Sequence of the *Oxytricha trifallax* macronuclear genome

Thomas G. Doak*, Glenn Herrick*, Laura F. Landweber†, Robert B. Weiss‡
*University of Utah; †Princeton University; ‡Utah Genome Center, University of Utah

Abstract

We propose complete sequencing of the macronuclear genome of the ciliated protozoan *Oxytricha trifallax* (Alveolate; class *Spirotrichea*). Ciliates have been important experimental organisms for over 100 years, contributing to the discovery and understanding of many essential cellular processes—including self-splicing RNA, telomere biochemistry, and transcriptional regulation by histone modification—with *Oxytricha* representing the lineage—the spirotrichs—with the very surprising discoveries of gene-sized macronuclear "chromosomes" and scrambled genes.

Of the 10 classes of ciliates, only oligohymenophorans and spirotrichs have been extensively studied by molecular geneticists, and inter-species comparisons between these two abound and drive the research community. Two oligohymenophorans, *Tetrahymena* and *Paramecium* (in Europe) are the subjects of current genome projects, while there is none for any spirotrich. *O. trifallax* is the most studied spirotrich; the production of a BAC library of its germline DNA has been funded by NHGRI. Its macronuclear genome sequence will elevate the molecular research effectiveness of *O. trifallax* and every other ciliate, so that the major ciliate systems can more efficiently cross-feed one another.

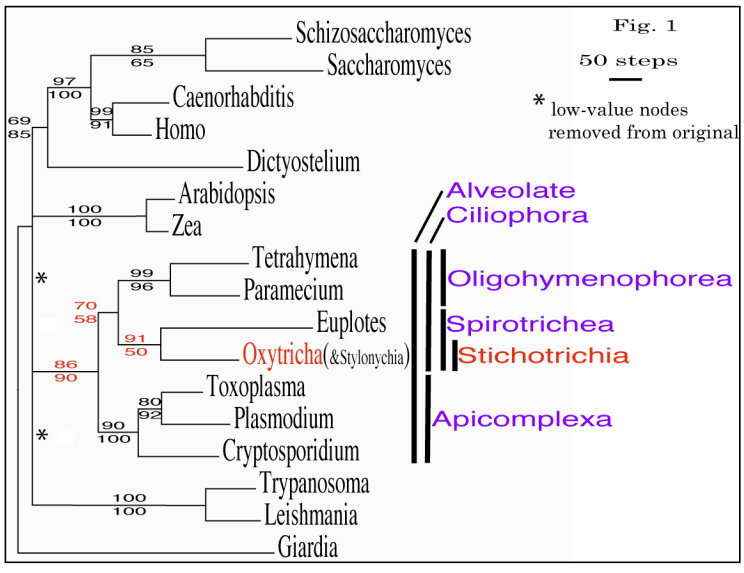
Thus, the *O. trifallax* macronuclear genome sequence will provide an important tool for ciliate biology, for basic eukaryotic cell biology—where ciliates excel as experimental systems—and for study of the related apicomplexan parasites, such as *Plasmodium*, which also have genomics projects underway. The *O. trifallax* macronuclear genome sequence is an essential companion to the ongoing *Tetrahymena* and *Paramecium* genome sequences.

The *O. trifallax* macronuclear genome has only 50 Mb of complexity, is nearly pure coding and regulatory sequences, and carries ~30,000 genes which run all cell processes; thus, to sequence the entire genome would be inexpensive and yield especially informative sequence. An initial pilot project has demonstrated that our approach is tractable and has already yielded considerable genomic information which we will use of in this proposal.

Attached are four letters of support, from Tom Cech, Larry Klobutcher, John Logsdon, and David Prescott; letters from Ed Orias and Jonathan Eisen will follow.

Biology of ciliates

Ciliates are single-celled eukaryotes of fundamental biological interest and hold a special place in modern molecular biology. Ciliates were the birthplace of telomere biochemistry (*e.g.*, Collins 1999). Cech (1990) discovered the first known self-splicing intron in *Tetrahymena*. Ciliates radically illustrate the C-value paradox (Gall 1981). Classically they were premier genetic eukaryotes in the hands of Tracy Sonneborn, his contemporaries, and his students (Sonneborn 1977, Nanney 1981).

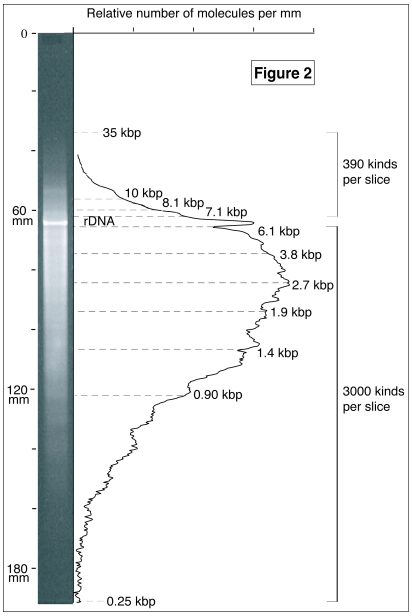


Ciliates diverged from other microbial eukaryotes quite recently, as part of a monophyletic lineage (alveolates) with apicomplexans (*e.g., Plasmodium*) and dinoflagellates (Wright & Lynn 1997; Baldauf *et al.* 2000). Therefore, as a phylogenetic outgroup, ciliates provide a foil to studies of the crown eukaryotes, plants, animals, and fungi (Figure 1, after Baldauf *et al.*

2000: bootstrap values >50% are shown above and below the lines, respectively, for amino acid parsimony and maximum likelihood analyses of second codon–position nucleotides). Ciliate molecular genetics has been concentrated in two of the ten ciliate classes (Fig. 1), the oligohymenophorans, with *Tetrahymena* and *Paramecium*, and the spirotrichs, with *Euplotes* ("hypotrichs," *sensu strictu*Bernhard *et al.* 2001) and several stichotrich species, of the genera *Stylonychia* and *Oxytricha*. Each is quite diverged from the next and has its specific virtues; this diversity fuels heuristic comparative biology by a vigorous research community, focused on either the workings of ciliate germline/soma nuclear dimorphism, or on fundamental eukaryotic biochemical pathways and cellular processes.

The ciliate macronuclear genome. Each ciliate carries in its single cell two kinds of nuclei: 1) a typical diploid, meiotic germ-line nucleus (the micronucleus or MIC), and 2) a macronucleus (MAC), a highly specialized expression organelle that provides all the transcripts for cell function. The MAC develops from a mitotic copy of the MIC, immediately after pair-wise cell mating, haploid gametic nuclear exchange, and zygosis. Stichotrich MAC differentiation is the premier showcase of somatic genome alterations (review: Jahn & Klobutcher 2002), and has been a major focus of ciliate research since the discovery of these dramatic processes in *Stylonychia* by Ammermann (1964). MIC chromosomes polytenize, non-coding DNA sequences interrupting MAC-destined sequences are spliced out, the coding pieces are unscrambled, polytene chromatids are fragmented and telomeres are polymerized onto the new ends by telomerase, and finally the resulting a-centric MAC "chromosomes" are highly amplified to a level scaled to the size of the ciliate. The old MAC is apoptotically destroyed as the new MAC differentiates.

MIC-limited and MAC-destined sequences are interspersed in the MIC genome. In *Oxytricha* >90% of the sequences are MIC-limited, being destroyed in the developing MAC. The new MAC is entirely responsible for vegetative, clonal growth of the exconjugant (Herrick 1994). Introns are effectively both rare and extremely short (avg.=118 bp in stichotrichs) and subtelomeric non-coding sequences are short (Hoffman *et al.* 1995). Thus, the MAC is nearly pure coding DNA.



Qualitatively, MAC differentiation is similar in most ciliates (Jahn & Klobutcher 2002, Yao *et al.* 2002), but the number of DNA splicing events, break sites, and chromosome kinds are orders of magnitude higher in spirotrichs than in oligohymenophorans. Oligohymenophoran MAC chromosomes contain hundreds of genes each, whereas spirotrich MAC chromosomes contain one or, in rare cases, a few genes each, making the spirotrich MAC genome ideal for gene discovery. This spirotrich pattern seems to be shared with other less-studied classes of ciliates (Riley & Katz 2001), suggesting that massive fragmentation may be both ancestral and more representative of ciliate biology than the modest fragmentation in *Tetrahymena* and *Paramecium*.

MAC sequence complexity is surprisingly constant across the range of ciliates (most have ~50 Mb, although *Tetrahymena* has ~200 Mb), while ploidy levels scale with the size of the ciliate (Soldo *et al.* 1981, Ammermann & Muenz 1982). The mature *Oxytricha* MAC genome consists of genes deployed on a collection of ~20,000 different mini-chromosomes amplified to a ploidy of ~1000/MAC. Their total sequence complexity is ~50 Mb. These linear "chromosomes" are tiny, with an average length of ~2400 bp, ranging from ~0.25 to ~40 kb (Maercker *et al.* 1999); Figure 2 shows FIGE analysis of *O. trifallax* MAC DNA, and the distribution of kinds of chromosomes across the size range. Estimates suggest that ciliates carry a large number of genes (>30,000), compared to gene counts for *Drosophila melanogaster* (~14,000) and *Caenorhabditis elegans* (~18,500) (Rubin *et al.* 2000).

Multi-gene MAC chromosomes and alternative fragmentation. While most spirotrich MAC chromosomes bear only one gene, the Herrick lab has found four that carry 2 genes and one that carries 3 (Seegmiller *et al*. 1997; unpubl.), and the Prescott lab has found other examples(D.M. Prescott, pers. comm.). Note that none of the ~400 large MAC chromosomes (≥7 kb; Fig. 1) has ever been studied; hence, we do not know if they carry gene clusters.

While some multi-gene MAC chromosomes are "simple," others are generated by alternative fragmentation of the polytene chromatids during differentiation, to form MAC chromosomes that partly overlap. Alternative processing is reproducible and common in stichotrich MAC development (Cartinhour & Herrick 1984; Klobutcher *et al*. 1988) as well as in other ciliates (*e.g*., Forney & Blackburn 1988). It is not known why particular genes are co-deployed (linked) on MAC chromosomes, or why alternative processing gives alternative groupings of genes, but patterns of linked genes and the alternative processing that gives rise to them are well conserved in the stichotrichs (K. Williams & G.H., unpubl.). This could indicate that the co-deployed genes are co-regulated and drive shared pathways; *e.g.*, the *Oxytricha* 81 locus generates a MAC chromosome that deploys two distantly related paralogs encoding mitochondrial solute carrier homologs (MSCs), which likely were brought together at this locus in the ciliate lineage (Seegmiller *et al*. 1997). The MAC genome sequence will allow us to learn if MAC-linked genes are functionally related: initial analyses of cloned MAC chromosomes from our pilot project (see B4) have identified 18 that carry two or more genes.

Alternative fragmentation seems to correlate with a variant form of telomere addition in stichotrichs. Sites of alternative fragmentation give rise to MAC ends that have telomeres added at many alternative sites within ~100 bp regions: *cis*-acting elements must influence the mechanism of telomere addition (Williams *et al*., submitted). The pilot project has yielded more such regions; the genome project should reveal many more.

Internally eliminated sequences (IESs) and scrambled genes. During polytenization of the MIC genome, many sequences internal to MAC-destined sequences are precisely excised and eliminated (Internal Eliminated Sequences, or IESs; Klobutcher & Herrick 1997). Almost all spirotrich and *Paramecium* genes are interrupted by multiple IESs. Many IESs are intact, cut-and-paste transposons (~4-5 kb long, with thousands of copies per haploid genome). Short IESs (< 0.5 kb) are believed to be degenerate, non-autonomous transposons that retain the *cis*-acting sequences required for precise excision. Excision of each IES precisely reverts the germline insertional mutation, enables the expression of the gene in the MAC, and allows the element to persist in the germline. The Landweber lab and collaborators are developing computational approaches to identify or predict sites of IES excision in the resulting MAC sequence.

A related phenomenon is the recent, exciting discovery of scrambled genes in the stichotrich MIC (Prescott 2000). Again genes are interrupted by many IESs, but contiguous gene segments are no longer adjacent, nor necessarily on the same strand or even at the same locus in the MIC (Landweber *et al*. 2000), and the order of gene segments is permuted in both random and strikingly nonrandom patterns, such as 1-3-5-7-2-4-6-8. IES removal permits linking of coding segments in the correct, translatable order and orientation. The DNA polymerase□□ gene is scrambled into 51 segments (Hoffman & Prescott 1997), dispersed on both strands of two unlinked MIC loci. As many as 20-30% of *Oxytricha* genes are estimated to be scrambled in the MIC genome.

The focus of this proposal is to sequence the MAC genome, but the availability of MAC sequence will allow the speedy retrieval of cognate MIC sequences and hence the identification of novel IESs and scrambled genes. We propose that the sequencing project choose a random set of MAC sequences and sequence the MIC equivalents, retrieved from the BAC library (see B5). In particular, large swaths of MIC sequence have never been examined to ascertain the gross structure of MAC-destined and MIC-limited sequences. For example, MAC-destined segments likely are clustered (Jahn *et al*. 1988), but how are

the other >90% organized? Where are the transposons? All known ciliate transposons display extremely novel features (see A8); new transposon types are sure to be discovered.

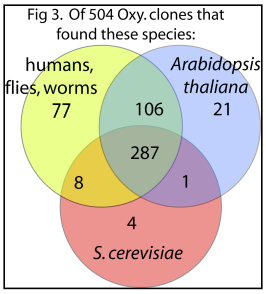
NHGRI "Points to be addressed."

Issues are addressed in groups to avoid redundancy and to provide coherent essays. Four primary questions are addressed: 1) Why sequence ciliate genomes; 2) Why sequence a another ciliate genome; 3) Why should this project be of a stichotrich; 4) Why is *O. trifallax* the best choice among the stichotrichs?

A. Biological rationales for determining the *O. trifallax* MAC genome sequence.

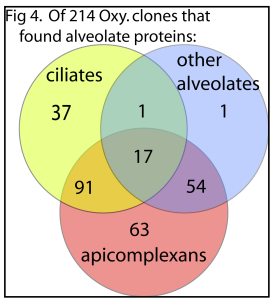
A1-6. How will the sequence inform our understanding of human health and disease, human biology, the human genome and biological processes relevant to human health, and provide surrogate systems for human experimentation?

• The *O. trifallax* predicted proteome sequence will be useful to human health and the human genome project. Accurate prediction of genes, spliced mRNAs, and the resulting proteins, is still problematic in higher metazoans, due to numerous large introns and alternative splicing. The identification of protein products often depends on the availability of orthologs in other genomes: these are produced both directly and economically in large numbers from spirotrich MAC DNA, because the DNA contains little more than coding regions plus associated regulatory DNA (introns are few and small, and alternative RNA splicing has not been observed). In our pilot project (see B4) 553 reads found homologs in database searches, of these, 9.2% found a human homolog, but not a yeast, fly, or worm homolog (BLAST E-value $\leq e^{-15}$); extrapolating, ~2500 human genes have ciliate—but not yeast, worm, or fly—homologs. These genes are cases where ciliates seem to be the only non-vertebrate model system available.



The presence of only very distant orthologs in the database reduces the probability of recognizing homology; *e.g.*, between *Plasmodium* and humans: thus it is always advantageous both to fill in the tree of phylogenetic diversity and to increase the number of orthologous sequences. Recent advances in bioinformatics—COGs, Psi-BLAST, and IMPALA—rely on this growing richness of ortholog diversity to identify homologs and to infer functions for unassigned ORFs in genome projects. Ciliates form a highly diverse clade of microbial eukaryotes that diverged just before the crown taxa of plants, animals, fungi, and stramenopile algae (Fig. 1), and their genes find homologs in crown eukaryotes metazoans, plants and fungi equally well. Figure 3 shows results from our pilot project, sorted to indicate numbers of hits to various sets of taxa, e.g. 77 hits (E-values $\leq e^{-15}$) to metazoa but not yeast nor plants; see B4; the exact values in these analyses are greatly influenced by the number of proteins available in the databases for each organism or clade.

Because of their diversity, members of different classes of ciliates provide different—though complementary—information; thus, the MAC sequence of a spirotrich will not be redundant, but will synergize strongly with the oligohymenophoran MAC sequences, illuminating all ciliate classes, their parasitic alveolate relatives, and crown eukaryotes. *Tetrahymena* has been proposed as an ideal organism for drug discovery and testing (Orias 2000); a better understanding of general ciliate biology can only add to this effort to develop unicellular eukaryotic models for human disease.



Ciliate sequences will serve as a closely related outgroup within the alveolates for inference of events during the evolution of apicomplexan parasitism (*e.g.*, *Plasmodium*, *Toxoplasma*, and *Theileria*). For example, *Plasmodium* likely lost genes as it became host-dependent, perhaps in parallel gaining pathogenesis genes. Knowing such host-pathogen relationships should aid in designing therapies. In our pilot project, 214 *Oxytricha* sequences found alveolate protein similarities in the databases ($\leq e^{-15}$), with considerable overlap between the ciliates and apicomplexans (91+17 in their intersection,

Fig. 4); these genes offer the immediate opportunity to examine the function of apicomplexans genes in ciliates.

• Basic research into the molecular genetics of ciliates has revealed a panoply of fascinating phenomena, the elucidations of which promise to reveal new biological processes that may function in a broad spectrum of biological systems.

As complex unicellular eukaryotes that employ elaborate cellular processes, the ciliates' strength as research organisms is most obvious in cellular physiology. Their cellular organization is far more representative of metazoan cells than is that of yeast (Orias 2000), and their large gene sets reflect these facts. Some of these functions are no doubt specific to their life-styles, most ciliates being large (often >100 μm), and free-living, but most functions are expected to be general to eukaryotes: 34% (553/1609 of our MAC chromosome clones find protein homologs in the databases (see B4).

Discoveries essential to understanding human biology and health have repeatedly been made in ciliates: *e.g.*, telomere structure and telomerase, histone modifications, and many aspects of cytoskeletal structure. In *Oxytricha* critical features of telomeres were first discovered: in particular, the identity of all ~40 million MAC telomeres (Herrick & Wesley 1978; Klobutcher *et al.* 1981), the 3' G-strand overhang (Pluta *et al.* 1982), and telomere-associated proteins and their interaction with the telomere sequence (Gottschling & Cech 1984; Gottschling & Zakian 1986; Classen *et al.* 2001). Multiple complete ciliate genomes will greatly facilitate identification of complete enzymatic pathways and systems of subcellular structure in ciliates, allowing more comprehensive studies. Ciliates will surely continue to be key unicellular model organisms for the study of development, differentiation (especially gene/genome rearrangements), the cytoskeleton and cell motility, electrophysiology, and chromatin structure and function. Given that telomeric repeats comprise ~2% of MAC DNA, spirotrich telomere biochemistry will continue to be particularly productive. MAC differentiation requires numerous biochemical steps relevant to eukaryotic biology in general and to human health in particular: cleavage of chromatids requires uncharacterized site-specific endonucleases; telomeres are added *de novo* to the resulting ends; and recombinational excision of IESs seems to employ a novel recombinase. Stichotrich development yields biochemical amounts of polytene chromosomes. Both genome rearrangements and telomere biology especially may inform our understanding of karyotype instability and oncogenic transformation in humans.

Epigenetic phenomena have a long history in ciliate biology. An example is the maintenance/replication of cell-surface (cortical) structures and patterns (Sonneborn 1975). Even aberrant cortical patterns created by microsurgery can be maintained. How are these patterns established? What are the organizing centers that nucleate patterning? How are patterns faithfully propagated through cell divisions and even encystment, when there is no ultrastructural sign of cytoskeletal patterning? These issues certainly are relevant to cytoskeletal function and the maintenance of cell differentiation in general.

Recent results have stimulated intense interest in Sonneborn's demonstrations that the genotype of the parental MAC can sequence-specifically guide the differentiation of the emerging MAC (Meyer & Garnier 2002). The epigenetic signals that pass from old to new MAC are unknown, but might be some form of RNA, possibly representing a novel effect of an RNAi-type pathway. RNAi-based genetic interference has recently been developed for a range of ciliates (Bastin *et al.* 2001; see B2).

Synergy between diverse ciliate molecular genetics experimental systems will be promoted by the knowledge of the *O. trifallax* MAC genome, as detailed below in B1. Especially, *Tetrahymena* genetic tools, coupled with a diverse set of ciliate genomic sequences, will allow the delineation and dissection of pathways shared with other organisms, most notably humans and their eukaryotic pathogens.

A7. How will the *O. trifallax* sequence facilitate experiments in other organisms?

The addition of ~30,000 more protein-coding sequences to public databases will obviously benefit researchers in many fields, including comparative genomics, evolution, bioinformatics, annotation, and eukaryotic biology quite broadly.

The most proximal effect will be on *Oxytricha*, and on ciliate molecular genetics. In turn, this will promote research outside the ciliates, enabling research of the biology of humans and especially our apicomplexan pathogens. For instance, since ciliate taxa are more closely related to each other than to other organisms of interest (*e.g.*, apicomplexans, yeast, metazoa, and metaphytes; Fig. 1), ciliate workers "searching outward" for orthologs in the sequence databases can use the ciliate ancestral state as a probe of the "outside" data set. A *Tetrahymena* and *Oxytricha* ortholog pair permits construction of a sequence profile (position-specific scoring matrix; Gribskov *et al.* 1990), which can be used to query the databases with a much higher likelihood of finding a strong match; the inclusion of a *Paramecium* ortholog allows reconstruction of a ciliate ancestral sequence. Of 553 *Oxytricha* clones that find database similarities, 26% have ciliate homologs (see B4), which is surprising, given the relatively small number of ciliate genes in the databases. This allows ciliate workers to discover homologs in other genomes (*e.g.*, see Doak *et al.* 1994). Reciprocally, the existence of these virtual ancestral sequences, represented in clusters of ciliate orthologs, will aid identification of ciliate homologs in searches with sequences from any other genome, such as the human genome. We like the idea of a searchable set of ciliate protein profiles, and envision it as a highly useful project for post-genomic informatics. Thus, the *Oxytricha* sequences will facilitate annotation of the unidentified ORFs in the human genome and in genomes of human pathogens, and domesticated animals and plants.

A8. How will the *O. trifallax* sequences expand our understanding of evolutionary processes in general, and human evolution in particular?

Knowledge of ciliate sequence characters allows one to use parsimony to infer events that occurred either within the ciliate lineage or in the lineage to other crown eukaryotes. Thus, for instance, certain introns of mitochondrial solute carrier protein (MSC) genes must have been inserted into the crown eukaryote ancestor after ciliates diverged from it, and at least one of the three introns in two *Oxytricha* MSC paralogs must have been inserted during ciliate evolution, even allowing for intron "sliding" (Seegmiller & Herrick, unpubl.).

Many eukaryotic nuclear or mitochondrial genomes use alternative genetic codes, but the greatest code diversity of all is seen in ciliates, where the use of either the standard code or one of three alternative codes is polyphyletic. For example, UAR encodes Gln in both oligohymenophorans and stichotrichs, while UGA encodes Cys in *Euplotes* but encodes Trp in two independent lineages; tracing coevolution of the code in ciliates and its associated translation components, such as release factors, currently offers the best model system for probing the selective and biochemical forces that drive code evolution (Lozupone *et al.* 2001).

The biology of ciliate transposons is especially rich. The relationship of Mariner/Tc1 and Pogo-like transposases to retroviral integrases was first recognized because of new transposons discovered in ciliates (Doak *et al.* 1994). But in addition to transposase, ciliate transposons encode novel protein kinases and tyrosine recombinases (Doak *et al.* 1997; Doak *et al.* in prep.;). Also, the population dynamics of ciliate transposons is novel: genes evolve under a purifying selection for protein function (Klobutcher & Herrick 1997). This could be unique to ciliates, but more likely it is an extreme on a continuum of different transposon-host relationships. We expect that understanding the ciliate case will illuminate the coevolution of hosts—such as humans—and of their transposon parasites. Ciliate transposons are also the likely ancestors of IESs (see above), an evolutionary transition analogous to that proposed for Group II self-splicing introns as ancestors of splicesomal introns (Cavalier-Smith 1985).

B. Strategic issues in acquiring *O. trifallax* macronuclear sequence

B1. Size and enthusiasm of the community that will use the *O. trifallax* sequence. Will the sequence stimulate the expansion of the research community?

O. trifallax research is currently being performed in the labs of Herrick, Prescott, Landweber, Hammersmith, Adl, Baroin, and W. F. Doolittle (another ~20 labs work on other stichotrichs). The *O. trifallax* genome sequences will immediately be important to not only stichotrich biologists, but to the full ciliate community. We all function within the ciliate research community, amongst whom research news circulates quite freely. Moreover, new and important findings in stichotrichs, such as the biochemical nature of telomeres, have spurred interest in ciliate research. Clearly the availability of the macronuclear genome sequence for *O. trifallax* would entice many new people to join this field, particular for comparative genomics. For example, the Lozupone *et al*. (2001) study of ciliate genetic codes led to several related publications around the world as people recognized the attractiveness of ciliates as a model system.

Ciliate biologists and molecular geneticists form an active and united community. A biennial meeting of this international group (US, Europe, Asia) has been held since 1982, recently as either Gordon or FASEB conferences, with an attendance of ~150. The entire community is looking forward to ciliate genome information. Major community efforts are in progress to advance the genomics of the oligohymenophorans *T. thermophila* (Orias 1998, 2000) and *P. tetraurelia* (in Europe; Dessen *et al*. 2001). An international group of 13 stichotrich researchers including biologists and computer scientists meets annually as part of Landweber's 5-year NSF ITR grant to study ciliate molecular genetics, evolution, and computational biology. We've received exuberant letters supporting the present proposal from Germany, France, Canada, and the Netherlands, as well as from researchers in this country, that go beyond the ciliate community, *e.g*. comparative genomics. Attached are four letters that make particularly salient points, from Tom Cech, Larry Klobutcher, John Logsdon, and David Prescott. An important letter, from Ed Orias and the *Tetrahymena* genomics Executive Committee will follow shortly, as will one from David Eisen.

• Why should another ciliate genome project be performed? Genome projects are ongoing for the oligohymenophorans *Paramecium* and *Tetrahymena*, which will build on their histories as intensely studied model organisms. Comparable histories of the spirotrichs begs that a spirotrich genome be sequenced.

Ciliate classes have been diverging for >1 billion years—since the ciliates diverged from other eukaryotes (Wright & Lynn 1997)—and thus are a very heterogeneous group. To understand "the ciliates" it is necessary to sample at least some of the breadth of this diversity; the genome of a second class of ciliates will thus be an essential start. The *Oligohymenophorea* and *Stichotrichia* are as distinct as any two ciliate classes—as well as being by far the best studied—and therefore approximate the diversity within the ciliate clade.

The *O. trifallax* sequences will serve both to complement and to augment the utility of the emerging sequences, and will also drastically enhance the capability to study *O. trifallax* and the other spirotrich systems. *Oxytricha* provides an "outgroup" to the oligohymenophorans. *Oxytricha* characters allow one to infer the direction and time of changes that occurred in the evolutionary history of *Paramecium* and *Tetrahymena*. As noted above, evolutionary reconstructions can locate on a phylogenetic tree the timing of events, such as horizontal gene transfers into one lineage, or clustering of genes in the germline (and their resulting co-deployment in the MAC). One can polarize a variety of protein gene changes, such as synonymous or nonsynonymous substitutions, gene duplication or loss, and accelerated or diminished rates of divergence in a lineage. For example, short IESs of *Paramecium* surprisingly resemble those of stichotrichs: thus the unique *Tetrahymena* IESs are not a character of all oligohymenophorans, but may have appeared in the *Tetrahymena* lineage (with the loss of basal stichotrich-like IESs).

In the spirotrichs, the stichotrichs and the early-diverging *Euplotes* have been intensely studied for their dramatic developmental chromosome fragmentation and elimination of MIC-limited sequences, with massive gene fragmentation and scrambling a curious innovation in spirotrichs. Interest is strong in the evolution of MAC structure during ciliate diversification: Riley & Katz (2000) suggest that the origin of MAC fragmentation is monophyletic but that the organization of genes into gene-sized pieces is polyphyletic.

Many ciliates have been studied, emphasizing the unique features of each lineage. Genome comparison will underscore the common features of ciliates, strengthening collaboration within the ciliate community, as well as highlighting universal eukaryotic features for non-ciliate researchers. The ability to identify one's favorite pathway in ciliates will encourage broader use of ciliates as an experimental system, and foster more collaborations between ciliate workers and other researchers. More generally, the availability of genomic resources for *O. trifallax* will attract young researchers to this exciting field, as stichotrichs have so many beguiling problems to offer, ranging from epigenetic phenomena to telomere biology. Traditionally, ciliate labs have trained numerous graduate students, who often left the field in search of more-tractable systems around which to build their careers. The proposed genomics project will make *Oxytricha*, and all ciliates, much more attractive to future students, as well as attract post-docs and young investigators trained in other fields.

B2. The suitability of *O. trifallax* for experimentation. What *O. trifallax* properties affect its ability to be studied in the laboratory?

•Why *O. trifallax*? Of the several spirotrichs being studied, we propose a genome project for *O. trifallax* (recently renamed *Sterkiella histriomuscorum*). The proposed work on *O. trifallax* represents the primary clade of focus in the spirotrichs and within the stichotrichs, as judged by Medline searches. That is, more work has been done on stichotrichs than on *Euplotes*, and on *Oxytricha* than on *Stylonychia*. Notably, stichotrichs and oligohymenophorans use the same derived UAR:Gln code, orthogonal to the *Euplotes* UGA:Cys code, permitting functional genomics experiments using stichotrich genes in *Tetrahymena*. Among the stichotrichs, many closely related species have been studied; they are sufficiently similar that the genome sequence of any one of them will be informative for the others. Comparative approaches among the various stichotrichs are already standard practice in various labs (*e.g.*, Ammermann, Herrick, Landweber, Prescott), including a range of species intermediate between *Euplotes* and *Stylonychia/Oxytricha*.

•*O. trifallax* is represented by many healthy strains that are being used in the 6 labs noted above. Unlike many stichotrichs, *O. trifallax* is easily stored as frozen cysts. Without such storage, vegetative lines senesce, leading to sterility and eventual death of the clone; clonal senescence has plagued much stichotrich research. But non-senescent *O. trifallax* cultures are fully capable of fertile mating, allowing Mendelian genetics, and mass synchronous matings to study biochemical intermediates and activities (*e.g.*, Williams *et al.* 1993; Adl & Berger 2000).

Are mutants available with defined phenotypes? While historically no efforts have been made to generate stichotrich mutants, many diverse wild isolates have been collected. The Ammermann, Herrick, Landweber, and Prescott labs, their collaborators, and taxonomists have each generated large stichotrich collections: both single isolates of novel species from diverse environments and multiple isolates of single species, including stored cysts of ~60 independent, mating-typed, wild isolates of *O. trifallax*; these collections offer a wide choice of diverse alleles and orthologs for loci of interest and permit divergence analyses that identify sequence features that survive purifying selection (*e.g.*, Seegmiller *et al.* 1996, 1997).

The phylogenetic relationships among stichotrich species and genera are being probed by a number of taxonomists (*e.g.*, Bernhard *et al.* 2001), allowing sequence and physiological differences to be placed in a firm phylogenetic context. The stichotrichs are all relatively closely related: hybridization probes and PCR primers developed in one genus can often be

used successfully for members of the other genera. This allows researchers to quickly access the genetic diversity represented by these isolate collections. RNAi interference should soon permit quick routes to null phenotypes (see below).

•What genomic resources and technologies are available that will allow the new sequence information to be effectively used? Ciliate MICs, MACs, and developing MACs can be transformed, RNAi technology works, gene replacements are routine, collections of mutants are available, genome maps are being developed, cloning by complementation works, and versatile transmission genetic tools are available.

The spirotrichs have traditionally been studied by structural biologists and biochemists, due to large cell size and the radical extent to which MIC genomes are processed during MAC development. The full MAC sequence will greatly further these biochemical studies, but will also stimulate development of molecular genetic techniques, often by borrowing methods from other ciliate systems. Such tools are readily adaptable, judging from the recent history of ciliate molecular genetics, starting with the transformation of *Tetrahymena* by Tondravi and Yao (1986), and the subsequent development of MAC gene "knock-out" techniques and MIC (germ line) transformation (Asai & Forney 1999). Transformation of *Paramecium* followed, as did GFP expression (Hauser *et al.* 2000) and transformation of the spirotrichs *S. lemnae* and *Euplotes crassus* (Ascenzioni & Lipps 1986; Bender *et al.* 1999, Erbeznik *et al.* 1999). *Paramecium* genes are being cloned by complementation (Kung *et al.* 2000). Antisense technology was first adapted for *Paramecium* (Hindrichsen *et al.* 1992; Galvani & Sperling 2002), then for *Tetrahymena* (Sweeney *et al.* 1998), and recently, RNAi has been demonstrated to inhibit expression of two proteins in both *S. lemnae* and *E. aediculatus* by feeding them double-stranded RNA-expressing bacteria (Lipps and Cech labs, in preparation). Lipotransfection has been developed for the distantly related ciliate *Blepharisma* (Matsuoka *et al.* 2000). An important aspect of a transformation system is understanding transcriptional regulation, so that suitable reporter genes and regulated promoters can be harnessed; similarly, a functional replication origin is required. The end sequences we have already generated are being analyzed for shared subtelomeric sequence motifs, which will include MAC chromosome replication origins (Murti & Prescott 1983; Skovorodkin *et al.* 2001), 5' transcriptional promoter elements, and 3' polyadenylation signals.

•How will the new sequence data enhance the experimental use of *O. trifallax*? Genome sequence will allow the rapid identification of all or most components of developmental and biochemical pathways, initially by sequence similarity, but in the future by using stage-specific cDNA libraries and chip technology to find suites of expressed genes. Interesting proteins in other organisms will be quickly identified in *O. trifallax.*

Genome sequences representing *Oligohymenophora* and *Spirotrichea* will aid in transferring results in one system to the other—allowing the use of unique advantages of each—and in distinguishing between functions that are unique to a particular ciliate clade, and those that are general. In particular, the knowledge of the genomes of *Tetrahymena* and *O. trifallax* will allow facile exchange of experimental leads: as interesting genes are discovered in *O. trifallax*, one can immediately identify the orthologs in *Tetrahymena* and perform "knock-out" transformations to learn the phenotypes; as comparable tools for *O. trifallax* mature, similarly, one can study the orthologs of *Tetrahymena* genes in the context of the *O. trifallax* genome. Genes found in other ciliates, *e.g.*, *Paramecium* and *S. lemnae*, can be brought into this dual-organism, structure-function analysis system. This stimulation of basic ciliate research will undoubtedly feed into studies of humans, other model systems, and human disease agents, and should lead to collaborations between ciliate labs and labs in these other fields, as the ease of experimental manipulation of ciliate genotype emerges.

B3. Rationale for the complete sequencing of the *O. trifallax* MAC genome.

We have developed this white paper after discussion with Dr. Eric Lander (Dir. Whitehead Institute). He felt that it was important to represent the spirotrichs with a

genome project, *O. trifallax* being the logical choice; and, given the evolutionary distance between *Oxytricha* and *Tetrahymena*/*Paramecium*, that the *Oxytricha* sequences would permit a very informative companion to the *Tetrahymena thermophila* MAC genome sequence, a high-priority project his sequencing center will do.

Our proposal is to obtain high-quality sequence of the MAC genome, a mere ~50 x 10^6 bp in unique sequence complexity, residing on mini-chromosomes that are trivial to clone. The MAC sequence will encode the entire proteome, and will contain all regulatory sequences, and minimal "junk" DNA sequences. MAC sequences contain splicesomal introns, but *Oxytricha* introns are few and small (~100 bp). They are easy to identify, having conventional conserved motifs surrounding GT...AG splice sites (Hoffman *et al.* 1995). The proposed sequence project has obvious parallels to complete sequencing of a cDNA library, but, by including regulatory DNA and not requiring copy normalization, the project provides much more information with greater ease, and a complete genome sequence without assembly.

We propose a 10x redundant library, in order to sample—and assure high quality sequence for—most MAC chromosomes. We hope for a complete genome sequence, especially since we and others are interested in the gain and loss of genes in eukaryotic lineages.

Beyond sequencing the MAC genome, we propose using selected MAC sequences to nucleate sequencing of cognate MIC regions. Much ciliate research focuses on genome reorganization during MAC differentiation, and this depends on comparison between MIC and MAC sequences. This work will be greatly furthered by the availability of extensive MIC sequence. Thus, we propose sequencing ~10 Mb of MIC regions obtained from a BAC library of MIC DNA, which will be constructed by P. de Jong's facility (NHGRI funded). MAC sequences probably are clustered in the MIC (*e.g.,* Jahn *et al.* 1988), so we hope to obtain a representative sampling of MAC-destined sequences in 10 Mb.

B4. Procedure for sequencing the *O. trifallax* MAC genome and the state of readiness of DNA for sequencing

The *Oxytricha* MAC genome size is ~50 Mb, deployed on ~20,000 MAC chromosomes, with sizes from ~0.5–45 kb (Fig. 2). Pure undamaged MAC DNA is readily obtained (Cartinhour & Herrick; 1984). The library will carry MAC DNA of a multiply-inbred lab strain generated by David Prescott from the original Indiana isolate, JRB310, of Hammersmith.

Since the MAC genome is already divided into easily cloned and informative pieces, we propose a library of intact MAC chromosomes, using plasmid vectors for MAC chromosomes <25 kb, and cosmids for MACs >25 kb. MAC chromosomes will first be separated on a preparative FIGE gel (Fig. 2), so that different size-classes can be cloned separately, to assure equal representation of all sized MACs in the final library. Sequencing would combine end-sequencing, primer walking, and transposon tagging, to be determined by the collaborating center.

Mitochondrial DNA copurifies with MACs in our preparations, and will also be sequenced; no spirotrich mtDNA has been sequenced. This will complete the sequencing of expressed *O. trifallax* genes, and will allow us to examine the mitochondrial genetic code, gene content, and mitochondrial telomeres (different than MAC telomeres, Morin & Cech 1986). Ciliate mtDNAs are large, 40-70 kb linears (review: Burger *et al.* 2000).

As a pilot to this project, R. Weiss and D. Dunn (Utah Genome Center), with the Herrick lab, cloned 2292 MAC chromosomes and sequenced both ends of each from vector primers; Table I summarizes the preliminary results (analyses in collaboration with the Landweber lab and associates). We've cloned and end-sequenced ~7% of the MAC chromosomes (2292 of ~30,000), for a total 2,292,558 nucleotides (phred >20), or ~5% of the genome; 437 chromosomes are completely sequenced; the largest is 1507 bp long; 1410 cloned chromosomes have reads into each end that don't overlap.

Table 1. Summary of cloning and sequencing pilot project.

Clone insert sizes, by agarose gel electrophoresis, 1st 384 clones	
Average (*cf.* genome avg, 2.4 kb)	1.9 kb
Maximum	7.0 kb
Minimum	0.3 kb
Clone inserts sequenced at one or both ends	2292 clones
Total sequence (≥phred 20 bases)	2,292,558 nt
Clones with completely sequenced inserts (largest, 1507 bp)	437 clones
Clones with *Oxytricha* inserts (telomere at 1 or 2 ends)	1609
Clones with intact MAC chromosomes (telomere at each end)	1185
Clones with BLASTx hits, scores $<e^{-15}$ ($<e^{-10}$)	553 (691)
Clones with BLASTx hits to ciliate proteins, scores $<e^{-15}$	146
Cloned MAC chromosomes with ≥2 gene	18
Number of probable *Oxytricha* protein genes identified	571
Number of probable *Oxytricha* ncRNA genes identified	50
Number of *Oxytricha* genes unidentified	≥1056

A relational MySQL database for this project has been built (Cavalcanti *et al.* unpublished). It contains, for each end-read: information about telomeres, GC content, size, BLAST hits (with full description and E-value), best BLAST hit, and taxonomic information for all hits. For each clone, an HTML page provides: 1.) this information for both ends, 2.) a graphical display of the best hits, 3.) start and stop codon positions in each of the six reading frames, 4.) a description of each of the 3 best BLAST hits, linked to their sequences in NCBI. A longer HTML page provides this information for all hits. We are working to submit the annotated sequences to the public databases.

In protein database searches (A. Cavalcanti & N. Stover, pers. comm.), 553 MAC clones (with 1 or 2 telomeres) find strong similarities to known or hypothetical proteins; of these 18 have two genes that give different hits. Thus, in this small pilot we have identified 571 *Oxytricha* genes with homologies in the databases (this number seems small, but in most cases we are searching with only genes fragments). There could easily be >1056 genes don't show database similarities; we will look for probable ORFs in clones and regions of clones that have not produced database hits. Among the protein-coding genes, at least 164 putative introns have been identified, on the basis of changes in reading frame to best BLAST hits; this method can only estimate 2/3 of the introns (although some clones may contain +1/-1 sequencing errors). These intron sequences, along with published sequences, can be used to build algorithms to more accurately find ciliate introns.

The clones also include (D. Ardell, pers. comm.) 22 unique tRNA genes (tRNAscan,-SE v.1.12, Lowe and Eddy 1997), representing 15 amino acids; two distinct LSU rDNA genes; and 26 putative 2'-O-methylation guide snoRNA genes (snoscan release 0.9, Lowe & Eddy 1999). Two of the predicted tRNAs have completely novel sequences, but were found by searching for tRNA secondary structural motifs; it remains to be seen—by finding orthologs in other stichotrichs and testing for sequence covariation—if these are functional. No genes were found for tmRNA, U-RNA, *xist*, *rox*, or *mei*.

Analyses of the subtelomeric sequences of the MAC clones extend the findings of Prescott & Dizick (2000), suggesting a special property of subtelomeric sequences that may act in *cis*, either in the MAC or in its development. A strong strand composition bias (there is a purine rich and a pyrimidine rich strand) extends ~100 bp internal to telomere sequences, with a 10 bp periodicity in information content (T.G.D. & A. Cavalcanti). More analyses are required, particularly to determine the variance among MACs in this pattern.

This pilot project also served to indicate a need for more care to eliminate food organism DNA (identified by high GC-content and absence of telomeres), and to prevent shearing of long MAC molecules, but both are easy to do, particularly with a library construction that demands the presence of two telomeres. In summary, we expect the MAC genome of *O. trifallax* to be very quick and inexpensive to sequence, given that the genome is very small, easily cloned, and that a pilot project was successful. At the same time this project will yield all the ~30,000 protein sequences encoded in the MAC genome.

B5. ADDITIONAL SUPPORT

The Landweber lab has an active 5-year NSF ITR grant to study the unscrambling of MIC genes during MAC development, using molecular biology, comparative biology and a range of computational approaches. This fascinating example of biocomplexity and biocomputation has stimulated both national and international collaboration of stichotrich and computational researchers.

In a pending NSF proposal, the Herrick lab has proposed a set of experiments that will: generate an archived and indexed 9.2-hit plasmid library of MAC chromosomes, specifically to be screened for examples of alternatively processed MAC chromosome families; it is proposed to obtain end sequence for these examples, and to completely sequence some of these clones. This is an inexpensive project, focused on a specific phenomenon, but will generate the high-quality library needed for the MAC genome sequence.

In a separate white paper, we proposed that NHGRI support the generation of a BAC clone library of the *O. trifallax* MIC genome. This project has been funded, to be done by the de Jong lab. It will greatly aid in the characterization of gross MIC structure, and the characterization of MAC differentiation.

Jonathan Eisen at TIGR has submitted a proposal to NIGMS (for autumn review) that would fund a database for the *Tetrahymena* genome, with the expectation that the database will grow to become the universal ciliate database. We've started discussing how we would include the *O. trifallax* genome in this project.

References

Adl SM, Berger JD. 2000. Timing of life cycle morphogenesis in synchronous samples of *Sterkiella histriomuscorum*. II. The sexual pathway. J Eukaryot Microbiol 47:443-449.

Ammermann D, Muenz A. 1982. DNA and protein content of different hypotrich ciliates. Eur J Cell Biol 27:22-24

Ammermann D, Steinbrück G, von Berger L, Hennig W. 1974. The development of the macronucleus in the ciliated protozoan *Stylonychia mytilus*. Chromosoma 45:401-429.

Ascenzioni F, Lipps HJ. 1986. A linear shuttle vector for yeast and the hypotrichous ciliate *Stylonychia*. Gene 46:123-126

Asai DJ, Forney JD, eds. 1999. *Tetrahymena thermophila*. Meth Cell Biol. 62 (Academic Press, NY).

Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972-977.

Bastin P, Galvani A, Sperling L. 2001. Genetic interference in protozoa. Res Microbiol 152:123-129.

Bender J, Kampfer M, Klein A. 1999. Faithful expression of a heterologous gene carried on an artificial macronuclear chromosome in *Euplotes crassus*. Nucleic Acids Res. 27:3168-3172.

Bernhard D, Stechmann A, Foissner W, Ammermann D, Hehn M, Schlegel M. 2001. Phylogenetic relationships within the class *Spirotrichea* (*Ciliophora*) inferred from small subunit rRNA gene sequences. Mol. Phylogenetics Evol. 21:86-92.

Cartinhour S, Herrick G. 1984. Three different macronuclear DNAs in *Oxytricha fallax* share a common sequence block. Mol Cell Biol 4:931-938.

Cavalier-Smith T. 1985. Selfish DNA and the origin of introns. Nature. 315:283-284.

Cech TR. 1990. Self-splicing of group I introns. Annu Rev Biochem. 59:543-568.

Classen S, Ruggles JA, Schultz SC. 2001. Crystal structure of the N-terminal domain of *Oxytricha nova* telomere end-binding protein alpha subunit both uncomplexed and complexed with telomeric ssDNA. J Mol Biol. 314:1113-1125.

Collins K. 1999. Ciliate telomerase biochemistry. Annu Rev Biochem 68:187-218.

Dessen P, Zagulski M, Gromadka R, Plattner H, Kissmehl R, Meyer E, Betermier M, Schultz JE, Linder JU, Pearlman RE, Kung C, Forney J, Satir BH, Van Houten JL, Keller AM, Froissard M, Sperling L, Cohen J. 2001. *Paramecium* genome survey: a pilot project. Trends Genet 17:306-8.

Doak TG, Witherspoon D, Doerder FP, Williams KR, Herrick G. 1997. Conserved features of ciliate TBE1 transposons. Genetica 101:75-86.

Doak TG, Doerder FP, Jahn C, Herrick G. 1994. A family of transposases genes in transposons found in prokaryotes, multicellular eukaryotes and ciliated protozoans. Proc Natl Acad. Sci USA 91:942-946.

Erbeznik M, Yao MC, Jahn CL. 1999. Characterization of the *Euplotes crassus* macronuclear rDNA and its potential as a DNA transformation vehicle. J Euk. Microbiol. 46:206-216.

Forney JD, Blackburn EH. 1988. Developmentally controlled telomere addition in wild-type and mutant *Paramecia*. Mol Cell Biol 8:251-258.

Gall JG. 1981. Chromosome structure and the C-value paradox. J Cell Biol 91:3-14.

Galvani A, Sperling L. 2002. RNA interference by feeding in Paramecium. Trends Genet 18:11-12.

Gottschling DE, Cech TR. 1984. Chromatin structure of the molecular ends of *Oxytricha* macronuclear DNA: phased nucleosomes and a telomeric complex. Cell 38:501-510.

Gottschling DE, Zakian VA. 1986. Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA. Cell 47:195-205.

Gribskov M, Lüthy R, Eisenberg D. 1990. Profile analysis. Methods Enzymol 183:146-159.

Hauser K, Haynes WJ, Kung C, Plattner H, Kissmehl R. 2000. Expression of the green fluorescent protein in *Paramecium tetraurelia*. Eur J Cell Biol. 79:144-149.

Herrick G. 1994. Germline-soma relationships in ciliated protozoa: the inception and evolution of nuclear dimorphism in one-celled animals. Sem Dev Biol 5:3-12.

Herrick G, Wesley RD. 1978. Isolation and characterization of a highly repetitious inverted terminal repeat sequence from *Oxytricha* macronuclear DNA. Proc. Nat. Acad. Sci. USA 75:2626-2630.

Hinrichsen RD, Fraga D, Reed MW. 1992. 3'-modified antisense oligodeoxyribonucleotides complementary to calmodulin mRNA alter behavioral responses in *Paramecium*. Proc. Nat. Acad. Sci. USA 89: 8601-8605.

Hoffman DC, Anderson RC, DuBois ML, Prescott DM. 1995. Macronuclear gene-sized molecules of hypotrichs. Nucleic Acids Res 23:1279-1283.

Jahn CL, Nilles LA, Krikau MF. 1988. Organization of the *Euplotes crassus* micronuclear genome. J Protozool. 35:590-601.

Klobutcher LA, Herrick G. 1997. Developmental genome reorganization in ciliated protozoa: the transposon link. Prog Nucleic Acid Res and Mol Biol 56:1-62.

Klobutcher LA, Huff ME, Gonye GE. 1988. Alternative use of chromosome fragmentation sites in the ciliated protozoan *Oxytricha nova*. Nucleic Acids Res 16:251-264.

Klobutcher LA, Swanton MT, Donini P, Prescott DM. 1981. All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. Proc Natl Acad Sci USA 78:3015-3019.
Jahn CL, Klobutcher LA. 2002. Genome remodeling in ciliated protozoa. Ann. Rev. Microbiol. 56:489-520.
Kung C, Saimi Y, Haynes WJ, Ling KY, Kissmehl R. 2000. Recent advances in the molecular genetics of *Paramecium*. J Eukaryot Microbiol 47:11-4.
Landweber LF, Kuo TC, Curtis EA. 2000. Evolution and assembly of an extremely scrambled gene. Proc Natl Acad Sci U S A. 97:3298-30.
Lauth MR, Spear BB, Heumann J, Prescott DM. 1976. DNA of ciliated protozoa: DNA sequence diminution during macronuclear development of *Oxytricha*. Cell 7:67-74.
Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955-964.
Lowe TM, Eddy SR. 1999. A computational screen for methylation guide snoRNAs in yeast. Science 283:1168-1171.
Lozupone CA, Knight RD, Landweber LF. 2001. The molecular basis of nuclear genetic code change in ciliates. Curr Biol. 11:65-74.
Maercker C, Kortwig H, Lipps HJ. 1999. Separation of micronuclear DNA of *Stylonychia lemnae* by pulsed-field electrophoresis and identification of a DNA molecule with a high copy number. Genome Res 9:654-61.
Matsuoka T, Moriyama N, Kida A, Okuda K, Suzuki T, Kotsuki H. 2000. Immunochemical analysis of a photoreceptor protein using anti-IP3 receptor antibody in the unicellular organism, *Blepharisma*. J Photochem Photobiol B 54:131-135.
Meyer E, Garnier O. 2002. Non-Mendelian inheritance and homology-dependent effects in ciliates. Adv. Genet. 46:305-337.
Morin GB, Cech TR. 1986. The telomeres of the linear mitochondrial DNA of *Tetrahymena thermophila* consist of 53 bp tandem repeats. Cell 46:873-883.
Murti KG, Prescott DM. 1983. Replication forms of the gene-sized DNA molecules of hypotrichous ciliates. Mol Cell Biol 3:1562-1566.
Nosek J, Tomaska L, Fukuhara H, Suyama Y, Kovac L. 1998. Linear mitochondrial genomes: 30 years down the line. Trends Genet 14:184-188.
Nanney DL. 1981. T. M. Sonneborn: an interpretation. Annu Rev Genet. 15:1-9.
Orias E. 1998. Mapping the germ-line and somatic genomes of a ciliated protozoan, *Tetrahymena thermophila*. Genome Res 8:91-99.
Orias E. 2000. Toward sequencing the *Tetrahymena* genome: exploiting the gift of nuclear dimorphism. J Eukaryot Microbiol 47:328-333.
Pluta AF, Kaine BP, Spear BB. 1982. The terminal organization of macronuclear DNA in *Oxytricha fallax*. NAR 10:8145-54.
Prescott DM. 2000. Genome gymnastics: unique modes of DNA evolution and processing in ciliates. Nat Rev Genet 1:191-8.
Prescott DM, Dizick SJ. 2000. A unique pattern of intrastrand anomalies in base composition of the DNA in hypotrichs. Nucleic Acids Res 28:4679-4688.
Riley JL, Katz LA. 2001. Widespread distribution of extensive chromosomal fragmentation in ciliates. Mol Biol Evol 18:1372-1377.
Rubin GM *et* ~49 other authors. 2000. Comparative genomics of the eukaryotes. Science 287:2204-2215.
Seegmiller A., Williams KR, Hammersmith RL, Doak TG, Messick T, Witherspoon D, Storjohann LL, Herrick G. 1996. Internal eliminated sequences interrupting the *Oxytricha* 81 locus: allelic divergence, conservation, conversions, and possible transposon origins. Mol Biol Evol 13:1351-1362.
Seegmiller A, Williams KR, Herrick G. 1997. Two two-gene macronuclear chromosomes of the hypotrichous ciliates *Oxytricha fallax* and *O. trifallax* generated by alternative processing of the 81 locus. Dev Genet 20: 438-357.
Skovorodkin IN, Zassoukhina IB, Hojak S, Ammermann D, Günzl A. 2001. Minichromosomal DNA replication in the macronucleus of the hypotrichous ciliate *Stylonychia lemnae* is independent of chromosome-internal sequences. Chromosoma 110:352-359.
Soldo AT, Brickson SA, Larin F. 1981. The kinetic and analytical complexities of the DNA genomes of certain marine and fresh-water ciliates. J. Protozool 28:377-383.
Sonneborn TM. 1975. *Tetrahymena pyriformis*. in Handbook of Genetics Vol 2 (RC King ed) Plenum Press, NY, pp 433-467 (Plenum, NY).
Sonneborn TM. 1977. Genetics of cellular differentiation: stable nuclear differentiation in eucaryotic unicells. Annu Rev Genet 11:349-367.
Sweeney R, Fan Q, Yao MC. 1998. Antisense in abundance: the ribosome as a vehicle for antisense RNA. Genet Eng. (N Y). 20:143-151.

Tondravi MM, Yao MC. 1986. Transformation of *Tetrahymena thermophila* by microinjection of ribosomal RNA genes. Proc Natl Acad Sci USA 83:4369-4373.
Williams K, Doak TG, Herrick G. 1993. Precise Excision of *Oxytricha trifallax* Telomere-Bearing Elements and formation of circles closed by a copy of the flanking target duplication. EMBO J 12:4593-4601.
Williams K, Doak TG, Herrick G. Chromatid breakage and telomere formation in ciliated protozoa: correlation between types of breakage and telomere-addition patterns in *Oxytricha fallax* and *O. trifallax*. Submitted to BMC Genetics, 4/02.
Wright ADG, Lynn DH. 1997. Maximum ages of ciliate lineages estimated using a small subunit rRNA molecular clock: Crown eukaryotes date back to the Paleoproterozoic. Archiv. Protistenkd. 148: 329-341.
Yao M-C, Duharcourt S, Chalker D. 2002. Genome-wide rearrangements of DNA in ciliates. In: Mobile DNA II. Craig NL, Craigie R, Gellert M, Lambowitz AM, eds. ASM Press, Washington, D.C., pp. 730-758.