other phylogenies (e.g. Figure 2), they may be presumed to be more distant from *T. thermophila*. Sequencing of these species will provide access to a more basal level of the tree, as well as data on their special characteristics (see below).

T. malaccensis, T. elliotti

T. malaccensis and *T. elliotti* were selected primarily because of their close relatedness to *T. thermophila*. Biogeographically, *T. elliotti* is the most cosmopolitan of the three, having been collected on several continents (http://www.life.uiuc.edu/nanney/tetrahymena /biogeography.html). Although *T. malaccensis* is *T. thermophila*'s closest known relative, the former has only been collected in Malaysia and the latter only in northeastern North America. Morphologically, and by most criteria of growth and nutrition, the three species are almost indistinguishable. The strains to be used are fertile and each has five MIC chromosomes, as does *T. thermophila* (Dr. H.S. Malik, unpublished data). Preliminary analysis of randomly selected and end-sequenced clones from genomic libraries of *T. malaccensis* and *T. elliotti* indicates a high degree of synteny between them and *T. thermophila* (Dr. C. Brunk, unpublished data).

T. mobilis, T. rostrata

Several tetrahymenine ciliate species are facultative or obligate parasites. The best studied of these is *Ichthyophthirius multifiliis*, a cause of "white spot" disease in fish, also known as "Ich" (Dickerson and Clark, 1998). Ich is a pest of major economic impact on the aquaculture industry. Substantial EST sequencing has been done with *I. multifiliis* and a whole genome shotgun sequencing proposal has been submitted to the NSF/USDA Joint Microbial Genome Sequencing Program (R. Coyne, T. Clark, and D. Cassidy-Hanley). The sequence of the histophagous (feeding on tissue) *T. mobilis* and *T. rostrata*, which are facultative parasite of various invertebrates (Corliss, 1960), would provide additional comparative data to evaluate how ciliates have adapted to this alternative lifestyle. Being relatively closely related to the genetically tractable model organism *T. thermophila* will be advantageous in this type of analysis. *T. rostrata* also undergoes morphological transformation during its life cycle and can inducibly form reproductive or resting cysts (Szablewski, 1993). During reproductive cyst formation, it undergoes autogamy (self-fertilization) (Gutierrez, 1985), which could potentially make classical genetic analysis feasible.

T. pyriformis

All *T. pyriformis* strains are amicronucleate and thus incapable of conjugation. The designation of species among asexual organisms is somewhat problematic, but based on conserved molecule sequence divergence, several other amicronucleate species related to *T. pyriformis* have been recognized; it is unclear how recently they diverged from micronucleate ancestors. As described above, induced loss of the MIC in *T. thermophila* and other ciliates is normally a lethal event. The only known amicronucleate *T. thermophila* strain has retained normally MIC-limited DNA in its MAC (Karrer et al., 1984). It will thus be of great interest to see whether the *T. pyriformis* MAC genome contains sequences related to *T. thermophila* MIC-limited DNA, and in particular to those sequences retained in the amicronucleate mutant.

T. pyriformis has also existed in nature for perhaps millennia with no apparent requirement for mitosis, meiosis, and the many functions associated with conjugation, such as cell fusion, nuclear exchange, nuclear determination, programmed genome rearrangement, and apoptotic nuclear death. Genes required only for these activities may have deteriorated by mutation or loss;

comparative genomic analyses are likely to facilitate their identification and provide insight into their function in *T. thermophila*, as well as the evolutionary consequences of irreversible transition to an asexual lifestyle.

Besides its genetic, developmental and evolutionary interest, *T. pyriformis* has been and remains a popular model organism in its own right. In fact, the strain selected for sequencing is the original *Tetrahymena* grown in laboratory culture by Andre Lwoff in 1923. *T. pyriformis* continues to be used for certain biochemical and toxicological studies, and is considered a natural reservoir for *Legionella pneumophila*, the causative agent of Legionnaire's disease, and a model for human infection (Cianciotto and Fields, 1992; Fields et al., 1984). These studies would be greatly facilitated by the availability of complete genome sequence information.

T. vorax, T. patula

T. vorax, also amicronucleate, is studied by several laboratories primarily as a model for the differentiation of specialized cell types, a process fundamental to the development of complex multicellular organisms (Ryals et al., 2002). When provided with bacterial prey or rich axenic medium, *T. vorax* resembles *T. thermophila* in its oral apparatus structure, suited for phagocytosis of relatively small particles; this cell form is known as a microstome (Figure 4, left). However, when starved and exposed to potential ciliate prey (*T. pyriformis* is commonly used in the laboratory) or an extract prepared from such prey, the cells undergo a dramatic, synchronous transformation to a carnivorous macrostome form (Figure 4, right). Most strikingly, the oral apparatus is greatly enlarged to accommodate the ingestion of prey in a volume range of about 4 orders of magnitude. This remodeling involves induction of macrostome-specific gene expression (Green et al., 2000), deposition in the oral apparatus of multiple novel polypeptides (Gulliksen et al., 1984), reorganization of cytoskeletal proteins (McLaughlin and Buhse, 2004), and modification of lipid composition (Ryals et al., 1989). The chemical signal that induces this transformation has been investigated and found to consist of a complex of iron and nucleic acid catabolites (Smith-Somerville et al., 2000). In addition to the polymorphic microstome/macrostome cell types, either form of *T. vorax* can undergo the formation of reproductive cysts; encased in a gelatinous membrane, binary fission occurs to produce 2 to 16 cells that rapidly assume the shape of either microstomes or macrostomes.

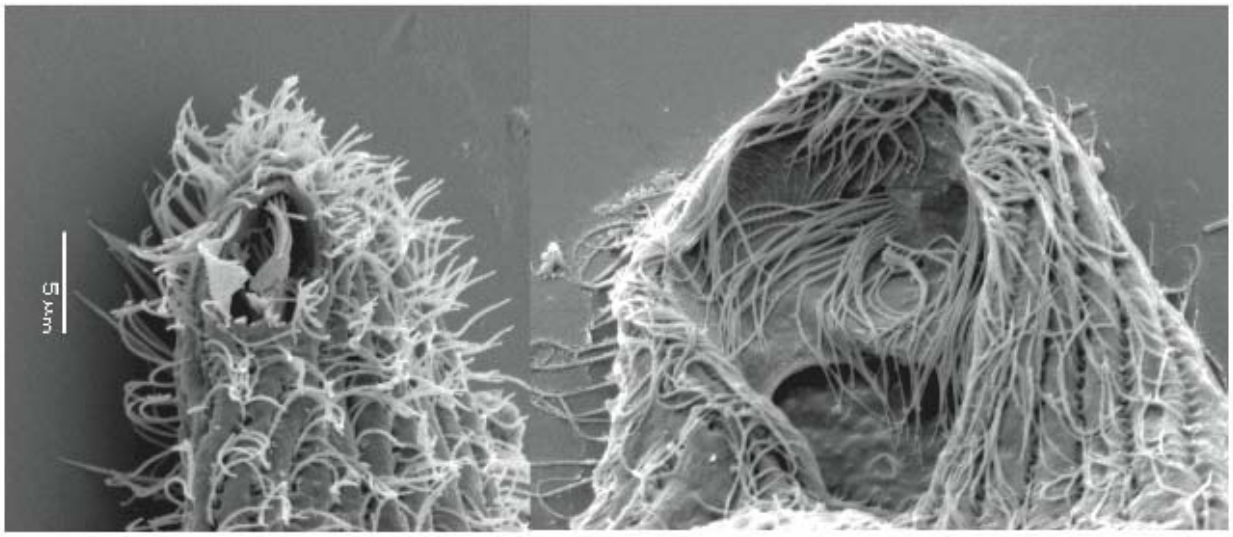


Figure 4: Scanning electron micrographs of the microstome (left) and macrostome (right) forms of the T. vorax oral apparatus (scale bar applies to both images; courtesy of H. Buhse).

Other *Tetrahymena* species, including *T. patula*, *T. caudata*, and *T. paravorax*, have evolved the microstome/macrostome transformation capability, apparently independently (Struder-Kypke et al., 2001). Interestingly, mutation of a single gene ("pseudomacrostome") in *T. thermophila* transforms its oral apparatus to a form that somewhat resembles that of the macrostomal form of *T. vorax* and others (Frankel et al., 1984). Genomic analysis will shed light on how the macrostomal form develops and how it arose during evolution and provide much-needed tools of analysis to the investigators studying this differentiation process. In addition, genome sequence data from *T. patula* will provide access to a more basal level of the Tetrahymena tree and open the australis group (the species of which appear to be distributed at convenient distances for comparative genomic purposes; see Figure 2) to easier analysis by degenerate PCR gene amplification and sequencing.

Strategy and practical considerations

All the species described above can be readily grown in simple axenic media and have no known symbionts. Methods for preparing MAC and/or MIC DNA are similar to those in use with *T. thermophila* (Gorovsky et al., 1975). Dr. Piroska Huvos, who has extensive experience purifying both MAC and MIC DNA from a wide variety of *Tetrahymena* species (Huvos, 1995), has agreed to provide the required samples. Of course, for the amicronucleate species, *T. pyriformis* and *T. vorax*, there will be no concerns about contamination of the MAC DNA with that of the MIC. We already have in hand a sample of highly pure *T. thermophila* MIC DNA prepared in the Gorovsky laboratory. Plasmid sequencing libraries of 2-4 kb and 4-6 kb insert size ranges have been prepared at TIGR and subjected to quality control sequencing that indicates a high success rate for each and the expected ratio of apparently MIC-limited to MAC-retained sequences. These libraries are available for full-scale shotgun sequencing.

The *T. thermophila* MAC WGS assembly was highly successful, even without large insert clones, which we were unable to obtain due to the high AT content of *Tetrahymena* DNA in general. Nevertheless, the availability of larger insert clones would be very valuable in future assemblies, especially of the MIC genomes, which have a much higher representation of repetitive sequences. Fortunately, we have recently been successful in obtaining high quality *T. thermophila* MAC libraries in the 12-20 kb insert range in a novel linear *E. coli* vector from Lucigen Corp. (http://www.lucigen.com/catalog/index.php?cPath=14_15_22). We expect that end-sequencing from a mixture of short, medium, and long insert libraries will be most efficient in producing high quality assemblies of both MAC and MIC genomes. Sequencing by newer technologies, such as the pyrosequencing method of 454 Life Sciences (Margulies et al., 2005; Ronaghi et al., 1998), or by a hybrid Sanger/454 approach (Goldberg et al., 2006), is another option, which we have proposed in the pending NSF/USDA *Ichthyophthirius* genome proposal. In addition, the Orias laboratory has been funded by the NCRR to perform HAPPY mapping to link the *T. thermophila* MAC chromosomes in their proper order and orientation in the MIC genome. The long-range linking information generated will be helpful in complete MIC genome assembly.

Because the MAC genome sequence is entirely represented in the MIC genome, sequencing of the *T. thermophila* MIC will provide additional coverage to help completely close the MAC. For *T. malaccensis* and *T. elliotti*, however, we believe the redundancy of high MAC coverage that would be obtained by, for example, 6X coverage of each (thus 12X for the MAC) would be impractical. Thus, we propose 3X coverage of each genome, which would provide 6X coverage of the MAC (which is more important for comparative genomic purposes in structural annotation of *T.*

thermophila) and also a wealth of information on MIC genome organization. We propose 6X coverage for the MAC genomes of the other species.

We believe a whole genome shotgun approach will be more practical and far more informative than a targeted approach for MIC genome sequencing because of the lower cost per unit of sequence data for the WGS approach and because, given our ignorance of MIC genome organization and its apparent high interspecific variability (Huvos, 1995) and in press), knowing *a priori* what to target would be highly problematic. In addition, tetrahymenine species are the best candidates among commonly studied ciliates for full MIC genome sequencing because their MIC genomes are not as large as some others (which may be over ten-fold the size of the MAC) and have a lower ploidy difference from the MAC (lowering the problem of cross-contamination).

The *T. thermophila* MIC genome has been estimated by reassociation kinetics to be 15-20% larger than that of the MAC (Yao and Gorovsky, 1974), i.e. about 120-125 Mb. Direct genome size estimates for the MACs of other species are somewhat uncertain because of the possibility that ploidy is significantly different than *T. thermophila*, but comparative genomic Southern analysis indicates that all species examined have roughly comparable genome size (P. Huvos; personal communication). For the micronucleate species described above, we are pursuing MIC genome size estimation relative to *T. thermophila* through the method of Feulgen image analysis densitometry (D.H. Lynn, T.R. Gregory; personal communication). We expect the genome sizes to be comparable.

Another practical consideration in genome assembly is the level of sequence polymorphism. The *T. thermophila* strain we propose sequencing is SB210, the same species used for MAC genome sequencing. It is derived from a highly inbred strain, and furthermore, SB210 underwent a recent "genomic exclusion" mating (Allen, 1967), which induces homozygosity of the MIC. Thus, polymorphism of the *T. thermophila* MIC will be minimal. *T. rostrata* has been documented to undergo autogamy (self-fertilization) (Gutierrez, 1985), which would also result in an inbred, low level of germline polymorphism. Because *Tetrahymena* MACs divide amitotically, they are also subject to another genetic phenomenon that minimizes lineage-specific polymorphism, called "phenotypic assortment". Due to stochastic fluctuations in allele segregation at each vegetative cell division, all, or nearly all, loci become homozygous in a given lineage within a few hundred cell generations. Prior to DNA extraction, a single cell from each species will be isolated and used as the starting point for the culture to minimize heterozygosity. Even without such precautions, polymorphism has not been detected in multiple sequenced clones from a variety of species, including the strains of *T. malaccensis*, *T. elliotti*, *T. mobilis*, and *T. rostrata* that we propose to sequence (N. Elde, H. Malik; personal communication). Polymorphism within the MICs of *T. malaccensis* and *T. elliotti* may be somewhat problematic for full assembly, but we cannot currently assess to what degree. In any case, the junctions between MIC-limited and MAC-retained sequences will not be difficult to recognize.

Plans for sequence analysis

A high priority for the *Tetrahymena* research community is improved annotation of the *T. thermophila* MAC genome sequence, for which comparative genomics would be the most efficient and effective approach. The community was consulted on the proposal described herein (through the Ciliate Molecular Biology list-serve group of 382 members), and we received universal support for the plan. Undoubtedly, individual investigators would make use of the comparative sequences made available to better annotate genes of interest. However, a more comprehensive genomic

analysis would be far more preferable. We believe this would be best achieved using the existing infrastructure and expertise available through the *Tetrahymena* Genome Database, TGD. The director of TGD, Dr. Mike Cherry of Stanford University, did his doctoral thesis research on *T. thermophila* and has a strong interest in maintaining and expanding the capability of this resource. Dr. Cherry also directs the *Saccharomyces* Genome Database and has access to exactly the bioinformatic expertise necessary to make optimal use of the additional *Tetrahymena* sequence data (Fisk et al., 2006). He has expressed his strong interest in attracting a graduate student or postdoc to perform a thorough comparative analysis on the data and making it available to the community through TGD's user-friendly web interface. He has colleagues at Stanford in the Computer Science department who would also likely be interested (personal communication). In addition, Dr. Nick Stover, the former head curator at TGD, has recently moved to a faculty position at Bradley University. Dr. Stover will remain active in TGD's development, and plans to involve graduate and undergraduate students in further genome annotation efforts (personal communication). The comparative analyses described here would make excellent projects for such students. The presenters of this proposal will also contribute to the comparative analysis to the greatest extent possible.

Sequencing priorities

1. MAC genomes of *T. malaccensis* and *T. elliotti*, because they will provide the most valuable comparative genomic information for improved annotation of the reference genome of *T. thermophila*. Note: coupling 3X MAC genome coverage with 3X MIC genome coverage (Priority #6) will essentially "kill two birds with one stone" in terms of the overall cost of sequencing.
2. MIC genome of *T. thermophila*, for greater understanding of programmed genome rearrangement and germline-specific chromosomal elements, as well as for strengthening the power of *Tetrahymena* forward genetics and advancing complete closure of the MAC genome.
3. MAC genomes of *T. rostrata* and *T. mobilis*, for comparative genomics at medium genetic distance from *T. thermophila* and because of these two species' adoption of obligate parasitism.
4. MAC genomes of *T. vorax* and/or *T. pyriformis* both for comparative genomic purposes and due to their inherent biological interest. We argue in favor of including both of these genomes because, apart from *T. thermophila*, they are the two species in most common use as experimental models and each offers unique traits of interest to biomedical science.
5. cDNA sequencing of each of the species selected (5 - 10,000 reads each) to provide greater power in accurate gene annotation.
6. MIC genomes of *T. malaccensis* and *T. elliotti* for comparative genomics that will greatly enhance the value of the *T. thermophila* MIC genome sequence.
7. An additional MAC genome from the more distant australis group, to provide broader taxonomic sampling of the genus and facilitate gene-targeted PCR amplification and sequencing from related species in this cluster. We propose *T. patula*, which undergoes microstome/macrostome transformation similar to *T. vorax* (although in a less controllable manner) and also forms reproductive cysts. These adaptations appear to have paraphyletic evolutionary origins.

Presenters:

Dr. Robert Coyne
The Institute for Genomic Research
rcoyne@tigr.org

Dr. Eduardo Orias
University of California, Santa Barbara
orias@lifesci.ucsb.edu

Additional Contributors:

Dr. Nels Elde
Fred Hutchinson Cancer Research Center

Dr. Harmit Malik
Fred Hutchinson Cancer Research Center

Dr. Clifford Brunk
University of California, Los Angeles

Dr. Howard Buhse
University of Illinois, Chicago

Dr. Piroska Huvos
University of Illinois, Carbondale

References:

Allen, S. L. (1967). Genomic exclusion: a rapid means for inducing homozygous diploid lines in Tetrahymena pyriformis, syngen 1. Science *155*, 575-577.
Asai, D. J., and Forney, J. D., eds. (2000). *Tetrahymena thermophila* (San Diego, CA: Academic Press).
Aury, J. M., Jaillon, O., Duret, L., Noel, B., Jubin, C., Porcel, B. M., Segurens, B., Daubin, V., Anthouard, V., Aiach, N., *et al.* (2006). Global trends of whole-genome duplications revealed by the ciliate Paramecium tetraurelia. Nature.
Barry, J. D., Ginger, M. L., Burton, P., and McCulloch, R. (2003). Why are parasite contingency genes often associated with telomeres? Int J Parasitol *33*, 29-45.
Bergman, C. M., Pfeiffer, B. D., Rincon-Limas, D. E., Hoskins, R. A., Gnirke, A., Mungall, C. J., Wang, A. M., Kronmiller, B., Pacleb, J., Park, S., *et al.* (2002). Assessing the impact of comparative genomic sequence data on the functional annotation of the Drosophila genome. Genome Biol *3*, RESEARCH0086.
Blackburn, E. H., and Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. J Mol Biol *120*, 33-53.
Blackburn, E. H., Greider, C. W., and Szostak, J. W. (2006). Telomeres and telomerase: the path from maize, Tetrahymena and yeast to human cancer and aging. Nat Med *12*, 1133-1138.

Blomberg, P., Randolph, C., Yao, C. H., and Yao, M. C. (1997). Regulatory sequences for the amplification and replication of the ribosomal DNA minichromosome in Tetrahymena thermophila. Mol Cell Biol *17*, 7237-7247.
Boldrin, F., Santovito, G., Gaertig, J., Wloga, D., Cassidy-Hanley, D., Clark, T. G., and Piccinni, E. (2006). Metallothionein gene from Tetrahymena thermophila with a copper-inducible-repressible promoter. Eukaryot Cell *5*, 422-425.
Bowman, G. R., Smith, D. G., Michael Siu, K. W., Pearlman, R. E., and Turkewitz, A. P. (2005). Genomic and proteomic evidence for a second family of dense core granule cargo proteins in Tetrahymena thermophila. J Eukaryot Microbiol *52*, 291-297.
Brownell, J. E., Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y., and Allis, C. D. (1996). Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell *84*, 843-851.
Brunk, C. F., Kahn, R. W., and Sadler, L. A. (1990). Phylogenetic relationships among Tetrahymena species determined using the polymerase chain reaction. J Mol Evol *30*, 290-297.
Cassidy-Hanley, D., Bisharyan, Y., Fridman, V., Gerber, J., Lin, C., Orias, E., Orias, J. D., Ryder, H., Vong, L., and Hamilton, E. P. (2005). Genome-Wide Characterization of Tetrahymena thermophila Chromosome Breakage Sites. II. Physical and Genetic Mapping. Genetics *170*, 1623-1631.
Cassidy-Hanley, D., Bowen, J., Lee, J. H., Cole, E., VerPlank, L. A., Gaertig, J., Gorovsky, M. A., and Bruns, P. J. (1997). Germline and somatic transformation of mating Tetrahymena thermophila by particle bombardment. Genetics *146*, 135-147.
Cervantes, M. D., Coyne, R. S., Xi, X., and Yao, M. C. (2006a). The condensin complex is essential for amitotic segregation of bulk chromosomes, but not nucleoli, in the ciliate Tetrahymena thermophila. Mol Cell Biol *26*, 4690-4700.
Cervantes, M. D., Xi, X., Vermaak, D., Yao, M. C., and Malik, H. S. (2006b). The CNA1 histone of the ciliate Tetrahymena thermophila is essential for chromosome segregation in the germline micronucleus. Mol Biol Cell *17*, 485-497.
Chalker, D. L., Ward, J. G., Randolph, C., and Yao, M. C. (2000). Microinjection of Tetrahymena thermophila. Methods Cell Biol *62*, 469-484.
Christensen, S. T., Guerra, C. F., Awan, A., Wheatley, D. N., and Satir, P. (2003). Insulin receptor-like proteins in Tetrahymena thermophila ciliary membranes. Curr Biol *13*, R50-52.
Cianciotto, N. P., and Fields, B. S. (1992). Legionella pneumophila mip gene potentiates intracellular infection of protozoa and human macrophages. Proc Natl Acad Sci U S A *89*, 5188-5191.
Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A., and Johnston, M. (2003). Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science *301*, 71-76.
Cole, E. S., and Bruns, P. J. (1992). Uniparental cytogamy: a novel method for bringing micronuclear mutations of Tetrahymena into homozygous macronuclear expression with precocious sexual maturity. Genetics *132*, 1017-1031.
Collins, K., and Gorovsky, M. A. (2005). Tetrahymena thermophila. Curr Biol *15*, R317-318.
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.
Coyne, R. S., Chalker, D. L., and Yao, M. C. (1996). Genome downsizing during ciliate development: nuclear division of labor through chromosome restructuring. Annu Rev Genet *30*, 557-578.

Coyne, R. S., and Yao, M. C. (1996). Evolutionary conservation of sequences directing chromosome breakage and rDNA palindrome formation in tetrahymenine ciliates. Genetics *144*, 1479-1487.
Davis, M. C., Ward, J. G., Herrick, G., and Allis, C. D. (1992). Programmed nuclear death: apoptotic-like degradation of specific nuclei in conjugating Tetrahymena. Dev Biol *154*, 419-432.
Dickerson, H., and Clark, T. (1998). Ichthyophthirius multifiliis: a model of cutaneous infection and immunity in fishes. Immunol Rev *166*, 377-384.
Doerder, F. P., and Berkowitz, M. S. (1987). Nucleo-cytoplasmic interaction during macronuclear differentiation in ciliate protists: genetic basis for cytoplasmic control of SerH expression during macronuclear development in Tetrahymena thermophila. Genetics *117*, 13-23.
Eddy, S. R. (2005). A model of the statistical power of comparative genome sequence analysis. PLoS Biol *3*, e10.
Eisen, J. A., Coyne, R. S., Wu, M., Wu, D., Thiagarajan, M., Wortman, J. R., Badger, J. H., Ren, Q., Amedeo, P., Jones, K. M., *et al.* (2006). Macronuclear Genome Sequence of the Ciliate Tetrahymena thermophila, a Model Eukaryote. PLoS Biol *4*.
El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., *et al.* (2005). Comparative genomics of trypanosomatid parasitic protozoa. Science *309*, 404-409.
Endoh, H., and Kobayashi, T. (2006). Death harmony played by nucleus and mitochondria: nuclear apoptosis during conjugation of tetrahymena. Autophagy *2*, 129-131.
Fields, B. S., Shotts, E. B., Jr., Feeley, J. C., Gorman, G. W., and Martin, W. T. (1984). Proliferation of Legionella pneumophila as an intracellular parasite of the ciliated protozoan Tetrahymena pyriformis. Appl Environ Microbiol *47*, 467-471.
Fisk, D. G., Ball, C. A., Dolinski, K., Engel, S. R., Hong, E. L., Issel-Tarver, L., Schwartz, K., Sethuraman, A., Botstein, D., and Cherry, J. M. (2006). Saccharomyces cerevisiae S288C genome annotation: a working hypothesis. Yeast *23*, 857-865.
Frankel, J., Nelsen, E. M., Bakowska, J., and Jenkins, L. M. (1984). Mutational analysis of patterning of oral structures in Tetrahymena. II. A graded basis for the individuality of intracellular structural arrays. J Embryol Exp Morphol *82*, 67-95.
Gaertig, J., Gao, Y., Tishgarten, T., Clark, T. G., and Dickerson, H. W. (1999). Surface display of a parasite antigen in the ciliate Tetrahymena thermophila. Nat Biotechnol *17*, 462-465.
Gaertig, J., Gu, L., Hai, B., and Gorovsky, M. A. (1994). High frequency vector-mediated transformation and gene replacement in Tetrahymena. Nucleic Acids Res *22*, 5391-5398.
Garcia, B. A., Joshi, S., Thomas, C. E., Chitta, R. K., Diaz, R. L., Busby, S. A., Andrews, P. C., Ogorzalek Loo, R. R., Shabanowitz, J., Kelleher, N. L., *et al.* (2006). Comprehensive phosphoprotein analysis of linker histone H1 from Tetrahymena thermophila. Mol Cell Proteomics *5*, 1593-1609.
Gibbons, I. R., and Rowe, A. J. (1965). Dynein: a protein with adenosine triphosphatase activity from cilia. Science *149*, 424-426.
Goldberg, S. M., Johnson, J., Busam, D., Feldblyum, T., Ferriera, S., Friedman, R., Halpern, A., Khouri, H., Kravitz, S. A., Lauro, F. M., *et al.* (2006). A Sanger/pyrosequencing hybrid approach for the generation of high-quality draft assemblies of marine microbial genomes. Proc Natl Acad Sci U S A *103*, 11240-11245.
Gorovsky, M. A., Yao, M. C., Keevert, J. B., and Pleger, G. L. (1975). Isolation of micro- and macronuclei of Tetrahymena pyriformis. Methods Cell Biol *9*, 311-327.

Green, M. M., LeBoeuf, R. D., and Churchill, P. F. (2000). Biological and molecular characterization of cellular differentiation in Tetrahymena vorax: a potential biocontrol protozoan. J Basic Microbiol *40*, 351-361.
Greider, C. W., and Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. Cell *43*, 405-413.
Gulliksen, O. M., Lovlie, A., and Kvammen, L. (1984). Different polypeptides in two homologous cellular structures: the microstomal and macrostomal oral apparatus of Tetrahymena vorax. Dev Biol *103*, 511-516.
Gutierrez, J. C. (1985). Microspectrophotometric study and kinetics of autogamy during encystment of Tetrahymena rostrata. Cell Biol Int Rep *9*, 169-173.
Hamilton, E. P., Dear, P. H., Rowland, T., Saks, K., Eisen, J. A., and Orias, E. (2006). Use of HAPPY mapping for the higher order assembly of the Tetrahymena genome. Genomics.
Hamilton, E. P., and Orias, E. (2000). Genetically mapping new mutants and cloned genes. Methods Cell Biol *62*, 265-280.
Haremaki, T., Sugai, T., and Takahashi, M. (1995). The vegetative micronucleus has a critical role in maintenance of cortical structure in Tetrahymena thermophila. Cell Struct Funct *20*, 239-244.
Haremaki, T., Sugai, T., and Takahashi, M. (1996). Involvement of active cellular mechanisms on the disorganization of oral apparatus in amicronucleate cells in Tetrahymena thermophila. Cell Struct Funct *21*, 73-80.
Hennessey, T. M., and Kuruvilla, H. G. (2000). Electrophysiology of Tetrahymena. Methods Cell Biol *62*, 363-377.
Howard-Till, R. A., and Yao, M. C. (2006). Induction of gene silencing by hairpin RNA expression in Tetrahymena thermophila reveals a second small RNA pathway. Mol Cell Biol.
Huvos, P. (1995). Developmental DNA rearrangements and micronucleus-specific sequences in five species within the Tetrahymena pyriformis species complex. Genetics *141*, 925-936.
Jacobs, M. E., Desouza, L. V., Samaranayake, H., Pearlman, R. E., Michael Siu, K. W., and Klobutcher, L. A. (2006). The Tetrahymena thermophila Phagosome Proteome. Eukaryot Cell.
Janke, C., Rogowski, K., Wloga, D., Regnard, C., Kajava, A. V., Strub, J. M., Temurak, N., van Dijk, J., Boucher, D., van Dorsselaer, A., *et al.* (2005). Tubulin polyglutamylase enzymes are members of the TTL domain protein family. Science *308*, 1758-1762.
Kaney, A. R. (1985). A transmissible developmental block in Tetrahymena thermophila. Exp Cell Res *157*, 315-321.
Kaney, A. R., and Speare, V. J. (1983). An amicronucleate mutant of Tetrahymena thermophila. Exp Cell Res *143*, 461-467.
Kapler, G. M. (1993). Developmentally regulated processing and replication of the Tetrahymena rDNA minichromosome. Curr Opin Genet Dev *3*, 730-735.
Karrer, K., Stein-Gavens, S., and Allitto, B. A. (1984). Micronucleus-specific DNA sequences in an amicronucleate mutant of Tetrahymena. Dev Biol *105*, 121-129.
Karrer, K. M. (2000). Tetrahymena genetics: two nuclei are better than one. Methods Cell Biol *62*, 127-186.
Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature *423*, 241-254.
Kirk, K. E., and Blackburn, E. H. (1995). An unusual sequence arrangement in the telomeres of the germ-line micronucleus in Tetrahymena thermophila. Genes Dev *9*, 59-71.

Li, S., Yin, L., Cole, E. S., Udani, R. A., and Karrer, K. M. (2006). Progeny of germ line knockouts of ASI2, a gene encoding a putative signal transduction receptor in Tetrahymena thermophila, fail to make the transition from sexual reproduction to vegetative growth. Dev Biol *295*, 633-646.
Liu, S., Hennessey, T., Rankin, S., and Pennock, D. G. (2005). Mutations in genes encoding inner arm dynein heavy chains in Tetrahymena thermophila lead to axonemal hypersensitivity to Ca2+. Cell Motil Cytoskeleton *62*, 133-140.
Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376-380.
McLaughlin, N. B., and Buhse, H. E., Jr. (2004). Localization by indirect immunofluorescence of tetrin, actin, and centrin to the oral apparatus and buccal cavity of the macrostomal form of Tetrahymena vorax. J Eukaryot Microbiol *51*, 253-257.
Mefford, H. C., and Trask, B. J. (2002). The complex structure and dynamic evolution of human subtelomeres. Nat Rev Genet *3*, 91-102.
Mochizuki, K., Fine, N. A., Fujisawa, T., and Gorovsky, M. A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. Cell *110*, 689-699.
Mochizuki, K., and Gorovsky, M. A. (2004). Small RNAs in genome rearrangement in Tetrahymena. Curr Opin Genet Dev *14*, 181-187.
Morrison, T. L., Yakisich, J. S., Cassidy-Hanley, D., and Kapler, G. M. (2005). TIF1 Represses rDNA replication initiation, but promotes normal S phase progression and chromosome transmission in Tetrahymena. Mol Biol Cell *16*, 2624-2635.
Mpoke, S., and Wolfe, J. (1996). DNA digestion and chromatin condensation during nuclear death in Tetrahymena. Exp Cell Res *225*, 357-365.
Nanney, D. L., Park, C., Preparata, R., and Simon, E. M. (1998). Comparison of sequence differences in a variable 23S rRNA domain among sets of cryptic species of ciliated protozoa. J Eukaryot Microbiol *45*, 91-100.
Nanney, D. L., and Simon, E. M. (2000). Laboratory and evolutionary history of Tetrahymena thermophila. Methods Cell Biol *62*, 3-25.
Orias, E. (1981). Probable somatic DNA rearrangements in mating type determination in *Tetrahymena thermophila*: a review and a model. Dev Genet *2*, 185-202.
Peterson, D. S., Gao, Y., Asokan, K., and Gaertig, J. (2002). The circumsporozoite protein of Plasmodium falciparum is expressed and localized to the cell surface in the free-living ciliate Tetrahymena thermophila. Mol Biochem Parasitol *122*, 119-126.
Preparata, R. M., Meyer, E. B., Preparata, F. P., Simon, E. M., Vossbrinck, C. R., and Nanney, D. L. (1989). Ciliate evolution: the ribosomal phylogenies of the tetrahymenine ciliates. J Mol Evol *28*, 427-441.
Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P., *et al.* (2005). Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. Genome Res *15*, 1-18.
Rogers, M. B., and Karrer, K. M. (1985). Adolescence in Tetrahymena thermophila. Proc Natl Acad Sci U S A *82*, 436-439.
Ronaghi, M., Uhlen, M., and Nyren, P. (1998). A sequencing method based on real-time pyrophosphate. Science *281*, 363, 365.
Ryals, P. E., Buhse, H. E., Jr., and Modzejewski, J. (1989). Lipid modification during cytodifferentiation of Tetrahymena vorax. Whole cell phospholipids and triacylglycerols of microstomal and macrostomal phenotypes. Biochim Biophys Acta *991*, 438-444.

Ryals, P. E., Smith-Somerville, H. E., and Buhse, H. E., Jr. (2002). Phenotype switching in polymorphic Tetrahymena: a single-cell Jekyll and Hyde. Int Rev Cytol *212*, 209-238.
Sadler, L. A., and Brunk, C. F. (1992). Phylogenetic relationships and unusual diversity in histone H4 proteins within the Tetrahymena pyriformis complex. Mol Biol Evol *9*, 70-84.
Shang, Y., Song, X., Bowen, J., Corstanje, R., Gao, Y., Gaertig, J., and Gorovsky, M. A. (2002). A robust inducible-repressible promoter greatly facilitates gene knockouts, conditional expression, and overexpression of homologous and heterologous genes in Tetrahymena thermophila. Proc Natl Acad Sci U S A *99*, 3734-3739.
Smith-Somerville, H. E., Hardman, J. K., Timkovich, R., Ray, W. J., Rose, K. E., Ryals, P. E., Gibbons, S. H., and Buhse, H. E., Jr. (2000). A complex of iron and nucleic acid catabolites is a signal that triggers differentiation in a freshwater protozoan. Proc Natl Acad Sci U S A *97*, 7325-7330.
Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., *et al.* (2003). The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol *1*, E45.
Stover, N. A., Krieger, C. J., Binkley, G., Dong, Q., Fisk, D. G., Nash, R., Sethuraman, A., Weng, S., and Cherry, J. M. (2006). Tetrahymena Genome Database (TGD): a new genomic resource for Tetrahymena thermophila research. Nucleic Acids Res *34*, D500-503.
Struder-Kypke, M. C., Wright, A. D., Jerome, C. A., and Lynn, D. H. (2001). Parallel evolution of histophagy in ciliates of the genus Tetrahymena. BMC Evol Biol *1*, 5.
Sweeney, R., Fan, Q., and Yao, M. C. (1996). Antisense ribosomes: rRNA as a vehicle for antisense RNAs. Proc Natl Acad Sci U S A *93*, 8518-8523.
Szablewski, L. (1993). The life cycle of Tetrahymena rostrata (Ciliata) in the laboratory. Acta Protozool *32*, 95-99.
Taverna, S. D., Coyne, R. S., and Allis, C. D. (2002). Methylation of histone h3 at lysine 9 targets programmed DNA elimination in tetrahymena. Cell *110*, 701-711.
Turkewitz, A. P. (2004). Out with a bang! Tetrahymena as a model system to study secretory granule biogenesis. Traffic *5*, 63-68.
Turkewitz, A. P., Orias, E., and Kapler, G. (2002). Functional genomics: the coming of age for Tetrahymena thermophila. Trends Genet *18*, 35-40.
Weide, T., Herrmann, L., Bockau, U., Niebur, N., Aldag, I., Laroy, W., Contreras, R., Tiedtke, A., and Hartmann, M. W. (2006). Secretion of functional human enzymes by Tetrahymena thermophila. BMC Biotechnol *6*, 19.
Yakisich, J. S., Sandoval, P. Y., Morrison, T. L., and Kapler, G. M. (2006). TIF1 Activates the Intra-S Phase Checkpoint Response in the Diploid Micronucleus and Amitotic Polyploid Macronucleus of Tetrahymena. Mol Biol Cell.
Yao, M. C., Fuller, P., and Xi, X. (2003). Programmed DNA deletion as an RNA-guided system of genome defense. Science *300*, 1581-1584.
Yao, M. C., and Gorovsky, M. A. (1974). Comparison of the sequences of macro- and micronuclear DNA of Tetrahymena pyriformis. Chromosoma *48*, 1-18.
Yao, M. C., and Yao, C. H. (1991). Transformation of Tetrahymena to cycloheximide resistance with a ribosomal protein gene through sequence replacement. Proc Natl Acad Sci U S A *88*, 9493-9497.
Yao, M. C., Yao, C. H., and Monks, B. (1990). The controlling sequence for site-specific chromosome breakage in Tetrahymena. Cell *63*, 763-772.

Yu, L., and Gorovsky, M. A. (2000). Protein tagging in Tetrahymena. Methods Cell Biol *62*, 549-559.
Zaug, A. J., and Cech, T. R. (1986). The intervening sequence RNA of Tetrahymena is an enzyme. Science *231*, 470-475.