Order Number 9126847

Transposons in the micronucleus of *Tetrahymena thermophila*

Wyman, Claire Louise, Ph.D.

University of California, Berkeley, 1990

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106

Transposons in the Micronucleus of *Tetrahymena thermophila*

By

Claire Louise Wyman

B.S. (The Johns Hopkins University) 1980
M.S. (The Johns Hopkins University) 1984

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

MOLECULAR BIOLOGY

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA at BERKELEY

Approved:

Elizabeth H. Blackburn ..... September 12, 1990
Chair **Date**

Michael R. Botchan ..... Sept. 11, 1990

William Rutter ..... Sept. 17, 1990

****

Abstract

Transposons in the Micronucleus of *Tetrahymena thermophila*

by

Claire Wyman

The *Tetrahymena thermophila* micronuclear genome is shown to contain repetitive DNA elements resembling transposons. Two types of elements are described, one type is structurally similar to elements that move via DNA intermediates and the other type is structurally similar to elements that move via RNA intermediates.

The existence of Tel-1 elements was inferred from previous studies of the *T. thermophila* micronuclear genome. The proposed structure of Tel-1 elements, short repeated sequences in inverted orientation at their ends, suggested that they were transposons of the class that move via a DNA intermediate. In order to determine if Tel-1 elements are mobile, the *T. thermophila* micronuclear genome was examined for molecular evidence of Tel-1 transposition. As assayed by Southern blotting, most standard laboratory strains of *T. thermophila* have indistinguishable distributions of Tel-1 elements in their micronuclear genomes. Differences in the micronuclear arrangement of Tel-1 elements are seen between a newly isolated *T. thermophila* strain and a standard laboratory strain, and also between laboratory strains that had been genetically manipulated to produce specific micronuclear genome anomalies. The observed rearrangements of Tel-1 elements are consistent with the hypothesis that these elements are mobile.

Elements from the other class, those that move via an RNA intermediate, have not previously been described in *T. thermophila*. Tmi elements are identified here as micronuclear limited middle repetitive elements that resemble RNA based mobile elements by sequence comparison. Several independent clones containing *T. thermophila* micronuclear DNA inserts included elements with a common restriction map. These elements are called Tmi for Tetrahymena micronuclear elements. The restriction map of the clones is shared by the majority of the genomic Tmi elements. From the sequence of cloned portions of Tmi elements, one end of the element is defined and several open reading frames (ORFs) are identified. Separate regions of the Tmi nucleotide sequence are similar to the human retrovirus HIV 2 and the Drosophila retrotransposon 17.6. Four amino acid motifs conserved among reverse transcriptases can be found with the correct spacing in the deduced amino acid sequence of one of the Tmi ORFs.

Elizabeth Blackburn
September 12, 1990

" The best thing to do is find what there is, and there's plenty."
(John Cage on NPR, May 7, 1989)

To those who are part of what there is and have helped me find much more especially my parents and Roland.

Acknowledgements

I would like to thank Liz Blackburn for her contributions to my scientific personality which are invaluable and innumerable. I have enjoyed working in her lab because it is a place where interesting ideas evolve.

Nick Cozzarelli has kindly provided space in his lab so that I could write my thesis here at Berkeley after Liz's lab had moved to San Francisco. I appreciate the hospitality he and the members of his lab have shown me for the last three months.

During his sabbatical year in Berkeley I learned a great deal about Tetrahymena genetics from Ed Orias.

I would like to acknowledge the influence Berkeley (not necessarily the university) has had on my life. This is a lively place to live and work where complacency is actively discouraged.

Several generations of Blackburn lab members have helped me along the way with ideas, techniques, encouragement and friendship:
John Bradley, Nicole Calakos (who helped enthusiastically on what turned out to be a hopeless project and sometimes even took my advice), Mike Cherry (who cured me of my obstinate objections to using computers), Phil Cohen (who reads interesting books, which I am not convinced proves that he is either not alive or not in the western world), Monica Flessel, Susan Fong, Jim Forney (who is the baghwan, who was the best darn post-doc named Jim, who cloned the DNA beyond the telomeres and who always asked the right questions), Carol Greider (who knows how I feel about her too. Thanks for sharing your opinions on experiments and seminars and life over afternoons of mini-

preps or evenings of beers), Eric Henderson (a few words, even the two he thought most common in the English language, could not do him justice), Drena Larson (who radiates kindness), Margaret Lee, Vicki Lundblad, Pan Wei-Chun, Huan Phan, Liz Robinson, Dan Romero, Janice Shampay, Dorothy Shippen-Lentz (who can turn a mean cartwheel if adequately inspired), Beth Spangler, Cathy Strahl, and Guoliang Yu.

The now disestablished Department of Molecular Biology and community of Stanley Hall was a near magical place to be a student. I did not imagine graduate school would involve working so hard and having so much fun. The people who worked in Stanley Hall were able to make it feel more like a family than a cog in the wheel of the university's bureaucracy. Andrea Jeremy was not only graduate assistant but a friend. We are all indebted to her for making a sport of cutting red tape and thereby making the life of a student much easier. Helen Wittmer graced the administrative office with her warmth and charm. Gary Madison kept mechanical things in the building and his sense of humor in working order, both against great odds at times. Don Berry could always be counted on for plates and media and a smile.

Thanks to the following people (and many more) for sharing the childlike joys of adventures in late night experimentation, "creative" uses of lab equipment and, on celebratory occasions, audio inspired kinetic body sculpture: Howard Benjamin, Mike Carey, Brian Condie, Jan Dungan, Carol Greider, Kathy Flynn, Kathy Hudson, Tom Jongens, and Jeff Reynolds.

Since they have yet to win the coveted Spoon Set and it is my fault, I would like to acknowledge those crazy Canadians, Paul MacPherson and Phil Cohen, for creating the best acts in the follies for the last two years.

The inhabitants of 2337 Grant St. provided a most excellent virtual experience of family here in Berkeley. The yard provided the site for the summer of endless bar-b-ques, which continued for several years in total disregard to calendar time. I enjoyed many a late night kitchen table conversation over a beer or three with Brian Condie and Kathy Hudson. Kathy became a friend forever. Her opinions and perceptions are always a treat. Roland and I can think of no one we'd rather spend a week or two walking around in the woods discussing bodily functions with. Besides you own Barbie.

I thank Barbara Sloan for shoes, fashion advise, travel adventures on both coasts and introducing me to the Madonna Inn.

Aan mijn Nederlandse familie, bedankt for het gevoel dat Nederland ook thuis is.

I have enjoyed the encouragement of my father Lyle, mother Carol, and brother Bruce in all my academic persuits. They have always been interested in what I'm up to.

I have too many things to thank Roland for than can be written here. I am grateful for his patience with my bad grammar while learning Dutch and my occasional bad humor while sequencing and writing. Thanks also for all the Mac Boxes featured in the figures here. I thank " my lucky stars" or "the powers that be" that we ever ended up in the same place. Roland deserves credit for coming up with what everyone else thought were good scientific reasons for returning to Berkeley. Roland, the rest does not require words.

Table of Contents

Abstract 1

Acknowledgements iii

List of Tables vii

List of Figures viii

List of Abbreviation x

Chapter 1. Introduction 1

Chapter 2.Materials and Methods 11

Chapter 3. Tel-1 Elements are Associated with Micronuclear Genome Rearrangements in Some *Tetrahymena thermophila* Strains 29

Introduction 29

Results 31

Discussion 38

Chapter 4.Identification of Tmi elements 63

Introduction 63

Results and Discussion 64

Chapter 5.Sequence Analysis of Tmi Elements 98

Introduction 98

Results 99

Discussion 108

Appendix. Analysis of Lambda Clones Containing Candidate Tel-1 Elements from the *Tetrahymena thermophila* Micronuclear Genome 142

Introduction 142

Results 143

Discussion 148

References 158

List of Tables

Table 3.1 Genotype and phenotype of *Tetrahymena thermophila* strains.....42

Table 3.2 Results of test crosses involving SB2040.1 and SB2040.2 cell stocks.....43

Table 5.1 Triplet rations of the Tmi ORFs and the corresponding +1 and +2 Reading Frames.....114

Table 5.2 Codon usage of the Tmi ORFs compared to that compiled from the sequences of other Tetrahymena genes.....115

Table 5.3 Reverse transcriptase motif comparison.....116

List of Figures

Figure 1.1 DNA processing events of macronulcear development 10

Figure 3.1 Restriction map of the *T. thermophila* micronuclear DNA insert in plasmid p17D59 44

Figure 3.2 Nuclear events in Tetrahymena conjugation 46

Figure 3.3 Southern blot of *BstXI* digested micronuclear and macronuclear DNA probed with Tel-1 51

Figure 3.4 Southern blot of *HindIII* digested micronuclear DNA from three standard laboratory strains probed with Tel-1 53

Figure 3.5 Southern blot of micronuclear DNA from a newly isolated *Tetrahymena thermophila* strain probed with Tel-1 and 5S 55

Figure 3.6 Southern blot of micronuclear DNA from nullisomic strains probed with Tel-1 57

Figure 3.7 Southern blot of micronuclear DNA from four identical whole-genome homozygote strains probed with Tel-1 59

Figure 3.8 Southern blot of micronuclear DNA from both stocks of cell line SB2040 probed with Tel-1 and 5S 61

Figure 4.1 Proposed scheme for amplification of Tel-1 elements using a primer from the inverted repeat end 75

Figure 4.2 The PCR 900 probe forms apparent half-length fragments upon heating and quick cooling 77

Figure 4.3 Sequence of the PCR 900 product 79

Figure 4.4 PCR 900 probed Southern blot of *ClaI* cut DNA 82

Figure 4.5 PCR 900 probed Southern blot of *HindIII* cut micronuclear DNA 84

Figure 4.6 PCR 900 probed Southern blot of whole cell DNA 86

Figure 4.7 Restriction maps of clones selected by hybridization to the PCR 900 probe 88

Figure 4.8 Southern blot of genomic DNA probed with common fragments from the PCR selected clones 90

Figure 4.9 Restriction maps of clones F08 F23 and D01 92

Figure 4.10 Hybridization of Tmi end and internal fragment probes to *Hin*dIII cut micronuclear DNA 94

Figure 4.11 Southern blot of genomic DNA from three Tetrahymena species probed with common Tmi fragments 96

Figure 5.1 Sequencing strategy for the three Tmi containing clones 117

Figure 5.2 F08 and F23 sequences aligned 119

Figure 5.3 Tmi sequence extended through clone D01 122

Figure 5.4 Alignment of Tmi end with P element end 124

Figure 5.5 Histogram of sequence comparison scores for the left part of Tmi and the invertebrate section of the GenBank database 126

Figure 5.6 Diagram of the overlaped similar regions of Tmi and retro-element sequences from GenBank 128

Figure 5.7 Histogram of sequence comparison scores from the right part of Tmi and the virus section on the GenBank database 130

Figure 5.8 Position of open reading frames in the Tmi sequence 132

Figure 5.9 Amino acid sequence of the proteins potentially encoded by the Tmi ORFs 134

Figure 5.10 Alignment of amino acid motifs of reverse transcriptases and the ORF 3 predicted protein 140

Figure A.1 Restriction maps of Tel-1 probe selected lambda clones 150

Figure A.2 Sequence around three of the *BstX*I sites of clone 22 152

Figure A.3 Restriction maps of C_4A_2 selected lambda clones 155

List of Abbreviations

EDTA	Ethylenediamine tetraacetate
kb	Kilobase pair
KOAc	Potassium Acetate
NH_4OAc	Ammonium Acetate
OD	Optical Density
PEG	Polyethylene Glycol
PIPES	1,4-Piperazinediethanesulfonic acid
rpm	revolutions per minute
SDS	Sodium Dodecylsulfate
Tris-HCl	Tris(hydroxymethyl)aminomethane titrated with HCl to the indicated pH
UV	Ultraviolet

"...Mondaugen set out for his turet and oscilograph, and the comforts of Science, which are glacial and few."

Thomas Pynchon, V.

Chapter 1

Introduction

Mobile genetic elements were first described by McClintock as the agents of unstable mutations in maize. They have now been described in every organism from which they have been actively sought. A large portion of the repetitive sequences in eukaryotic genomes are believed to be mobile elements or their remnants. Mobile genetic elements account for up to 10% of the Drosophila genome (Manning et al., 1975). As much as 10% of the mouse genome consists of a single type of element, L1Md (Hutchinson et al., 1989). Speculation over the function of mobile elements varies from accusations that they are merely parasites (Doolittle and Sapienza, 1980; Orgel and Crick, 1980) to proposals that they are the origin of introns (Cech, 1985), agents that regulate gene expression (Britten and Davidson, 1969; Davidson and Britten, 1979) or major factors in genome evolution (Rogers, 1985).

The subject of this thesis is the description of two types of mobile elements that inhabit the *Tetrahymena thermophila* micronuclear genome. Rather than trying to determine the larger issue of whether mobile elements are selfish hitchhikers or slaves to the functional and evolutionary survival of their hosts, this study has two relatively modest aims that pertain to *T. thermophila*. First, it is hoped that mobile elements can be domesticated in *T. thermophila* as they have been in other species, e.g., the P elements of Drosophila (Rubin and Spradling, 1982). *T. thermophila* mobile elements could be used in the construction of vectors for transforming the *T. thermophila* germline. Currently there are no methods available for transformation of the *T. thermophila* micronucleus. Second, understanding how transposons move in *T.*

thermophila might help in understanding the mechanism of the dramatic genome rearrangements that occur in the development of the Tetrahymena somatic nucleus. Some of these genome rearrangement reactions appear to be mechanistically similar to the reactions that RNA-based mobile elements use for replication or those that DNA-based mobile elements use for movement.

Genome Organization and Reorganization in *Tetrahymena thermophila*

T. thermophila, although unicellular, like other ciliated protozoa has two functionally distinct nuclei. These nuclei differ in physical size, function, DNA content, genome structure, time of replication and mode of division. The germline micronucleus contains five pairs of chromosomes, divides mitotically, replicates early in the cell cycle and is almost entirely transcriptionally silent. The somatic macronucleus contains a highly processed and amplified version of the micronuclear genome. It replicates late in the cell cycle, divides amitotically and provides most if not all expression of the nuclear genes.

A new macronucleus is formed from a division product of the micronucleus following conjugation, the sexual phase of *T. thermophila* development (see chapter 3 for a detailed description of the nuclear events of conjugation). There are several different aspects to the developmentally controlled genome restructuring required to produce a functional macronucleus. These processes are diagramed in Figure 1 (the diagram is not meant to imply temporal order, for a review see Gall, 1986). The micronuclear chromosomes are fragmented into hundreds of pieces resulting in macronuclear DNA molecules whose average size is 600 kb. DNA rearrangement events remove sequences from both within and between the future macronuclear DNA molecules. In all, 10-20% of the micronuclear genome complexity is lost by these sequence elimination events.

Telomeres are added to the ends of the macronuclear-destined DNA molecules. All macronuclear DNA molecules are amplified to about 45C with the exception of the ribosomal RNA genes. These genes occur on a palindromic DNA molecule that is present in about 10,000 copies in the macronucleus.

The deletions of internal sequences from the macronuclear destined regions are estimated to occur at 5,000 sites during macronuclear development in *T. thermophila* (Howard and Blackburn, 1985; Yao et al., 1984). DNA rearrangements with some similar features are used more sparingly in the mammalian immune system to produce functional immunoglobulin and receptor molecules (Lutzker and Alt, 1989). In ciliates, these DNA rearrangements are required, in some cases, to produce functional genes. Examples of the generation of coding sequence by precise DNA deletions have been described in *Oxytricha nova* (Klobutcher et al., 1984), *Oxytricha fallax* (Herrick et al., 1987) and *Euplotes crassus* (Baird et al., 1989).

The rearranged genome contained in the macronucleus seems to be required to achieve adequate and efficient gene expression in ciliates. Sequences eliminated from the macronucleus possibly served functions specific to the micronucleus that are no longer needed or are detrimental to the macronucleus. They could for instance be micronuclear specific origins of replication that direct synthesis of micronuclear DNA early in the cell cycle. While another possibility is that some of them are centromeres that are needed for mitotic division. Micronuclear limited sequences may specifically prevent transcription in the micronucleus or make the genome inaccessible to the transcription apparatus by favoring compact forms of DNA and chromosomes.

The sequences removed from within macronuclear destined DNA are called internal eliminated sequences or IESs. Some IESs are unique sequences and others are

repetitive. The sequence structure of both repetitive and unique IESs suggests that they are removed in a process that resembles transposon excision. A study of nine unique IESs from *Oxytricha nova* (Ribas-Aparicio et al., 1987) and three unique IESs from *Euplotes crassus* (Baird, et al., 1989) revealed that they were all bounded by short direct repeats (2-6 bp) usually surrounding (sometimes imperfect) inverted repeats. One copy of the direct repeats was retained in the macronuclear sequence after the IES had been removed. The three sequenced IESs from *T. thermophila* share the characteristic of short direct repeats one of which is maintained in the processed macronucleus (Austerberry and Yao, 1987; Austerberry and Yao, 1988). This arrangement of sequence elements resembles a transposon with inverted repeats at its ends surrounded by a direct repeat of its integration site. Of the four models for IES removal proposed (Ribas-Aparicio, et al., 1987) two share mechanistic features with transposition events. One of these presents IES removal as analogous to precise excision of a transposon that moves via a DNA intermediate, such as P elements of Drosophila or Tc1 elements of *Caenorhabditis elegans*. The other model proposes that IESs are spliced from RNA copies of micronuclear sequences that are then reverse transcribed to produce macronuclear DNA molecules. This mechanism resembles steps in the movement of mobile elements that have an RNA intermediate; the retroviruses, retrotransposons and retroposons. Aside from one unconfirmed report of reverse transcriptase activity in a ciliate undergoing macronuclear development (Lipps, 1985), there is currently no evidence to support this RNA-based model for macronuclear development.

Some repetitive IESs more clearly resemble transposons. In *Euplotes crassus*, Tec1 elements are repetitive micronuclear limited sequences that were commonly found in association with macronuclear destined sequences (Jahn et al., 1988). In at least one case a Tec1 element must be specifically removed to create a macronuclear gene (Baird, et al., 1989). The terminal portions of two Tec1 elements have been sequenced and

shown to consist of long terminal inverted repeats (Baird, et al., 1989; Jahn et al., 1989). It was shown that all Tec1 sequences are eliminated from the developing macronucleus at about the same time. Extrachromosomal circular copies of Tec1 appear at this time (Jahn, et al., 1989). Extrachromosomal circular forms have been described for transposable elements of *Caenorhabditis elegans* (Ruan and Emmons, 1984), maize (Sudaresan and Freeling, 1987), and *Escherichia coli* (Morisato and Kleckner, 1984). The telomere bearing elements (TBEs) of *Oxytricha fallax* are another example of micronuclear limited elements with inverted repeat ends (Herrick et al., 1985). Some TBEs are also specifically removed during macronuclear development (Hunter et al., 1989). The structure of Tec1 and TBE1 suggests they are transposons whose removal is mechanistically similar to precise excision of well characterized transposons. Tausta and Klobutcher (1989) suggest that the unique IESs may be lower affinity sites for the same DNA cutting and ligation machinery responsible for excision of the repetitive transposon like IESs.

Characteristics of Mobile Genetic Elements

Mobile genetic elements are generally middle repetitive in their resident genome, and due to their ability to move show polymorphic distributions in the genome of different strains or isolates of a species. Mobile genetic elements can be divided into two broad classes based on their molecular characteristics. Those that employ a DNA intermediate in transposition share a collection of structural features and those that employ an RNA intermediate have a different set of shared characteristics. Well studied examples of transposons with DNA intermediates include the P elements of *Drosophila melanogaster*, Ac and Ds elements of maize and Tc1 elements of *Caenorhabditis elegans* (all reviewed in Berg and Howe, 1989). For ease of identification I will refer to this class of elements as the P-like elements. The other broad class of mobile elements that move

via an RNA intermediate includes retroviruses, retrotransposons such as Ty1 elements of yeast, copia and related elements of Drosophila, and retroposons such as mammalian LINEs (all reviewed in Berg and Howe, 1989). I will refer to these elements collectively as retro-elements.

P-like elements have short inverted repeated sequences at their ends. They create a direct duplication of their target site sequence upon insertion such that copies of the target site border the element. The length of the target site duplication is characteristic of the particular element. Element copies capable of promoting their own movement encode a required protein called a transposase. Transposition reactions include insertion of DNA elements into the genome and excision of elements from the genome.

The steps required for movement of a retro-element include synthesis of an RNA copy of the element, reverse transcription of the RNA to make a DNA copy and insertion of the DNA copy into the host genome. Removal of DNA copies of the element from the genome is not a normal part of the mobile lifestyle of these elements. Retro-elements include retroviruses, non-viral retrotransposons, and retroposons. Retrotransposons and retroposons preserve the features of retroviruses to lesser degrees.

The features of the generic retroviral genome used in the identification of retro-elements discussed in this thesis are: (1) the presence of long direct repeat sequences at their ends (LTRs), (2) the capacity to encode a reverse transcriptase necessary to convert the RNA intermediate into a DNA copy, (3) a characteristic order of genes needed to replicate the viral genome, make viral particles and integrate the viral genome into the host genome. A typical viral genome encodes polypeptides with several functions, often as polyproteins, that must be processed into individual gene products. The characteristic arrangement of genes in order of transcription is; gag (encoding the group

specific antigen and a protease function), pol (encoding the reverse transcriptase, RNase H, and integrase functions), and env (encoding structural proteins of the viral particles). The functions listed here are common to all retroviruses. In addition many viruses encode gene products with specific regulatory functions. Integration of the DNA copy of a retrovirus or other retro-element also results in duplications of the target sequence flanking the integrated viral genome.

Retroposons also move via an RNA intermediate and share some of the structural features of retroviruses. They have LTRs, an open reading frame for a reverse transcriptase, and sometimes coding potential for other proteins similar to retroviral proteins, often in the same order as in a retrovirus genome. Ty1 (Eichinger and Boeke, 1988; Garfinkel et al., 1985) of yeast and copia (Shiba and Saigo, 1983) of Drosophila even form intracellular virus-like particles (VLPs) but lack extracellular forms. Reverse transcriptase activity has been demonstrated to be associated with the VLPs (Garfinkel, et al., 1985). Other retrotransposons such as 17.6 of Drosophila have LTRs and open reading frames for several retroviral-like proteins, gag, pol and env in the order expected of a retrovirus (Kugimiya et al., 1983; Saigo et al., 1984), but have not been shown to form VLPs.

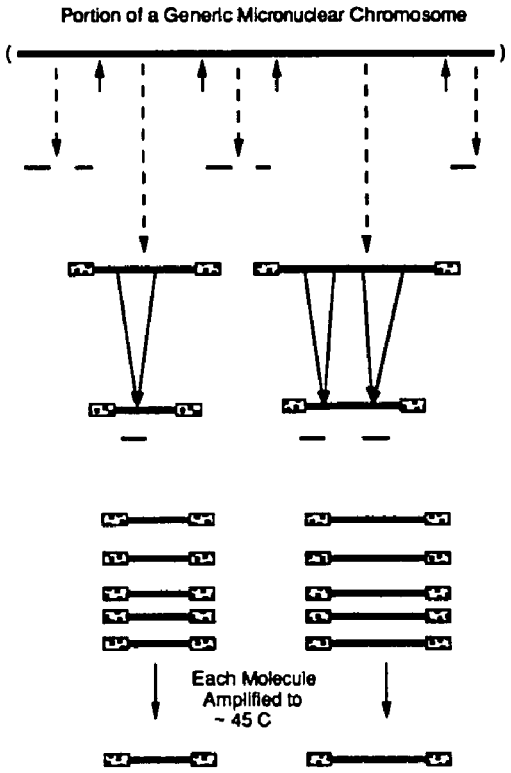
Retro-elements lacking LTRs have been called retroposons by Rogers (1985). They include dispersed repetitive sequences such as LINEs (Hutchinson, et al., 1989). Some have coding capacity for more than one retroviral-like protein but all complete elements share a coding capacity for a reverse transcriptase-like protein. There are no characteristic sequence repeats at their termini. Some such elements have poly A stretches at one end presumably originating from the poly A tails of the RNA copy of the element.

This thesis presents evidence that mobile elements of both the P-like and retro-element classes exist in the *Tetrahymena thermophila* micronuclear genome. Elements of the P-like class, Tel-1 elements, were previously described (Cherry and Blackburn, 1985) and believed to end in short inverted repeats. They are like TBE1 of Oxytricha (Herrick, et al., 1985) by virtue of having telomeric repeats at their ends even though they were embedded in internal regions of the micronuclear chromosomes. The Tel-1 ends identified in micronuclear DNA were eliminated from macronuclear sequences but were not themselves the sites of DNA breakage and rejoining (Cherry and Blackburn, 1985). The previous work did not address the mobility of Tel-1 elements within the micronucleus. Demonstrating that a particular element is mobile and identifying conditions which promote its mobility are among the fist steps in identifying candidate elements for development into transformation vectors. Chapter 3 presents evidence that Tel-1 elements are mobile in the micronuclear genome of cells that have experienced unusual nuclear events.

Another class of middle repetitive micronuclear limited sequences, that resemble retro-elements, is described in chapter 4. These elements are named Tmi for Tetrahymena micronuclear elements. Chapter 5 presents the results of sequencing the portions of Tmi elements that have been cloned. This revealed that they share regions of similarity with both a retrovirus and a retrotransposon. One Tmi open reading frame can encode a polypeptide which contains sequence motifs common to reverse transcriptases. It is also possible to develop transformation vectors from retro-elements as has been done for mammalian cells using retroviruses (Gilboa et al., 1986) and seems imminent for yeast using Ty1 (Eichinger and Boeke, 1990; Jacobs et al., 1988). In addition one of the models for macronuclear genome formation involved reactions similar to those of retro-element replication (Ribas-Aparicio, et al., 1987). A simple version of this model involving transcription of micronuclear sequences,

splicing out the micronuclear limited regions, and reverse transcription to make the macronuclear DNA molecules is unlikely to account for IES removal in *T. thermophila*. The micronuclear sequences to be transcribed would be expected to be larger than the final macronuclear DNA molecules which in Tetrahymena are on average 600 kb. Nevertheless, a repetitive sequence resembling a reverse transcriptase gene that is limited to the micronucleus is very tantalizing in this regard. A complete description of a functional Tel-1 or Tmi element has not yet been achieved. However, the evidence suggests that functional versions of these elements exist and searching for them is likely to be fruitful.

Figure 1.1 DNA Processing Events of Macronuclear Development



DNA Rearrangement Events

Chromosome Fragmentation

Micronuclear Limited Sequences Degraded

Telomeres Added to Macronuclear Destined DNA

Internal Micronuclear Limited Sequences Removed

Micronuclear Limited Sequences Degraded

Mature Macronuclear DNA Moleucles Amplified

Chapter 2

Materials and Methods

Nucleic Acid Isolation Procedures

A. Rapid Boiling Method for Plasmid DNA Isolation

This protocol is essentially as described by Holmes and Quigley (1981). Bacterial cells harboring plasmids were grown at 37°C with shaking in LB (1% Bacto-tryptone, 0.5% Bacto-yeast extract, 1% NaCl) containing the appropriate antibiotics. Cells were collected from 1 ml of an overnight culture by centrifugation in an Eppendorf microcentrifuge for 2 minutes and the supernatant was discarded. The pellet was resuspended in 100 µl of 8% sucrose, 5% Triton X-100, 50 mM EDTA, 50 mM Tris-HCl (pH 8.0), and 5 µl 10 mg/ml lysozyme. Samples were boiled in a water bath for 40 seconds and then centrifuged for 10 minutes. The pellet was removed and discarded. The DNA was precipitated form the supernatant by adding 120 µl of isopropanol, incubating at -20°C for 5 minutes and centrifuging for 5 minutes. The pellet was dried and resuspended in 30 µl TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA).

B. Rapid Alkali-SDS Plasmid Preparation from 10 ml Overnight Cultures

This protocol is a modification of that described by Birnboim and Doly (1979). 10 ml of an overnight culture of bacterial cells, grown in LB with antibiotics at 37°C with shaking, was centrifuged in a Falcon 2006 tube for 5 minutes at 5,000 rpm in a Sorvall HB-4 rotor. After discarding the supernatant the pellet was resuspended in 400 µl of 20 mM EDTA, 50 mM Tris-HCl (pH 8.0), 50 mM glucose. After incubated at room temperature for 5 minutes and incubation on ice for 10 minutes, 800 µl of 0.2 M NaOH,

1% SDS was added and mixed by pipetting. This mixture was incubated on ice for 10 minutes after which 600 µl of 3 M KOAc (pH 4.8) was added, mixed by pipetting and further incubated on ice for 15 minutes. This solution was centrifuged for 10 minutes at 10,000 rpm in a Sorvall HB-4 rotor. The supernatant was removed to a clean Falcon 2006 tube and 0.6 volumes of isopropanol was added. After a 30 minute incubation on ice this was centrifuged for 10 minutes at 10,000 rpm in a Sorvall HB-4 rotor. The pellet was washed with 70% ethanol, resuspended in 200 µl TE and transferred to a 1.5 ml Eppendorf tube. The Falcon tube was washed with an additional 100 µl TE. This combined 300 µl of resuspended DNA was extracted twice with an equal volume of phenol:chloroform (1:1) buffered with 100 mM Tris-HCl [pH 8.0], and precipitated with an equal volume of 5 M NH_4OAc and two volumes of isopropanol. After incubation on ice for 15 minutes the precipitate was collected by centrifugation in an Eppendorf microcentrifuge for 20 minutes. This pellet was resuspended in 200 µl of 2.5 M NH_4OAc and reprecipitated with 2.5 volumes of 95% ethanol. After incubation on ice for 15 minutes the precipitate was collected by centrifugation in an Eppendorf microcentrifuge for 20 minutes. The DNA pellet was dried and resuspended in 100 µl TE.

C. Preparation of Lambda Phage DNA

For a 50 ml culture,100 µl of recombinant phage in lambda EMBL4 vector from a plaque plug in SM (0.1M NaCl, 0.01M $MgSO_4$, 0.05M Tris-HCl [pH 7.5], 0.01% gelatin) or 10 µl of a phage lysate stock was added to 200 µl of a stationary culture of *E. coli* strain NM538. The phage were allowed to adsorb to the cells for 30 minutes at 37°C without shaking and then diluted to 50 ml with LB supplemented with $MgCl_2$ to 10 mM. Following overnight growth at 37°C with shaking, well lysed cultures were treated with 2-3 ml of chloroform for 20 minutes at room temperature. The cell debris was removed by centrifugation in 50 ml tubes for 10 minutes at 8,000 rpm in a Sorvall SS 34 rotor. The

supernatant was transfered to a 50 ml screw cap plastic Corning tube and treated with DNase I and RNase A each at 1 µg/ml for 30-60 minutes at 37°C. This was then centrifuged at 26,000 rpm for 2 hours in a Ti 35 rotor in a Beckman ultracentrifuge. The supernatant was discarded and the phage pellet resuspended in 2 ml SM and transfered to a 15 ml Falcon 2006 tube. DNA was extracted from the phage by the addition of 0.4 ml 0.25 M EDTA, 0.5 M Tris-HCl (pH 9.0), 2.5 % SDS and incubation at 65°C for 30 minutes. Then 0.5 ml 8 M KOAc was added and the mixture chilled on ice for 15 minutes. The DNA remained in solution upon centrifugation for 15 minutes at 10,000 rpm in a Sorvall SS 34 rotor. The supernatant was decanted into a clean Falcon 2006 tube and 5.6 ml of 95% ethanol was added. After incubation at -20°C for 1 hour the DNA was precipitated by centrifugation for 30 minutes at 16,000 rpm in a Sorvall SS 34 rotor. The pellet was dissolved in 0.4 ml 2 M NH_4OAc and transfered to a 1.5 ml microfuge tube. This was extracted once or twice with phenol:chloroform (1:1) equilibrated with 100 mM Tris-HCl (pH 8.0). The DNA was precipitated by adding 0.8 ml 95% ethanol, chilling at -20°C for 15 minutes and centrifuging for 5 minutes in an Eppendorf microcentrifuge. After drying the pellet was resuspended in 50-100 µl TE.

D. Preparation of Single-Stranded DNA from Cultures of Cells Harboring pUC118 and pUC119 Plasmids

Recombinant pUC118 or pUC119 plasmids to be used to produce single-stranded DNA were maintained in *E. coli* MV1193. The plasmids, host strain, helper phage and protocol for their use were obtained form Dr. J. Vieira (University of Minnisota). Either a single colony or 10-20 µl of an overnight culture of MV1193 containing the plasmid of interest was inoculated into 2 ml 2X YT broth (1% Bacto-yeast extract, 1.6% Bacto-tryptone, 1% NaCl) supplemented with 100 µg/ml ampicillin, 10^7 M13K07 phage/ml and, 0.001% thiamine. After 2 hours growth at 37°C with shaking, kanamycin was added to 70 µg/ml and growth was continued overnight.

The cells were removed from 1.5 ml of overnight culture by centrifuging it twice for 2 minutes in a microcentrifuge removing only part of the supernatant each time. Phage were precipitated from 1-1.2 ml of supernatant by the addition of 300 µl 20% PEG 6000, 2.5M NaCl, incubation at room temperature for 1-2 hours and centrifugation for 2 minutes. The PEG supernatant was carefully removed with a drawn out Pasteur pipette. The centrifugation and supernatant removal were repeated once. The phage pellet was extracted 3 or 4 times with phenol:chloroform (1:1) equilibrated with 100 mM Tris-HCl (pH 8.0). The DNA was precipitated by the addition of an equal volume of 5 M NH_4OAc and 2.2 volumes of 95% ethanol, incubation on ice for 15-20 minutes and centrifugation for 20 minutes. The pellet was dried and resuspended in 20 µl of TE.

E. Preparation of RNA from *T. thermophila*

RNA was prepared by one of two methods from 200 ml cultures of logarithmically growing or starved *T. thermophila* cells at a density of approximately 2.5×10^5 cells/ml. The cells were pelleted in 50 ml screw cap Corning plastic tubes by centrifugation for 4 minutes at approximately 1,500 rpm in an IEC bench top centrifuge. The supernatant was removed by aspiration. At this point total RNA was isolated by the Guanidium Method as described in Current Protocols in Molecular Biology (Ausubel et al., 1988) starting with step 2 or by a phenol extraction method essentially as described by Martindale and Bruns (1983) with the modifications noted below.

RNA was prepared by phenol extraction of total *T. thermophila* nucleic acids. The cell pellets were harvested as described above from 200 ml of culture and pooled by resuspension in 6 ml lysis buffer (0.1 M NaCl, 10 mM EDTA, 10 mM Tris-HCl [pH 7.4], 1% SDS, 0.5 mg/ml heparin, 2 mM aurintricarboxylic acid). The cells were lysed in this

buffer by incubation for 5 minutes at room temperature. The cell lysate was extracted with phenol:chloroform (1:1, equilibrated with 100 mM Tris HCl, pH 7.4) four times or more until there was no material visible at the interface. Nucleic acids were precipitated by adding 1/10 volume of 3M NaOAc and 2 volumes 95% ethanol, incubating at -20°C for several hours and centrifuging for 10 minutes at 10,000 rpm in a Sorvall HB-4 rotor. The pellet was washed with 70% ethanol and dried. The nucleic acids were resuspended in 100-300 µl RNase free distilled water and stored at -80 °C.

Poly-A^+ RNA was purified from total cellular RNA by the following method. Up to 1 mg of RNA in 0.5 ml distilled water was denatured by adding 10 µl of 0.5 M EDTA, 100 µl of 0.2 M PIPES (pH 6.4), 300 µl of formamide and heating to 65°C for 5 minutes. The denatured RNA was diluted immediately into binding buffer (0.4 M NaCl, 20 mM Na-citrate, 0.05% SDS, 5 mM EDTA), such that the concentration of RNA was less than 250 µg/ml, and transferred to a Corning 15 or 50 ml screw cap tube. 100-200 µl (settled volume) of oligo dT cellulose (Collaborative Research type III), stored in binding buffer at 4°C, was added to the RNA in binding buffer and the suspension was rotated or rocked at room temperature for 30 minutes. The resin was pelleted by centrifugation at 2,000 rpm for 2 minutes in an IEC benchtop centrifuge. The supernatant was discarded and the resin was washed with a tubefull of binding buffer. After centrifugation as before, the supernatant was again discarded. The washed resin was transfered to a 1 ml syringe pluged with siliconized glass wool. The column was washed with 2 ml binding buffer followed by 2 ml of 1/2 X binding buffer and 50 µl of elution solution (1 mM EDTA). The RNA was eluted with elution solution pre-warmed to 65°C. Eight 0.5 ml fractions were collected. RNA was precipitated by adding 50 µl 3M NaOAc and 1 ml 95% ethanol to each fraction and freezing. After centrifugation for 15 minutes the supernatants were discarded and the pellets were resuspended in a combined volume of 450 µl water. Undissolved material was removed by centrifuging for 5 minutes in an Eppendorf

microcentrifuge followed by transfer of the supernatant to a new tube. After determining the concentration, the RNA was ethanol precipitated again and resuspended to the desired concentration in RNase free distilled water. RNA samples in distilled water were stored at - 80°C.

F. Preparation of Purified Micronuclear and Macronuclear DNA from *T. thermophila*

Micronuclei and macronuclei were prepared from two liters of *T. thermophila* in 2% PPYS at a density of approximately 2.5 X 10^5 cells/ml. The cells were pelleted in 300 ml bottles centrifuged at 2,800 rpm for 4 minutes in a Sorval GSA rotor. The top 3/4 of the supernatant was removed by aspiration untill the total volume would fit into 4, 50 ml plastic centrifuge tubes. The 50 ml tubes were centrifuged at 2,800 rpm for 4 minutes in a Sorvall HB-4 rotor. The top 3/4 of the supernatant was aspitated off and the cells resuspended in the remaining liquid. The centrifuge tubes were washed with 100 ml of cell wash solution (290 mM sucrose, 10 mM Tris-HCl [pH 7.5], 1.5 mM $MgCl_2$). The cells in wash solution were again pelleted by spinning at 2,800 rpm for 4 minutes in a Sorvall HB-4 rotor and the supernatant removed by aspiration. These washed cells were resuspended in 180 ml of blendor solution (4% gum arabic, 88 mM sucrose, 10 mM Tris-HCl [pH 7.5], 1.5 mM $MgCl_2$, 5 mM spermidine, 1 mM spermine) and pooled in a waring blendor cup. Before blending 1.8 ml n-octanol was added to the blendor cup. The cells were disrupted by blending at low speed for 90 seconds in 30 second bursts interspensed with 30 seconds for cooling in an ice water bath. EDTA was added to the disrupted cell solution in the blendor cup to a final concentration of 8 mM. One more 30 second burst of blending was done before differential centrifugation of the lysed cells to seperate the micronuclei and macronuclei.

The macronuclei were pelleted by centrifugation of the lysed cells in 50 ml centrifuge tubes for 5.5 minutes at 3,000 rmp in a Sorvall HB-4 rotor. The supernatant was poured back

into the blendor cup and the pelleted macronuclei were resuspended in a total of 25 ml gradient solution (5 mM spermidine, 5 mM EDTA) plus 4% sucrose in a corex centrifuge tube and saved on ice. Two more rounds of blending and centrigfugation were done, the second of these centrifugation steps was at 3,500 rpm for 5.5 minutes. The pellets from both of these centrifugation steps were discarded.

The micronuclei were pelleted after one final 30 second pulse of blending. The disrupted cells were centrifuged in 50 ml centrifuge tubes for 10 minutes at 10,000 rpm in a Sorvall HB-4 rotor. The micronuclear pellets were resuspended in a total of 25 ml of gradient solution with 4% sucrose and transfer to a 30 ml Corex centrifuge tube. This suspension was balanced against the macronuclei from the previous step and both tubes were centrifuged for 10 minutes at 5,500 rpm in a Sorvall HB-4 rotor. The macronuclear pellet was resuspended in 2 ml of final buffer (60 mM KCl, 5 mM NaCl, 0.5 mM spermidine, 0.15 mM spermine, 15 mM Tris-HCl [pH 7.5]) in a 14 ml Falcon tube. The micronuclei were resuspended in 0.5 ml of final buffer in a 14 ml falcon tube. At this point the purity of the preperation was checked by counting a sample of the micronuclei and macronuclei in a hemocytometer after methyl green staining. A field of more than 500 micronuclei was examined for macronuclear contamination and a field of more than 500 macronuclei was axamined for micronuclear contamination. A preperation was deemed pure if no contaminating nuclei were seen in the field.

The micronuclear and macronuclear preperations were lysed by the addition of NDS (0.9 ml to the micronuclei and 3.6 ml to the macronuclei) and incubation at 50°C for 20 minutes. NDS consists of 0.5 M EDTA, 10 mM Tris-HCl (pH 7.5), and 1% SDS. The protein in the lysed nuclei samples was digested by adding HET-pronase (0.5 M EDTA, 10 mM Tris-HCl [pH 7.5], and 0.2 mg/ml pronase), 0.5 ml to the micronuclei and 2 ml to the macronuclei, and incubating at 50°C for 1 hour. Digestion was stopped by adding 5 ml of

0.1 X SSC (1X SSC = 150 mM NaCl, 15 mM Na-citrate, pH 7.0) to the micronuclei and 20 ml of 0.1X SSC to the macronuclei.

DNA was purified from these preperations by extraction with equal volumes of phenol:chloroform (1:1) until there was no material at the interface followed by three ether extractions. The DNA was precipitated by adding 2 volumes of 95% ethanol and chilling at -20°C for at least 2 hours. The DNA pellet was collected by centrifugation at 10,000 rpm for 10 minutes in a Sorvall HB-4 rotor. This DNA pellet was washed with 70% ethanol, dried briefly under vacuum, and resuspended in TE.

G. Preperation of Whole-Cell DNA from *T. thermophila*

T. thermophila cells were grown in 2% PPYS at 30°C with gently agitation to a density of ~ $2.5X10^5$ cells/ml. Cells were pelleted from 15-20 ml of culture by centrifugation for 5 minutes at 1,500 rpm in an IEC benchtop centrifuge. The volume of the cell pellet was estimated and 2 volumes of NDS added. The suspension was transferred to a 1.5 ml Eppendorf tube and 1/3 volume of the original cell pellet of 0.2 mg/ml pronase was added and mixed well. This was incubated for 4 hours or more at 55°C. Incubation was stopped by the addition of 3/4 volume of 0.1 X SSC. The digested sample was extracted two or three times with phenol:chloroform (1:1). DNA was precipitated by adding 2 volumes of 95% ethanol and chilling for 2 hours or more at -20°C. the DNA pellet was collected by centrifugation for 15 minutes in an Eppendorf microcentrifuge. The DNA pellet was washed with 70% ethanol, and resuspended in 100-500 µl TE by gently rocking overnight at room temperature. RNA in the sample was digested by the addition of 1/10 volume 1 mg/ml RNase A and incubation at 37°C for 2 hours. The RNase was removed by twice extracting with an equal volume of phenol:chloroform (1:1). The DNA was again precipitated by adding 1/10 volume of 3 M NaOAc and 2.5 volumes of 95% ethanol and chilling at -20°C for 2 hours or more. After centrifugation

for 15 minutes in an Eppendorf microcentrifuge the pellet was washed with 70% ethanol, briefly dried under vacuum and resuspended in TE.

Enzymes

Restriction enzyems, T4 DNA ligase, and Klenow DNA polymerase were obtained from New England Biolabs. T4 DNA polymerase was obtained from IBI. Calf intestinal phosphatase was obtained from Pharmacia. Terminal deoxynucleotidly transferase was obtained from Pharmacia. Taq DNA polymerase was obtained from Perkins Elmer/Cetus. All enzymes were used according to the suppliers instructions.

Agarose Gel Electrophoresis of Nucleic Acids

Agarose gel electrophoresis was carried out in standard horizontal gel boxes. Agarose was obtained from Sigma or Sea Kem. Sea Kem GTG grade agarose was used for preperative gels. DNA in restriction digestion reactions was loaded directly onto agarose gels after the addition of 1/5 volumes loading dye (50% [v/v] glycerol, 100 mM EDTA, 0.05% [w/v] Bromophenol Blue). DNA was typically seperated on gels of 0.7-1.2% agarose run in TAE buffer (40 mM Tris, 20 mM Acetic Acid, 2 mM EDTA, pH 8.0). Small DNA fragments such as those produced by polymerase chain reaction (PCR) amplification procedures were separated on gels of 2-4% New-Sieve agarose (Sea Kem) run in TAE buffer.

RNA was denatured before loading onto gels by mixing 5 µl of RNA in water with 10 µl dimethylsulfoxide, 5 µl 40% glyoxal, 1 µl 0.2 M Na_2HPO_4/NaH_2PO_4 [pH 7.4], and

heating to 50°C for 1 hour. The sample was chilled on ice then allowed to warm to room temperature before adding loading dye and loading onto the gel. RNA was seperated on 1.0-1.4 % agarose gels run in 10 mM Na_2HPO_4/NaH_2PO_4 (pH 7.0), the buffer was recurculated during electrophoresis.

Blotting Nucleic Acids to Membranes

A. Southern Blots Made by Capilary Transfer of DNA to Nitrocellulose

Following electrophoresis the DNA in the gel was depurinated by soaking the gel in 0.25 M HCl for 20 minutes followed by soaking in water for 10 minutes. The DNA in the gel was then denatured by soaking the gel in 0.5 M NaOH, 1.5 M NaCl for 20 minutes followed by soaking in water for 10 minutes. The gel was neutralized by soaking in 0.5 M Tris-HCl (pH 7.0), 3 M NaCl for 15 minutes, changing to fresh solution and soaking for an additional 15 minutes. The gel was placed on top of blot block paper (Schleicher & Schuell), 3 or 4 sheets of GB004 and one sheet GB002 next to the gel, cut to the same size as the gel and soaked in 20X SSC. The nitrocellulose filter (Schleicher & Schuell) was placed on top of the gel followed by 1 sheet of GB002 blot block paper soaked in 20X SSC, and a stack of 2-3 inches of GB004 blot block paper. The stack was weighted with a thick glass plate and the blot allowed to transfer overnight. After dismanteling the blot the filter was washed in 20X SSC, dried and baked at 80°C under vacuum for 2 hours.

B. Southern Blots Made by Capilary transfer of DNA to Nytran

Following electrophoresis the DNA in the gel was depurinated and denatured as described above for transfer to Nitrocellulose except that the gel was not soaked in water after denaturation. DNA was transferred to Nytran (Schleicher & Schuell) in the denaturation solution (0.5 M NaOH, 1.5 M NaCl). The blots were set up as described

above except that the blot block sheets that were soaked in 20X SSC were soaked in denaturation solution instead. The transfer was allowed to proceed overnight. After diamanteling the blot the filter was neutralized by soaking for 10-15 minutes in 2X SSC. DNA was fixed to Nytran filters by UV light crosslinking. The damp filter was placed on a Fotodyne UV300 UV transiluminator with the DNA side down and the UV light was turned on for 10 seconds. Alternatively the damp filter was placed DNA side up in a UV Stratalinker 1800 (Stratagene) and cross linked using the auto cross link setting provided by the manufactured (1200 μJ).

C. Southern Blots Made by Vacuum Transfer of DNA to Nytran

DNA was also transfered to Nytran (Schleicher & Schuell) membranes using a VacuGene apparatus and VacuGene model 2016 vacuum pump (Pharmacia/LKB). The Nytran membrane was placed inside a screen cut to be slightly smaller than the gel. The gel was placed on top of the filter overlaping the screen at the edges. A vacuum of 40 cm H_2O was applied. Depurination solution (0.25 M HCl) was added to cover the gel and removed after 4 minutes. Denaturation solution (0.5 M NaOH, 1.5 M NaCl) was added to twice the depth of the gel and transfer from this solution continued for 1 hour. The denaturation solution was removed and the gel discarded. The Nytran filter was neutralized by soaking in 2X SSC for 10-15 minutes. After briefly drying the damp filter was treated as described above to UV cross link the DNA.

D. Northern Blots Made by Vacuum Transfer of RNA to Nytran

The filter and gel were arranged in the VacuGene apparatus as described above for DNA transfer. A vacuum of 40 cm H_2O was applied and 7.5 mM NaOH was added to twice the gel depth. Transfer from this solution was allowed to continue for 1 hour or more. After dismanteling the blot the filter was neutralized by soaking in 2X SSC for 10-15 minutes. The filter was not further processed before prehybridization.

E. Colony Lifts for Screening of Recombinant Plasmids in Bacteria

Nitrocellulose filters circles (BA45 from Schleicher & Schuell) were placed on an fresh agar plate until they were completely wet. The wet filter was then placed on a plate with colonies, marked unambiguously with India ink and allowed to sit for 5 minutes. The filter was then removed from the plate and placed, colonies side up, on a piece of Watmann 3MM paper soaked with 0.5 M NaOH for 10 minutes. The filter was then placed on a piece of Watmann 3MM paper soaked in 1 M Tris-HCl (pH 7.4) for 10 minutes. And finally placed on a piece of Watmann 3 MM paper soaked in 1.5 M NaCl, 0.5 M Tris-HCl (pH 7.4) for 10 minutes. All filters were air dried and baked at 80°C under vacuum for 2 hours.

F. Plaque Lifts for Screening of Recombinant Lambda Phage

Plates with recombinant plaques to be screened were chilled at 4°C for at least 1 hour before plaque lifts were done. Nytran filters (Schleicher & Schuell) that had been cut to circles of the appropriate size were placed on a fresh agar plate and allowed to become completely wet. They were then placed on plates with plaques. The filters were unambiguously marked with India ink and allowed to sit on the plates with plaques for 1 minute. The filters were removed from the plates and soaked in 0.5 M NaOH for 30 seconds then transfered to 1.5 M NaCl, 0.5 M Tris-HCl (pH 7.4) for 30 seconds, then allowed to air dry. This procedure could be repeated for up to 7 filters per plate as long as subsequent filters were allowed to sit on the plates for 3 minutes. DNA was fixed to the filters by one of two methods. DNA was fixed to the filters by either autoclaving for 2 minutes on a fast exhaust cycle or UV cross linking using the Stratalinker 1800 as described for Southern blots of DNA.

Recovery of DNA from Agarose Gels

Specific fragments or size selected regions of DNA were recovered from agarose gels as follows. Fragments to be used as probes of blots, colony lift filters, or plaque lift filters were first cut out of Sea Kem GTG agarose gels. This gel slice was then imbeded in a Sea Kem low melt agarose gel and again electrophoresed. A gel slice containing the fragment of interest was cut out of the low melt gel and stored at -20°C. DNA fragments to be used in cloning procedures were purified from Sea Kem GTG agarose after electrophoresis. The gel slice was disrupted by forcing it through a 20 gague needle or smashing it in an Eppendorf tube. Phenol (equilibrated with 0.1 M Tris-HCl [pH 7.4]) was added to the smashed gel slice and vortexed for 2 minutes. This was placed at -80°C untill frozen then allowed to thaw at room temperature. After centrifugation at 4°C for 15 minutes in an Eppendorf microfuge, the aqueous phase was removed. The aqueous phase was extracted with phenol:chloroform (1:1) twice. The DNA was precipitated from the solution by adding 1 µl 10 mg/ml carrier *E.coli* tRNA, 1/10 volume 3M NaOAc, 2.5 volumes 95% ethanol and chilling at -20°C for at least 2 hours. After collecting the DNA by centrifugation in an Eppendorf microcentrifuge for 10 minutes the pellet was resuspended in TE or an appropriate reaction buffer.

Labeling DNA Fragments with ^{32}P

DNA fragments to be used as probes were labeled with $\alpha^{32}P$ labeled nucleotide triphosphates (400 Ci/mMol, Amersham) by one of two procedures. Nick translations of fragments using 2 labeled nucleotides were preformed as described in Maniatis et al. (1982). Better results were obtained when fragments were labeled by the random primer method (Feinberg and Vogelstein, 1983) Equivalent probes were produced with either

the Random Primed DNA Labeling Kit from Boehringer Mannheim or the Multi-Prime kit from Ammersham. Typically a reaction included 10 µl of DNA in low melt agarose, two labeled nucleotide triphosphates and two unlabeled nucleotide triphosphates in a total reaction volume of 50 µl (buffer, primers, enzyme and unlabeled nucleotides from the kit used as per the suppliers instructions). The DNA was denatured by boiling the low melt gel slice for 2 minutes before adding the DNA in low melt agarose to the reaction. The reaction was allowed to proceed at room temperature overnight. The reaction mixture was diluted with 100 µl distilled water and 5 µl 10 mg/ml carrier *E.coli* tRNA was added before disolving the agarose by heating to 72°C. The reaction mix was then extracted twice with phenol:chloroform (1:1). The labeled DNA was precipitated from the aqueous phase by adding 200 µl 5M NH_4OAc and 1 ml 95% ethanol, incubating at room temperature for 10 minutes and centrifuging for 25 minutes in an Eppendorf microfuge. The probe was briefly dried before being resuspended in 50-200 µl of distilled water. Probes in distilled water were boiled for 2-3 minutes before being diluited into hybridization solution and added to filters to be probed.

Hybridization of ^{32}P labeled Probes to Nucleic Acids on Membranes

All blots, colony and plaque lifts were hybridized to probes and washed using the same conditions. Filters were pre-pre hybridized in 10X Denhardts solution (1X Denhardts = 0.1% [w/v] ficol, 0.1% [w/v] polyvinlypyrolidine [MW 400,000, Sigma], 0.1% [w/v] Bovine Serum Albumin), 4X SSC and 0.1% SDS at 65°C for 1-4 hours. The filters were then prehybridized in 1X Denhardts, 4X SSC for 4 hours or more at 65°C. Denatured ^{32}P labeled probe was added and hybridized to the filters for at least 8 hours at 65°C. Following hybridization the filters were washed first in two changes of 4X SSC, 0.1% SDS for 15 minutes at room temperature, then 30 minutes to 1 hour in 4X SSC, 0.5%

SDS at 65°C and finally for 20 min to 1 hour in 0.1X SSC, 0.1% SDS at 42°C. Blots were air dried and exposed to Kodak XAR-5 film. Most exposures of genomic blots were overnight to 6 days using intensifying screens at -80°C. Plaque lifts and blots of cloned DNA were usually exposed for 1-6 hours using intensifying screens at -80°C.

Subcloning

Insert and vector DNAs to be subcloned were processed with restriction enzymes, T4 DNA polymerase, calf intestinal phosphotase and Klenow using conditions recommended by the suppliers. Ligations were carried out using conditions recommended by the supplier at 14-16°C overnight. Recombinant plasmids in pBR322, pUC18 and pUC19 vectors were maintained in *E. coli* strain DH5α. Recombinant plasmids in pUC118 and pUC119 vectors were maintained in *E. coli* strain MV1193 (J. Miller, Cold Spring Harbor Laboratory).

Bacterial cells to be used for transformations were grown in LB at 37°C with shaking to an OD_{600} of 0.2-0.4. Cells from 40 ml of culture were collected by centrifugation at 5,000 rpm for 5 minutes in a Sorvall HB-4 rotor. The cells were resuspended in 20 ml 0.05 M $CaCl_2$ and incubated on ice for 2 hours or more. The cells were then pelleted by centrifugation at 5,000 rpm for 5 minutes in a Sorvall HB-4 rotor resuspended in 2 ml 0.05M $CaCl_2$ and incubated on ice for 15 minutes. 10 µl of DNA from a ligation mix was added to 200 µl of the cells and incubated on ice for 25 minutes. The cells with DNA were heat-shocked at 37°C for 2 minutes and returned to room temperature for 10 minutes or more before plating. Transformed cells, 20-200 µl, were spread on LB agar plates containing antibiotics. For blue/white selection of recombinant plasmids that have a disruptable beta-galactosidase gene the cells were spread on LB agar plates covered with

3 ml of top agar supplemented with 10 µl 100 mM IPTG, 50 µl X-gal (40 mg/ml in dimethylformamide) and antibiotics.

Sequencing

All sequencing was done from single stranded-templates prepared as described above from recombinant plasmids in either pUC118 or pUC119 vectors. Sequencing reactions were done using a sequenase^{TM} kit and the sequenase^{TM} version of modified T7 DNA polymerase (United States Biochemical Corporation) following the instructions of the supplier. Sequencing primers were obtained from New England Biolabs or synthesized on an Applied Biosience DNA synthesizer. Sequencing products were labeled with ^{35}S dATP and seperated on 6% acrylamide 8M urea gels run in 0.6X TBE buffer (1X TBE = 0.1 M Tris, 0.1M Boric Acid, 1.2 mM EDTA, pH 8.3). The gels were dried before exposure to Kodak XAR-5 film. The accumulated sequence data was stored and analyzed using the Bionet programs of Inteligenetics (Mountain View, California) run on a SUN computer.

Tetrahymena thermophila Strains

Tetrahymena thermophila strains SB2040, SB2044, SB2045 and SB2046 were the gift of Dr. S.-G. Zhang and Dr. E. Orias (University of California, Santa Barbara). They were genetically constructed so as to be identical whole-genome homozygotes. ME44C was the gift of Dr. E. Simon (University of Illinois). This strain was isolated from the wild in 1985. Other strains were from among our laboratory collection. The genotypes and phenotypes of all *T. thermophila* strains used is presented in Table 3.1. *T. thermophila* strains were mainteained in continuous vegetative growth in 5 ml of 1% PPYS (1%

proteose peptone, 1% yeast extract, 0.0015% Sequestrine) in 10 ml screw cap tubes at room temperature. Stock cultures were transferred bimonthly.

Mating of *T. thermophila*

For mating experiments 50 ml cultures of *T. thermophila* were grown in 2% PPYS at 30°C with gently swirling to a density of 2.5 X 10^5 cells/ml. The cells were harvested by centrifuging at 1,500 rpm for 5 minutes in an IEC benchtop centrifuge. Most of the supernatant was removed by aspiration and the cell pellet was washed with 100 ml Dryl's solution (1.7 mM Na-citrate, 1.2 mM NaH_2PO_4, 1 mM Na_2HPO_4, 2 µM $CaCl_2$). The cells were again pelleted and washed with Dryl's solution. After two washings the cells were counted and resuspended to 2.5 X 10^5 cells/ml in Dryl's solution. The cells were starved by incubating in Dryl's solution at 30°C with gentle swirling for 20 hours. Starved cells were counted and an equal number of cells of each mating type were mixed in a total volume of 40 ml of Dryl's solution in a petri dish. The cells were allowed to pair by incubating the the mixure of mating types for 3 hours at 30°C without shaking. After three hours individual pairs were picked and placed into 30 µl drops of 1% PPYS suplimented with 250 µg/ml Penecillin and Streptomycin. After 18 hours at room temperature the four individual karyonide cells from each pair were removed to seperate wells of a microtiter plate. Each well contained 100 µl 1% PPYS suplimented with 250 µg/ml Penecillin and Streptomycin. The microtiter plates were incubated for three days at 30°C in a moist chamber.

After three days of growth the phenotype of the karyonidal cell cultures was tested. In the mating done here the karyonidal descendents were tested for the expression of genes for resistence to cycloheximide and 2-deoxygalactose. To do this the cultures in the

microtiter plates were replica plated into three new sets of microtiter plates, one containing 1% PPYS control media, one containing 1% PPYS with 15 μg/ml cycloheximide and one containing 1% PPYS with 2.5 mg/ml 2-deoxygalactose. After incubation for three days at 30°C in a moist chamber each well was scored for growth of cells.

Pre-recycling Solid Waste Volume Reduction

Previous to the return for reuse program, the volume of styrofaom boxes recieved with enzyme orders was reduced in one of two precedures. As a combined solid waste volume reduction and stress reduction program styrofaom boxes were used as targets for martial arts practice or simply stomped upon. More esthetically pleasing results were achieved by presure and temperature induced polystyrene de-expansion protocols. The polystyrene object in question was sometimes decorated with magic markers or sharpies prior to de-expansion. Polystyrene boxes were placed open end down in an autoclave along with their lids if desired. The autoclave was set to a cycle of 20 minutes with fast exhaust and started. When the autoclave cycle was completed the box was turned open side up and returned to the autoclave for an additional 20 minute cycle with fast exhaust. The box lid was also returned to the autoclave for the second cycle if its de-expansion after the first cycle was not satisfactory. Following this procedure the size of the boxes, as well as their contribution to the volume of solid waste from the laboratory, was greatly reduced. They were no longer suitable for refrigerated transport but could be used as part of a desk organization system, as radiation shields or presented to friends and family as gifts.

Chapter 3

Tel-1 Elements are Associated with Micronuclear Genome Rearrangements in some *Tetrahymena thermophila* Strains

Introduction

The *Tetrahymena thermophila* micronuclear genome contains middle repetitive DNA elements designated Tel-1. The structure of these elements was deduced previously from studies of blocks of non-telomeric micronuclear repeated C_4A_2 sequence, mic C_4A_2, (Cherry and Blackburn, 1985). All cloned examples of mic C_4A_2 were associated with a 30 bp conserved sequence. One clone, p17D59, contained two blocks of mic C_4A_2 and their adjacent 30 bp conserved sequences in inverted orientation 4.3 kb apart (Figure 1). Other middle repetitive DNA elements with inverted repeats at their ends, for example P of *Drosophila melanogaster* and Tc1 of *Caenorhabditis elegans*, are known to be mobile genetic elements or transposons. This chapter presents molecular evidence that Tel-1 elements are mobile. The molecular evidence for Tel-1 transposition is most striking in Tetrahymena strains that have undergone unusual genomic events. McClintock has suggested that transposons are activated in genomes that have undergone genomic "shock" (McClintock, 1984). She defines genomic shock as conditions that "force the genome to restructure itself in order to overcome a threat to it's survival" (McClintock, 1984). The genetic history of the *T. thermophila* strains exhibiting evidence of Tel-1 transposition is consistent with periods of such genome restructuring.

The genetic organization of *T. thermophila* allows transposition events in the micronuclear genome to occur without immediate phenotypic consequences. Genetic

analysis in other organisms has shown that movement of transposable elements is often detected by their causing frequent, unstable mutations (numerous examples reviewed in (Berg and Howe, 1989). However, any such mutation that is lethal will not be recoverable for analysis. In *T. thermophila* changes in the micronucleus are phenotypically silent as long as the cells are maintained in vegetative growth and do not undergo mating. *T. thermophila* have two functionally distinct nuclei. The germline micronuclei contains five diploid pairs of chromosomes and divides mitotically. There is only one reported case of a gene being transcribed from the micronuclear genome (Stein-Gavens et al., 1987) and all other data suggest it is transcriptionally silent (Gorovsky, 1973; Gorovsky, 1980). The somatic macronucleus is transcriptionally active and responsible for the phenotype of the cell. The macronucleus is derived from a division product of the zygotic micronucleus after mating (Figure 2A outlines the nuclear events in a normal mating). Following macronuclear differentiation, a process involving DNA elimination, chromosome fragmentation and amplification, the resulting macronucleus contains linear, acentric subchromosomal DNAs in high copy number. Table 1 contains a list of all strains and cell lines used in this chapter, their micronuclear genotype and macronuclear phenotype.

Within a species, different strains or isolates typically exhibit distinct genomic distributions of transposons because of the ability of transposons to move within their resident genomes. The *T. thermophila* micronuclear genome was examined for molecular evidence of Tel-1 transposition. The arrangement of Tel-1 elements in the micronuclear genomes of several *T. thermophila* strains and cell lines was assayed by Southern blotting. Rearrangements due to transposition will result in restriction fragment length polymorphisms linked to the element and detectable by an element probe. By this analysis, most inbred laboratory strains tested had indistinguishable

distributions of Tel-1 elements. However, differences in the micronuclear arrangement of Tel-1 elements were seen between a newly isolated *T. thermophila* strain and common laboratory strains, and also between laboratory strains that had been genetically manipulated to produce specific micronuclear genome anomalies. Most striking was the change over time in the distribution of Tel-1 elements in one stock of a cell line that had originally been constructed to be a whole-genome homozygote. This stock of cells is also shown to be genetically impaired, as it failed to produce cross fertilized progeny in a test mating. The observed rearrangements of Tel-1 sequences are consistent with the hypothesis that these elements are mobile.

Results

Complete Tel-1 elements are limited to the micronucleus

The Tel-1 end 30 bp conserved sequence contains the recognition site for *Bst*XI (see Figure 1), a restriction enzyme that cuts the *T. thermophila* genome infrequently. Tel-1 elements are operationally defined here as conserved DNA elements, bounded by *Bst*XI sites, that are similar in sequence to p17D59 as judged by cross hybridization at high stringency. The 4.3 kb *Bst*XI fragment from p17D59 was used as a probe and will be called the Tel-1 probe in all subsequent experiments. In experiments where *Bst*XI was used to digest *T. thermophila* DNA the possibility of incomplete digestion by the enzyme was ruled out by mixing lambda DNA with the genomic DNA and observing its complete digestion as judged by the banding pattern on an ethidium bromide stained agarose gel. In all laboratory strains examined, digestion of micronuclear DNA with *Bst*XI released a prominent 9.7 kb fragment homologous to the Tel-1 probe (Figure 3 lane 1). Thus many genomic copies of Tel-1 elements appear to be 9.7 kb long. A 4.3 kb

*Bst*XI fragment homologous to the Tel-1 probe was not detected in this experiment or similar experiments, even after long autoradiographic exposures. The p17D59 clone was obtained from a *T. thermophila* strain which during long term vegetative propagation had changed, so that it was no longer possible to purify micronuclei from it at the time these studies were initiated. The 4.3 kb *Bst*XI fragment from p17D59 is a rare or aberrant form, possibly particular to the strain from which it was cloned. Blots displaying macronuclear DNA digested with *Bst*XI and hybridized with the Tel-1 probe showed hybridization to limit mobility material only (Figure 3 lane 2). This implies that although sequences between the inverted repeats in p17D59 are retained in the macronucleus there do not appear to be *BstXI* sites separated by conserved distances, as would be expected if whole elements were present in the macronuclear genome.

Distribution of Tel-1 elements is similar among standard laboratory strains but differs in a strain newly isolated from the wild

Available laboratory strains of *T. thermophila* were screened for the distribution of Tel-1 elements in their micronuclei. Purified micronuclear DNA was digested with restriction enzymes, displayed on Southern blots and probed with the Tel-1 probe. Figure 4 shows such a blot of *Hin*dIII-digested DNA from three representative laboratory strains. The pattern of bands is identical among them. Similar results were obtained with four other restriction enzymes (data not shown); however *Hin*dIII digestion produced the broadest and best resolved size distribution of fragments. Other laboratory strains tested showed distributions of Tel-1 elements indistinguishable from those shown here (data not shown). Although this analysis was not exhaustive, no differences in the distribution of Tel-1 elements in standard laboratory strains of *T. thermophila* were observed.

In contrast, a newly isolated *T. thermophila* strain, ME44C, showed a distribution of Tel-1 elements in its micronucleus that was distinct from the laboratory strains. Strain ME44C was isolated from the wild in 1985 (generous gift of Dr. E. Simons, University of Illinois). The micronuclear DNA of this strain is compared to one of the standard laboratory strains as exemplified by SB2046 in Figure 5A. There are several obvious differences between these strains. There are more Tel-1-homologous *Bst*XI fragments in the ME44C micronuclear DNA (lane 3) and they are larger than those in the micronuclear DNA of the laboratory strain SB2046 (lane 4). Comparison of *Hin*dIII digested DNA from ME44C (lane 5) and SB2046 (lane 6) provides further evidence that these strains differ noticeably in the distribution of Tel-1 sequences in their micronuclear DNA.

The differences in the organization of Tel-1 sequences in the micronuclear genomes of these strains could have resulted from their genomes being drastically rearranged relative to each other, or from rearrangements specifically involving of Tel-1 sequences, against a background of a relatively stable genome. To determine whether the genome of ME44C is in general grossly different from that of SB2046, the probe was stripped from this blot and it was reprobed with a portion of the cloned 5S ribosomal RNA gene. 5S ribosomal RNA genes are dispersed and repetitive in the *T. thermophila* micronucleus and occur on four of the five chromosomes (Allen et al., 1984). Thus 5S sequences are a good probe for sampling diverse genomic locations. It can be seen that all of the 5S-homologous *Hin*dIII fragments in the SB2046 micronuclear DNA (Figure 5B lane 2) have exact size counterparts in the ME44C micronuclear DNA (Figure 5B lane 1). There are three or four additional 5S-homologous *Hin*dIII fragments in the ME44C micronuclear DNA which may have arisen by gene duplication or some other mechanism. Nevertheless the pattern of 5S genes in these two strains is overall very similar. In contrast it is difficult to find more than three or four out of more than

twenty Tel-1-homologous *Hin*dIII fragments that are identical in size between these two strains (compare lanes 5 and 6 Figure 5A).

Rearrangement of Tel-1 sequences in strains that have experienced genomic "shock"

It has been shown that transposable elements in maize are activated in genomes that have undergone what McClintock refers to as genomic "shock" (reviewed in Döring and Starlinger, 1986). In the manipulations used to generate two types of *T. thermophila* strains in the laboratory, nullisomics and whole-genome homozygotes, their genomes have been forced to respond to conditions that could be perceived as genomic "shock", in at least one sense. The conditions imposed in the construction of these types of strains are not ones for which the *T. thermophila* genome has a controlled response and may have induced abnormal genome reorganization events, including transposition.

Nullisomic strains are missing part or all of one or more of the micronuclear chromosomes. The critical step in the construction of such strains involves forcing a cell with a haploid micronucleus to undergo meiosis (Bruns and Brussard, 1981; Bruns et al., 1983). This results, in meiosis I, in some of the five chromosomes segregating to one spindle pole and some to the other, so that neither resulting meiotic product has a full complement of chromosomes. The chromosomes can also break and in some cases are maintained as incomplete chromosomes. A subsequent specialized cross induces homozygosis via diploidization of the meiotic pronuclei to produce an aneuploid micronucleus missing whole chromosomes or chromosome ends in a cell that retains its parental macronucleus (see diagram in Figure 2C).

The micronuclear arrangement of Tel-1 elements in three nullisomic stains was compared to a standard laboratory strain by Southern blotting. In the micronuclear

genomes of the three nullisomic strains, new as well as missing Tel-1 homologous sequences were evident, as shown in the Southern blot of their micronuclear DNA (Figure 6, lanes 3,4 and 5). As expected, some Tel-1-homologous *Hin*dIII fragments were missing compared to a standard laboratory strain, B1868 (Figure 6, lanes 1 and 2). Therefore Tel-1 sequences, or sequences closely linked to them, are also rearranged relative to B1868 as indicated by the arrows to the right of the blot in Figure 6.

A set of four identical whole-genome homozygotes provides a better-defined collection of abnormally manipulated micronuclei in which to assay Tel-1 transposition. The critical step in the crosses undertaken to produce identical whole-genome homozygotes is the diploidization of a single haploid gametic nucleus (diagramed in Figure 2B) (Allen, 1967a; Bruns, 1986). Identical whole-genome homozygotes are expected to show no micronuclear restriction fragment length polymorphisms. As shown in Figure 7 the four strains had nearly identical micronuclei with respect to the Tel-1 probe. Careful inspection revealed one unexpected difference, as indicated by the arrow to the right of the blot, pointing out the position of an additional Tel-1-homologous *Hin*dIII fragment in the SB2040 micronuclear DNA (lane 3). An additional Tel-1-homologous *Hin*dIII fragment was seen in micronuclear DNA from strain SB2040 in several experiments. *Eco*RI digested DNA from the four whole-genome homozygotes also showed an extra Tel-1-homologous fragment in the SB2040 DNA sample (data not shown).

All four of these cell lines were maintained in continuous vegetative growth in duplicate stock tubes. One pair of duplicates were designated SB2040.1 and SB2040.2. Micronuclear DNA prepared six months after that shown in Figure 7 indicates that the population of cells in stock SB2040.1 had undergone drastic micronuclear genome rearrangements with respect to Tel-1 sequences (Figure 8A). Specifically, some Tel-1-homologous *Hin*dIII fragments were missing, and there were also new Tel-1

homologous fragments, indicated by arrows beside the blot in Figure 8A. Although the SB2040 cell line was a clonal line from a single founder micronucleus these stocks were started from a minimum population of ~10^3 cells. The restriction fragment length polymorphisms seen on the Southern blots could therefore have resulted from either a different distribution of several cell types in the populations of each stock, or from one or a few cells whose progeny came to predominate in one stock after six months. Isolation of 24 single cells from each of the SB2040.1 and SB2040.2 stocks showed that within each stock the cells were identical, as assayed by Southern blotting of whole cell DNA samples using a probe that hybridized only to micronuclear sequences and recognized genome rearrangements between the SB2040 stocks (data not shown, a description of the PCR 900 probe used can be found in chapter 4). This implies that the SB2040.1 stock of cells with rearranged micronuclei relative to SB2040.2 resulted from a single cell or small number of cells, and that the progeny of these cells with such rearrangements took over the population.

A similar Southern blot probed with the 5S ribosomal RNA gene (Figure 8B) shows that although SB2040.1 had lost some 5S-homologous fragments, none were rearranged relative to SB2040.2. Therefore, these restriction fragment length polymorphisms are linked to Tel-1 sequences and not another middle repetitive sequence, the 5S gene. As noted above, SB2040 was genetically constructed so as to be a whole-genome homozygote. Hence the restriction fragment length polymorphisms observed could not have arisen by gene conversion or mitotic cross-over events during vegetative division of the micronuclei.

Genetic consequences of Tel-1-associated micronuclear rearrangements

Probing with both the Tel-1 and 5S rRNA probes shows the SB2040.1 micronuclear genome to be greatly simplified compared with that of SB2040.2 stock cells. The blots in Figure 8 were prepared from gels in which equal amounts of DNA were loaded in each lane. Reduced sequence complexity in the micronuclear DNA of a sample resulted in fewer bands for a given probe and those remaining represented a larger proportion of the total DNA and appeared darker on the blot. Similar micronuclear genome simplification has been observed in *T. thermophila* "star" strains (Yao, 1982). "Star" strains are genetically defined by the loss of their micronuclei in meiosis and consequent sterility; they are unable to contribute micronuclear genetic markers to their progeny (Allen, 1967b; Orias, 1986)(see diagram Figure 2B, first round cross). In order to determine whether the observed genomic rearrangements in the SB2040.1 stock had equivalent genetic consequences test matings were performed. Both SB2040.1 and SB2040.2 were crossed with another strain, SB210 (see Table 1 for strain genotypes and phenotypes). SB2040.1 and .2 are phenotypically resistant to cycloheximide and were genetically constructed to be homozygous for resistance in their micronuclei. SB210 is a heterokaryon; it is phenotypically sensitive to 2-deoxygalactose but homozygous for resistance in its micronucleus. A normal mating between strains with these genetic markers will result in inheritance of 2-deoxygalactose resistance from the SB210 parent and resistance to cycloheximide from the SB2040 parent; i.e.; progeny that are simultaneously resistant to both drugs.

In separate crosses, SB210 cells were mixed with either SB2040.1 or SB2040.2 cells. Single mating pairs were isolated into media drops. After division the four karyonides were separated into individual microtiter wells, grown for three days, and tested for drug sensitivity. As can be seen in Table 2, the SB2040.2 cells, which did not have any

Tel-1-associated micronuclear rearrangements, produced almost exclusively cross fertilized progeny. In contrast, no cross fertilized progeny were recovered from the SB2040.1 cross. Furthermore, the micronuclear rearrangements in SB2040.1 did not simply result in the loss of the cycloheximide resistance marker. If this had been the case, the progeny would have become resistant to 2-deoxygalactose as a result of forming a new macronucleus containing this resistance gene from the SB210 parent. However, none of the exconjugants from the SB2040.1 cross expressed the 2-deoxygalactose resistance marker. Therefor, it appears that SB2040.1 cells behaved in this cross like a "star" strain . The progeny of a mating involving a "star" strain retain their old macronuclei and thus the parental phenotype (see Figure 2B, first round cross, as an example)(Allen, 1967b; Orias, 1986). As judged by the cross shown in Table 2 the genomic rearrangements of SB2040.1 cells have rendered them sterile.

Discussion

The evidence presented in this chapter supports the hypothesis that Tel-1 elements are mobile in the *T. thermophila* micronuclear genome. The number and location of Tel-1 elements are polymorphic between a newly isolated strain, ME44C, and standard laboratory strains. Nullisomic strains have not only lost Tel-1-homologous sequences but some are rearranged. Most convincingly, one stock of one of four cell lines that were identical whole-genome homozygotes showed rearrangements of Tel-1 sequences over time.

Tel-1 associated changes seen in the SB2040.1 stock are the strongest evidence for transposition. Because these cells are whole-genome homozygotes, processes such as gene conversion or homologous recombination, that might have occurred during

vegetative growth, could not result in the observed Tel-1 homologous restriction fragment length polymorphisms. The mating experiment demonstrating the sterility of SB2040.1 is reminiscent of the sterility associated with dysgenic crosses of Drosophila where extensive P transposition has occurred (Engels, 1989).

The SB2040.1 stock has a simplified micronuclear genome as evident from the missing Tel-1 and 5S rRNA fragments. Loss of sequences can be initiated by transposition. Movement of some elements, such as Ac and Ds of Maize, is often accompanied by chromosome breakage, inversions and deletions (Fedoroff, 1989). Deletions clearly result in loss of DNA. Broken chromosomes are additionally unstable and can be lost or initiate recombination events. Although Tel-1 transposition could have caused the observed genome rearrangements the data presented here do not prove this. It is possible that Tel-1 sequences are not themselves mobile, but are closely linked to mobile elements in the genome and thus fortuitously detect the restriction fragment length polymorphisms observed here. If the Tel-1 rearrangements observed were due to chance the probability of observing no changes with either the 5S probe or another repetitive micronuclear sequence probe, the 22A fragment of a lambda clone containing *T. thermophila* micronuclear DNA described in the appendix of this thesis, (data not shown) can be estimated. Accounting for only the Tel-1, 5S, and 22A hybridizing bands still present in SB2040.1, there were 3 Tel-1 hybridizing bands that were different out of 13 and no 5S or 22A hybridizing bands that were different out of 14. Assuming that the 3 out of 13 are chance observations and using this probability of an event occuring in a binomial equation for other events, the chance of seeing no changes in 14 is <0.03. Therefore the changes in the genome are unlikely to be random and more likely to be linked to Tel-1.

While the conditions that activate Tel-1 transposition are not known, the nullisomic and whole-genome homozygote strains have genetic histories that, based on the notion of genomic shock, may be consistent with transposon activation. Other transposons, such as those of maize, are activated in a nucleus that has inherited broken chromosomes. Specifically, previously inactive Ac and Spm elements transposed in nuclei when a broken chromosome was introduced. Similar activation of maize transposons occurred following ionizing radiation, a process known to produce chromosome breaks (reviewed in Döring and Starlinger, 1986). In generating those nullisomic strains that are missing parts of chromosomes, chromosome breakage must also have occurred. In other studies middle repetitive sequences were found to be missing from nullisomic strains but none were rearranged (Howard and Blackburn, 1985; Yao, 1982). The appearance of new Tel-1-homologous fragments in the nullisomic strains is therefore consistent with the possibility that Tel-1 transposition occurred. However, some recombination events may have occurred along with chromosome loss in the nullisomic strains. Those missing chromosome arms are known to have fused the broken ends of different chromosomes together (Bruns, 1986). Therefore the cause of the Tel-1 rearrangements seen among the nullisomics can not be unequivocally attributed to transposition alone.

It is interesting that sterility and rearrangements occurred in only one of the four identical whole-genome homozygote karyonidal lines examined. The process of genomic exclusion that produces whole-genome homozygotes involves as a first step mating to a "star" strain. After this cross it has been noted that the clone derived from the "star" conjugant, that is having the cytoplasm of the "star" parent, can lose its fertility rapidly (Weindruch and Doerder, 1975). There has been no molecular analysis of this phenomenon. However it has been speculated that products produced by the retained macronucleus of the "star" parent result in DNA damage and possibly

transposition (Bruns, 1986). Cytoplasmic inheritance of traits relating to DNA processing has been documented in the related ciliate, Paramecium (Epstein and Forney, 1984; Forney and Blackburn, 1988). The four cell lines discussed here have undergone a second round cross (diagramed in Figure 2B) and it is not known which of them inherited the "star" parent cytoplasm.

As shown here, most 5S rRNA sequences were in identical local genomic environments in the newly isolated and laboratory strains, the only exceptions being a few additional 5S homologous bands in the ME44C micronuclear DNA. Similarly, in another study using a different repetitive micronuclear sequence to probe Southern blots no differences between laboratory and newly isolated strains were observed (Yao, 1982). Therefore repetitive sequence probes do not always show differences in the genomic organization of newly isolated strains when compared to laboratory strains. Molecular evidence such as that presented here has been used to identify or support the mobility of a number of well characterized elements. Tc1 elements of *C. elegans* were initially identified as the agents responsible for restriction fragment length polymorphisms observed between the common laboratory strains Bristol and Bergerac (Moerman and Waterson, 1989). The movement of Mu elements in Maize has been inferred from their non-Mendelian segregation of copy number and genomic location following crosses (Allemen and Freeling, 1986). Restriction fragment length polymorphisms detected by a DRE-1 probe are associated with movement of these elements in *Dictyostelium discoideum* (Marschalek et al., 1989). The variation in Tel-1 arrangement in ME44C relative to laboratory strains is analogous to these examples of elements known to be mobile.

Table 3.1 Genotypes and Phenotypes of Strains Used in This Work

Strains	Micronuclear Genotype	Macronuclear Phenotype	Mating Type
ME44C	wild type	wild type	not determined
SB805	*rmm1/rmm1,Mpr1/Mpr1*	6mp-R, pm-R	VII
SB813	*rmm1/rmm1,Mpr1/Mpr1,Pmr1/Pmr1*	6mp-R,pm-(R/S)	IV
SB210	*galA1/galA1*	2dgal-S,cycl-S	VI
B1868	wild type	wild type	not determined
SB2040	*ChxA2/ChxA2*	cycl-R,2dgal-S	IV
SB2044	*ChxA2/ChxA2*	cycl-R,2dgal-S	IV
SB2045	*ChxA2/ChxA2*	cycl-R,2dgal-S	V
SB2046	*ChxA2/ChxA2*	cycl-R,2dgal-S	VII
CU371	nulli 1R,2R		not determined
CU358	nulli 3,4,5	Cycl-R,6mp-R	not determined
CU359	nulli 2,3,5	Cycl-R,6mp-R	IV

Locus names are as follows: ChxA2, cycloheximide (cycl) resistance (Roberts and Orias, 1973, Byrne et al., 1978); Mpr1, 6-methylpurine (6mp) resistance (Byrne et al., 1978); Pmr1, paromomycin (pm) resistance (Bruns et al., 1985, Spangler and Blackburn, 1985); rmm1, ribosomal DNA maturation and maintenance (Larson et al., 1986); galA1, 2-deoxygalactose (2dgal) resistance (Roberts and Morse, 1980). Strain ME44C was obtained from Dr. E. Simon, University of Illinois. SB805 is described in Larson et al., 1986. SB813 is described in Yaeger,1987,Thesis, University of California, Santa Barbara. Strain SB210 was obtained from Dr. E. Orias, University of California, Santa Barbara. Strain B1868 was obtained from Dr. D. Nanney, University of Illinois. Strains SB2040, SB2044, SB2045, and SB2046 were obtained from Dr. S.-G. Zhang and Dr. E. Orias, University of California, Santa Barbara. Strain CU371 is described in Allen et al., 1984. Strains CU358 and CU359 are described in Bruns et al., 1983.

Table 3.2 Results of Test Crosses Involving SB2040.1 and SB2040.2 Cell Stocks

Cross	Karyonidal Lines Scored	Observed Phenotype		
		Cross Fertilized[1]	SB210 Parent-Like[2]	SB2040 Parent-Like[3]
SB210 X SB2040.1	50	0	14	36
SB210 X SB2040.2	49	41	2	5

1. Cells were scored as cross fertilized if they were resistant to both cycloheximide (15 μg/ml) and 2-deoxygalactose (2.5 mg/ml).
2. Cells were scored as having a SB210 parental-like phenotype if they were sensitive to both 2-deoxygalactose (2.5 mg/ml) and cycloheximide (15 μg/ml).
3. Cells were scored as having a SB2040 parental-like phenotype if they were sensative to 2-deoxygalactose (2.5 mg/ml) and resistant to cycloheximide (15 μg/ml).

Figure 3.1

Restriction map of the *T. thermophila* micronuclear DNA insert in plasmid p17D59. The C_4A_2 repeats are indicated by black boxes, the inverted repeats, containing the *BstXI* restriction sites, are indicated by arrowheads under the map. The fragment used as the Tel-1 probe is indicated. Sites for the restriction enzymes *BstXI* (B), *Eco*RI (E), and *Hin*dIII (H) are indicated.

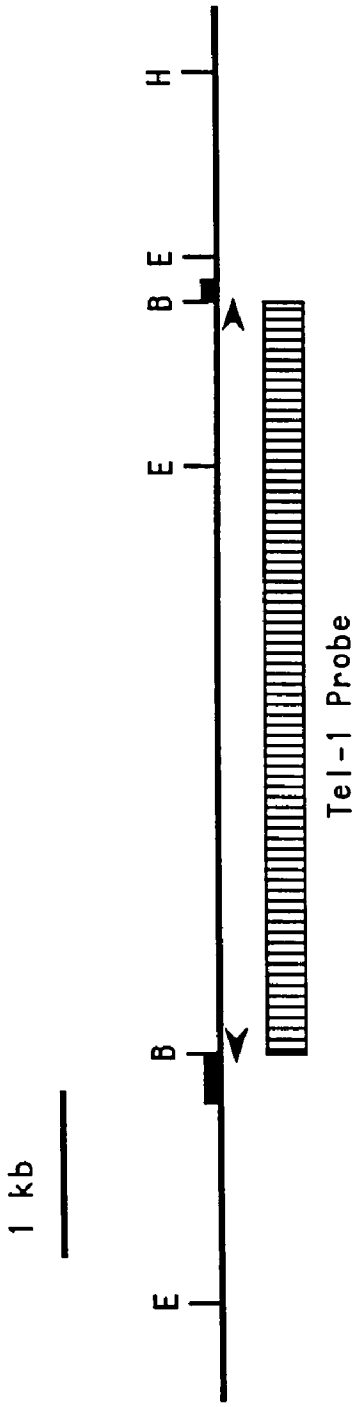

H
E
B
E
B
E
1 kb
Tel-1 Probe

Figure 3.2

Nuclear events in Tetrahymena Conjugation.

A. Normal Mating.

(I) Vegetative cells. (II) Cells pair and the micronucleus in each conjugant undergoes meiosis, loss of three meiotic products, and mitosis of the remaining one. The result of the micronuclear divisions is shown. There are two identical haploid gametic pronuclei in each cell. The arrows indicate reciprocal exchange of pronuclei. (III) In each cell the migratory pronucleus has fused with the stationary pronucleus forming the diploid fertilization nucleus. (IV) Fertilization nuclei undergo two post-zygotic mitotic divisions. The anterior products begin to develop into new macronuclei. (V) The exconjugants separate. The old macronucleus and one of the new micronuclei are destroyed. (VI) The first post-zygotic cell division results in four karyonides, each with an independently differentiated new macronuclei and a mitotic copy of the fertilization micronucleus from the exconjugant.

B. Genomic Exclusion.

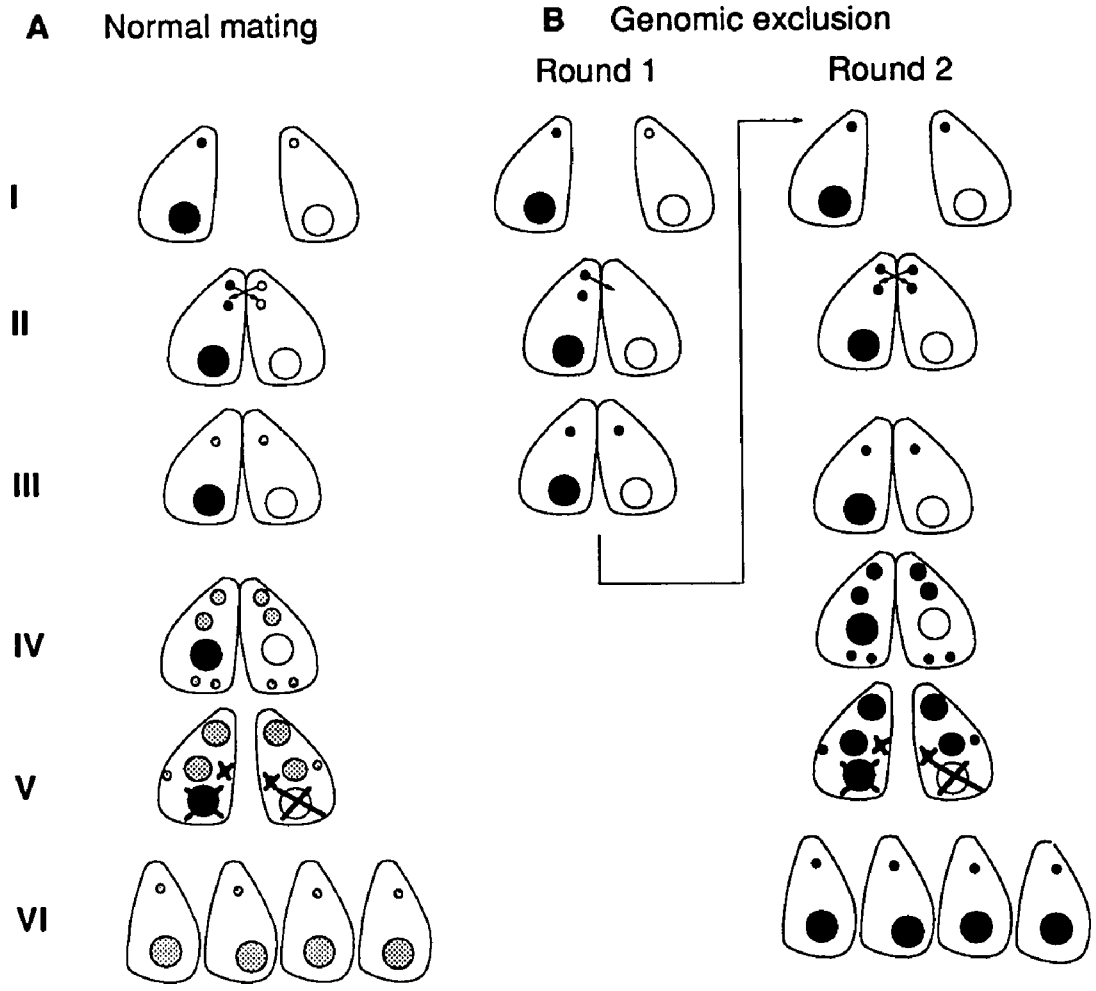
Cells are aligned so that the stages I-VI as explained in (A) are equivalent. Round 1. (I) vegetative cells. The cell to the right with white nuclei is a "star" strain. "Star" strains lose their micronuclei at meiosis. (II) Cells pair and conjugants undergo meiosis. At this stage the "star" strain cell loses its micronucleus and has no gametic pronucleus to exchange. The migratory pronucleus from the normal cell is transferred to the "star" strain cell. The haploid pronuclei in each cell are identical. These pronuclei each diploidize to make a whole-genome homozygous micronucleus and the cells maintain their old macronuclei. Round 2. (I) Vegetative cells from round 1 mating that have identical micronuclei and have retained their old macronuclei. (II) Cells pair and conjugants undergo meiosis and mitosis as already described. Note that the cells had identical whole-genome homozygous diploid micronuclei as a result of the round 1 cross. Now they exchange genetically identical pronuclei. (IV-VI) These stages proceed

as described in (A) Karyonides receive identical micronuclei that are homozygous throughout their entire genome and independently differentiated micronuclei.

C. Construction of Nullisomic Strains.

(i) during a rare triple conjugation event all cells produce two haploid gametic pronuclei as described before. (ii) Pronuclear exchange results in unequal distribution of haploid gametic nuclei. The left conjugant received two haploid pronuclei that can fuse to produce a diploid fertilization nuclei as in a normal mating. The center conjugant received three haploid pronuclei that can fuse to make a triploid fertilization nuclei. The conjugant of interest for the rest of this mating scheme is the rightmost one. This conjugant is left with only one copy of its haploid pronuclei. This micronucleus remains haploid and the old macronucleus remains functional. (I) The cell with the haploid micronucleus is to be mated with a normal cell (dark nuclei). Stages I-VI are analogous to those described in (A). (II) Micronuclear meiosis and mitosis ensues to produce the gametic pronuclei. For the left cell with the haploid micronucleus this is often fatal. However, some cells successfully go through the reductive divisions of meiosis and produce aneuploid pronuclei missing whole chromosomes or parts of chromosomes. (III) Pronuclear fusion in the conjugants results in fertilization nuclei that are monosomic for some chromosomes. Nevertheless, the nuclear events proceed as for stages IV-VI of a normal mating (described in part A) and functional new macronuclei are formed. (I*) A vegetative cell with a monosomic micronucleus from the previous cross is mated to a "star" strain. This is the same type of mating as diagramed in part B, genomic exclusion round 1. (II*) Upon meiosis and mitosis the "star" strain loses its micronucleus. The strain with a monosomic micronucleus produces aneuploid gametic pronuclei missing chromosomes or their parts. (III*) After pronuclear migration each cell has a pronucleus that is haploid for some chromosomes and missing others. As in the genomic exclusion round 1 cross the conjugants diploidize their pronuclei and retain their old macronuclei. Diploidization of

aneuploid pronuclei results in micronuclei missing both pairs of particular chromosomes. The resulting strains are nullisomic for the missing chromosomes. The presence or absence of each of the five micronuclear chromosomes in the exconjugants is determined by cytology and further test crosses.


A Normal mating
B Genomic exclusion
Round 1
Round 2
I
II
III
IV
V
VI

C Construction of Nullisomics

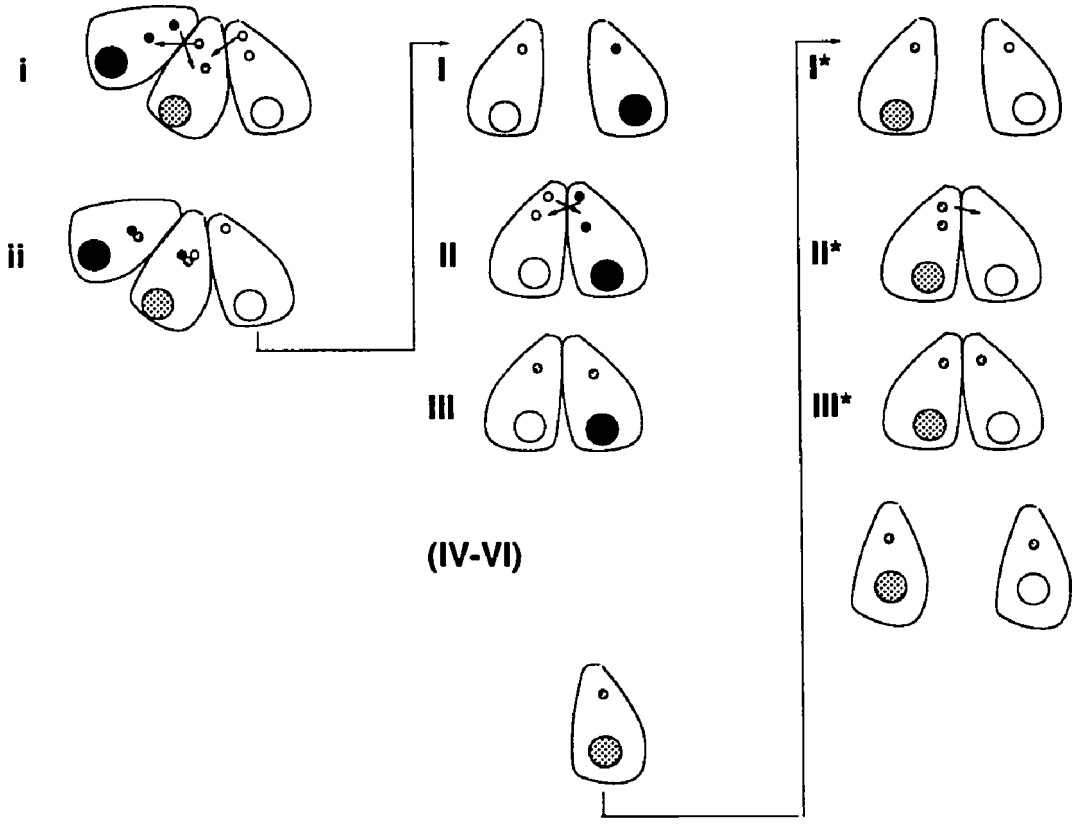
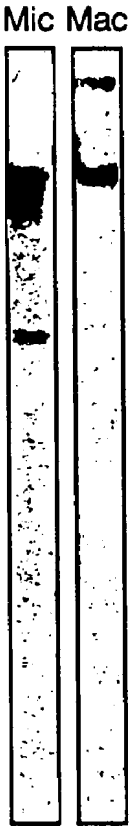


Figure 3.3

Southern blot of *BstXI* digested micronuclear and macronuclear DNA probed with Tel-1. Micronuclear (mic) and macronuclear (mac) DNA from strain SB2045, 10 μg of each sample, was digested with *BstXI*, run on a 0.8% agarose gel, blotted to nitrocellulose and hybridized to a nick translated Tel-1 probe.



*Bst*XI digest

Tel-1 probe

Figure 3.4

Southern blot of *Hin*dIII digested micronuclear DNA from three standard laboratory strains probed with Tel-1. 10 μg of micronuclear DNA from each strain (lane 1= SB2045, land 2= SB805, lane 3= SB813) was digested with *Hin*dIII, run on a 0.8% agarose gel, blotted to nitrocellulose, and hybridized to a nick translated Tel-1 probe.

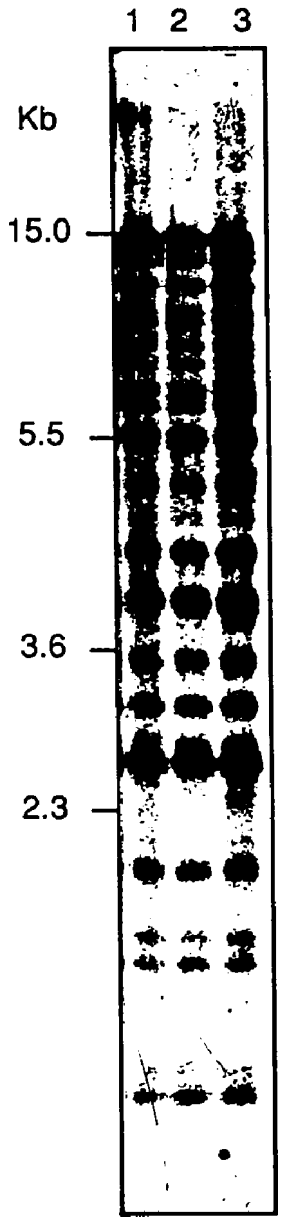

1 2 3
Kb
15.0
5.5
3.6
2.3
Tel-1 probe

Figure 3.5

Southern blot of micronuclear genome of a newly isolated *T. thermophila* strain probed with Tel-1 and 5S.

A. Southern blot of micronuclear DNA from the newly isolated strain ME44C (lanes 1,3, and 5) and a standard laboratory strain, SB2046 (lanes 2,4, and 6). Each lane contains 10 μg of DNA, undigested in lanes 1 and 2, *BstXI* digested in lanes 3 and 4, and *Hin*dIII digested in lanes 5 and 6. The DNA was run on a 0.8% agarose gel, blotted to nitrocellulose, and hybridized to a nick translated Tel-1 probe.

B. The *Hin*dIII lanes from the blot shown in A were stripped of probe and rehybridized to a nick translated 5S rRNA gene probe. (lane 1= ME44C, lane 2= SB2046).

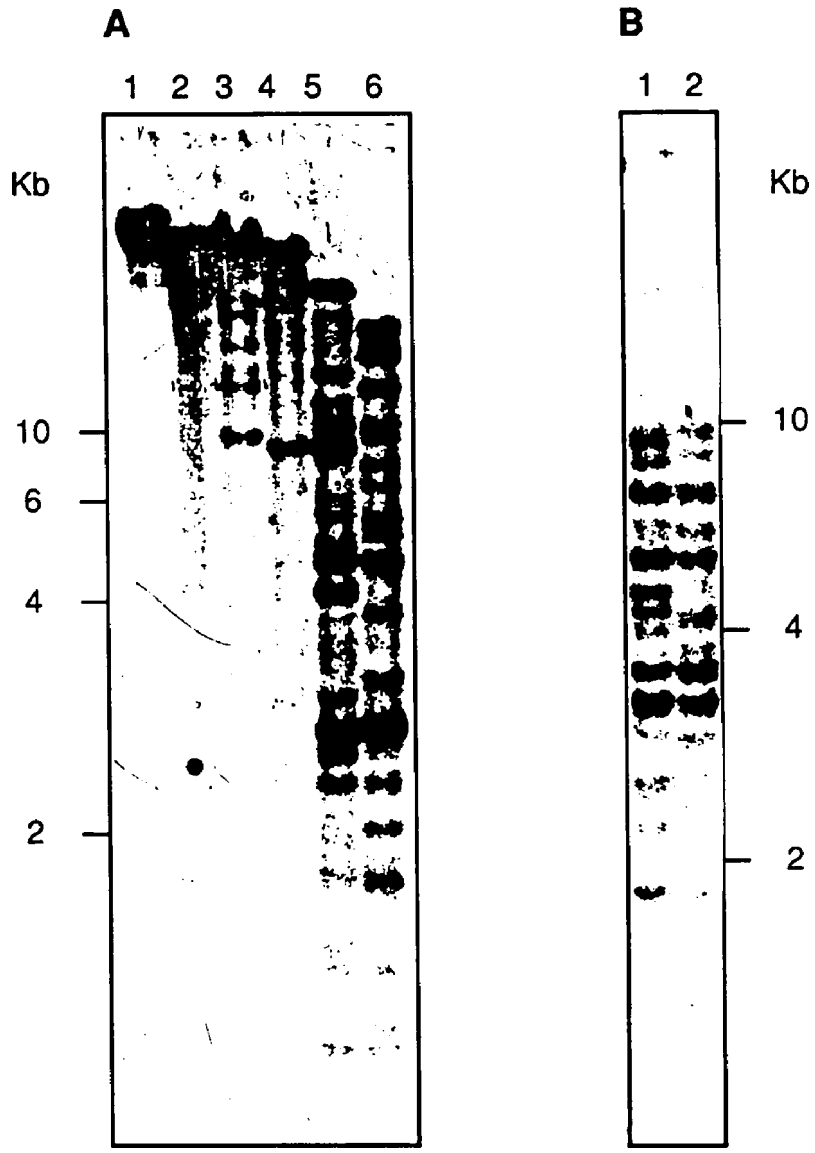
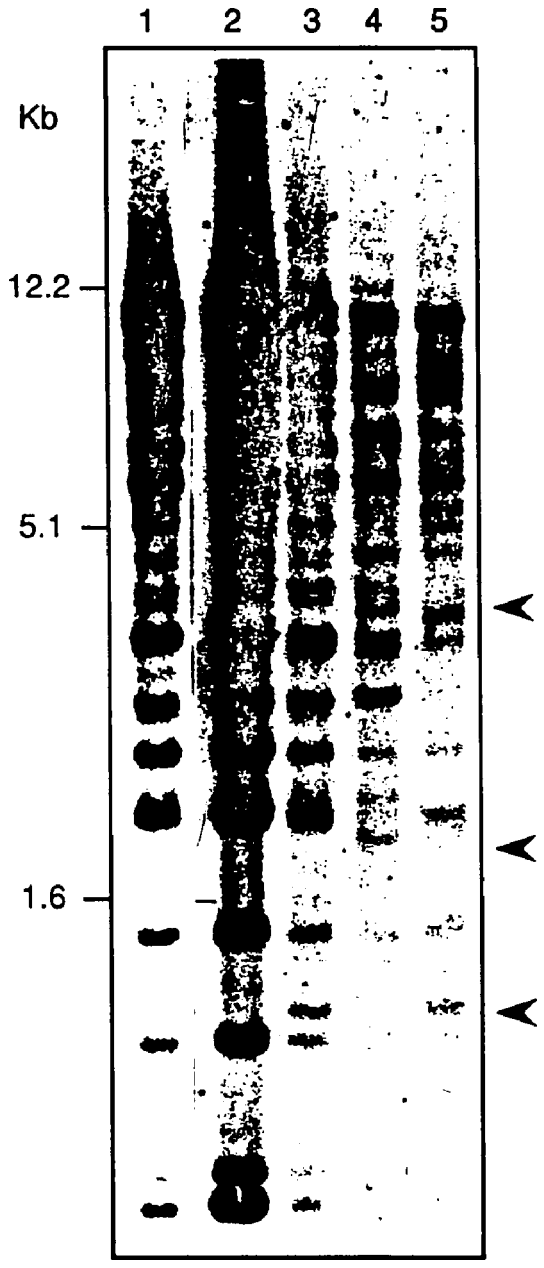

A
1 2 3 4 5 6
Kb
10
6
4
2
Tel-1 probe
B
1 2
Kb
10
4
2
5S probe

Figure 3.6

Southern blot of micronuclear DNA from nullisomic strains probed with Tel-1. The blot was prepared from *Hin*dIII digested micronuclear DNA from a standard laboratory strain, B1868 (lane 2 and a lighter exposure of the same sample in lane 1) and three Nullisomic strains (lane 3= Nulli 1R,2R, lane 4= Nulli 3,4,5, lane 5= Nulli 2,3,5). The chart below the lanes indicates which chromosomes are present in the various nullisomic strains. The DNA (not equal amounts in each lane) was separated on a 0.8% agarose gel, blotted to nitrocellulose, and hybridized to a nick translated Tel-1 probe. The arrows to the right indicate the position of Tel-1 homologous bands that are rearranged relative to B1868.

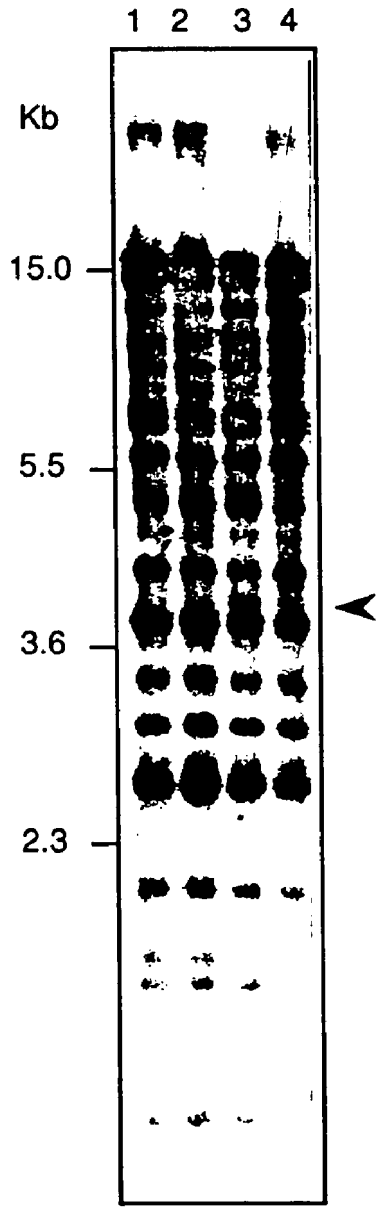


Tel-1 probe

Mic	Chromosomes Present			
1	+	L	+	+
2	+	L		
3	+	+	+	
4	+	+		+
5	+	+		

Figure 3.7

Southern blot of micronuclear DNA from four identical whole-genome homozygotes probed with Tel-1. 10 µg of *Hin*dIII digested micronuclear DNA from each cell line (lane 1= SB2044, 2=SB2046, 3=SB2040, 4=SB2045) was separated on a 0.8% agarose gel, blotted to nitrocellulose and hybridized to a nick translated Tel-1 probe. The arrowhead to the right indicates the position of the additional band in the SB2040 lane.



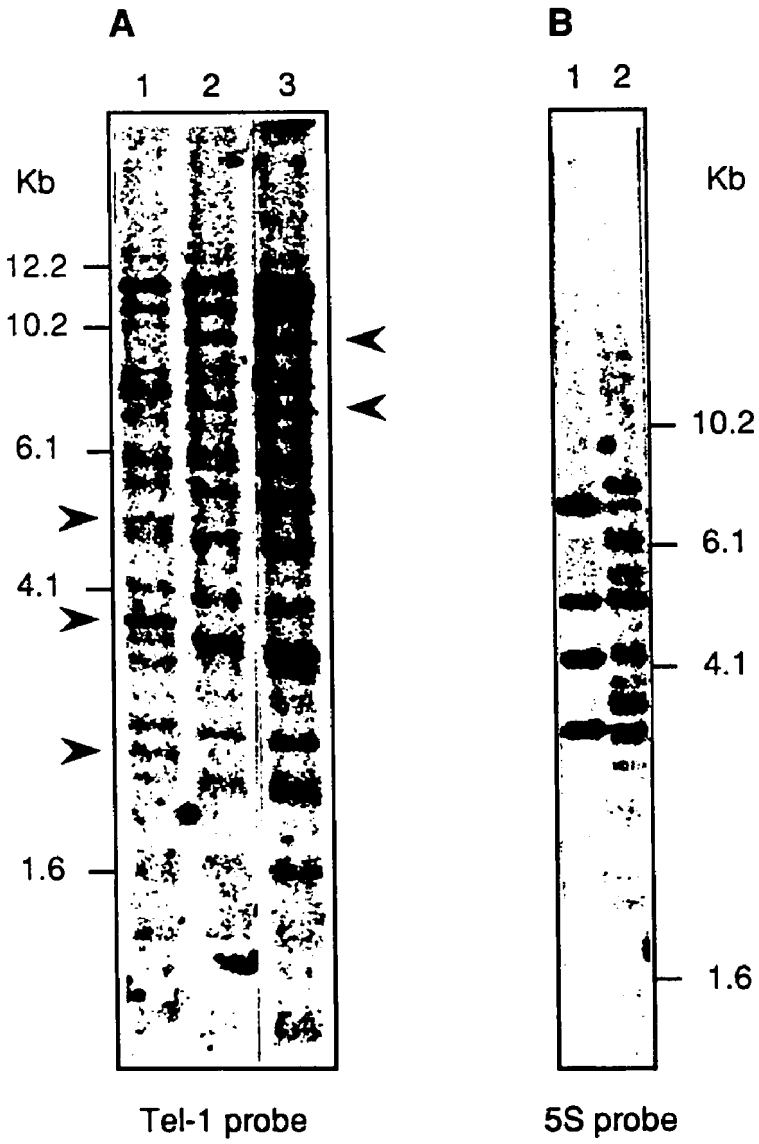
Tel-1 probe

Figure 3.8

Southern blot of micronuclear DNA from both stocks of SB2040 probed with Tel-1 and 5S.

A. Southern blot of micronuclear DNA prepared six months after that shown in figure 6. Micronuclear DNA from SB2040 stocks (lane 1=SB2040.1, lane 2= SB2040.2) and SB2046 (lane 3), as a representative standard strain, was digested with *Hin*dIII, separated on a 0.8% agarose gel, blotted to nytran and hybridized to a nick translated Tel-1 probe. Arrowheads to the left of the lanes indicate new Tel-1 homologous *Hin*dIII fragments appearing in SB2040.1. Arrowheads to the right indicate Tel-1 homologous *Hin*dIII fragments now missing from SB2040.1.

B. The SB2040.1 and SB2040.2 lanes from A stripped of probe and rehybridized to a nick translated 5S rRNA gene probe.


A
1 2 3
Kb
12.2
10.2
6.1
4.1
1.6
Tel-1 probe
B
1 2
Kb
10.2
6.1
4.1
1.6
5S probe

Chapter 4

Identification of Tmi Elements

Introduction

In this and the following chapter the *Tetrahymena thermophila* micronuclear genome is shown to contain retrovirus-like elements or retrotransposons designated Tmi (Tetrahymena micronuclear) elements. Tmi elements are moderately repetitive, have a common genomic restriction map and are limited to the micronuclear genome. They were associated with some micronuclear genome rearrangements in stock SB2040.1. Several DNA clones containing one end of the conserved element and variable lengths of flanking DNA were isolated and analyzed. The restriction map of the clones was colinear with the majority of elements in the micronuclear genome. Sequences from the end of the Tmi element are about twice as abundant in the genome as internal Tmi sequences. Sequencing (to be presented in the following chapter) of the cloned micronuclear loci revealed that they include conserved DNA elements unrelated to Tel-1. Rather, by sequence comparison Tmi elements are similar to a retrovirus and a retrotransposon. Other Tetrahymena species showed cross hybridization to Tmi probes but did not show conservation of the Tmi restriction map as determined for *Tetrahymena thermophila.*

One stock of the whole-genome homozygote cell line SB2040 has been shown to have undergone micronuclear genome rearrangements linked to Tel-1 elements (chapter 3 of this thesis). An attempt to amplify Tel-1 elements from Tetrahymena micronuclear DNA by the polymerase chain reaction was unsuccessful for this purpose but did

produce a fragment that was useful in the work described here as a micronucleus specific probe. This was a micronuclear amplification product generated with Tel-1 end primers, designated the PCR 900 probe, that detects micronuclear rearrangements in the SB2040.1 stock relative to the SB2040.2 stock. However, unlike the entire Tel-1 probe the PCR 900 probe was specific for micronuclear loci and hybridized to apparently multiple copy genomic restriction fragments generated by several different enzymes. The PCR 900 probe and genomic clones containing similar micronuclear elements selected by hybridization to it, will be described in this chapter.

Results and Discussion

PCR amplification of a 900 bp fragment from *T. thermophila* using Tel-1 end primers

A synthetic oligonucleotide designed to prime synthesis in from both ends of the Tel-1 inverted repeats was used in PCR amplification of Tetrahymena DNA (diagramed in Figure 1). One µg of micronuclear DNA was mixed with the Tel-1 end primer in a Gene Amp kit (Perkins Elmer/Cetus) reaction mix and amplified for 40 cycles in a Perkins Elmer thermal cycler with the following settings for a simple step program: denaturation at 95°C for 2 minutes, annealing at 55°C for 4 minutes and polymerization at 72°C for 8 minutes. A prominent 900 bp amplification product resulted and was shown on Southern blots to hybridize to the Tel-1 probe (data not shown). The Tel-1 end oligomer used as a primer contains a *Bst*XI restriction site and the amplified product contains *Bst*XI sites near both ends. Southern blots failed to detect a 0.9 kb genomic *Bst*XI fragment homologous to either this PCR 900 fragment or to the Tel-1 probe in micronuclear or macronuclear DNA of the *T. thermophila* strain used as a template for PCR amplification. Therefore this specific fragment was not

detectable in the *T. thermophila* genome from which it was amplified. Either the amplified 900 bp piece is rare in the genome or it was generated in this form during the amplification reactions. The primer used in the amplification reaction was 31 bp long and as little as 16 bp of homology is needed to start synthesis using this protocol (C. Wyman unpublished observation). Synthesis from a genomic template could have begun with hybridization of only the 3' 16 nucleotides of the primer that do not include the *BstXI* site. The genomic locus amplified in this reaction may not have *BstXI* restriction sites.

The primer used insured that the PCR 900 product had at least 31 bp inverted repeats at its ends. However, when the PCR 900 product was boiled and quickly cooled before loading onto an agarose gel it migrated as if it was half of its original size (Figure 2). In such a heating and quick cooling protocol extensive inverted repeats from one strand of a DNA fragment will hybridize to each other creating "snap back" molecules that appear shorter than the original double stranded DNA. The PCR 900 product appears to have significant inverted repeat sequences in order to show complete reduction to a half length "snap back" molecule. The PCR 900 product was cloned in two pieces and completely sequenced (Figure 3A and 3B). The sequence revealed the ends to be long inverted repeats, 240 bp long, related to the previously sequenced regions adjacent to $micC_4A_2$ repeats (Figure 3C) presumed to be Tel-1 ends (Cherry and Blackburn, 1985). However, as shown below, the genomic clones isolated by hybridization to the PCR 900 probe share sequence similarity to the middle, Tel-1 unrelated, segment of this fragment and are not in any obvious way related to Tel-1 elements. It is likely that the Tel-1-like sequences at the inverted repeat ends of the PCR probe were not efficiently labeled in the procedure used to make hybridization probes. In the random oligomer primed synthesis of labeled probes (see chapter 2) the DNA fragment used as a template (in this case the PCR 900 product) is first boiled to denature it and allow hybridization

of oligomeric primers. Intramolecular annealing of the inverted repeats of the PCR 900 product, as discussed above, probably made them less available for synthesis of labeled probes than the internal region.

This apparent association of Tel-1 sequences with a another type of element may be similar to the association of different types of mobile elements known in other species. Some elements apparently insert preferentially into copies of themselves, other elements, or repetitive sequences. Some yeast Ty elements have been shown to insert preferentially into delta sequences, which are solo Ty LTRs (Warmington et al., 1986). One type of element in Dictyostelium, DRE2, is always found associated with another type of element, DRE1 (Marschalek, et al., 1989). Composite elements of Xenopus, with internal sequences similar to retroviral open reading frames and inverted repeat ends, may have resulted from one type of element inserting into the other (Garrett et al., 1989).

The PCR 900 probe recognized only micronuclear sequences and some of them were rearranged between SB2040.1 and SB2040.2

The PCR 900 fragment was labeled and used as a probe of *T. thermophila* genomic Southern blots. The PCR 900 hybridizing genomic loci were all limited to the micronuclear genome. Figure 4 shows a Southern blot of purified macronuclear DNA (lane 1), purified micronuclear DNA (lane 2), and whole cell DNA (lane 3), all digested with *ClaI*. Equal amounts of DNA were loaded in each lane as judged by ethidium bromide staining (data not shown). In the micronuclear DNA the two most prominent PCR 900 homologous *ClaI* fragments were approximately 2.8 kb and 3.8 kb. These bands were also evident in the whole cell DNA sample. The faint hybridization to the macronuclear DNA sample was most likely due to slight contamination of the

macronuclei preparation with micronuclei. The macronucleus has only 80-90% of the sequence complexity of the micronucleus (Karrer, 1986). Hence any sequence that is retained in the macronucleus would be expected to be at least slightly over represented in macronuclear DNA and show stronger hybridizing bands if, as in this case, equal amounts of micronuclear and macronuclear DNA were loaded on the gel. In contrast to the results seen here with the PCR 900 probe, the Tel-1 probe hybridized to repeated sequences in macronuclear DNA as well as micronuclear DNA.

As shown previously, the Tel-1 probe recognized micronuclear loci that had rearranged in stock SB2040.1 relative to stock SB2040.2 or other standard laboratory strains. Figure 5 shows that the PCR 900 probe also recognized multiple *Hin*dIII fragments in *T. thermophila* micronuclear DNA and some of them were rearranged (indicated by arrows to the right of the figure) between SB2040.1 (lane 2) and a standard laboratory strain SB2046 (lane 1). There was no *Hin*dIII site within the PCR 900 probe. Therefore the number of *Hin*dIII fragments that hybridize to the PCR 900 probe is an estimate of the number of genomic copies of PCR 900 similar sequences. Based on the blot in Figure 5 there are approximately 20 copies sequences similar to the of PCR 900 probe in the genomes of these strains.

Common restriction sites in micronuclear loci homologous to the PCR 900 probe

The blot of *Hin*dIII digested micronuclear DNA indicated that the PCR 900 homologous loci were repetitive, with approximately 20 copies (Figure 5) in the micronuclear genome. The blot of *Cla*I digested micronuclear DNA indicated that most of these PCR 900 homologous loci had common *Cla*I sites 2.8 or 3.8 kb apart (Figure 4). Because the PCR 900 probe hybridizes to micronuclear limited DNA, for the analysis of PCR 900 homologous loci, whole cell DNA samples can be used on Southern blots. Whole cell

DNA was cut with several restriction enzymes, displayed on Southern blots, and hybridized to the PCR 900 probe. This revealed several common restriction sites in the PCR 900 homologous loci. There was one prominent 6 kb *Xmn*I fragment (Figure 6 lane 1) and one prominent *Bst*XI fragment of 4.3 kb (Figure 6 lane 2). This implies that the common restriction map of the micronuclear PCR 900 similar loci extends for a minimum of 6 kb. For both enzymes these fragments represented the majority of hybridization to the probe. Although the ends of the PCR 900 probe were similar to Tel-1 ends, for reasons mentioned above they were probably not represented in the probe made from this fragment. The hybridization is to the non-Tel-1 sequences from the middle of the PCR 900 probe. The Tel-1 probe, in contrast, hybridized mainly to genomic locations without common *Bst*XI sites as evident by the intensity of the smear near limit mobility in Figure 3.3 (this thesis chapter 3). The PCR 900 probe recognized micronuclear limited loci with common sites for at least three restriction enzymes, *Bst*XI, *Cla*I, and *Xmn*I. It detected micronuclear rearrangements between the SB2040.1 and SB2040.2, suggesting the possibility that the PCR 900 cross hybridizing loci include mobile genetic elements.

The PCR 900 probe selected clones containing similar micronuclear limited elements designated Tmi

In order to identify different examples of the common micronuclear PCR 900 homologous loci two clone libraries were constructed by inserting HindIII digested SB2040.1 micronuclear DNA into pUC19 plasmid vector. The clone libraries differed in the size of *Hin*dIII fragments selected for inserts, one containing inserts of 3-5 kb and the other inserts of 5-8 kb. Clones were selected by colony hybridization to the PCR 900 probe. Eleven independent clones were so identified and further analyzed. Restriction maps of these clones are shown in Figure 7. As can be seen all 11 shared a similar

arrangement of 3 *Eco*RI sites and one *BstX*I site. In addition 9 of the clones could be aligned at one of their HindIII ends. The PCR 900 probe hybridized to the common 300 bp *Eco*RI fragment of these clones. Subsequent sequence analysis (see following chapter) has shown a 250 bp region overlapping the middle common *Eco*RI site in the clones to be almost identical to the middle 250 bp of the PCR 900 probe (Figure 2 indicates this segment of the PCR 900 fragment and the sequence of two of the clones is reported in the following chapter). The 9 clones that can be aligned by their *Hin*dIII sites at one end included about 2.5 kb of a common micronuclear element and a variable amount of flanking sequence.

Southern blots were prepared to see if the common restriction map of the clones was representative of the genome. The restriction map of the clones suggests common *Eco*RI fragments of 300 and 700 bp. Figure 8 shows a blot of micronuclear DNA (lane 1) and macronuclear DNA (lane 2) digested with *Eco*RI and probed with either the 300 bp (panel A) or 700 bp (panel B) EcoRI fragments from clone F08. The arrangement of *Eco*RI sites that produced these fragments in the clones is clearly colinear in the genome. The sequences contained in these two *Eco*RI fragments were also limited to the micronuclear genome. Assuming the *BstX*I site in the clones is one fixed end of the prominent 4.3 kb genomic *BstX*I fragment that hybridizes to the PCR 900 probe (Figure 6, lane 2) or the *Eco*RI 300 or 700 probes (data not shown) the genomic elements continue to a *BstX*I site 1.5 kb past the *Hin*dIII end of the examples cloned.

Clones F08 and F23 were more extensively restriction mapped, as shown in Figure 9. A *Cla*I and an *Xho*I site are present in identical positions in these two clones. Based on genomic blots of *Cla*I digested DNA hybridized to the PCR 900 probe (Figure 4) (the *Eco*RI 300 or 700 probes from clone F08 gave the same results) the genomic elements have another *Cla*I site either 0.6 or 1.6 kb beyond the cloned right *Hin*dIII end. Genomic DNA

doubly digested with *Xho*I and *Bst*XI and displayed on a Southern blot shows a 1.2 kb fragment that hybridizes to the *Eco*RI 300 bp probe (Figure 8C). This is the size of the fragment predicted from the F08 and F23 restriction maps and confirms that this arrangement of a *Bst*XI site and an *Xho*I site is also common in the genomic elements. Genomic Southern blots probed with all tested fragments from the non-conserved end of the clones (left side in Figure 7) show the elements to be embedded in repetitive micronuclear limited sequences (data not shown). Thus the PCR 900 selected clones are 9 examples of one end of a conserved micronuclear limited elements now designated Tmi elements.

One additional genomic clone likely to represent a more complete Tmi element was selected for study. Genomic Southern blots probed with the conserved 1.4 kb *Eco*RI to *Hin*dIII fragment (right side in Figure 7) from F08 showed strong hybridization to a 2.5 kb genomic *Eco*RI fragment (data not shown). This indicates that the genomic elements have an common *Eco*RI site 1.1 kb to the right of the cloned *Hin*dIII end. By hybridization to the PCR 900 probe, clone D01 was selected from a genomic library of *Cla*I cut micronuclear DNA cloned into pUC118. Its restriction map overlapped the previous clones from the conserved *Cla*I to *Hin*dIII sites and continued to include the *Eco*RI and *Bst*XI sites predicted from the genomic mapping (see Figure 9).

Clone D01 did not have an internal *Cla*I site either 0.6 kb or 1.6 kb away from the *Hin*dIII site as predicted from the genomic mapping presented above. However, other restriction sites in the portion of clone D01 not shared by clones F08 and F23 are common in the genome. A genomic Southern blot of *Eco*RI digested DNA was probed with the 1.7 kb *Eco*RI fragment of D01 that is to the right of the *Hin*dIII site (see Figure 9) that marks the end of the overlap with clone F08 and F32. This probe hybridized to a similarly sized genomic *Eco*RI fragment (data not shown) indicating that the additional

sequences in clone D01 are representative of common micronuclear elements. The probe D01 1.7 kb *Eco*RI sequences were also limited to the micronucleus. Clone D01 therefore appears to contain a copy of a Tmi element lacking a *Cla*I site. Clones F08, F23 and D01 were sequenced in order to determine the type of genomic elements they represented. The sequence and its analysis are presented in the following chapter.

Tmi end sequences are more common in the genome than Tmi internal sequences

Southern blots of *Hin*dIII digested micronuclear DNA showed more than 20 bands that hybridize to the PCR 900 probe (Figure 5) or to either of the conserved *Eco*RI fragments (300 or 700 bp) from F08. Figure 10 shows that the pattern of bands hybridizing to the two *Eco*RI fragment probes (lane 1 was probed with the 700 bp *Eco*RI fragment and lane 2 with the 300 bp *Eco*RI fragment) was similar but not identical. In contrast, similar blots hybridized to a probe from a more internal position of the element (the 1.0 kb *Eco*RI to *Hin*dIII fragment from the right end of clone D01, hereafter referred to as the internal fragment probe) showed hybridization to only about 8 fragments in *Hin*dIII cut micronuclear DNA (Figure 10, lane 3). As the bands hybridizing to the internal probe are of different intensities, the more intense bands may represent multiple copies of this sized fragment. From these results it is concluded that there are a minimum of 8 genomic loci homologous to the internal probe. Furthermore, there are more representatives of Tmi end sequences than Tmi internal sequences in the genome.

The blots shown in Figure 10 are of *Hin*dIII digested micronuclear DNA. All but one or two of the *Hin*dIII fragments that hybridize to the internal probe also hybridize to the *Eco*RI 700 end probe (compare lanes 3 and 1, Figure 10). Recall that 9 of 11 of the clones analyzed had a conserved *Hin*dIII site that indicated the 700 bp and 300 bp *Eco*RI probe sequences would be on separate *Hin*dIII fragments from the internal probe sequences.

Therefore one would not expect genomic *Hin*dIII fragments to hybridize to both internal and end probes unless either, (1) the *Hin*dIII site present in the cloned versions of the element is not common in the genome or, (2) these *Eco*RI 700 and 300 bp sequences are repeated at the other end of the element.

The additional fragments hybridizing to Tmi end probes could have resulted from these sequences being repeated at the ends of the element. End sequences from complete elements would be twice as frequent in the genome as internal sequences. This would be true for elements with either inverted or directly repeated ends. Mobile DNA elements with inverted repeats at their ends typically have less than 100 bp of repeated sequence. Recently transposon-like element from the micronucleus of another ciliate , *Euplotes crassus*, have been described (Jahn, et al., 1989) which have long inverted repeats of about 700 bp at their ends. It would be unlike typical transposons but not unprecedented that the *Eco*RI 300 bp fragment, that is separated from the end by 700 bp, is included in such an inverted repeat. The directly repeated ends of retroviruses and retroposons are typically longer and could include the 1 kb of sequence sampled here.

The additional end sequences could alternatively be due to solitary element ends in the genome. In this case the end sequences would not always be linked to internal sequences in the genome There are several *Hin*dIII fragments that hybridize to the *Eco*RI 700 and 300 bp probes that do not hybridize to the internal probe (compare lanes 1 and 2 to 3 in Figure 10). Solo ends can be generated by homologous recombination between direct repeats at the ends of an element. This is believed to be the source of delta sequences in yeast. They are known to be single copies of Ty element LTRs and are more common in the genome than complete Ty elements (Boeke, 1989). Over representation of one end of an element has also been described for retroposons. In this case there are frequent elements truncated at one end (for example LINEs, (Hutchinson,

et al., 1989)). From the data presented here an end repeat would have to extend into the *Eco*RI 300 bp fragment for the solo end sequences generated by homologous recombination to be detected by this probe.

A definitive description of the end structure of Tmi elements awaits the cloning of a complete element. However it seems likely that the sequences at the end of what has been cloned, the *Eco*RI 700 and 300 bp fragments, are repeated in some form. There are also possible examples of Tmi end sequences in the genome that are not part of complete elements.

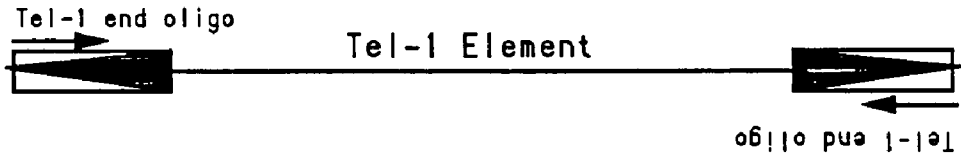
Other species of Tetrahymena had Tmi cross hybridizing sequences

The genomes of other species of Tetrahymena, *T. hegewischi*, *T. pigmentosa*, and *T. malaccensis*, were screened for Tmi elements. Phylogenetic trees based on either macronuclear rRNA sequence comparisons (Sogin et al., 1986) or mitochondrial sequence comparison (Morin and Cech, 1988) place *T. malaccensis* close to *T. thermophila*, *T. hegewischi* farthest from *T. thermophila* and *T. pigmentosa* at an intermediate distance from *T. thermophila*. Whole cell DNA from these three species was digested separately with *Eco*RI and *Hin*dIII and displayed on a Southern blot. This blot was probed with a mixture of the *Eco*RI 300 and 700 bp fragments from clone F08. Hybridization and blot washing conditions were at the same high stringency as for Southern blots of *T. thermophila* DNA (see chapter 2 for specific conditions used). As can be seen in Figure 11 there are several *Eco*RI and *Hin*dIII fragments from each species that hybridize to the probe. The *Eco*RI fragments predicted from the Tmi clones and *T. thermophila* genomic map were not detected in DNA from any of the other Tetrahymena species. These results indicate that there are repetitive Tmi related

sequences in other Tetrahymena species however, elements having the same restriction map as those from *T. thermophila* were not detected.

Figure 4.1.

Proposed scheme for amplification of Tel-1 elements using a primer from the inverted repeat end. A Tel-1 end primer was synthesized such that it would function as a primer for synthesis in from both ends of a Tel-1 element in a polymerase chain reaction. A theoretical Tel-1 element with inverted repeats at its ends is diagrammed. The synthetic oligonucleotide is drawn over the position where it will hybridize to the Tel-1 end and the direction of DNA synthesis indicated by the arrowhead. The sequence of the Tel-1 end oligomer used is shown.



Tel-1 End Oligonucleotide

5'CGGGTTCCCCATTGAGTTGGGGTTAGTATAA 3'

Figure 4.2.

The PCR 900 probe forms apparent half length fragments upon heating and quick cooling. Two samples of the PCR 900 probe were loaded onto the agarose gel shown here. In lane 1 the sample was loaded directly onto the gel after adding loading dye. The sample in lane 2 was first boiled for 2 minutes then immersed in an ice water bath before adding loading dye and loading onto the gel. Lane 3 has the 1 kb ladder (Bethesda Research Labs) DNA size markers.

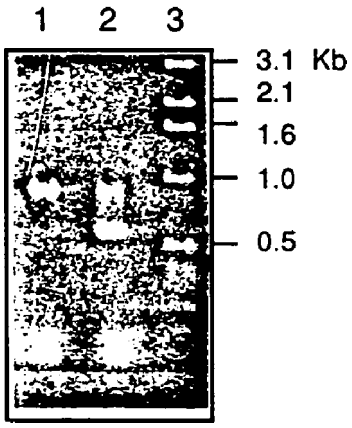
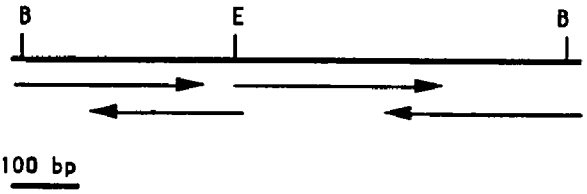

1
2
3
3.1 Kb
2.1
1.6
1.0
0.5

Figure 4.3.

The sequence of the PCR 900 product. The prominent Tel-1 hybridizing PCR product was subcloned in two pieces from an internal *EcoRI* site. (A) Diagram of the sequencing strategy. (B) Complete sequence of the PCR product. The inverted repeat ends related to Tel-1 elements are underlined. The central portion that is almost identical in sequence to portions of the clones F08 and F23 is underlined with stars. (C) The ends of the PCR 900 probe sequence aligned with previously sequenced mic C_4A_2 adjacent regions believed to be Tel-1 ends (Cherry and Blackburn, 1985).

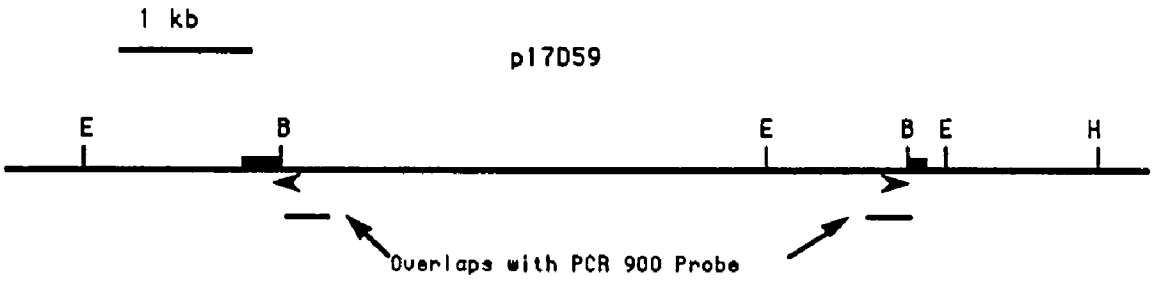
A



B

```
        10         20         30         40         50         60         70
CGGGTTCCCC ATTGAGTTGG GGTTAGTATA ATTTTAATAT TTTATAATTT TTAATAATTA ATTAATTTTA

        80         90        100        110        120        130        140
AATATTGAAT AAAAATATTT GAGTTATCAA TTTCCCCTTT TCTTATTGGT TACTTTAAAT TTTAATTTGA

       150        160        170        180        190        200        210
GCTCACAATA CCAAAAACCA TAAATACAAC TCAGATGGTT AGCAAGCAAA AAGTTTATAT ACATCTACTA

       220        230        240        250        260        270        280
TTATGAAATA ATTAAAAATC AAATTTTAAT GGAGATGCTC ATTTTAAGCG CATTCATCTA TTTTCTATGC

       290        300        310        320        330        340        350
CCAATGAATT ACCATTTTCC ATAGTTTTAC CGAATACTGA GTTATTCATC AATTTGTAAA AATCTTTTTC
           ********** ********** ********** ********** ********** **********
       360        370        380        390        400        410        420
GAATTCGTTT CTTGCTTGCT GTCTCATTTG AGTGTTTAAA TTAATACACT CTTCCATGAA ATTTGATTAT
********** ********** ********** ********** ********** ********** **********
       430        440        450        460        470        480        490
CGTATTTGAT AATTCTATGT ATTTTACCAA TTTTAATCCT TACTCTAAAT ACAATTTTTA ATTTCTGATA
********** ********** ********** ********** ********** ********** **********
       500        510        520        530        540        550        560
TTGCTAACGT ATCTCTCTTT GTTGTTAAGA TTTGGAGTTA ATTTCTTTGG TAATTCATTG GGCATCCCCG
********** ********** ********** ********** ********
       570        580        590        600        610        620        630
TCACCGATCA TAAAATTGAA CCCCAGGCCA TAAGGAAAAT AAGGTTGGAA CGCCCCAAAA GGAACCCACC

       640        650        660        670        680        690        700
TCCATAAAAT TGAATTTTTT AATTTTTTAA TTTCTTTGAT AATAGAAGAT GTATATAAAT TTTTTGCTTG

       710        720        730        740        750        760        770
CTAATCAATC TGAGTTATAT TTATGGTTTT TGGTATTATG AGCTCAAATT AACATTTAAA GTAACCATAA

       780        790        800        810        820        830        840
GAAAAGGGGA AATTGATAAC TCAAATATTT TTATTCAATA TTTAAAATTT AATTTAACTA TTAAAAATTA

       850        860        870        880        890
AATTATGAAA TATAAAATAT CAAAATTATA CTAACCCCTT CTCAATGGGG AACCCG
```

C



Alignment of the left end of the PCR 900 probe with the left side conserved sequence from p17D59

```
                                                                  10
PCR 900                                                           CGGGTTCCCCATTGA
                                                                  ||||||||||||||||
p17D59      CCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACTCAACGGGTTCCCCATTGA

               20        30        40        50        60        70
PCR 900     GTTGGGGTTAGTATAATTTTAATATTTTATAATT-TTTAAT-AATTAATTAATTTTAAA-
            ||||||||||||||||||||||||||| |||| ||  | ||| ||||  |  || |||||
p17D59      GTTGGGGTTAGTATAATTTTAATATTGTATATTTCATAAATGAATTTTTATATATTAAAT

                  80        90           100       110       120
PCR 900     -TATTGAATAAAAATATTTGAG-----TTATCAATTTCCCCTTTTCTTATTGGTTACTTT
             ||||  ||  |   | | |||     ||||   |||   ||||| |  ||  ||  | |
p17D59      TTATTATATTTATGGAATAGAGCCTATTTATGTTTTTATACTTTTATGCTT-CTTTGTAT

             130       140       150       160       170       180
PCR 900     AAATTTTAATTTGAGCTCACAATACCAAAAACCATAAATACAACTCAGATGGTTAGCAAG
            |  || ||| ||    |    || |  |||    |  ||||  |    |  | |  ||
p17D59      ATGTTATAAATTCTATTTTTTATTCATAAACATTTGGTTACATTTTTTTTTTTCATAAAT

               190       200       210       220       230       240
PCR 900     CA-AAAAGTTTATATACAT--CTACTATTATGAAATAATTAAAAATCAAATTTTAATGGA
             | | || ||  |||| ||   |||| || | | |  ||  |   | | ||   | |
p17D59      AATATAAATTACTATAAATAAATACTTTTTTAATAATATCTATTGTTAGATAAGATTTAT

                  250       260       270       280       290       300
PCR 900     GATGCTCATTTTAAGCGCATTCATCTATTTTCTATGCCCAATGAATTACCATTTTCCATA
             || |   | | || |   || || ||||
p17D59      AATACATGTATAAAACATCTTTAT-TATT
```

Alignment of the right end of the PCR 900 probe with the right side conserved sequence from p17D59

```
                  670       680       690       700       710       720
PCR 900     TTTCTTTGATAATAGAAGATGTATATAAATTTTTTGCTTGCTAATCAATCTGAGTTATAT
                 ||| |  |  | |  || ||||          | ||   | |||  || ||
p17D59      AGAAAATTTTAAAAATATTTTTTAATGAATTGAAATTGAAATTATTTTTTTGACATAGAT

                  730       740       750       760       770       780
PCR 900     TTATGGTTTTTGGTATTATGAGCTCAAATTAACATTTAAAGTAACCATAAGAAAAGGGGA
            || |   |||     | ||      | |  ||  |||| | | |  |   |   ||
p17D59      TTTTAAATTTCTCATTCATTTTTGAATAAAAAATTTTACATTCAAAAATCAAGTCGGTTT

                  790       800       810       820       830       840
PCR 900     AATTGATAACTCAAATATTTTTATTCAATATTTAAAATTTAATTTAACTATTAAAAATTA
              |  || || |   | |||||||  ||||||||||| |||||||| ||||||||||||
p17D59      CTTCTATGACCCTCCT-TTTTTATCAAATATTTAAAA-TTAATTTAATTATTAAAAATTA

                  850       860       870       880       890
PCR 900     AATTATGAAATATAAAATATCAAAATTATACTAACCCCAAAACTCAATGGGGAACCCG
            |||||  |||||||||||||| ||||||| ||||||||||||||||||||||||||
p17D59      AATTAAGAAATATAAAATATTAAAATTATGCTAACCCCAACTCAATGGGGAACCCGTTGG
```

Figure 4.4.

PCR 900 probed Southern blot of *ClaI* cut DNA. Each lane was loaded with 10 µg of DNA that had been digested with *ClaI*, purified macronuclear DNA from strain SB2044 (lane 1), purified micronuclear DNA from strain SB2046 (lane 2), and whole cell DNA from strain SB2040. The DNA was electrophoresed on a 1% agarose TAE gel. The gel was blotted to Nytran and probed with the PCR 900 fragment that had been ^{32}P labeled by the random primer method.

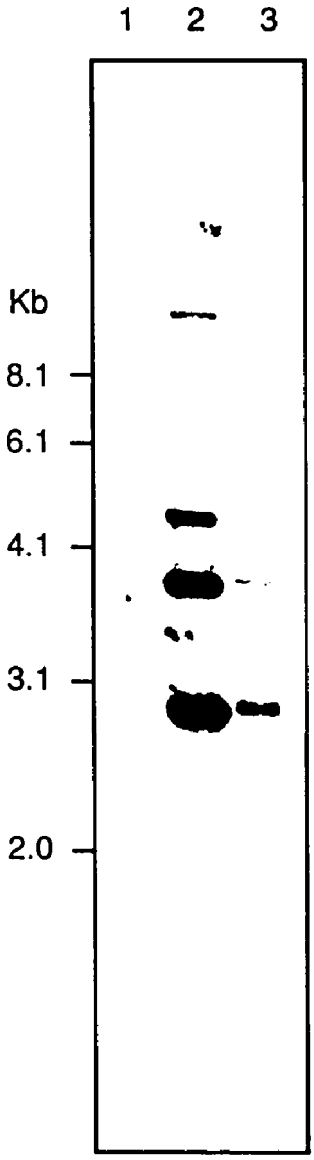

1
2
3
Kb
8.1
6.1
4.1
3.1
2.0

Figure 4.5.

Southern blot of *Hind* III cut micronuclear DNA probed with the PCR 900 fragment. 10 µg of micronuclear DNA from each strain, SB2046 (lane 1) and SB2040 stock .1 (lane 2), was digested with *Hind*III and electrophoresed on a 1% agarose TAE gel. The gel was blotted to nytran and probed with the PCR 900 fragment which had been ^{32}P labeled by the random primer method. Arrowheads to the right of lane 2 indicate PCR 900 hybridizing bands that differ from those in lane 1.

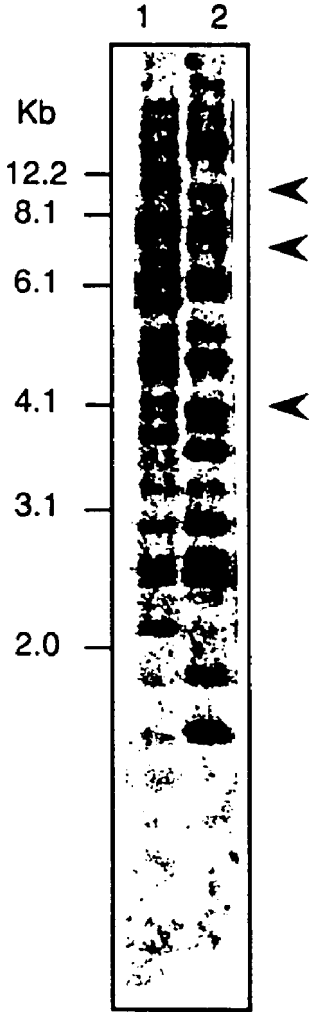

1
2
Kb
12.2
8.1
6.1
4.1
3.1
2.0

Figure 4.6.

PCR 900 probed Southern blot of whole cell DNA. 10 µg of whole cell DNA from strain SB20240.1 was digested with either *Xmn*I (lane 1) or *Bst*XI (lane 2) and electrophoresed on a 1% agarose TAE gel. The gel was blotted to Nytran and probed with the PCR 900 fragment which had been ^{32}P labeled by the random primer method.

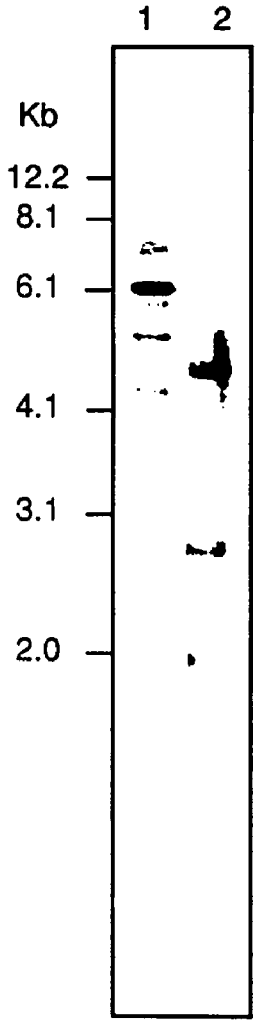

1
2
Kb
12.2
8.1
6.1
4.1
3.1
2.0

Figure 4.7.

Restriction maps of clones selected by hybridization to the PCR 900 probe. Clones containing micronuclear *Hin*dIII fragments from strain SB2040.1 were selected by colony hybridization to the PCR 900 probe. The restriction map for the inserts of 11 independent clones is shown. A bar at the bottom of the figure indicates 1 kb. Positions of *Hin*dIII (H), *Eco*RI (E), and *Bst*XI (B) were mapped. Restriction sites in brackets indicate that there is a site at one of the two positions.

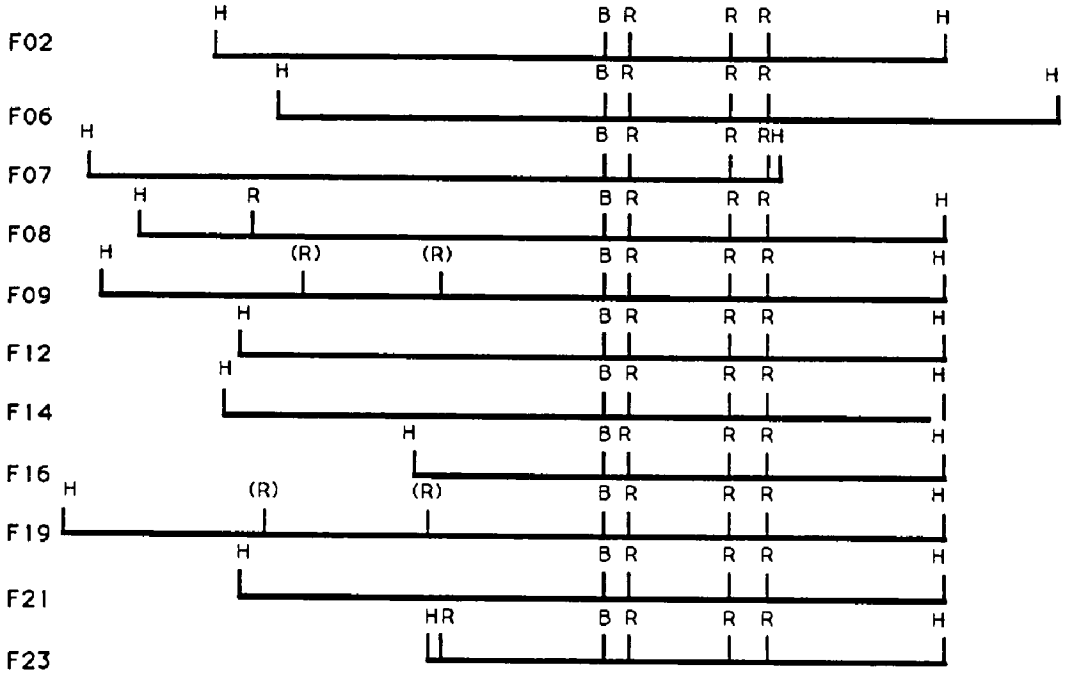

F02
F06
F07
F08
F09
F12
F14
F16
F19
F21
F23
H
B
R
(R)
HR
RH

Figure 4.8.

Southern blots of genomic DNA probed with common fragments from the PCR 900 selected clones. Purified micronuclear DNA (lane 1 in each paned) and macronuclear DNA (lane 2 in each panel) from strain SB2046 was digested with *Eco*RI (10 μg of DNA per lane) and electrophoresed on a 1% agarose TAE gel. The gels were blotted to Nytran and probed with either the 300 bp *Eco*RI fragment (panel A) or the 700 bp *Eco*RI fragment (panel B) from clone F08 which had been ^{32}P labeled by the random primer method.

(C) Conservation of F08 and F23 map in the genome. 10 μg of whole cell DNA from strain SB2040.2 was digested with *Xho*I and *Bst*XI and electrophoresed on a 1% agarose TAE gel. The gel was blotted to Nytran and probed with the *Eco*RI 300 bp fragment from F08 which had been ^{32}P labeled by the random primer method.

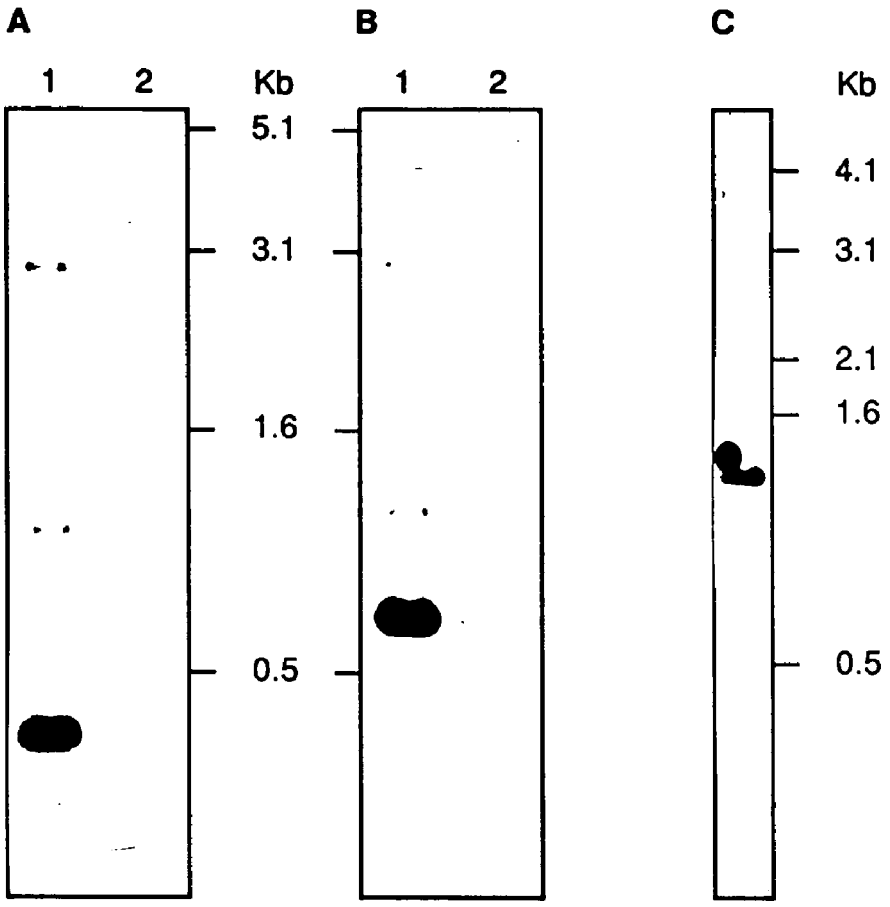

A
1
2
Kb
5.1
3.1
1.6
0.5
B
1
2
C
Kb
4.1
3.1
2.1
1.6
0.5

Figure 4.9.

Restriction maps of clones F08, F23 and D01. Clones F08 and F23 containing micronuclear DNA from strain SB2040.1 and clone D01 containing micronuclear DNA from strain SB2046 were selected as examples of conserved genomic elements designated Tmi. Restriction maps for the inserts from these clones are presented for *BstX*I (B), *Cla*I (C), *Bgl*II (G), *Eco*RI (E), *Hin*dIII (H), and *Xho*I (X). The scale bar at the bottom of the Figure indicates 1 kb. Prominent genomic fragments hybridizing to probes from these clones are shown below the maps. The *Eco*RI 300 and 700 bp probes and the *Eco*RI to *Hin*dIII 1.4 kb fragments used as probes are indicated below the F08 map. The 1 kb *Eco*RI to *Hin*dIII fragment used as a probe is indicated below the D01 map.

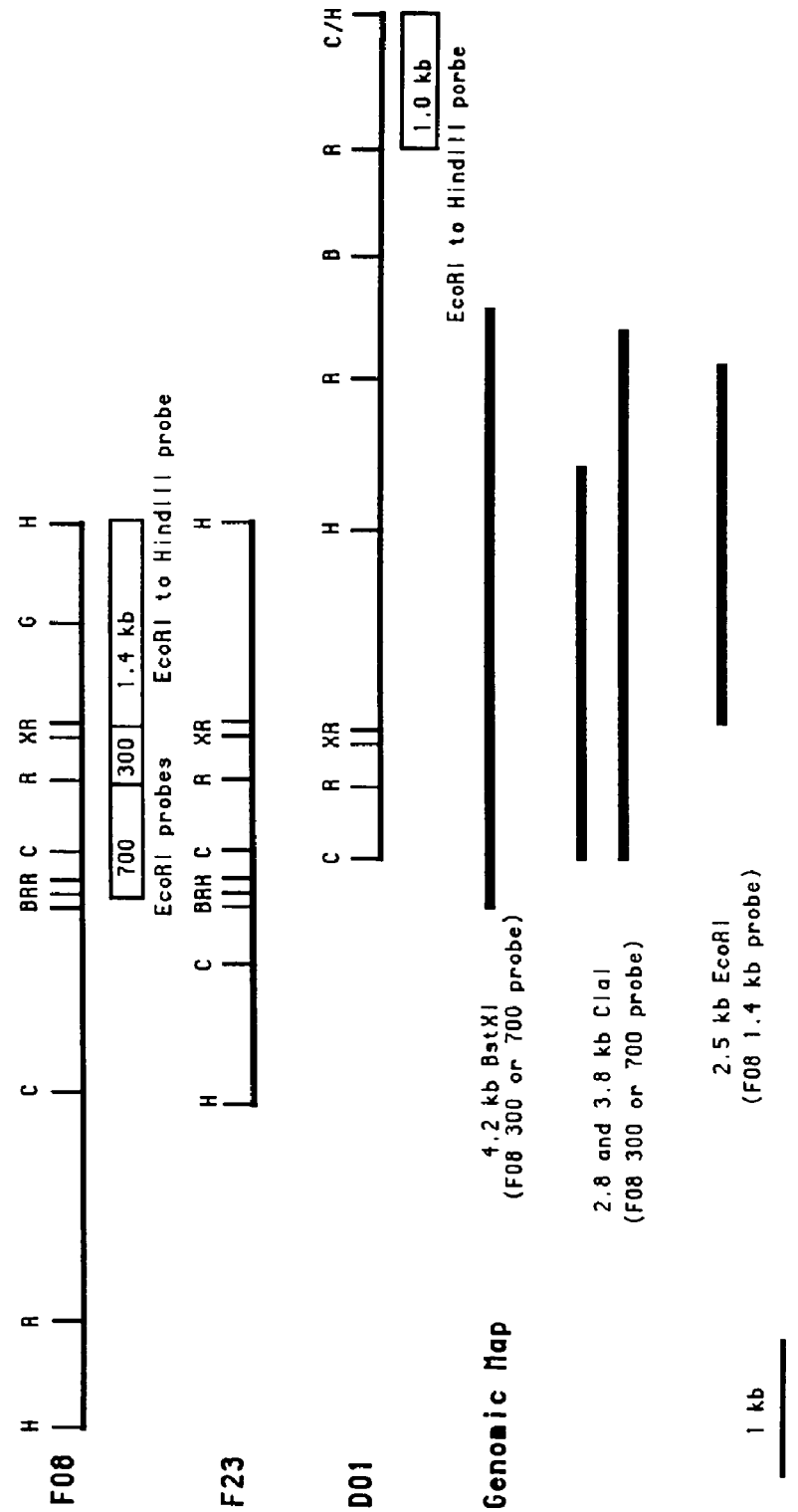
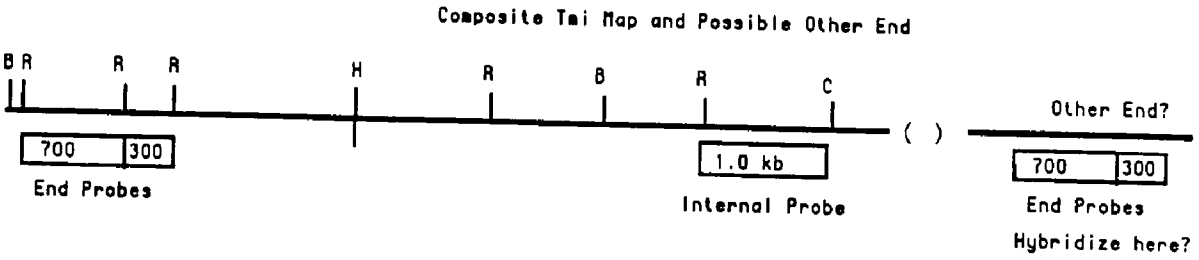

F08
H
R
C
BRH C
R XR
G
H
700
300
1.4 kb
EcoRI probes
EcoRI to HindIII probe
F23
H
C
BRH C
R XR
H
D01
C
R XR
H
R
B
R
C/H
1.0 kb
EcoRI to HindIII probe
Genomic Map
4.2 kb BstXI
(F08 300 or 700 probe)
2.8 and 3.8 kb ClaI
(F08 300 or 700 probe)
2.5 kb EcoRI
(F08 1.4 kb probe)
1 kb

Figure 4.10.

Hybridization of Tmi end and internal fragment probes to *Hin*dIII cut micronuclear DNA. The diagram represents a Tmi element and indicates which portion of the element the probes used would detect. The *Eco*RI 700 fragment would detect the end and the *Eco*RI 300 fragment is immediately internal to that. The D01 *Eco*RI to *Cla*I 1 kb fragment is distal from the end, separated form the other probes by a *Hin*dIII site predicted from the cloned examples, and presumable represents an internal position of the element. Purified micronuclear DNA from strain SB2046 was digested with *Hin*dIII and electrophoresed on a 1% agarose TAE gel, 10 μg DNA per lane. The gel was blotted to Nytran and separate strips of the blot were probed with either the *Eco*RI 700 fragment (lane 1), the *Eco*RI 300 fragment (lane 2) or the *Eco*RI to *Hin*dIII 1 kb fragment (lane 3). All probes were ^{32}P labeled by the random primer method.


Composite Tai Map and Possible Other End
B R
R
R
H
R
B
R
C
( )
Other End?
700
300
End Probes
1.0 kb
Internal Probe
700
300
End Probes
Hybridize here?

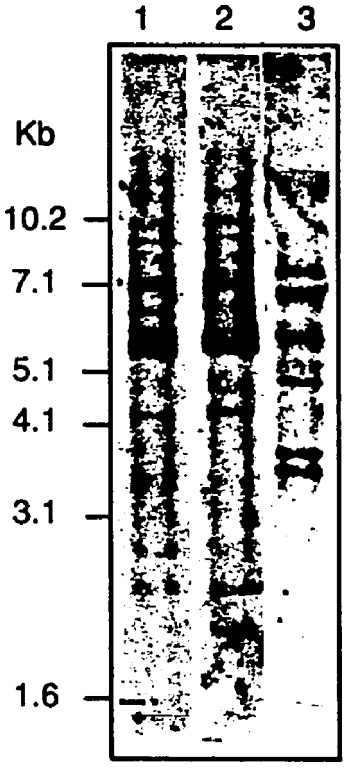
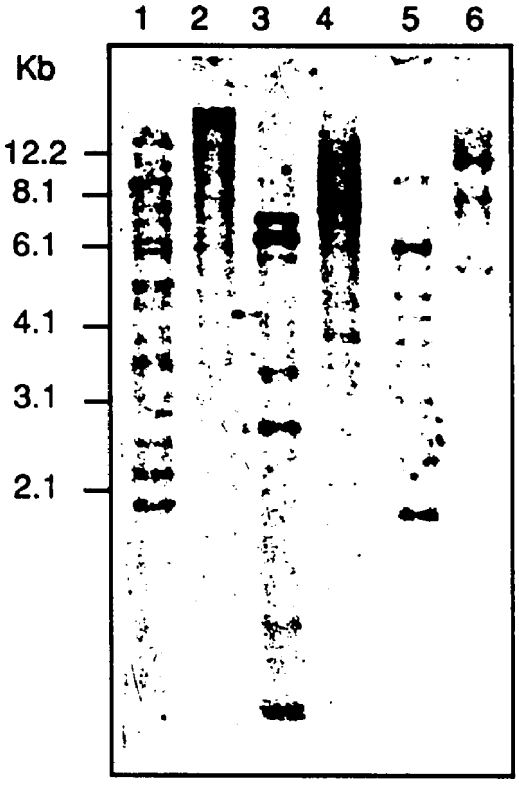

1
2
3
Kb
10.2
7.1
5.1
4.1
3.1
1.6

Figure 4.11.

Southern blot of genomic DNA from three Tetrahymena species probed with common Tmi fragments. Whole cell DNA from *Tetrahymena hegewischi* (lanes 1 and 2), *Tetrahymena malaccensis* (lanes 3 and 4) and *Tetrahymena pigmentosa* (lanes 5 and 6) was digested with *Eco*RI (lanes 1,3 and 5) or *Hin*dIII (lanes 2,4 and 6) and electrophoresed on a 1 % agarose TAE gel. There was approximately 10 μg of DNA per lane. The gel was blotted to Nytran and probed with a mixture of the *Eco*RI 300 and 700 bp fragments which had been ^{32}P labeled by the random primer method.


1 2 3 4 5 6
Kb
12.2
8.1
6.1
4.1
3.1
2.1

Chapter 5

Sequence Analysis of Tmi Elements

Introduction

The cloned portions of Tmi elements were sequenced in order to determine if they were similar to known mobile elements. One end of the Tmi element was defined and several open reading frames were found. Although a complete element was not sequenced all evidence suggests that Tmi elements are related to retroviruses or retrotransposons.

Mobile elements can be divided into two major types based on their mechanism of transposition and structure. Short inverted terminal repeats are characteristic of elements that move via DNA intermediates (P of Drosophila, Ac of maize, Tc1 of *C.elegans*, all reviewed in Berg and Howe, 1989). Long terminal direct repeats are characteristic of elements, included in the class called retro-elements, that move via RNA intermediates (retroviruses, Ty1 of yeast, copia of Drosophila, reviewed in Berg and Howe, 1989). In addition there are several elements which lack apparent terminal repeats (LINEs of mammals (Hutchinson, et al., 1989), CRE1 (Gabriel et al., 1990), ingi (Kimmel et al., 1987) and SLACs (Aksoy et al., 1987) of parasitic protozoans, and Jockey of Drosophila (Finnegan, 1989)). These resemble retro-elements in other ways. The available Tmi clones are unlikely to include both ends of the element, and as the sequence of only one end was determined, classification of Tmi elements based simply on the presence and type of repeated ends was not possible.

Global nucleic acid sequence comparison showed Tmi elements to be most similar to a retrotransposon and a retrovirus. On this basis, there were two regions of Tmi that were similar to regions of the Drosophila retrotransposon 17.6 and one region of Tmi was similar to the human retrovirus HIV2. The open reading frames were analyzed for their potential to code for proteins. Open reading frames were predicted based on the Tetrahymena genetic code, a variant of the universal genetic code that uses UAA and UAG as glutamine codons and not as stop codons (Martindale, 1989). One reading frame had potential transcription and translation control sequences. This open reading frame codes for a polypeptide that shares amino acid motifs with retroviral reverse transcriptases.

Results

Sequence of the Cloned Tmi Elements

Three of the recombinant plasmids containing portions of Tmi elements were selected for sequencing. Based on restriction mapping and hybridization to common genomic restriction fragments (presented in the previous chapter), plasmids F08 and F23 both included one end of Tmi and different flanking sequences. The restriction map of plasmid D01 indicated that 1.4 kb of the left side of the insert was the same as the right 1.4 kb of the F08 and F23 inserts. Hybridization to common genomic restriction fragments indicated that the D01 insert also contained additional internal portions of a Tmi element (see chapter 4 for details). Figure 1 shows the strategy used to sequence these clones. F08 and F23 were sequenced from the *Hin*dIII site at the right side of the insert through their common regions. About 100 base pairs into their divergent regions to the left of their *Bst*XI sites were also sequenced. One 100 bp section of F23, an *Eco*RI

fragment that was not recovered in a subclone although it was present in *Eco*RI digests of F23, was not sequenced. Clone D01 was sequenced rightward from the *Hin*dIII site it shared with F08 and F23 to the end of the insert. These three sequences provide a combined 6.3 kb of Tmi sequence. In most positions only one strand was sequenced, but most regions were sequenced more than once. All ambiguities were resolved by comparing the sequence in question on different sequencing gels, including one where the questionable bases could be read from an easily interpretable portion of a gel.

The sequences of F08 and F23 are aligned in Figure 2. To the right of their shared *Bst*XI site, the two sequences are almost identical, with occasional base changes, small deletions an additions at 75 of ~2700 positions. The Tmi end is shown in Figure 2 as the point where the two sequences diverge. This position is numbered 1 in the sequence and the numbering continues based on the F08 sequence. The Tmi sequence was extended through those regions represented in plasmid D01. Figure 3 shows the sequence from the *Hin*dIII site in D01 to the right end of the insert. Numbering of the sequence begins at 2784 to indicate that it is a continuation of the Tmi sequence from F08 in Figure 2.

Nucleic Acid Sequence Analysis of Tmi

A. The Terminal Region

While the long terminal direct repeats (LTRs) characteristic of retroviruses and retrotransposons do not share common primary sequence, they do usually have short inverted repeats at their ends (Varmus and Brown, 1989). These usually begin with TG and end with CA, but there are examples of LTRs with imperfect inverted repeats at the ends of their LTRs and LTRs that do not begin with TG and end with CA. Short repeats can be found in the Tmi sequence but without another end for comparison their significance is unknown. The Tmi end sequence shares some similarity with the

inverted repeat end of Drosophila P elements (Engels, 1989). The alignment is shown in Figure 4. The Tmi end is as similar to the P end as are the ends of other P-like elements to each other; P, hobo, 1723, Tam3, and Ac/Ds (Streck et al., 1986). No repeat of this end in inverted or direct orientation was found in the ~6,500 nucleotides of Tmi sequence determined.

Retroviruses and mobile elements that have a similar mode of replication have characteristic sites just inside their LTRs used to prime synthesis of the DNA copy of the element from the RNA copy. A tRNA binds to one end of the retrovirus or retrotransposon RNA copy and serves as a primer for DNA synthesis. Sequences complementary to tRNAs, including their terminal CCA, are found just inside one LTR. There are several occurrences of TGG within 1 kb of the Tmi end that could be the end of a tRNA primer binding site. The nucleic acid comparison searches performed did not reveal any significant similarity of Tmi to tRNAs in the database. However, the number of Tetrahymena tRNAs in the database (GenBank, June 1989) was not large (two from nuclear genes and six from mitochondrial genes). Therefore, analysis of the Tmi end region did not provide a conclusive assignment of Tmi elements to the retro-element or P-like element class.

B. Comparison of the Tmi Sequence to Nucleic Acid Databases

The Tmi sequence was compared in parts to the GenBank data base (June 1989) using the program fastn. It should be noted that the Tmi sequence is relatively A+T rich, as is the *T. thermophila* genome overall (75-80%). Some very A+T rich sequences from the database showed similarity to Tmi that was limited to dispersed matches of A's and T's in the sequences. Such alignments were discarded as insignificant, even if they had relatively high scores in a fastn comparison.

The part of the sequence from the Tmi end rightward to the *Hin*dIII site (Figure 2) showed significant similarity to the Drosophila retrotransposon 17.6 (Kugimiya, et al., 1983; Saigo, et al., 1984). The histogram of similarity scores for the invertebrate section of GenBank, in which the 17.6 retrotransposon sequence listed, is shown in Figure 5. The score for 17.6, at greater than 160 on the histogram, is clearly outside the normal distribution of scores. The Tmi sequence was compared to all sections of the Genbank database and the score for 17.6 was the highest overall score. The region of overlap, diagramed in Figure 6, between Tmi and 17.6 includes the 17.6 LTR. However the end of the 17.6 LTR did not coincide with the Tmi end in the alignment. The other half of Tmi, from the *Hin*dIII site rightward, also showed some similarity to the Drosophila 17.6 element (diagramed in Figure 6). While the comparison score (154, mean score was 52.4 with a standard deviation of 12.89) was not outside of the normal distribution of scores, the similarity is noted here because these elements showed two regions of similarity. In this second region, the overlap included a long stretch of A's that precedes an open reading frame (ORF) of Tmi (to be discussed below) and the gag gene of 17.6 (Saigo, et al., 1984). Both the Tmi ORF and the 17.6 gag gene were included in the overlap region.

The region of Tmi to the right of the *Hin*dIII site (sequenced from clone D01 as shown in Figure 4) also showed significant similarity to the human retrovirus HIV2 (Clavel et al., 1986; Guyader et al., 1987). The region of overlap (diagramed in Figure 6) included a predicted open reading frame from Tmi and the polymerase gene from HIV2. The histogram of scores comparing Tmi to the viral portion of GenBank is shown in Figure 7. The HIV2 score is outside the normal distribution of scores. The Tmi ORF present in the region of similarity between Tmi and HIV2 is the same Tmi ORF mentioned above in the region of overlap between 17.6 and Tmi. In conclusion, Tmi sequences are significantly and consistently similar to retro-elements.

C. Analysis of Open Reading Frames and Their Protein Coding Potential

T. thermophila, like many ciliates, use UAA and UAG for glutamine codons and not for stop codons. (Hanyu et al., 1986; Horowitz and Gorovsky, 1985; Martindale, 1989). Allowing only UGA as a stop codon the entire 6.3 kb or Tmi sequence was searched for open reading frames. The ORFs were based on the F08 sequence. The differences between the F08 and F23 sequences did not significantly altered the predicted ORFs. The F23 sequence predicted amino acid differences at some positions but did not change the length of the ORFs. There are five ORFs of 200 amino acids or more. They all occur in the same orientation for transcription from left to right in the Tmi sequence. There were no significant ORFs in the opposite orientation. The position and transcriptional orientation of these ORFs are diagramed in Figure 8. The deduced amino acid sequence of peptides encoded by these ORFs is shown in Figure 9.

Certain sequence characteristics distinguish DNA regions likely to code for proteins from those fortuitously lacking stop codons. It has been noted that coding frames have distinct patterns of triplets relative to non-coding frames. The over representation of RNY triplets over YNR triplets (R=purine, N=any base, Y=pyrimidine) is predictive of coding frames (Shepherd, 1981). Coding frames typically have a RNY:YNR ratio of 2:1 (Shepherd, 1981). Analysis of triplet frequency based on these ratios has been helpful in identifying the proper coding frames for genes in Paramecium (Preer et al., 1985). The abundance of RNY and YNR triplets in the five Tmi ORFs and the +1 and +2 frames from the same regions were determined and are presented in Table 1. Both ORF1 and ORF3 had patterns of nucleotide triplets consistent with the hypothesis that they are true coding reading frames.

The *T. thermophila* genome is in general A+T rich (75-80%). The coding regions of genes so far sequenced have a lower A+T content of 51-70% (compiled in Martindale, 1989). The data available for Tetrahymena are almost exclusively from genes coding for very abundant proteins (e.g. actins and histones). Of the sequenced genes, *cnj*B, the one with the highest A+T content (70%) encodes a protein that is under developmental control and only expressed for a brief period of time during conjugation (Martindale and Bruns, 1983). Therefore, the G+C richness of genes coding for abundant proteins may not reflect the G+C content of Tetrahymena genes in general.

The A+T content of the Tmi ORFs is: 68% for ORF1, 75% for ORF 2, 79% for ORF3, 80% for ORF4 and 80% for ORF5. Only ORF1 is within the range of A+T content for known Tetrahymena coding sequences. The relatively low A+T content of ORF1 does supports the hypothesis that it is a protein coding sequence. The Tmi elements are all limited to the micronucleus of the strains investigated. The micronucleus is largely transcriptionally inactive. There is one report of transcription of a micronuclear limited region (Stein-Gavens, et al., 1987). The sequence of the transcribed micronuclear region was not reported. The rules for coding sequences from the micronucleus are not known and may be different from those of highly expressed macronuclear genes. The observation that these micronuclear limited ORFs in Tmi do not follow the norm for macronuclear coding sequences does not eliminate the possibility that they code for proteins under some circumstances. In addition the particular Tmi copy sequenced may not now be expressed. Without the selective pressure to produce functional messages the sequence may be drifting toward A+T richness.

There are limited data on consensus promoter elements and translational start sequences from Tetrahymena. Sequence comparison from 61 ciliate genes reveals a

translation initiation consensus of AAAATGG (Brunk and Sadler, 1990) Analysis of the Histone H3 and H4 intergenic regions from 22 Tetrahymena species identified only a single putative promoter element, TATCCAATCARA, located 70-170 nucleotides upstream of the start codon. It was also noted that the intergenic region had a high A+T content, biased such that within 100 nucleotides of the start codon the coding strand was up to 75% A (Brunk and Sadler, 1990). Of the Tmi ORFs only ORF3 had sequence elements similar to these at its beginning. There is a sequence 100 nucleotides upstream from the initiation codon that matches the promoter element at 8 of twelve positions (**TTTCCAAA**TC**AA**, underlined in Figure 9C). The start codon in ORF3 also lies in a region similar to the consensus translation start (AAAATGT, only the last T being different from the consensus). In addition this start codon is preceded on the coding strand by an A rich sequence including a stretch of 20 consecutive A's. Although the overall A+T content of ORF3 is not within the range of known Tetrahymena genes the presence of these transcriptional and translational control consensus elements support the hypothesis that ORF3 could function as a gene in Tetrahymena.

The codon usage of the five Tmi ORFs was tabulated and compared to the codon usage of Tetrahymena genes known to code for proteins. The codon usage compiled from 13 Tetrahymena genes (Martindale, 1989) is compared to the codon usage for the five Tmi ORFs in Table 2. It can be seen that the codon usage for the Tmi ORFs sometimes reflects the same bias as Tetrahymena genes in general. In other cases the codon usage of the Tmi ORF's reflects the A+T richness of the sequence. The data for codon usage in Tetrahymena genes was compiled mostly from genes, 12 of the 13, coding for abundant proteins. Presumably these genes are under selective pressure to maintain codons matching abundant tRNAs in order to assure adequate protein expression. One gene included in the compiled totals for codon usage did not share the same codon bias as the genes coding for abundant proteins. This is the *cnjB* gene (encoding a developmentally

controlled gene product that is only expressed for a brief time during conjugation, (Martindale and Bruns, 1983; Martindale and Taylor, 1988)). The codon usage for *cnjB* is also listed separately in Table 2. The codon usage of the Tmi ORFs are similar to that of *cnjB*. The similarity to *cnjB* can be most easily seen by comparing the Asn, Lys, Tyr and Pro codons. The codon usage of the Tmi ORFs implies that they do not encode abundantly expressed proteins. Because these ORFs are limited to the micronucleus they are unlikely to produce abundant protein products. The codon usage of the Tmi ORFs is similar to at least one Tetrahymena gene.

Amino Acid sequence analysis

The deduced amino acid sequences of the polypeptides encoded by the Tmi ORFs (see Figure 9) were compared to the NBRF-PIR database (June 1989 release) using the fasta program. No striking similarities to amino acid sequences in the data base were produced. However, OFR3 did show fairly high scores for a DNA-directed RNA polymerase (83, mean score 28 with a standard deviation of 7.88) and a retroviral related reverse transcriptase (83, mean score 28 with a standard deviation of 7.88). These were not the highest scores but are mentioned here because they showed similarity to proteins with functions expected to be relevant to a retro-element. In addition, the amino acid sequences were compared to the GenBank database using the tfastn. No striking similarities between the deduced Tmi polypeptides and sequences in the database were seen. However, it is possible to miss similar features of proteins by this sort of global amino acid sequence comparison.

An amino acid motif common to reverse transcriptases was noticed in the deduced amino acid sequence from ORF3. The motif YXDD (Kamer and Argos, 1984), here YYDD, is underlined in Figure 9C. This is the most constant of four amino acid motifs

conserved among reverse transcriptases so far sequenced (Poch et al., 1989). Using the alignment of reverse transcriptases and their shared motifs presented in Poch et al. (1989), Lundblad and Blackburn (1990) deduced a consensus for the four motifs. There are reasonable matches to all four of these motifs in the deduced ORF3 polypeptide. A sample of reverse transcriptases and RNA directed RNA polymerases (references for the sequences can be found in, Poch et al., 1989) are aligned with the ORF3 polypeptide in Figure 10. Stars above the sequence indicate identity between amino acids of ORF3 and the consensus motifs. Amino acid residues among the other RNA dependent polymerases that are identical to those in the ORF3 polypeptide are underlined. A quantitative form of this comparison is presented in Table 3. The ORF3 polypeptide matches the consensus fairly well and at least as well as some of the known reverse transcriptases. The ORF4 polypeptide also has a YXDD motif but has no similarity to the other reverse transcriptase motifs. The definition of the four motifs is discriminating enough not to include by chance all amino acid sequences with a YXDD motif. In conclusion, it is likely that the Tmi ORF3 codes for a reverse transcriptase-like enzyme or did so in a functional version of the element.

Tmi ORFs 1,2,4 and 5 were examined for similarity to other retroviral proteins. The organization of the genome of a typical retrovirus or retrotransposon suggests that ORF1 or 2 would likely be similar to the gag polypeptide and ORF4 or 5 would be similar to either the RNase H or integrase domain of the *pol* gene product. Extensive amino acid motifs conservation has not been defined for the gag, RNase H and integrase polypeptides. The hallmark of gag polypeptides is a Cys-His motif (Covey, 1986) commonly, $CX_2CX_4HX_4C$ (Xiong and Eickbush, 1988). Neither ORF1 or 2 coded for such a motif. Although ORF1 coded for more cysteine and histidine residues than the other ORFs, they did not occur in a pattern typical of known Cys-His repeats. Comparison of the RNase H domains of several retroviral polypeptides and *E.coli* RNase H has shown

some shared amino acid sequences (Johnson et al., 1986). There is only limited similarity between ORF4 and one of the regions of amino acid sequence typical of RNase H. There is no apparent similarity between the ORF4 or 5 amino acid sequences and RNase H. There are no well defined integrase motifs beyond a presumed DNA binding Cys-His repeat of the type that can form a zinc binding domain (Johnson, et al., 1986). No Cys-His repeats occur in either the ORF4 or 5 predicted polypeptides. Therefore, aside from ORF3, the Tmi ORFs do not encode polypeptides with identifiable similarity to retroviral proteins.

Discussion

The accumulated sequence analysis presented here consistently suggests that Tmi elements are related to retro-elements. The nucleic acid sequence of Tmi was similar to the Drosophila retrotransposon 17.6 in two regions. Tmi also showed sequence similarity to the human retrovirus HIV2. The amino acid sequence of Tmi ORF3 polypeptide contained motifs similar to those of reverse transcriptases.

For the purpose of this discussion retro-elements are defined, using the terminology of Rogers (1985), as including retroviruses, retrotransposons (not viruses but possessing LTRs) and retroposons (apparently mobile non LTR elements). All of these share a method of dispersal that involves making an RNA copy of the element, reverse transcribing it into DNA, and inserting the DNA copy into the host genome. Most retroviruses encode gene products to accomplish reverse transcription, integration and production of viral particles. Other retro-elements lack gene products to accomplish one or more of the these reactions. Presumably the less autonomous elements rely more heavily on host functions to help in their replication and recombination. For

instance the only common feature of LINEs and related retroposons is a capacity to encode a reverse transcriptase-like protein (Aksoy, et al., 1987; Gabriel, et al., 1990; Hutchinson, et al., 1989; Xiong and Eickbush, 1988). Although a complete Tmi has not been defined, the sequenced portion shares with retroposons the potential to code for a reverse transcriptase-related protein.

The predicted ORF3 polypeptide is similar to reverse transcriptases in all four of the defined motifs. The predicted polypeptide is only about 200 amino acids and thus about 22 kD for an average amino acid composition. As known reverse transcriptases are between 60 and 95 kD (Varmus and Brown, 1989), the ORF3 polypeptide is probably too small to be a reverse transcriptase. However, retroviral genes are often expressed by frame shifts during translation (Varmus and Brown, 1989). Translational frame shifting could result in much larger proteins produced from the Tmi ORFs. There is no evidence suggesting that the Tmi copy cloned is functional. If it is a non-functional element, it may have accumulated mutations resulting in stop codons that shortening the predicted polypeptide. The additive evidence, including the presence of consensus transcriptional and translational sequences and the RNY:YNR triplet ratio, implies that ORF3 is a protein coding sequence.

Many other retroposons (CRE1 (Gabriel, et al., 1990), SLACs (Aksoy, et al., 1987),and R1Bm (Xiong and Eickbush, 1988)) have additional open reading frames that code for polypeptides similar to retroviral proteins. It is not known if proteins are produced from these ORFs or if they are necessary for movement of these elements. The polypeptide predicted from ORF1 did not show amino acid similarity to known retroviral proteins. The criteria for amino acid similarity were not as well defined for these retroviral proteins compared to those for reverse transcriptases. It is possible that a protein that is related but does not share large segments or amino acid similarity

would not be recognized. Both the G+C content and the RNY:YNR triplet ratio imply that ORF1 is a protein coding sequence or the remnant of a protein coding sequence.

The work described here is the first identification of a retro-element in *T. thermophila*. It is unlikley that *T. thermophila* could have easily acquired retro-elements from other species. In order for retro-elements in *T. thermophila* to be mobile they need to be able to express their gene products. Particularly, element functions must be encoded in the same variation of the universal genetic code used by *T. thermophila*, where UAA and UAG are not termination codons but glutamine codons. This suggests that the introduction of retro-elements into *T. thermophila* either preceded the drift in the genetic code or was accompanied by rapid mutation of element coding sequences. In the first case, the coding potential of retro-elements present in the genome before the drift in the genetic code would have evolved concomitantly with *T. thermophila* genes in general. Alternatively, retroposons could have evolved from within the *T. thermophila* genome if segments of DNA acquired the ability to copy and disseminate themselves. However recently introduced retroviruses or retrotransposons would need to have accumulated mutations rapidly so that their gene products could be efficiently expressed in *T. thermophila*.

Nucleic acid sequence comparison also indicated that Tmi elements are similar to retro-elements. The left LTR of 17.6 (Kugimiya, et al., 1983; Saigo, et al., 1984) overlaps Tmi in a region that begins 300 nucleotides in from the defined Tmi end (diagramed in Figure 6). In the other region of overlap between Tmi and 17.6 the gag gene of 17.6 is juxtaposed with ORF3 of Tmi. The amino acid sequence derived from ORF3 does not resemble a gag gene product and the nucleic acid sequences of the two coding regions are not strikingly similar. The similarity in this region is strongest in sequences preceding the open reading frames. However, the relative orientation of these two

elements is maintained in these two regions of alignment. Tmi similarity with HIV2 (Clavel, et al., 1986; Guyader, et al., 1987) overlaps ORF3 with part of the HIV2 pol gene. The ORF3 polypeptide does resemble reverse transcriptases, one of the functions encoded by retroviral pol genes. The overlapped region of the pol gene may encode the carboxy terminal portion of the reverse transcriptase or the amino terminal portion of the integrase function. The nucleotide position of the boundry between these two functions was not specified in the reference to the sequence (Guyader, et al., 1987). However, assuming an average size of 80 kd for reverse transcriptase its coding sequence would overlap ORF3 in the area of similarity as shown in Figure 6. It is possible that the different retroviral genes arose by duplication and diversification. The sequence similarities noted here could reflect such an ancestral relationship. Aside from functional relationships between the regions of overlap, their linear arrangement suggests a polarity for the sequenced Tmi element. This orientation indicates that the Tmi end sequences is the left end of an element and copying or transcription of the element proceeds to the right. All of the Tmi ORFs are also oriented in this direction.

Tmi sequences were found only in the micronuclear genome of the strains tested. The micronucleus functions as the germline in *T. thermophila*, and it is not generally transcriptionally active. Sequestering potentially mobile elements in the micronucleus could be advantageous for the element, the cell, or both. It is an advantage for Tmi elements that they remain in the germline. This assures that they will be transferred to progeny and maintained in the population. Conversely, the presence of mobile genetic elements in the macronucleus could be deleterious to *T. thermophila*. A micronuclear element that is not eliminated during macronuclear development would be expected to be amplified to at least ~45C as are all other macronuclear destined sequences. Once in the macronucleus, which is

transcriptionally active, protein coding sequences in the element could also be expressed. The presence of Tmi encoded enzymes such as reverse transcriptase could disrupt cellular functions. The reverse transcriptases of retroviruses, Ty1 of yeast and copia of Drosophila are all sequestered in virus or virus-like particles (Varmus and Brown, 1989; Garfinkel, et al., 1985; Shiba and Siago, 1983). It has been suggested that unsequestered reverse transcriptase copying other cellular mRNAs is deleterious to cells (Garfinkel, et al., 1985). The inability to detect Tmi elements in macronuclear DNA could be due to a selective disadvantage to cells harboring macronuclear elements.

Attempts to identify RNA messages homologous to Tmi probes were unsuccessful. RNA was isolated from vegetatively growing and starved cells. The latter are competent for mating. No Tmi homologous RNA was detected in either total RNA or poly A^+ selected RNA. Control experiments indicated that the RNA samples used were not degraded as intact abundant messages were easily detected in these samples (data not shown). *T. thermophila* cell extracts that were active in assays for a DNA processing events associated with macronuclear development (Robinson et al., 1989) showed no reverse transcriptase activity (assayed in the same manner as AMV reverse transcriptase using the protocol of the supplier, IBI). Given the micronuclear location of the Tmi coding sequences, their expression may be linked to function of the micronucleus. The only micronuclear sequence that has been shown to be expressed is only detected from starved cells (Stein-Gavens, et al., 1987). It has been suggested that micronuclear specific sequences could be important for micronuclear functions, including mating and meiosis (Martindale and Bruns, 1983; Rogers and Karrer, 1989). They would likely be expressed during meiosis and or mating. It will be necessary to look carefully for Tmi homologous RNAs during a mating time course in order to determine if these sequences are transcribed.

The DNA elimination events in ciliates resemble transposon excision in some ways. DNA sequences to be eliminated are bounded by short direct repeats (Austerberry and Yao, 1987; Austerberry and Yao, 1988; Ribas-Aparicio, et al., 1987). Tmi elements described here were all embedded in micronuclear limited DNA. Their ends did not coincide with sites of DNA breakage and rejoining that occur during macronuclear differentiation. Hence Tmi elements may be removed from the macronucleus passively along with their surrounding sequences. In contrast, micronuclear transposon like elements are precisely excised in two other ciliates. Tec1 elements of *Euplotes crassus* are precisely and completely removed during macronuclear development (Jahn, et al., 1989). In addition in *Oxytricha fallax* some TBE1 elements are also removed during macronuclear development (Hunter, et al., 1989). These are examples of P-like transposable elements with inverted repeats at their ends surrounded by target site direct duplications.

There are possible roles for retro-elements in macronuclear development. One model for macronuclear differentiation (Ribas-Aparicio, et al., 1987) involves reverse transcription of spliced micronuclear RNA transcripts to produce macronuclear DNA molecules. While Tmi elements themselves may not be the DNA targets for the rearrangement reactions of macronuclear development, mobile elements do encode proteins involved in nucleic acid synthesis and recombination reactions that could play a part in macronuclear development. Proposing a specific role for Tmi in macronuclear development awaits information concerning the expression and function of Tmi gene products.

Table 5.1. Triplet ratios of the Tmi ORFs and the corresponding +1 and +2 reading frames.

	RNY	YNR
ORF 1	**127**	**72**
+1	101	118
+2	92	112
ORF 2	70	68
+1	64	91
+2	79	61
ORF 3	**92**	**53**
+1	67	70
+2	50	85
ORF 4	88	71
+1	72	98
+2	92	83
ORF 5	68	56
+1	53	61
+2	65	68

Table 5.2. Codon usage of the Tmi ORFs compared to that compiled from the sequences of other Tetrahymena genes.

Amino Acid	Codon	Tet [1]	orf1	orf2	orf3	orf4	orf5	cnjB [1]
Ala	GCT	139	5	1	1	6	2	11
	GCC	68	0	0	1	0	1	3
	GCA	9	5	4	3	3	2	5
	GCG	0	1	0	0	0	0	0
Arg	CGT	7	0	1	1	2	0	0
	CGC	0	0	0	1	0	0	0
	CGA	0	2	2	0	0	0	0
	CGG	0	2	1	0	0	0	0
	AGA	148	14	3	2	5	9	18
	AGG	4	0	4	1	1	0	2
Asn	AAT	41	29	22	28	26	19	24
	ACC	80	13	11	8	4	1	17
Asp	GAT	87	16	2	19	13	13	12
	GAC	40	6	4	1	1	3	7
Cys	TGT	11	4	4	1	1	5	5
	TGC	32	6	1	1	1	0	8
Gln	CAA	41	15	15	9	6	5	11
	CAG	5	4	0	0	1	3	3
	TAA	33	6	9	15	17	14	19
	TAG	19	5	1	0	4	1	14
Glu	GAA	152	18	2	11	15	11	31
	GAG	16	3	0	1	3	1	10
Gly	GGT	140	4	0	4	1	2	1
	GGC	9	0	0	0	1	0	1
	GGA	14	7	1	6	5	3	4
	GGG	1	2	0	0	0	0	0
His	CAT	14	5	4	1	3	1	8
	CAC	42	4	0	2	1	2	2
Ile	ATT	91	14	15	15	20	12	30
	ATC	86	10	7	5	3	1	13
	ATA	16	9	23	12	22	12	14
Leu	TTA	49	12	12	15	17	20	14
	TTG	53	11	2	1	4	1	10
	CTT	44	2	5	1	2	2	16
	CTC	50	6	3	3	1	3	5
	CTA	9	5	6	22	6	2	8
	CTG	4	0	4	0	2	0	3
Lys	AAA	58	26	16	34	41	27	22
	AAG	235	12	2	1	3	5	15
Phe	TTT	27	5	20	18	34	14	17
	TTC	85	13	8	5	6	5	9
Pro	CCT	60	3	0	2	3	1	12
	CCC	59	2	1	0	0	0	0
	CCA	8	3	3	2	3	3	5
	CCG	2	1	2	0	1	0	0
Ser	TCT	83	1	6	6	3	1	16
	TCC	69	1	7	0	1	0	2
	TCA	21	10	8	8	9	4	8
	TCG	1	5	1	1	0	3	0
	AGT	10	4	4	0	2	1	4
	AGC	16	2	1	1	3	0	12
Thr	ACT	64	8	5	2	4	7	7
	ACC	69	2	1	2	0	0	2
	ACA	9	10	10	6	10	6	4
	ACG	1	2	2	1	2	0	1
Tyr	TAT	46	13	5	15	18	12	25
	TAC	57	18	8	5	5	4	10
Val	GTT	79	1	3	3	3	6	12
	GTC	77	3	0	2	1	0	1
	GTA	11	8	4	1	1	2	8
	GTG	3	2	1	1	1	0	3
Met	ATG	61	20	4	6	10	4	10
Trp	TGG	23	0	1	0	1	0	4

1) Martindale, 1989.

Table 5.3. Reverse transcriptase motif comparison.

Matches Seqence	out of 8 Motif A		out of 16 Motif B		out of 8 Motif C		out of 5 Motif D
ORF 2	5	(46)	6	(14)	3	(11)	3
HepB	8	(71)	8	(10)	8	(19)	4
HepBWo	8	(69)	6	(10)	8	(19)	4
HepDu	7	(17)	9	(9)	7	(19)	5
HIV 2	8	(24)	15	(10)	8	(20)	4
17.6	7	(17)	8	(6)	7	(19)	4
copia	6	(25)	4	(21)	8	(7)	2
Ingi	6	(42)	10	(9)	6	(25)	3
BNYVV	3	(40)	9	(11)	4	(12)	1
AaMV	5	(42)	7	(13)	5	(11)	2
TYMV	2	(36)	5	(10)	5	(10)	2

The amino acid motifs are as shown in Figure 10. The numbers in brackets between the columns are the number of amino acid residues between the motifs. The references for the reverse transcriptases can be found in Poch et al., (1989). The RNA dependent DNA polymerases listed are; HepB, Hepatitis B human; HepWo, Woodchuck hepatitis B; HepBDu, Duck hepatitis B; HIV2, Human immunodeficiency type 2; 17.6, Drosophila 17.6 element; copia, Drosophila copia element; ingi, Trypanosoma ingi element. The RNA dependent RNA polymerases listed are; BNYVV, Beet necrotic yellow vein virus; AaMV, Alfalfa mosaic virus; TYMV, Turnip yellow mosaic virus.

Figure 5.1.

Sequencing strategy for the three Tmi containing clones. The regions sequenced are indicated by an arrow below the maps of clone F08, F23 and D01. The direction of the arrow indicates the direction of sequencing.

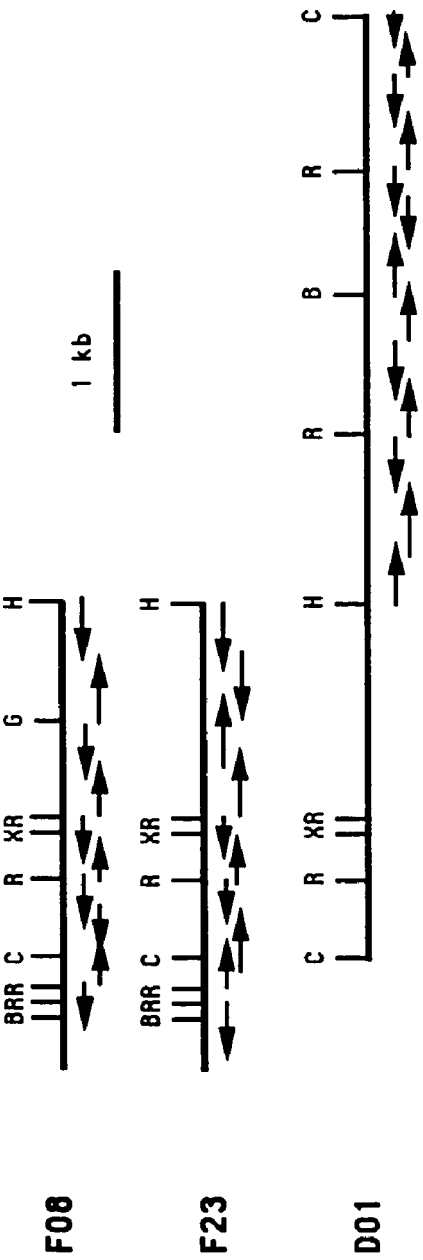

1 kb
F08
F23
D01
BRR C R XR G H
BRR C R XR H
C R XR H R B R C

Figure 5.2.

F08 and F23 sequences aligned. The regions of these clones sequenced were aligned. The top line of sequence is from clone F08 and the bottom line is from clone F23. The end of the Tmi element was defined as the point where the sequences diverged and is numbered one above the sequence. Numbering of the sequence continues from there based on the F08 sequence. Mismatches between the two sequences are indicated by stars. The N's in the F23 sequence indicate the position of an *Eco*RI fragment that is present in the clone but was not recovered as a subclone for sequencing.

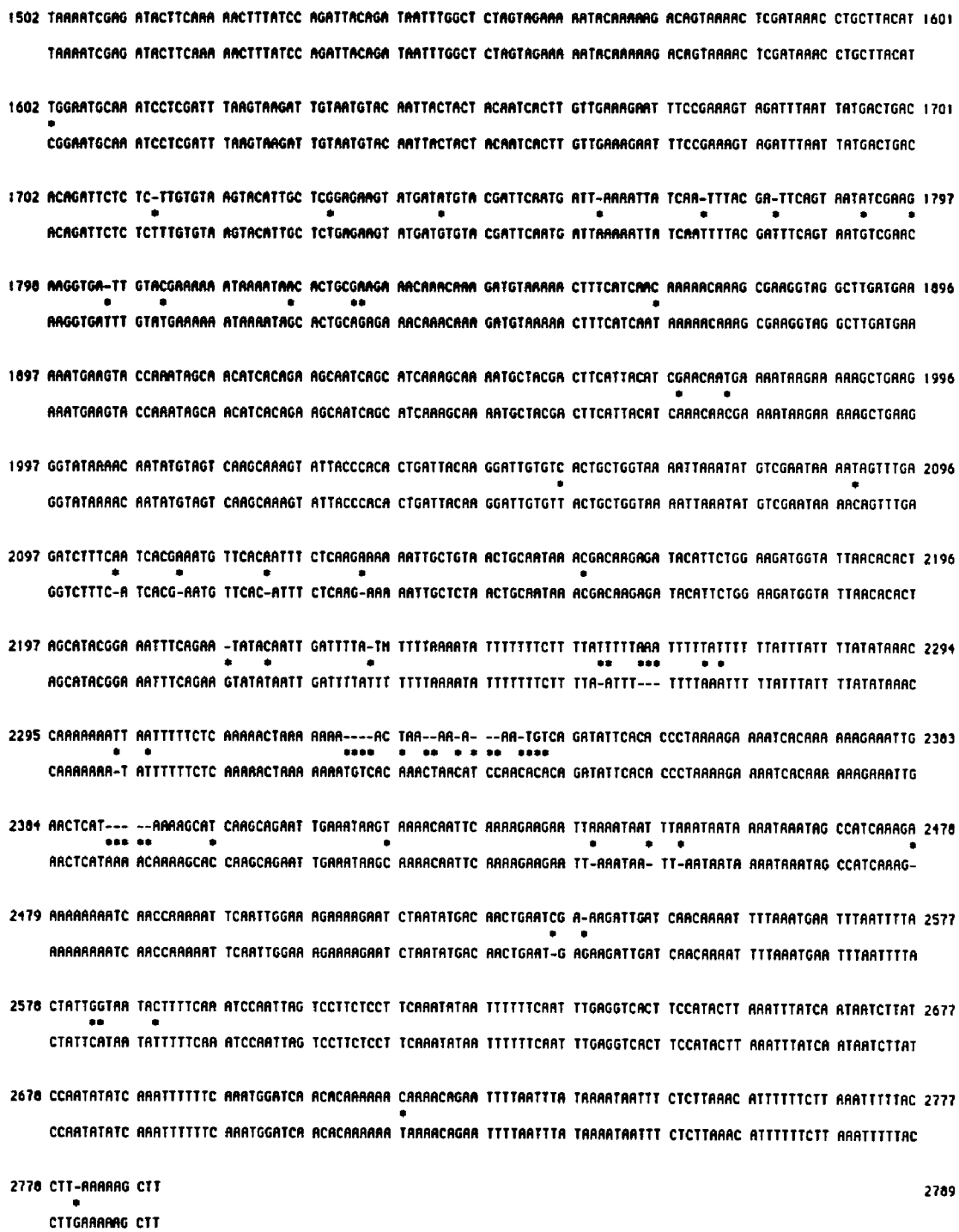


Figure 5.3.

Tmi sequence extended through clone D01. The sequence from clone D01 is presented numbered as a continuation of the Tmi sequence from clone F08 in Figure 2.

2701 AAGCTTTGTT TTAAACTTTT TACTGTACTG TTTACTTAGA AATTACATTC GACAACGGGT ATTAAGGCTA AATCGTTGCA CAATCATCGC AGCAAAGCCA 2800

2801 TTCCTCATAA TAAGACAAAT TGTAACTATT ACAAAATTCA CAAATAATCA TTTTTTTAAA ATATTTATTT TTAATTGTTG TATATATTAA CAAAAAATAA 2900

2901 TTTTAATGTC TGACATCAAT TACATTTTCA TAGACAGTTC CATACTAAAC GTGCCGAATA ATTACAATTT TCGAGTTATT CTCAACGAAA AGATAAAAAT 3000

3001 AAAAAAATAC ATAAAACTAA TTCAAGCAAA TATACCATTT GACGATTACC TAATCGACAA TTACATAACA CATTTTACAT AAATGGCAAA ATGTACACAA 3100

3101 TAACAAACGG TATTTACGAT ATCCAGTCTG ATAAATCAAA TGTAATAAAT AATCAGGACA GATATTTCAT CTTTTTCGAT TACTCTGTCT CAACTTACTT 3200

3201 ATAGGATTAC ATTTCAAAAT ACTCAAAATT TTTCACTGAA CCTAAACAGA GTTTATCAAA ATATAGGATC ACCGATACAA ATTACTCAAC AATAATCAAT 3300

3301 ACATAACAAC ATAAACTCCT CTCATCAATC TCCCTCAAGT AGTGCTCTTA AATATTGAAG AAATCCAGAT ATTCAAATCG TTCAATCTCA AAAATATTCT 3400

3401 GTCGACTACT CATTTCTCAT TCTCAATACC GTCAATAGGG GATCAAACAT TCAATATTAA AATCAATTTG GTGAAAATAA AGTGGATTGT TCTTGCGATT 3500

3501 TACGCCAAAT GACGATATAA TTATATAAAA ACAATGGATA CTTTTTCCAA ATCAACTCAG AATTCAACTC TTCGATATCA ATGAAAAATT TTTATTTTTT 3600

3601 ATCAAATTAT ATAACTAAAA AAAAAAAAAA AAAAAATCAA AAAAATTTAA AAATGTAATA TTTATAAAAA TAATCAGATC CTTTTACAAA ACCTTAATTT 3700

3701 TATTTTGATG GTGCTCGTAT CTTGGATGGA TTAGCACTTT TTTATCATAA AAAAATAAAT ATAGATGAAA ATTTTAAAGA TTCATTAGAA TTAATTTATA 3800

3801 ATACAAATGA TGAAAAACTA AAAGATATTA AAAATATTTA AATATTTAAA GAAGAAATAG GTACAGATTT CATTGAAAAA CAAATGTTCA GATAATTTAC 3900

3901 TGATAACAAT ATAGTTGGAG AAAATGAAAT AAAAAATATT TTTTAAAAAA ACCTCTACAC AGATATATAA AAATAAAATA ATTTATTAGC ATCTATATTT 4000

4001 TTATATTTTA TTCACAAGTT CTACTATGAT GATGAGAATA GCGATGATAT CAATAATATT AAAAATTTTA AAAATTCAAA AGATATGATT TAAATTTATA 4100

4101 AAAATATGAA AACAAACGGA CTACCATATA CCAAATTTAT CAATAAAATT AAATCTGTTT ACACAGATAT TCACAACTTA TTAGTAGAAA TATTATAATT 4200

4201 AAAATATGGT TAATAAGCAA ATATTGCCTT TTATGGATAT TGAAGAAGAA AAATAATCTC CATAAAAAAA ATTATTTTAA TTTATGGAAG GAAATAACCT 4300

4301 AAAAAAATTC ACTGGAATTA TTTTAAAACT CGAGTAAAAT TTATTTATTT AATTTTATTT TTATTTTTTT CTTTATGGAT ACCTTTTTTC CAAAATTATG 4400

4401 GTTGAAATAA AATTTTTATT TTTTTATCAA ATTATATAAC TAAAAAAATA ATTTAAAAAA TAATTAAAAA TGTTTTTATA AAAATAATCA AAATCATTCA 4500

4501 ATACACCTAA ATTTTATTTT GATGGTGCTC GTATTTTGGA TGGATTAACA TTATATTATC ATAAAAAAAT AAAAGAAAAT TTTGCTAATA CATTAGAATT 4600

4601 GAATTATATT GTAAATAACA CTGATGAAAT AAAAGATATT GAAAATATTA AAATAATTGA AAGATTAGGC ACAGATTTTA TTGAAATCAA ATGTATAAAT 4700

4701 AATTCATCAG GAATAATAAT GCTGGAGAGA TTTTAATAAA AAATAATTTT TAAAGAAAAC TTTTATGAAG ATTTGTAAAA ATAAAATAAT TTTTTAGCAT 4800

4801 CTATATTTTT ATATTTTTTC CACAAGTACT ACTATGATGA TGAGAATAGT GCTGATTTTA ATGATATTAC AAAATTTAAA AATGTGGCTT CTATAATACC 4900

4901 AGTTTACAAA GATATGAAAT CAAATGCACT ACCTACACGT AAAATTAGAA TAAATATGCT AGCAAGCAAT ATCGCTCTAA TGATATTCAA GATGCAAAGG 5000

5001 ATCACTCAAA GAAGCAGAGC TAACTGGTCA ATTACAAAAA TTAAAGTCAA AAAACCGCAA AGCTAGAATT AGCTAAAAAG TGATGAAAAA ACACATAATT 5100

5101 ATTTTTTTTA AATAAATTTA TTTTTATTTT TTCTTTATTT TATATAAAAA ATATATATTT TTTTATCATG CTCAATTTTC AAAATAGAAA TTACGTATTG 5200

5201 AAGGTTAAAG TATTTTTTAC TATGATAGGT ATAATTCTTT ACATGAAATT AAATATTTTT CAATTGTATA CATCAGTAAT GAAGGAATAT TTAGGAATTC 5300

5301 AACAAGAAAT TAGACTTCGT GAAAAATTAT CCAAATGCAG TAACGAACGA TCAAAAGAGG TACAGAGAAA TAAAAATTCA CAGGATTAAT ATGTTTTATG 5400

5401 ATTATTATTT AGATTAAGTT TTGTAGATGA TGCAAATGCT ATATTATGAA AATTTTATCC TAATAAATAT AAATGAAGAA GATAGTGAAG ATTTTATGAC 5500

5501 AGAATTTTAA AAATTAAAAT AAACAGGTTT AAAATTGTAA CACTGTTTCG ACAAAGCTCT TAAATTGTGA AGATATTAAG CTAATGATTG AGAAGTAAAT 5600

5601 TAAGGAAGAA TATCCAGTTC AAACATTTCT CAGATAAATA TTGTAAAAAA TGAAAAAAAT AATTTTTATT TTTTTAAAAT ATTATATATA CAAAAAAAGA 5700

5701 ATAAAATAAT TAATGTTACC ATTCACTCAA CAAAAATTAA AATTACAAGG AAGAGATATT TTTTACTATA ATAGAGGATG TCACATTCCT AAAATTTATT 5800

5801 TATTTTCAAA ATAAATTTTT GACGATCTAT AAAATTATTT AGGAATTTAA TTCGAATTAA ATCAGAGATC GAAAATATCA TAAGCAGTAG ACAATGATGG 5900

5901 TAATTAAAAA TTATATGATA GAATAAAAAT TGATAGAATA TCGTATTTAT ATGATTTTTA TGTTGACACT ATTTTAAATT TAACTCAATT ACTTTACTAT 6000

6001 GAAATATTCG ATCATAGAAA TCTCTACTTA TTTTCTACTG TAGATACAGC TTTATAATAA TTTTTCATAG CAAATAATAA TGATTTATTT GTTAAAAATA 6100

6101 CAAAAGCTAC AATATAGAAT GAAATATATA AAAAATTAGA TGAAACTGAA AATGATGAAA AAACATTTTT TAAACACTTT GATAAAAATG TTAAAAAATC 6200

6201 AGGTAGAATA TTAAGTGAAT GTTTCAAATC AGCCCTTGAA AAGTGTTAAC CATTAAAAAC TATGATC 6350

Figure 5.4.

Alignment of Tmi end with P element end. The first 32 nucleotides of the F08 Tmi end are aligned with the inverted repeat end of a P element from Drosophila (Engels, 1989 #53). For this alignment the Tmi end is presented as starting with CA. The T preceding the CA may be part of the end or may be a part of a preferred insertion site.

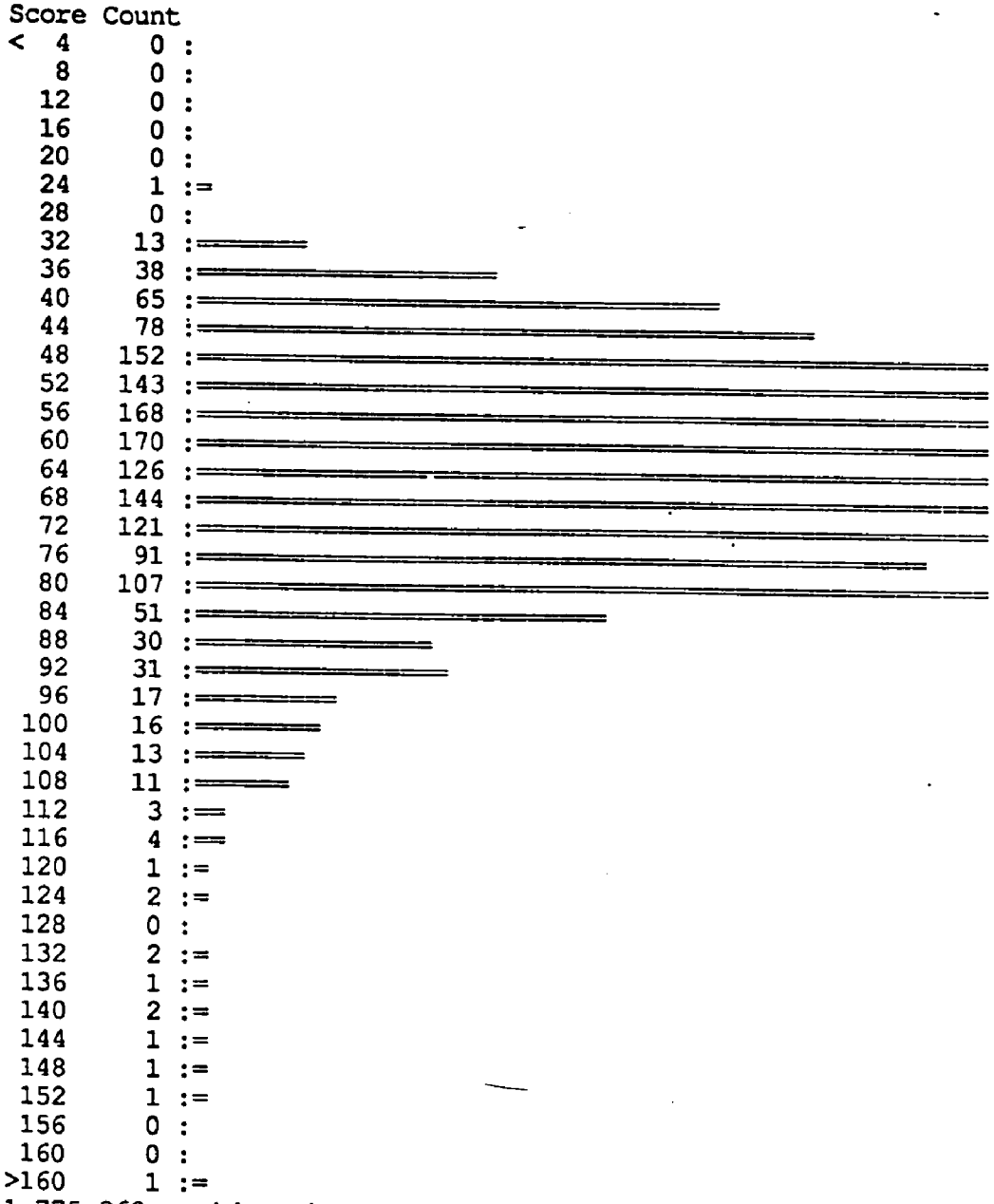
```
Tmi  CAAGAAGAAAAAGATAAGTAATTGATCACGAC
     || || |||||  | |     | |  |
P    CATGATGAAAATAACATAAGGTGGTCCCGTCG
```

The two sequences match at 14 of 32 positions for 42% identity

Figure 5.5.

Histogram of sequence comparison scores for the left part of Tmi and the Invertebrate section of the GenBank database (June 1989 release). The vertical axis indicates the comparison score. The number of sequences in this section of the database having a given score are noted with (=) . The mean score, standard deviation, and top scoring sequence are listed at the bottom of the histogram. The fastn search done to generate this comparison data used a ktup value of 4. The top scoring sequence had 46.4 % identity with the test region of Tmi over a 1800 nucleotide overlap, from base number 360 to 2160 in the Tmi sequence and base number 0 to 1800 in the 17.6 sequence.

4.17F08seq, 2092 NT vs. INVERTEBRATE database (Upper Strand)



1,775,369 residues in 1,605 sequences. Mean score (s.d.): 56.0 (10.49)

>DROTN176 D.melanogaster copia-like element 17.6, complete <U> 190.

Figure 5.6.

Diagram of the overlaped similar regions of Tmi and retro-element sequences from GenBank. The Tmi sequence is shown with the ORFs positioned above. The thick lines below Tmi represent the region of overlap with the 17.6 or HIV2 sequences. The top numbers at the end of the overlap regions are the boundaries of the overlap in the Tmi sequence. The bottom numbers at the ends of the overlap regions are the boundaries of the overlap in the 17.6 or HIV2 sequence. The position of the 17.6 LTR and gag gene in the overlap region are shown. The gag gene extends beyond the end of the overlap region. The HIV2 overlap region is within the HIV2 pol gene which extends from position 1829 to 4939. The arrows indicate the approximate positions of the reverse transcriptase (RT) and integrase (IN) coding sequences based on the average size of these proteins. The nucleotide numbering for 17.6 and HIV2 is as presented in their GenBank entries (June 1989 release).

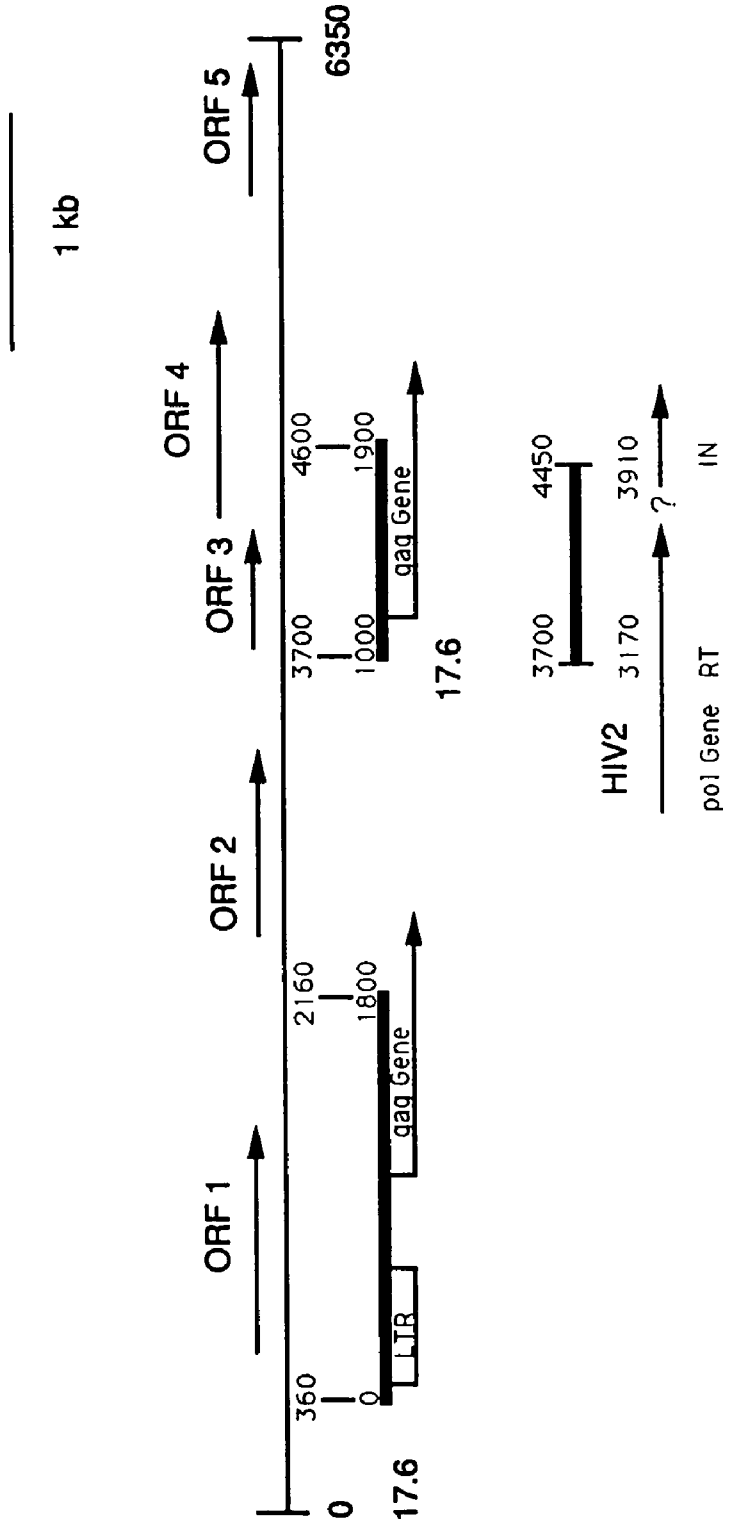
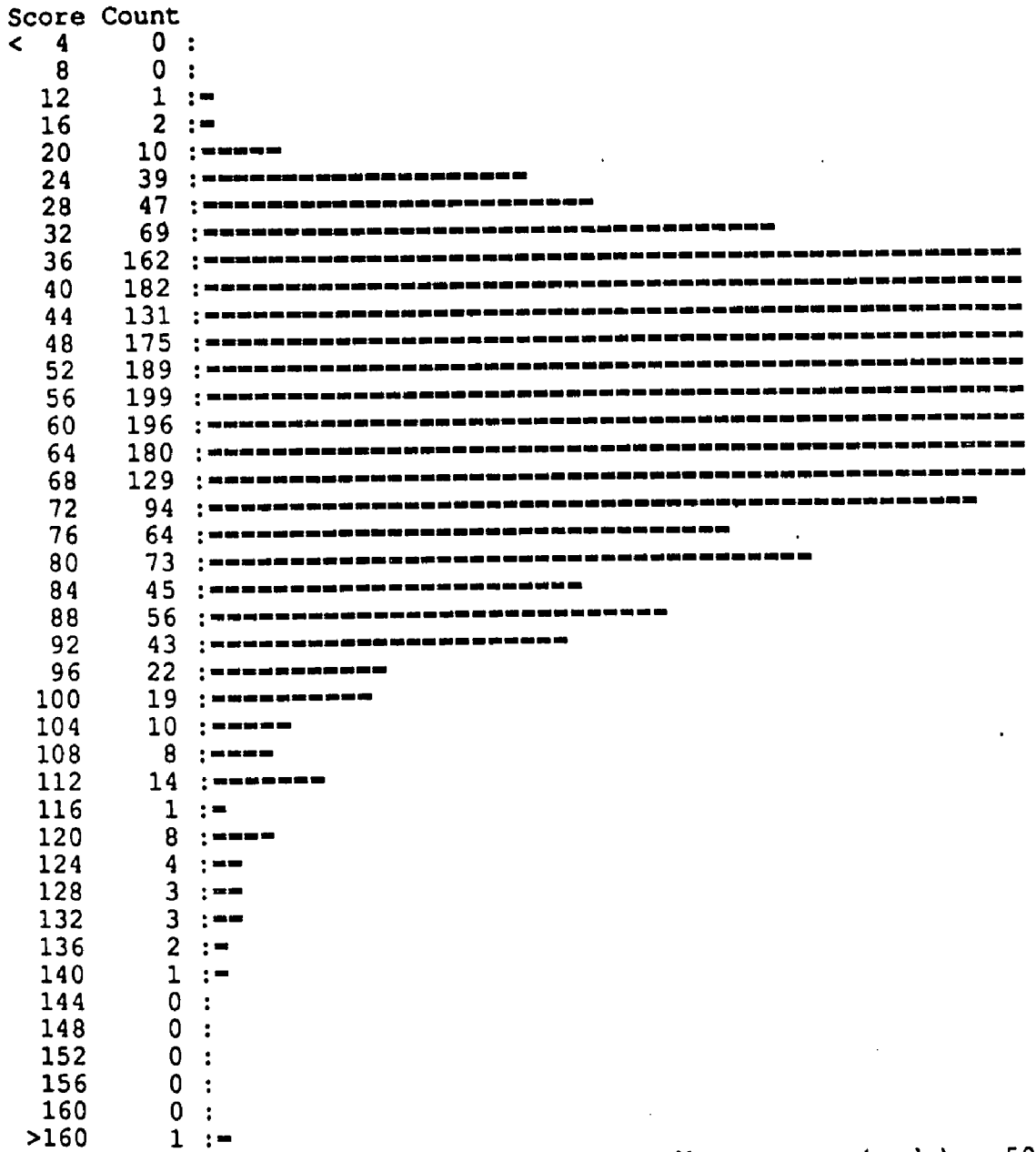

1 kb
ORF 1
ORF 2
ORF 3
ORF 4
ORF 5
6350
4600
1900
3700
1000
17.6
2160
1800
360
0
0
17.6
qag Gene
LTR
qag Gene
3700
4450
HIV2
3170
3910
?
pol Gene
RT
IN

Figure 5.7.

Histogram of sequence comparison scores for the right part of Tmi sequenced and the Virus section of the GenBank database. The vertical axis indicates the comparison score. The number of sequences in this section of the database having a given score are noted with (=). The mean score, standard deviation, and top scoring sequence are listed at the bottom of the histogram. The fastn search done to generate this comparison data used a ktup value of 4. The top scoring sequence had 45.0 % identity with the test region of Tmi over a 750 nucleotide overlap, from nucleotide 3700 to 4450 in the Tmi sequence and from nucleotide 3160 to 3910 in the HIV2 sequence.



```
Score Count
<  4      0 :
   8      0 :
  12      1 :=
  16      2 :=
  20     10 :=====
  24     39 :====================
  28     47 :=========================
  32     69 :====================================
  36    162 :=========================================================
  40    182 :=========================================================
  44    131 :=========================================================
  48    175 :=========================================================
  52    189 :=========================================================
  56    199 :=========================================================
  60    196 :=========================================================
  64    180 :=========================================================
  68    129 :=========================================================
  72     94 :================================================
  76     64 :=================================
  80     73 :======================================
  84     45 :=======================
  88     56 :=============================
  92     43 :======================
  96     22 :===========
 100     19 :==========
 104     10 :=====
 108      8 :====
 112     14 :=======
 116      1 :=
 120      8 :====
 124      4 :==
 128      3 :==
 132      3 :==
 136      2 :=
 140      1 :=
 144      0 :
 148      0 :
 152      0 :
 156      0 :
 160      0 :
>160      1 :=
3,501,718 residues in 2,182 sequences.  Mean score (s.d.):  50.4 (12.86)
```

>HIV2ROD Human immunodeficiency virus type 2, ROD isolate <U> 162,

Figure 5.8.

Position of open reading frames (ORFs) in the Tmi sequence. The position and orientation of the five ORFs encoding polypeptides of 200 amino acids or more are shown as arrows above a composite Tmi. The nucleotide positions of the composite Tmi are those for F08 and the D01 continuation of the sequence from Figures 2 and 3. The positions of the first nucleotide of the start codon and the last nucleotide before the stop codon for each ORF are indicated below the map. The *Hin*dIII site separating F08 from D01 derived positions is shown for orientation. ORF5 does not have a stop codon within the region sequenced.

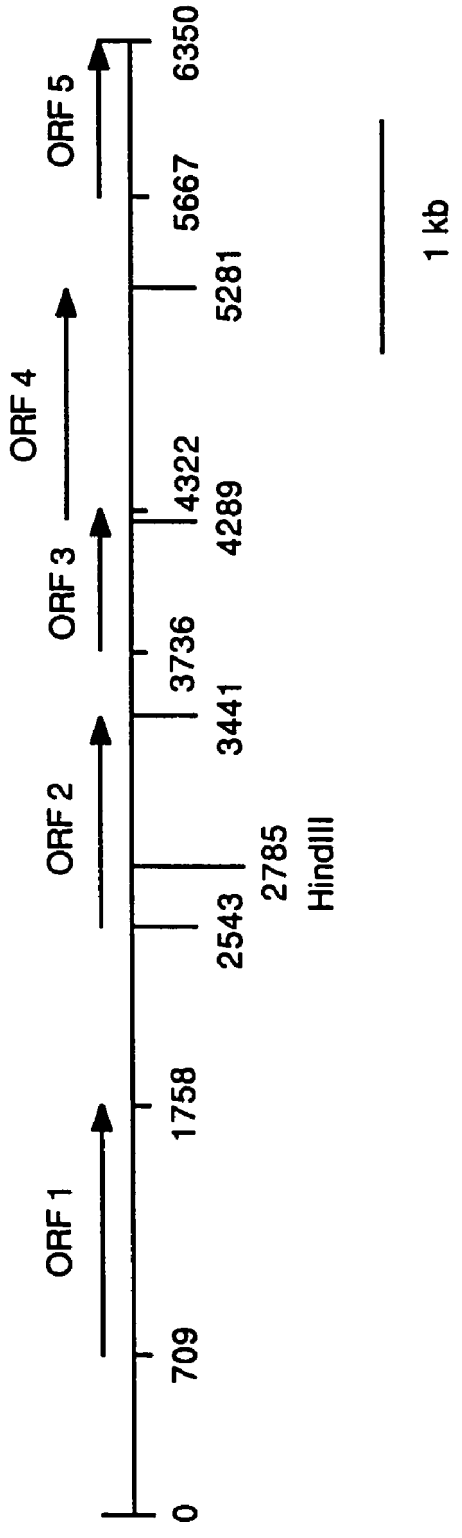

ORF 1
ORF 2
ORF 3
ORF 4
ORF 5
0
709
1758
2543
2785
HindIII
3441
3736
4289
4322
5281
5667
6350
1 kb

Figure 5.9.

Amino acid sequence of the proteins potentially encoded by the Tmi ORFs. The five ORFs are presented along with the sequence of proteins they can encode starting with the first methionine except in the case of ORF3. Translation of ORF3 starts with the third methionine codon because it occurs within a constellation of sequence elements characteristic of the beginning of expressed genes in Tetrahymena. The consensus translational start sequence and potential upstream transcriptional control sequence are underlined. (A) Translation of ORF1. (B) Translation of ORF2. (C) Translation of ORF3. The YYDD motif is underlined. (D) Translation of ORF4. (E) Translation of ORF5.

(A) ORF 1

```
CGA ATG AAG CGA TCG ATG ATA AAA AAT ACA ACA ATG CAG TAG ACT TAT TCA ACA
                                        M  Q  Q  T  Y  S  T   0

AAT TAG GAT TTA AAT CAT TCA AGT AAT GGT TGC AAT TGT ATC AAA CAC TCG ATA
N  Q  D  L  N  H  S  S  N  G  C  N  C  I  K  H  S  I  26

TAA AAT TGC TTA CAG ATG TGT GGA TTA ATT TCA GAA ATA TTT GTC AAT AAA CTT
Q  N  C  L  Q  M  C  G  L  I  S  E  I  F  V  N  K  L  44

ACG GAT TAG AAC CGG GAC ACT ATT TCA CTT CAC CAG GTC TCG CAT GGG ATG CAA
T  D  Q  N  R  D  T  I  S  L  H  Q  V  S  H  G  M  Q  62

TGC TCA AAT AAC TAG AAT AAG ATT GGA CTC ATC ACT AAC CCC GAT ATG TAT TTG
C  S  N  N  Q  N  K  I  G  L  I  T  N  P  D  M  Y  L  80

ATG TTC GAA AGA CAA TCG AGA GGA GGA ATC TCA ACA ACT GGT AAT TTA AGA TAT
M  F  E  R  Q  S  R  G  G  I  S  T  T  G  N  L  R  Y  98

GCT CAA GCG AAC AAC AAA TAC TGC TCA AAT TAT GAT GAG TAA AAA CCT TAC ACT
A  Q  A  N  N  K  Y  C  S  N  Y  D  E  Q  K  P  Y  T  116

CAA TTG ATG TAT TTC GAT GCA AAT AAT CTA TAC GGG CAA GCA ATG TCA CAA AAA
Q  L  M  Y  F  D  A  N  N  L  Y  G  Q  A  M  S  Q  K  134

TTA CCC TAT AGA AAT TTC ACT TTC ATA AGA GAA GAA AAG TTT AAC GAA TTC AAT
L  P  Y  R  N  F  T  F  I  R  E  E  K  F  N  E  F  N  152

TAG GAT TTC ATT ATG AAA TAC AAT AAT ACA AAA TTC GGT TAT GTG TAC GAA GTA
Q  D  F  I  M  K  Y  N  N  T  K  F  G  Y  V  Y  E  V  170

GAT TTG GAA TAT CCA TTC GAA TTG CAT GAT TTA CAC AAT GCA TAT CCT CTA GCA
D  L  E  Y  P  F  E  L  H  D  L  H  N  A  Y  P  L  A  188

CCA GAA TAA ATC ACC ATT CAA TAT GAT TAA CTA TCC AAT GAC AAT CAA GAC ATG
P  E  Q  I  T  I  Q  Y  D  Q  L  S  N  D  N  Q  D  M  206

ATT CAA AAG CTA AAG GGA AGT AAC GAC ATC AAA TAT ATA AGC GGA AAG TAA AAG
I  Q  K  L  K  G  S  N  D  I  K  Y  I  S  G  K  Q  K  224

AAA TTA ACT CCA AAT CTC AAC GAC AAG GAG AGA TAC GTT TGC AAT ATC AGA AAT
K  L  T  P  N  L  N  D  K  E  R  Y  V  C  N  I  R  N  242

TTA AAA TTG TAT CTC GAG AAG GAT TAA ATT TTG GTA AAA ATA CAT AGA ATT ATC
L  K  L  Y  L  E  K  D  Q  I  L  V  K  I  H  R  I  I  260

AAA TAT GAA CAA TCA AAT TTC ATG GAA GTG TAC ATC CAT TTA AAC ACA CAA ATG
K  Y  E  Q  S  N  F  M  E  V  Y  I  H  L  N  T  Q  M  278

AGA CAA CAA GCA ACA AAC GAA TTC GAA AAA GAT TTT TAC AAA TTA ATG AAT AAC
R  Q  Q  A  T  N  E  F  E  K  D  F  Y  K  L  M  N  N  296

TCA GTA TTC GGT AAA ACC ATG GAA AAT GTC AGA AAC AGA GCT TGC TAC GAA TTG
S  V  F  G  K  T  M  E  N  V  R  N  R  A  C  Y  E  L  314

GTA TTC AGC AAA TAC AGA ATG GAA AAA ATA ATA AAA TCG AGA TAC TTC AAA AAC
V  F  S  K  Y  R  M  E  K  I  I  K  S  R  Y  F  K  N  332

TTT ATC CAG ATT ACA GAT AAT TTG GCT CTA GTA GAA AAA TAC AAA AAG ACA GTA
F  I  Q  I  T  D  N  L  A  L  V  E  K  Y  K  K  T  V  350

AAA CTC GAT AAA CCT GCT TAC ATT GGA ATG CAA ATC CTC GAT TTA AGT AAG ATT
K  L  D  K  P  A  Y  I  G  M  Q  I  L  D  L  S  K  I  368

GTA ATG TAC AAT TAC TAC TAC AAT CAC TTG TTG AAA GAA TTT CCG AAA GTA GAT
V  M  Y  N  Y  Y  Y  N  H  L  L  K  E  F  P  K  V  D  386

TTA ATT ATG ACT GAC ACA GAT TCT CTC TTG TGT AAG TAC ATT GCT CGG AGA AGT
L  I  M  T  D  T  D  S  L  L  C  K  Y  I  A  R  R  S  404

ATG ATA TGT ACG ATT CAA TGA
M  I  C  T  I  Q   . 450
```

(B) ORF 2

```
ATG AAT TTA ATT TTA CTA TTG GTA ATA CTT TTC AAA TCC AAT TAG TCC TTC TCC
M   N   L   I   L   L   L   V   I   L   F   K   S   N   Q   S   F   S   18

TTC AAA TAT AAT TTT TTC AAT TTG AGG TCA CTT CCA TAC TTA AAT TTA TCA ATA
F   K   Y   N   F   F   N   L   R   S   L   P   Y   L   N   L   S   I   36

ATC TTA TCC AAT ATA TCA AAT TTT TTC AAA TGG ATC AAC ACA AAA AAC AAA ACA
I   L   S   N   I   S   N   F   F   K   W   I   N   T   K   N   K   T   54

GAA TTT TAA TTT ATA AAA TAA TTT CTC TTA AAC ATT TTT TCT TAA ATT TTT ACC
E   F   Q   F   I   K   Q   F   L   L   N   I   F   S   Q   I   F   T   72

TTA AAA AGC TTT GTT TTA AAC TTT TTA CTG TAC TGT TTA CTT AGA AAT TAC ATT
L   K   S   F   V   L   N   F   L   L   Y   C   L   L   R   N   Y   I   90

CGA CAA CGG GTA TTA AGG CTA AAT CGT TGC ACA ATC ATC GCA GCA AAG CCA TTC
R   Q   R   V   L   R   L   N   R   C   T   I   I   A   A   K   P   F  108

CTC ATA ATA AGA CAA ATT GTA ACT ATT ACA AAA TTC ACA AAT AAT CAT TTT TTT
L   I   I   R   Q   I   V   T   I   T   K   F   T   N   N   H   F   F  126

AAA ATA TTT ATT TTT AAT TGT TGT ATA TAT TAA CAA AAA ATA ATT TTA ATG TCT
K   I   F   I   F   N   C   C   I   Y   Q   Q   K   I   I   L   M   S  144

GAC ATC AAT TAC ATT TTC ATA GAC AGT TCC ATA CTA AAC GTG CCG AAT AAT TAC
D   I   N   Y   I   F   I   D   S   S   I   L   N   V   P   N   N   Y  162

AAT TTT CGA GTT ATT CTC AAC GAA AAG ATA AAA ATA AAA AAA TAC ATA AAA CTA
N   F   R   V   I   L   N   E   K   I   K   I   K   K   Y   I   K   L  180

ATT CAA GCA AAT ATA CCA TTT GAC GAT TAC CTA ATC GAC AAT TAC ATA ACA CAT
I   Q   A   N   I   P   F   D   D   Y   L   I   D   N   Y   I   T   H  198

TTT ACA TAA ATG GCA AAA TGT ACA CAA TAA CAA ACG GTA TTT ACG ATA TCC AGT
F   T   Q   M   A   K   C   T   Q   Q   Q   T   V   F   T   I   S   S  216

CTG ATA AAT CAA ATG TAA TAA ATA ATC AGG ACA GAT ATT TCA TCT TTT TCG ATT
L   I   N   Q   M   Q   Q   I   I   R   T   D   I   S   S   F   S   I  234

ACT CTG TCT CAA CTT ACT TAT AGG ATT ACA TTT CAA AAT ACT CAA AAT TTT TCA
T   L   S   Q   L   T   Y   R   I   T   F   Q   N   T   Q   N   F   S  252

CTG AAC CTA AAC AGA GTT TAT CAA AAT ATA GGA TCA CCG ATA CAA ATT ACT CAA
L   N   L   N   R   V   Y   Q   N   I   G   S   P   I   Q   I   T   Q  270

CAA TAA TCA ATA CAT AAC AAC ATA AAC TCC TCT CAT CAA TCT CCC TCA AGT AGT
Q   Q   S   I   H   N   N   I   N   S   S   H   Q   S   P   S   S   S  288

GCT CTT AAA TAT TGA
A   L   K   Y   .  292
```

(C) ORF 3

```
AGA AAT CCA GAT ATT CAA ATC GTT CAA TCT CAA AAA TAT TCT GTC GAC TAC TCA

TTT CTC ATT CTC AAT ACC GTC AAT AGG GGA TCA AAC ATT CAA TAT TAA AAT CAA

TTT GGT GAA AAT AAA GTG GAT TGT TCT TGC GAT TTA CGC CAA ATG ACG ATA TAA

TTA TAT AAA AAC AAT GGA TAC TTT TTC CAA ATC AAC TCA GAA TTC AAC TCT TCG

ATA TCA ATG AAA AAT TTT TAT TTT TTA TCA AAT TAT ATA ACT AAA AAA AAA AAA

AAA AAA AAT CAA AAA AAT TTA AAA ATG TAA TAT TTA TAA AAA TAA TCA GAT CCT
                            M   Q   Y   L   Q   K   Q   S   D   P   10

TTT ACA AAA CCT TAA TTT TAT TTT GAT GGT GCT CGT ATC TTG GAT GGA TTA GCA
F   T   K   P   Q   F   Y   F   D   G   A   R   I   L   D   G   L   A   28

CTT TTT TAT CAT AAA AAA ATA AAT ATA GAT GAA AAT TTT AAA GAT TCA TTA GAA
L   F   Y   H   K   K   I   N   I   D   E   N   F   K   D   S   L   E   46

TTA ATT TAT AAT ACA AAT GAT GAA AAA CTA AAA GAT ATT AAA AAT ATT TAA ATA
L   I   Y   N   T   N   D   E   K   L   K   D   I   K   N   I   Q   I   64

TTT AAA GAA GAA ATA GGT ACA GAT TTC ATT GAA AAA CAA ATG TTC AGA TAA TTT
F   K   E   E   I   G   T   D   F   I   E   K   Q   M   F   R   Q   F   82

ACT GAT AAC AAT ATA GTT GGA GAA AAT GAA ATA AAA AAT ATT TTT TAA AAA AAC
T   D   N   N   I   V   G   E   N   E   I   K   N   I   F   Q   K   N  100

CTC TAC ACA GAT ATA TAA AAA TAA AAT AAT TTA TTA GCA TCT ATA TTT TTA TAT
L   Y   T   D   I   Q   K   Q   N   N   L   L   A   S   I   F   L   Y  118

TTT ATT CAC AAG TTC TAC TAT GAT GAT GAG AAT AGC GAT GAT ATC AAT AAT ATT
F   I   H   K   F   Y   Y   D   D   E   N   S   D   D   I   N   N   I  136

AAA AAT TTT AAA AAT TCA AAA GAT ATG ATT TAA ATT TAT AAA AAT ATG AAA ACA
K   N   F   K   N   S   K   D   N   I   Q   I   Y   K   N   N   K   T  154

AAC GGA CTA CCA TAT ACC AAA TTT ATC AAT AAA ATT AAA TCT GTT TAC ACA GAT
N   G   L   P   Y   T   K   F   I   N   K   I   K   S   V   Y   T   D  172

ATT CAC AAC TTA TTA GTA GAA ATA TTA TAA TTA AAA TAT GGT TAA TAA GCA AAT
I   H   N   L   L   V   E   I   L   Q   L   K   Y   G   Q   Q   A   N  190

ATT GCC TTT TAT GGA TAT TAG
I   A   F   Y   G   Y   . 196
```

(D) ORF 4

```
AAA CAA ACG GAC TAC CAT ATA CCA AAT TTA TCA ATA AAA TTA AAT CTG TTT ACA

CAG ATA TTC ACA ACT TAT TAG TAG AAA TAT TAT AAT TAA AAT ATG GTT AAT AAG
                                                    M   V   N   K    4

CAA ATA TTG CCT TTT ATG GAT ATT GAA GAA GAA AAA TAA TCT CCA TAA AAA AAA
Q   I   L   P   F   M   D   I   E   E   E   K   Q   S   P   Q   K   K  22

TTA TTT TAA TTT ATG GAA GGA AAT AAC CTA AAA AAA TTC ACT GGA ATT ATT TTA
L   F   Q   F   M   E   G   N   N   L   K   K   F   T   G   I   I   L  40

AAA CTC GAG TAA AAT TTA TTT ATT TAA TTT TAT TTT TAT TTT TTT CTT TAT GGA
K   L   E   Q   N   L   F   I   Q   F   Y   F   Y   F   F   L   Y   G  58

TAC CTT TTT TCC AAA ATT ATG GTT GAA ATA AAA TTT TTA TTT TTT TAT CAA ATT
Y   L   F   S   K   I   M   V   E   I   K   F   L   F   F   Y   Q   I  76

ATA TAA CTA AAA AAA TAA TTT AAA AAA TAA TTA AAA ATG TTT TTA TAA AAA TAA
I   Q   L   K   K   Q   F   K   K   Q   L   K   M   F   L   Q   K   Q  94

TCA AAA TCA TTC AAT ACA CCT AAA TTT TAT TTT GAT GGT GCT CGT ATT TTG GAT
S   K   S   F   N   T   P   K   F   Y   F   D   G   A   R   I   L   D 112

GGA TTA ACA TTA TAT TAT CAT AAA AAA ATA AAA GAA AAT TTT GCT AAT ACA TTA
G   L   T   L   Y   Y   H   K   K   I   K   E   N   F   A   N   T   L 130

GAA TTG AAT TAT ATT GTA AAT AAC ACT GAT GAA ATA AAA GAT ATT GAA AAT ATT
E   L   N   Y   I   V   N   N   T   D   E   I   K   D   I   E   N   I 148

AAA ATA ATT GAA AGA TTA GGC ACA GAT TTT ATT GAA ATC AAA TGT ATA AAT AAT
K   I   I   E   R   L   G   T   D   F   I   E   I   K   C   I   N   N 166

TCA TCA GGA ATA ATA ATG CTG GAG AGA TTT TAA TAA AAA ATA ATT TTT AAA GAA
S   S   G   I   I   M   L   E   R   F   Q   Q   K   I   I   F   K   E 184

AAC TTT TAT GAA GAT TTG TAA AAA TAA AAT AAT TTT TTA GCA TCT ATA TTT TTA
N   F   Y   E   D   L   Q   K   Q   N   N   F   L   A   S   I   F   L 202

TAT TTT TTC CAC AAG TAC TAC TAT GAT GAT GAG AAT AGT GCT GAT TTT AAT GAT
Y   F   F   H   K   Y   V   Y   D   D   E   N   S   A   D   F   N   D 220

ATT ACA AAA TTT AAA AAT GTG GCT TCT ATA ATA CCA GTT TAC AAA GAT ATG AAA
I   T   K   F   K   N   V   A   S   I   I   P   V   Y   K   D   M   K 238

TCA AAT GCA CTA CCT ACA CGT AAA ATT AGA ATA AAT ATG CTA GCA AGC AAT ATC
S   N   A   L   P   T   R   K   I   R   I   N   M   L   A   S   N   I 256

GCT CTA ATG ATA TTC AAG ATG CAA AGG ATC ACT CAA AGA AGC AGA GCT AAC TGG
A   L   M   I   F   K   M   Q   R   I   T   Q   R   S   R   A   N   W 274

TCA ATT ACA AAA ATT AAA GTC AAA AAA CCG CAA AGC TAG AAT TAG CTA AAA AGT
S   I   T   K   I   K   V   K   K   P   Q   S   Q   N   Q   L   K   S 292

GAT GAA AAA ACA CAT AAT TAT TTT TTT TAA ATA AAT TTA TTT TTA TTT TTT CTT
D   E   K   T   H   N   Y   F   F   Q   I   N   L   F   L   F   F   L 310

TAT TTT ATA TAA AAA ATA TAT ATT TTT TTA TCA TGC TCA ATT TTC AAA ATA GAA
Y   F   I   Q   K   I   Y   I   F   L   S   C   S   I   F   K   I   E 328

ATT ACG TAT TAG
I   T   Y   .  323
```

(E) ORF 5

```
CAG AAT TTT AAA AAT TAA AAT AAA CAG GTT TAA AAT TGT AAC ACT GTT TCG ACA

AAG CTC TTA AAT TGT GAA GAT ATT AAG CTA ATG ATT GAG AAG TAA ATT AAG GAA
                                    M   I   E   K   Q   I   K   E    8

GAA TAT CCA GTT CAA ACA TTT CTC AGA TAA ATA TTG TAA AAA ATG AAA AAA ATA
E   Y   P   V   Q   T   F   L   R   Q   I   L   Q   K   M   K   K   I   26

ATT TTT ATT TTT TTA AAA TAT TAT ATA TAC AAA AAA AGA ATA AAA TAA TTA ATG
I   F   I   F   L   K   Y   Y   I   Y   K   K   R   I   K   Q   L   M   44

TTA CCA TTC ACT CAA CAA AAA TTA AAA TTA CAA GGA AGA GAT ATT TTT TAC TAT
L   P   F   T   Q   Q   K   L   K   L   Q   G   R   D   I   F   Y   Y   62

AAT AGA GGA TGT CAC ATT CCT AAA ATT TAT TTA TTT TCA AAA TAA ATT TTT GAC
M   R   G   C   H   I   P   K   I   Y   L   F   S   K   Q   I   F   D   80

GAT CTA TAA AAT TAT TTA GGA ATT TAA TTC GAA TTA AAT CAG AGA TCG AAA ATA
D   L   Q   M   Y   L   G   I   Q   F   E   L   M   Q   R   S   K   I   98

TCA TAA GCA GTA GAC AAT GAT GGT AAT TAA AAA TTA TAT GAT AGA ATA AAA ATT
S   Q   A   V   D   M   D   G   M   Q   K   L   Y   D   R   I   K   I  116

GAT AGA ATA TCG TAT TTA TAT GAT TTT TAT GTT GAC ACT ATT TTA AAT TTA ACT
D   R   I   S   Y   L   Y   D   F   Y   V   D   T   I   L   M   L   T  134

CAA TTA CTT TAC TAT GAA ATA TTC GAT CAT AGA AAT CTC TAC TTA TTT TCT ACT
Q   L   L   Y   Y   E   I   F   D   H   R   M   L   Y   L   F   S   T  152

GTA GAT ACA GCT TTA TAA TAA TTT TTC ATA GCA AAT AAT AAT GAT TTA TTT GTT
V   D   T   A   L   Q   Q   F   F   I   A   M   M   M   D   L   F   V  170

AAA AAT ACA AAA GCT ACA ATA TAG AAT GAA ATA TAT AAA AAA TTA GAT GAA ACT
K   M   T   K   A   T   I   Q   M   E   I   Y   K   K   L   D   E   T  188

GAA AAT GAT GAA AAA ACA TTT TTT AAA CAC TTT GAT AAA AAT GTT AAA AAA TCA
E   M   D   E   K   T   F   F   K   H   F   D   K   M   V   K   K   S  206

GGT AGA ATA TTA AGT GAA TGT TTC AAA TCA GCC CTT GAA AAG TGT TAA CCA TTA
G   R   I   L   S   E   C   F   K   S   A   L   E   K   C   Q   P   L  224

AAA ACT ATG ATC TGA
K   T   M   I    . 228
```

Figure 5.10.

Alignment of amino acid motifs of reverse transcriptases and the ORF3 polypeptide. Amino acid sequences are presented in standard three letter code in capital letters. The consensus motifs represent hydrophobic amino acids as (h),and polar amino acids as (p). Positions where the ORF3 polypeptide matches the motif are starred over the sequence. Amino acid residues among the reverse transcriptases that are identical to the ORF3 polypeptide are underlined. The references for the reverse transcriptases can be found in Poch et al., (1989). The RNA dependent DNA polymerases listed are; HepB, Hepatitis B human; HepWo, Woodchuck hepatitis B; HepBDu, Duck hepatitis B; HIV2, Human immunodeficiency type 2; 17.6, Drosophila 17.6 element; copia, Drosophila copia element; ingi, Trypanosoma ingi element. The RNA dependent RNA polymerases listed are; BNYVV, Beet necrotic yellow vein virus; AaMV, Alfalfa mosaic virus; TYMV, Turnip yellow mosaic virus.

Motif A

```
                    GFF
          h   hDh  AYY h
          *   **    ** *
ORF3      YFDGARILDGLALFYHK

HepB      SNLSWLSLDVSAAFYHL
HepBWo    TDLQWLSLDVSAAFYHI
HepBDu    VQMPRISLDLSQAFYHL
HIV2      KKRRITVLDVGDAYFSI
17.6      RCNYFTTIDLAKGFHQI
copia     YNLKVHQMDVKTAFLNG
INGI      YRTGAVFVDYEKAFTDV
BNYVV     DSAINGVIDAAA.CDSG
AaSMV     ASFHFKEIDFSK.FDKS
TYMV      HSTPKIANDYTA.FDQS
```

Motif B

```
          RF  R G           Q
          KY hK.VhPQG   SP h N  h  hh
          *         *     * *  *  *
ORF3      RQFTD.NNIVGENEIKNIFQKNLYTDI

HepB      ILGFR.KIMPGVGLSPFLLAQFTSAIC
HepBWo    IMGFR.KLPMGVGLSPFLLAQFTSALA
HepBDu    VYYFR.KAPMGVGLSPFLLHLFTTALG
HIV2      RYIYK.VLPQGWKGSPAIFQHTMRQVL
17.6      HYEYL.RMPFGLKNAPATFQRCMNDIL
copia     KAIYG.LKQAARCWFEVFEQALKECEF
INGI      RTFER.GVPQGTVPGSIMFIIVMNSLS
BNYVV     RAHMSYVKTSGEPGT LLGNTILMGAM
AaSMV     FFNVDFQRRTGDALT YLGNTIVTLAC
TYMV      FGPLTCMRLTGEPGT YDDNTDYNLAV
```

Motif C

```
            F
          h YhDDhhh
            * **
ORF3      HKFYYDDENSD

HepB      AFSYMDDVVLG
HepBWo    VFAYMDDLVLG
HepBDu    TFTYMDDFLIC
HIV2      IIQYMDDILIA
17.6      CLVYLDDIIVF
copia     VLLYVDDVVIA
INGI      HGFFADDLTLL
BNYVV     MAMKGDDGFKR
AaMV      VVASQDDSLIG
TYMV      IMUSGDDSLID
```

Motif D

```
                    D
          Gh-h.--EK
            * *   *
ORF3      KDMIQIYKNMKTN

HepB      SLGIHL.NPNKTK
HepBWo    DLGIHL.NVNKTK
HepBDu    ELGIRI.NFDKTT
HIV2      GLGFST.PDEKFQ
17.6      KNKLKL.QLDKCE
copia     NNFKRY.LMEKFR
INGI      EYFMSV.NVAKTK
BNYVV     ILKKETVLDFKLD
AaMV      LFTTLFNLEAKFP
TYMV      DTGVIW.NASKHK
```

Appendix

Analysis of Lambda Clones Containing Candidate Tel-1 Elements from the *Tetrahymena thermophila* Micronuclear Genome

Introduction

A genomic library of *Tetrahymena thermophila* micronuclear DNA was constructed in phage lambda and screened for examples of Tel-1 elements. Genomic Southern blots indicated that the *T. thermophila* micronucleus contained *Bst*XI fragments approximately 10 kb long that hybridized to a Tel-1 probe. Given the assumption that Tel-1 elements have inverted repeat ends including a *Bst*XI restriction site this implies that complete Tel-1 elements in *T. thermophila* are at least 10 kb long. Several lambda clones that hybridized to the Tel-1 probe and had *Bst*XI sites within their inserts were identified. However the Tel-1 cross hybridizing regions of these clones were not otherwise similar, as would be expected of conserved genomic elements. Nor did the regions around the *Bst*XI sites that were sequences resemble Tel-1 ends. A molecular description of complete Tel-1 elements is not yet available.

The lambda clones examined did have some interesting features and will be described here. The lambda clones contain DNA from the micronuclear genome and all appear to represent middle repetitive sequences. Middle repetitive sequences occur on all five micronuclear chromosomes (Allen, et al., 1984; Howard and Blackburn, 1985; Karrer, 1983). All copies of some micronuclear repetitive sequences are entirely eliminated during macronuclear development (Howard and Blackburn, 1985; Karrer, 1983; Yao, 1982). Some copies of other micronuclear repetitive sequences are apparently retained

in the macronucleus and can be rearranged during macronuclear development (Allen, et al., 1984; Allitto and Karrer, 1986; Howard and Blackburn, 1985). The arrangement of repetitive sequences in the micronucleus may be important for micronuclear function. Alternatively repetitive micronuclear sequences may have functions in genome rearrangement events required for macronuclear development. The description of the the lambda clones of micronuclear DNA presented here may be of interest for future studies of *T. thermophila* micronuclear genome organization.

Based on the original observation that mic C_4A_2 sequences were adjacent to Tel-1 ends, the lambda library was also screened for C_4A_2 containing clones. None of the eight clones selected on the basis of hybridization to a C_4A_2 repeat probe had inserts containing the large *Bst*XI fragments expected of complete Tel-1 elements. Therefore, intact Tel-1 elements were not apparently tightly linked to mic C_4A_2 sequences in these clones, or by implication, in the genome. However, every sequenced mic C_4A_2 had a Tel-1 end associated with it.

Results

Library construction

A recombinant library was made by ligating purified micronuclear DNA from strain SB2040.2 that was partially digested with the restriction enzyme *Sau*3AI into lambda EMBL4 arms with *Bam*HI ends (Frischauf et al., 1983). Micronuclear DNA was prepared as described in the materials and methods chapter. The DNA had to be additionally purified by cesium chloride equilibrium density gradient centrifugation before it would produce recombinant DNA that could be packaged efficiently into phage. Test ligations

and packaging of macronuclear DNA from the same preparation produced efficiently packaged phage without extra purification.

EMBL4 lambda DNA (Promega Biotech) was prepared to accept *Sau*3AI insert fragments by digestion with *Bam*HI and *Sal*I. This leaves *Bam*HI sticky ends on the vector arms which can be ligated to *Sau*3AI cut insert fragments. Precipitation of the digested DNA with 0.15 volumes of 3M sodium acetate and 0.6 volumes of isopropyl alcohol leaves the small lambda polylinker fragment with *Bam*HI and *Sal*I ends in solution. The lambda arms do not have to be purified away from the internal lambda fragment as it has *Sal*I ends which cannot efficiently religate to the BamHI ends of the arms.

Micronuclear DNA from strain SB2040.2 was purified by cesium chloride density gradient centrifugation. 1.53 g of cesium chloride and 100 µl of 10 mg/ml ethidium bromide were added to approximately 50 µg of micronuclear DNA dissolved in 1.615 ml of TE and 100 µl 10 mg/ml ethidium bromide in a 1.5 ml centrifuge tube. This was centrifuged for 13 hours at 100,000 rpm in a TLA-100.2 rotor in a Beckman model TL-100 ultracentrifuge. After centrifugation the top of the tube was cut off and 150 ul fractions removed with a cut off eppendorf pipette tip. Fractions containing DNA were pooled and the DNA precipitated by adding 3 volumes of 70% ethanol and chilling for 2 hours at -20°C. The DNA was pelleted by centrifugation in an eppendorf microfuge for 15 minutes and then washed twice with 70% ethanol. The pellet was briefly dried under vacuum before being resuspended in TE.

The gradient pure micronuclear DNA was partially digested with *Sau*3AI. 10 µg of DNA was incubated at 37°C with 1 U of *Sau*3AI in a total volume of 100 ul. 25 µl samples were removed after 1.5, 3, 5, and 7 minutes. After digestion the reaction mixture was extracted twice with phenol:chloroform (1:1) and the DNA was ethanol precipitated.

The average size of the DNA after this digestion was between 10 and 20 kb as assayed by agarose gel electrophoresis. 2 µg of this partially digested micronuclear DNA was ligated to 5 µg of *Bam*HI-cut EMBL4 vector arms (this was a 2:1 molar excess of insert DNA) using T4 DNA ligase (conditions specified by the supplier). The ligation mixture was packaged in vitro in four separate packaging reactions using 5 µl of the ligation mix in each (packaging extracts were the generous gift of Kathy Hudson and Kathryn Anderson, UC Berkeley). Recombinant phage were plated on *E. coli* strain NM538 (obtained from Promega Biotech).

Analysis of Tel-1 selected recombinant phage

Tel-1 hybridizing phage were identified from duplicate plaque lift filters of more than 30,000 phage. Given the *T. thermophila* genome size of 10^8 bp (Karrer, 1986) and the average insert size of 17 kb screening 28,000 plaques gives a 99% chance of detecting a single copy sequence (Maniatis et al., 1982). 43 plaques that were positive for hybridization to the Tel-1 probe on duplicate filters were picked. DNA was isolated from 11 of these that had been twice plaque purified. All of these 11 clones contained inserts that hybridized to the Tel-1 probe on Southern blots. However, only 5 (#10, 19, 22, 31, and 36) also had *Bst*XI fragments of approximately 10 kb included in their inserts (Figure 1). The Tel-1 cross hybridizing sequences were within the large *Bst*XI fragments in clones 22, 31 and 36. Clone 36 appeared to be undergoing rearrangements as restriction digests indicated a mixture of phage even after two rounds of plaque purification. More detailed restriction mapping, subcloning and sequencing efforts concentrated on clones 22 and 31.

Clone 22 was the most likely to contain a complete Tel-1 element. The Tel-1 probe hybridized to a 10 kb *Bst*XI fragment in the insert. There were, despite their overall

rarity in the micronuclear genome, clusters of *BstXI* sites in this and the other clones (discussed below). Subclones were constructed for sequencing across both *BstXI* sites from one end of the insert and one *BstXI* site from the other end of the insert. The regions sequenced are indicated in Figure 2A and the sequence is shown in Figure 2C. None of the *BstXI* sites sequenced is embedded in sequence resembling the predicted Tel-1 end (Figure 2B). Hence it seems unlikely that this clone represents a Tel-1 element as previously defined. The Tel-1 probe is 4.3 kb long and, based on Southern blotting, clearly contains repetitive sequences that are not bounded by *BstXI* sites and are not limited to the micronucleus. Clones identified by hybridization to this probe could have hybridized to repetitive sequences in the probe that are not part of Tel-1 elements.

Analysis of C_4A_2 selected recombinant phage

Because Tel-1 ends were associated with micC_4A_2 sequences in previous clones (Cherry and Blackburn, 1985) this lambda library was also screened for C_4A_2 hybridizing clones. 40,000 plaques from the amplified library were screened by hybridization of a C_4A_2 repeat probe to duplicate plaque lift filters. DNA from 8 phage that were positive for hybridization to the C_4A_2 probe and subsequently twice plaque purified was analyzed. Only three out of eight micC_4A_2 containing clones also contained Tel-1 cross hybridizing sequences. Clones 1CA and 5CA had the relative arrangement of Tel-1 and C_4A_2 repeats expected from previous analysis of micC_4A_2 repeats. Only one of the 11 clones described above selected by hybridization to the Tel-1 probe also hybridized to a C_4A_2 probe. This indicates that complete Tel-1 elements and micC_4A_2 sequences are not always associated with each other as previously proposed (Cherry and Blackburn, 1985). The position of C_4A_2 sequences in these clones was not precisely mapped. It is possible that the C_4A_2 sequences in all of these clones are adjacent to a *BstXI* site. The C_4A_2 hybridizing regions of these clones were not sequenced so it is not known if they

abut the same 30 bp conserved sequence reported by Cherry and Blackburn (1985) Of the clones whose inserts hybridized to both a Tel-1 and a C_4A_2 probe none had large *Bst*XI fragments in their inserts (Figure 3). As there were no easily identified complete Tel-1 elements among these clones they were not analyzed further.

Clustering of restriction sites and common fragments among Tel-1 selected recombinant phage

Although an identifiable example of a complete Tel-1 element was not found in this lambda library, restriction analysis of the Tel-1 hybridizing lambda clones revealed some unexpected features. Restriction maps of clones 19, 22, 31 and 36 are shown in Figure 1. These clones all contained *Bst*XI restriction sites, sometimes several clustered within one kb of each other. This is unusual because *Bst*XI sites are rare in the *T. thermophila* micronuclear genome, presumably because the recognition sequence for *Bst*XI is G+C rich and the *T. thermophila* genome, with the exception of coding sequences (Martindale, 1989), is A+T rich (Karrer, 1986). *Bst*XI digested micronuclear DNA is indistinguishable from undigested DNA on ethidium stained agarose gels, with *Bst*XI digested fragments migrating at limit mobility (greater than 15 kb). Figure 2 shows the regions of clone 22 that were sequenced, the complete sequence of these regions and an alignment of the sequenced *Bst*XI sites with the Tel-1 end. The *Bst*XI sites in clone 22 were not similar to Tel-1 ends. The 0.5 kb *Bst*XI fragment from the left end of clone 22 (contained within the subclone p22A shown in Figure 2) did not cross hybridize with similarly sized *Bst*XI fragments from the other clones. Southern blots of micronuclear DNA cut with *Bst*XI and probed with the p22A insert fragment showed hybridization to multiple bands. The signal from hybridization with the genomic band of about 0.5 kb was not more intense than from the other genomic bands. In

conclusion, sequences in the p22A probe are repetitive in the micronuclear genome but the arrangement of *Bst*XI sites in clone 22 is not common in the micronuclear genome.

Most (10 of 11) of the Tel-1 selected lambda clones from which DNA was prepared also shared a 5.5 kb *Hin*dIII fragment in their inserts. This fragment did not cross hybridize to the Tel-1 probe. The 5.5 kb *Hin*dIII fragments did cross hybridize to each other (data not shown). This *Hin*dIII fragment is unlikely to be a contaminating lambda DNA fragment. The *Hin*dIII and *Bst*XI sites mapped within this fragment do not correspond to mapped *Hin*dIII and *Bst*XI sites in lambda. Probes made from this region of the lambda clones did not hybridize to lambda size markers on Southern blots. When the 5.5 kb *Hin*dIII fragment from one of the lambda clones was used as a probe of *Hin*dIII digested micronuclear DNA displayed on a Southern blot it hybridized to multiple fragments. The 5.5 kb band is only slightly more intense than the others. Therefore, this 5.5 kb *Hin*dIII fragment is not common in the micronuclear genome and is apparently over represented in the lambda clones.

Discussion

The clustering of *Bst*XI sites that are rare in the genome and the over representation of a 5.5 kb *Hin*dIII fragment that is not common to the genome indicate that this lambda clone library was not a good representation of the *T. thermophila* micronuclear genome. In a library of random fragments covering the entire genome one would expect to see a close correspondence between common features of the clones and the genome. The construction of this library may have inadvertently favored cloning particular regions of the micronuclear genome and excluding others. The *T. thermophila* genome is in general A+T rich ,75% (Karrer, 1986), and some regions of the genome have even higher A+T content. Sequences internal to macronuclear telomeres are up to 90% A+T

(Spangler et al., 1988) and 83-87% A+T near micC_4A_2, (Cherry and Blackburn, 1985). An increased A+T content in non-coding regions is implied by the lower A+T content of coding regions of approximately 60% (Martindale, 1989). There is obviously a paucity of restriction enzyme recognition sites that are G+C rich or even include G's and C's in regions of the genome that consist mostly of A's and T's. The choice of partial *Sau*3AI digestion of micronuclear DNA may not have produced random pieces of micronuclear DNA in a size range that would be accepted by the vector (10-20 kb). Rather the recombinant clones produced may have favored regions of the genome that were not A+T rich and therefore had relatively more recognition sites for restriction enzymes such as *Bst*XI. Regions that were very A+T rich would have been excluded from the library because partial *Sau*3AI digestion did not reduce them to fragments smaller than 20 kb. The association of the 5.5 kb *Hin*dIII fragment with regions of the genome that hybridize to the Tel-1 probe is intriguing but was not further investigated here.

Figure A.1.

Restriction maps of Tel-1 probe selected lambda clones. The maps of the inserts of 5 of the lambda clones are shown oriented with the insert end adjoining the long lambda arm to the left. The position of recognition sites for *Bgl*II (G), *Bst*XI (B), *Hin*dIII (H), and *Sal*I (S) are indicated above the linear maps. The dashed line above the maps indicates the 5.5 kb *Hin*dIII fragment that was shared by 10 of 11 of the clones analyzed. The thin line below the maps indicates the approximate position of a *Bst*XI site or sites that were not precisely mapped.

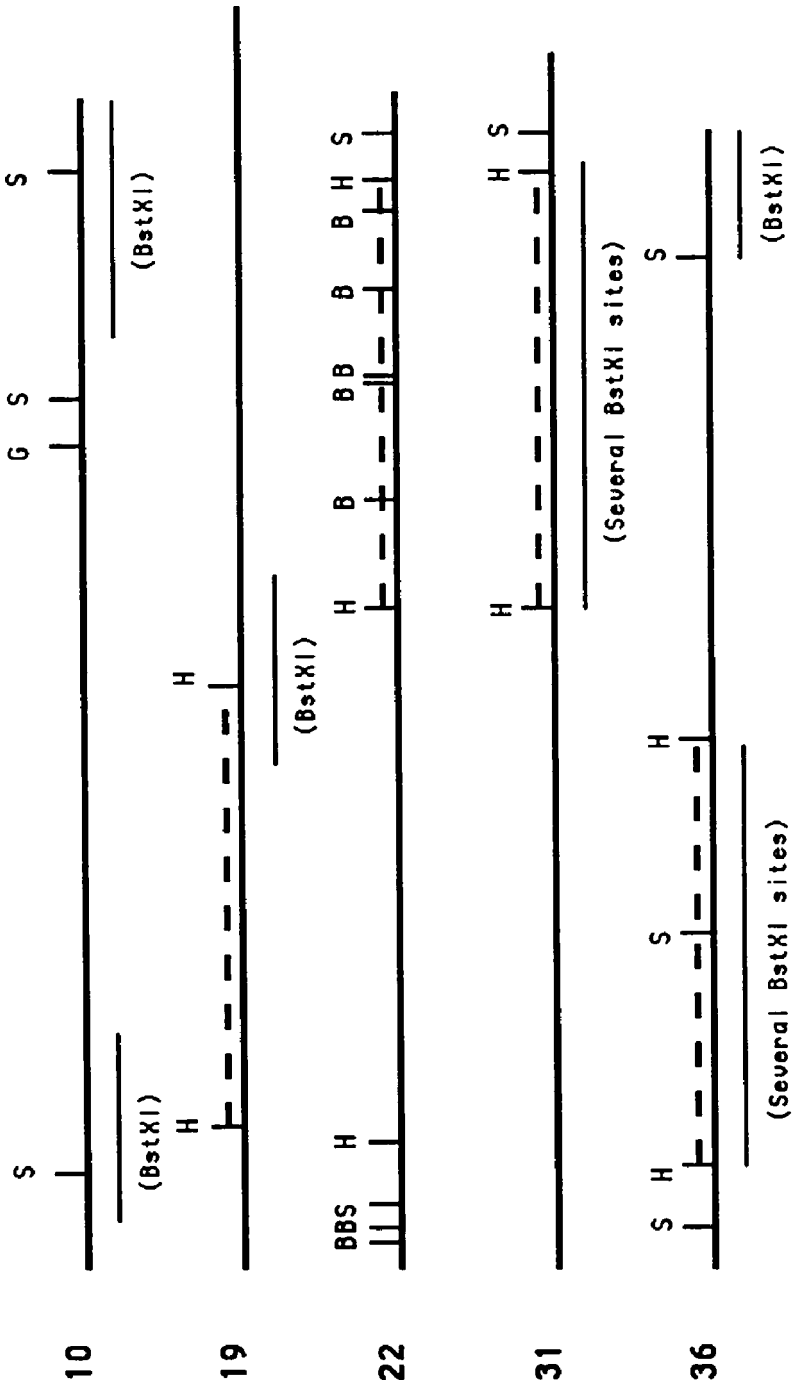
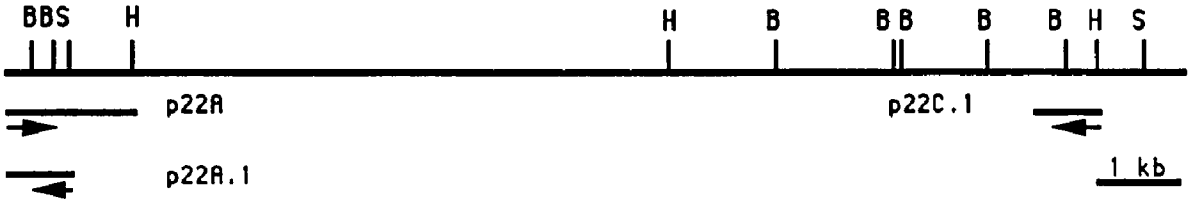

10
19
22
31
36
S
(BstXI)
H
G S
S
(BstXI)
H
BBS H
H B BB B B H S
(BstXI)
H
H S
(Several BstXI sites)
S H
S
H
(Several BstXI sites)
S
(BstXI)
1 kb

Figure A.2.

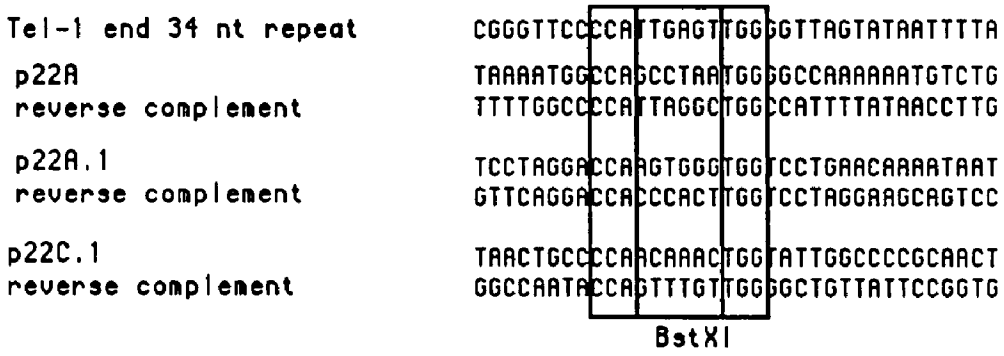
Sequence around three of the *BstX*I sites of clone 22. (A) The restriction map of lambda clone 22 insert (key to enzyme abbreviations is the same as in Figure 1). The subclones constructed to sequence across the *BstX*I sites are drawn below the map. The direction of sequencing from the subclones is indicated by the small arrow. (B) The sequences surounding the *BstX*I sites of clone 22 are lined up with the Tel-1 inverted repeat end by fixing them with respect to their *BstX*I sites. Both the sequence determined from the subclones and an overlaping but not identical portion of their reverse complements are shown. In no case does a sequence around a clone 22 *BstX*I sites closely resemble the Tel-1 end. In contrast the proviously sequenced Tel-1 ends associated with mic C_4A_2 repeats (Cherry and Blackburn, 1985) differed from each other at no more than a few sites within this conserved sequence. (C) Sequence of clone 22 regions that were determined. For each subclone the sequence is written 5' to 3' in the direction of the arrows below the map in part (A).

A

22


BBS
H
H
B
BB
B
B
H
S
p22A
p22C.1
p22A.1
1 kb

B


Tel-1 end 34 nt repeat
CGGGTTCCCCATTGAGTTGGGGTTAGTATAATTTTA
p22A
TAAAATGGCCAGCCTAATGGGGCCAAAAAATGTCTG
reverse complement
TTTTGGCCCCATTAGGCTGGCCATTTTATAACCTTG
p22A.1
TCCTAGGACCAAGTGGGTGGTCCTGAACAAAATAAT
reverse complement
GTTCAGGACCACCCACTTGGTCCTAGGAAGCAGTCC
p22C.1
TAACTGCCCCAACAAACTGGTATTGGCCCCGCAACT
reverse complement
GGCCAATACCAGTTTGTTGGGGCTGTTATTCCGGTG
BstXI

C

p22A

AATTAATTCA ATAATAAAAG CAAAATTAAA AANAAAACAT AAAAAAAATA TTTAATTCCT 60

CCTACCCTCT AATTCAAACT GATAGAAGAC AGTTTGAATT AGATTCTTTG CTTTTTGATG 120

GCAGTGCTAC CTTGGCACTC TTTTAGGCAA TGCAAATAAA TTTTTGACAT GCTCAAGGAC 180

CCACTCCCTC ATCTTTTCAA AAATATCTAT TTGGAAAAAT CCTATAGGAA GGGGTGTCTC 240

TTCAAGGTTA TAAAATGGCC AGCCTAATGG GGCCAAAAAA TGTCTGGGCC AGCTGGACTC 300

CTAGGGACTT TAATCTCGAC ACCTTGNGCC TGCTCTATTC TAGAGTTAAG GGCATCTTAT 360

CTTGGTGACT CACTGCAAGA TGAGGCTCTT CTAGGAGGCT ATCTGAGGCA TCTCAGCTCA 420

TTATGTAATA TT 432

p22A.1

TCAAGGAGGA GCAAGAGAAG AACTTGAAAC AGATGGAGGA GCTTAAGAAG AAGATGTGAC 60

AAGCCTAGGA CTGCTTCCTA GGACCAAGTG GGTGGTCCTG AACAAAATAA TATCTTATCT 120

GAATGTTTTT ATTTGTATTT TATTTGAAAT GTTTTATTTG TATAGAAATA ATGAAAATAT 180

TTTATCTTTT ACTTTAGTGT AAAATTTTAT ATATGTCGGC TCATAGAGCA GAGGCCTGTG 240

AGAAAGATAT 250

p22C.1

AGCACCACCA CCGGAATAAC TTGCCCCAAC AAACTGGTAT TGGCCCCGCA ACTTAGGTGC 60

GTCTGTAATG AGAATAAAAC GCCGTTACGC CAATCTCTGG AGCGCGAATC GCGAAGAAAG 120

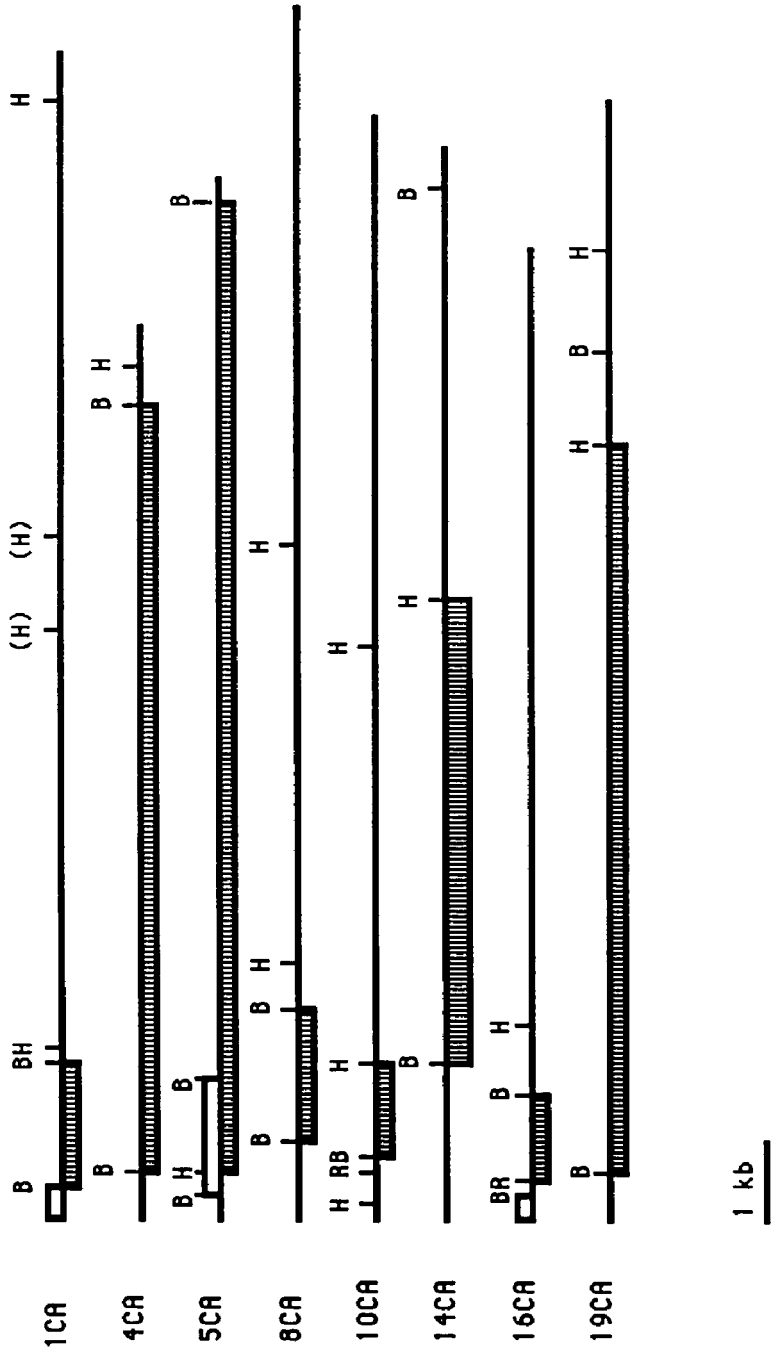
CGAAGGGGCA ATGCGTAACG CCCATTCCCT CCCCAAATTT GCCGATATAG CCCAACCTAG 180

CAACGTAATC AGTTTCGTGT TTCCTTTCGC GACGTCACTG TGTGTTGCTG CTCTTTCGGA 240

GATAGCTTAC 250

Figure A.3.

Restriction maps of C_4A_2 selected lambda clones. The maps of the inserts from 8 lambda clones are shown aligned so that the left end of the insert would adjoin the long lambda arm. The positions of recognition sites for the enzymes *Bst*XI (B), *Eco*RI (R), and *Hin*dIII (H) are indicated. Parenthesis indicate that there is a restriction site at one of the two positions shown. The fragment from each clone that hybridized to the C_4A_2 repeat probe is indicated by a striped box below the map. In those cases where there was also a fragment that hybridized to the Tel-1 probe that fragment is indicated by an open box above the map line.


1CA
4CA
5CA
8CA
10CA
14CA
16CA
19CA
1 kb

"'Where the tree of knowledge stands is always paradise': thus speak the oldest and youngest serpents."

Nietzsche, Beyond Good and Evil

References

Aksoy, S., Lalor, T. M., Martin, J., Van der Ploeg, L. H. T. and Richards, F. (1987). Multiple copies of a retroposon interrupt spliced leader RNA genes in the African trypanosome, *Trypanosoma gambiense*. EMBO J. *6*, 3819-3826.

Allemen, M. and Freeling, M. (1986). The *Mu* transposable elements of maize: Evidence for transposition and copy number regulation during development. Genetics *112*, 107-119.

Allen, S. L. (1967a). Genomic exclusion: A rapid means for inducing homozygous diploid lines in *Tetrahymena pyriformis* syngen 1. Science *155*, 575-577.

Allen, S. L. (1967b). Cytogenetics of genomic exclusion in *Tetrahymena*. Genetics 55, 797-822.

Allen, S. L., Ervin, P. R., McLaren, N. C. and Brand, R. E. (1984). The 5S ribosomal RNA gene clusters in *Tetrahymena thermophila*: Strain differences, chromosal localization and loss during micronuclear aging. Mol. Gen. Genet. *197*, 244-253.

Allitto, B. A. and Karrer, K. M. (1986). A family of DNA sequences is reproducibly rearranged in the somatic nucleus of Tetrahymena. Nucl. Acids Res. *14*, 8007-8025.

Austerberry, C. F. and Yao, M.-C. (1987). Nucleotide sequence structure and consistency of a developmentally regulated DNA deletion in *Tetrahymena thermophila*. Mol. Cell Biol. *7*, 435-443.

Austerberry, C. F. and Yao, M.-C. (1988). Sequence structures of two developmentally regulated, alternative DNA deletion junctions in *Tetrahymena thermophila*. Mol. Cell. Biol. *8*, 3947-3940.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1988). Current Protocols in Molecular Biology. New York, New York: John Wiley and Sons.

Baird, S. E., Fino, G. M., Tausta, S. L. and Klobutcher, L. A. (1989). Micronuclear genome organization in *Euplotes crassus*: A transposon-like element that is excised during macronuclear development. Mol. Cell. Biol. *9*, 3793-3807.

Berg, D. E. and Howe, M. M. (1989). Mobile DNA. Washington, D.C.: American Society for Microbiology.

Boeke, J. D. (1989). Transposable elements in Saccharomyces cerevisiae. in Mobile DNA. Ed. D. E. Berg and M. M. Howe. Washington, D.C.: American Society for Microbiology. 335-374.

Britten, R. J. and Davidson, E. H. (1969). Gene regulation for higher cells: A theory. Science *165*, 349-357.

Brunk, C. F. and Sadler, L. A. (1990). Characterization of the promoter region of Tetrahymena genes. Nucl. Acids Res. *18*, 323-329.

Bruns, P. J. (1986). Genetic organization of Tetrahymena. in The Molecular Biology of Ciliated Protozoa. Ed. J. G. Gall. Orlando, Florida: Accademic Press, Inc. 27-44.

Bruns, P. J. and Brussard, T. B. (1981). Nullisomic Tetrahymena: Eliminating germinal chromosomes. Science *213*, 549-551.

Bruns, P. J., Brussard, T. B. and Merriam, E. V. (1983). Nullisomic Tetrahymena. II. A set of nullisomics define the germinal chromosomes. Genetics *104*, 257-270.

Cech, T. R. (1985). Self-splicing RNA: Implications for evolution. in Genome Evolution in Prokaryotes and Eukaryotes. Ed. D. C. Reanney and P. Chambon. Orlando, Florida: Academic Press, Inc. 3-22.

Cherry, M. C. and Blackburn, E. H. (1985). The internally located telomeric sequences in the germ-line chromosomes of Tetrahymena are at the ends of transposon-like elements. Cell *43*, 747-758.

Clavel, F., Guyader, M., Guetard, D., Salle, M., Montagnier, L. and Alizon, M. (1986). Molecular cloning and polymorphism of the human immunodeficiency virus type 2. Nature *324*, 691-695.

Covey, S. N. (1986). Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. Nucl. Acids Res. *14*, 623-633.

Davidson, E. H. and Britten, R. J. (1979). Regulation of gene expression: possible role of repetitive sequences. Science *204*, 1052-1059.

Doolittle, W. F. and Sapienza, C. (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature *284*, 601-603.

Doring, H.-P. and Starlinger, P. (1986). Molecular genetics of transposable elements in plants. Ann. Rev. Genet. *1986*, 175-200.

Eichinger, D. J. and Boeke, J. D. (1988). The DNA intermediate in Yeast Ty1 element transposition copurifies with virus-like particles: Cell-free trnsposition. Cell *54*, 955-966.

Eichinger, D. J. and Boeke, J. D. (1990). A specific terminal structure is required for Ty1 transposition. Genes & Development *4*, 324-330.

Engels, W. R. (1989). P elements in *Drosophila melanogaster*. in Mobile DNA. Ed. D. E. Berg and M. M. Howe. Washington, D.C.: American Society for Microbiology. 437-484.

Epstein, L. M. and Forney, J. D. (1984). Mendelian and non-Mendelian mutations affecting surface antigen expression in *Paramecium tetraurelia*. Mol. Cell. Biol. *4*, 1583-1590.

Fedoroff, N. V. (1989). Maize transposable elements. in Mobile DNA. Ed. D. E. Berg and M. M. Howe. Washington, D.C.: American Society for Microbiology. 375-411.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. *132*, 6-13.

Finnegan, D. J. (1989). F and related elements in *Drosophila melanogaster*. in Mobile DNA. Ed. D. E. Berg and M. M. Howe. Washington, D.C.: American Society for Microbiology. 519-529.

Forney, J. D. and Blackburn, E. H. (1988). Developmentally controlled telomere addition in wild-type and mutant Paramecia. Mol. Cell. Biol. *8*, 251-258.

Frischauf, A.-M., Lehrach, H., Poustka, A. and Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. Mol. Biol. *170*, 827-842.

Gabriel, A., Yen, T. J., Schwartz, D. C., Smith, C. L., Boeke, J. D., Solner-Webb, B. and Cleveland, D. W. (1990). Rapidly rearranging retrotransposon within the miniexon gene locus of *Crithidia fasciculata*. Mol. Cell. Biol. *10*, 615-624.

Gall, J. G. (1986). The Molecular Biology of Ciliated Protozoa. Orlando, Florida: Academic Press, Inc.

Garfinkel, D. J., Boeke, J. D. and Fink, G. R. (1985). Ty element transposition: Reverse transcriptase and virus-like particles. Cell 42, 507-517.

Garrett, J. E., Knutzon, D. S. and Carroll, D. (1989). Composite transposable elements in the *Xenopus laevis* genome. Mol. Cell. Biol. *9*, 3018-3027.

Gilboa, E., Eglitis, M. A., Kantoff, P. W. and Anderson, W. F. (1986). Transfer and expression of cloned genes using retroviral vectors. BioTechniques 4, 504-512.

Gorovsky, M. A. (1973). Macro- and micronuclei of *Tetrahymena pyriformis*: A model system for studying the structure and function of eukaryotic nuclei. J. Protozool. *20*, 19-25.

Gorovsky, M. A. (1980). Genome organization and reorganization in Tetrahymena. Ann. Rev. Genet. *14*, 203-239.

Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L. and Alizon, M. (1987). Genome organization and transactivation of the human immunodeficiency virus type 2. Nature *326*, 662-669.

Hanyu, N., Kuchino, Y., Nishimura, S. and Beier, H. (1986). Dramatic events in ciliate evolution: Alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two Tetrahymena $tRNAs^{Gln}$. EMBO J. 5, 1307-1311.

Herrick, G., Cartinhour, S., Dawson, D., Ang, D., Sheets, R., Lee, A. and Williams, K. (1985). Mobile elements bounded by C_4A_4 telomeric repeats in Oxytricha fallax. Cell 43, 759-768.

Herrick, G., Cartinhour, S. W., Williams, K. R. and Knotter, K. P. (1987). Multiple sequence versions of the *Oxytricha fallax* 81-MAC alternate processing family. J. Protozool. 34, 429-434.

Horowitz, S. and Gorovsky, M. A. (1985). An unusual genetic code in nuclear genes of Tetrahymena. Proc. Natl. Acad. Sci. USA 82, 2452-2455.

Howard, E. A. and Blackburn, E. H. (1985). Reproducible and variable genomic rearrangements occur in the developing somatic nucleus of the ciliate *Tetrahymena thermophila*. Mol. Cell. Biol. 5, 2039-2050.

Hunter, C. H., Williams, K., Cartinour, S. and Herrick, G. (1989). Precies excision of telomere-bearing transposons during *Ocytricha fallax* macronuclear development. Genes & Development 3, 2101-2112.

Hutchinson, C. A., III, Hardies, S. C., Loeb, D. D., Shehee, W. R. and Edgell, M. H. (1989). LINEs and related retroposons: Long interspersed repeated sequences in the eucaryotic genome. in Mobile DNA. Ed. D. E. Berg and M. M. Howe. Washington, D.C.: American Society for Microbiology. 593-618.

Jacobs, E., Dewerchin, M. and Boeke, J. D. (1988). Retrovirus-like vectors for *Saccharomyces cerevisiae*: Integration of foreign genes controlled by effecient promoters into yeast chromosomal DNA. Gene 67, 259-269.

Jahn, C. L., Krikau, M. F. and Shyman, S. (1989). Developmentally coordinated en masse excision of a highly repetitive element in *E. crassus*. Cell 59, 1009-1018.

Jahn, C. L., Niles, L. A. and Krikau, M. F. (1988). Organization of the *Euplotes crassus* micronuclear genome. J. Protozool. *35*, 590-601.

Johnson, M. S., McClure, M. A., Feng, D.-F., Gray, J. and Doolittle, R. F. (1986). Computer analysis of retroviral *pol* genes: Assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc. Natl. Acad. Sci. USA *83*, 7648-7652.

Kamer, G. and Argos, P. (1984). Primary structure comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. Nucl. Acids Res. *12*, 7269-7282.

Karrer, K. (1983). Germline-specific DNA sequences are present on all five micronuclear chromosomes in *Tetrahymena thermophila*. Mol. Cell. Biol. *3*, 1909-1919.

Karrer, K. M. (1986). The nuclear DNAs of holotrichous ciliates. in The Molecular Biology of Ciliated Protozoa. Ed. J. G. Gall. Orlando, Florida: Acad. Press, Inc. 85-110.

Kimmel, B. E., Ole-Moiyoi, O. K. and Young, J. R. (1987). Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. Mol. Cell. Biol. *7*, 11465-1475.

Klobutcher, L. A., Jahn, C. L. and Prescott, D. M. (1984). Internal sequences are eliminated from genes during macronuclear development in the ciliated protozoan *Oxytricha nova*. Cell *36*, 1045-1055.

Kugimiya, W., Ikenaga, H. and Saigo, K. (1983). Close relationship between the long terminal repeats of avial leukosis-sarcoma virus and *copia*-like movable genetic elements of *Drosophila*. Proc. Natl. Acad. Sci. USA *80*, 3193-3197.

Lipps, H. J. (1985). A reverse transcriptase like enzyme in the developing macronucleus of the hypotrichous ciliate Stylonychia. Curr. Genet. *10*, 239-243.

Lutzker, S. G. and Alt, F. W. (1989). Immunoglobulin heavy-chain class switching. in Mobile DNA. Ed. D. E. Berg and M. M. Howe. Washington, D.C.: American Society for Microbiology. 693-714.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular Cloning a Laboratory Manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Manning, J. E., Schmid, C. W. and Davidson, N. (1975). Interspersion of repetitive and non-repetitive DNA sequences in the *Drosophila melanogaster* genome. Cell 4, 141-155.

Marschalek, R., Brechner, T., Amon-Böhm, E. and Dingerman, T. (1989). Transfer RNA genes: Landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. Science *244*, 1493-1496.

Martindale, D. W. (1989). Codon usage in Tetrahymena and other ciliates. J. Protozool. *36*, 29-34.

Martindale, D. W. and Bruns, P. J. (1983). Cloning abundant mRNA species present during conjugation of *Tetrahymena thermophila*: Identification of mRNA species present exclusively during meiosis. Mol. Cell. Biol. *3*, 1857-1865.

Martindale, D. W. and Taylor, F. M. (1988). Multiple introns in a conjugation-specific gene from *Tetrahymena thermophila*. Nucl. Acids Res. *16*, 2189-2201.

McClintock, B. (1984). The significance of responses of the genome to challenge. Science *226*, 792-801.

Moerman, D. G. and Waterson, R. H. (1989). Mobile elements in *Caenorhabditis elegans* and other nematodes. in Mobile DNA. Ed. D. E. Berg and M. M. Howe. Washington, D.C.: American Society for Microbiology. 537-556.

Morin, G. B. and Cech, T. R. (1988). Phylogenetic relationships and altered genome structures among Tetrahymena mitochondrial DNAs. Nucl. Acids Res. *16*, 327-346.

Morisato, D. and Kleckner, N. (1984). Transposase promotes double strand breaks and single strand joints at Tn10 termini *in vivo*. Cell *39*, 181-190.

Orgel, L. E. and Crick, F. H. C. (1980). Selfish DNA: The ultimate parasite. Nature *284*, 604-607.

Orias, E. (1986). Ciliate conjugation. in The Molecular Biology of Ciliated Protozoa. Ed. J. G. Gall. Orlando, Florida: Academic Press, Inc. 45-84.

Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. (1989). Identification of four conseved motifs among the RNA-dependent polymerase encoding elements. EMBO J. *8*, 3867-3874.

Preer, J. R., Jr., Preer, L. B., Rudman, B. M. and Barnett, A. J. (1985). Deviation from the universal code shown by the gene for surface protein 51A in Paramecium. Nature *314*, 188-190.

Ribas-Aparicio, R. M., Sparkowski, J. J., Proulx, N. E., Mitchell, J. D. and Klobutcher, L. A. (1987). Nucleic acid splicing events occur frequently during macronuclear development in the protozoan *Oxytricha nova* and involve the elimination of unique DNA. Genes & Development *1*, 323-336.

Robinson, E. K., Cohen, P. D. and Blackburn, E. H. (1989). A novel DNA deletion-ligation reaction catalyzed *in vitro* by developmentally controlled activity from Tetrahymena cells. Cell *58*, 887-900.

Rogers, J. H. (1985). Origin and evolution of retroposons. in Genome Evolution in Prokaryotes and Eukaryotes. Ed. D. C. Reanney and P. Chambon. Orlando, Florida: Academic Press, Inc. 271.

Rogers, M. B. and Karrer, K. M. (1989). Cloning of Tetrahymena genomic sequences whose message abundance is increased during conjugation. Dev. Biol. *131*, 261-268.

Ruan, K.-S. and Emmons, S. W. (1984). Extrachromosomal copies of transposon Tc1 in the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA *81*, 4018-4022.

Rubin, G. M. and Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable elcment vectors. Science *218*, 348-353.

Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. and Yuki, S. (1984). Idintification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. Nature *312*, 659-661.

Shepherd, J. C. W. (1981). Method to determine the reading frame of a protein from the purine/pyrimidine genome sequence and its possible evolutionary justification. Proc. Natl. Acad. Sci. USA*78*, 1596-1600.

Shiba, T. and Saigo, K. (1983). Retrovirus-like particles containing RNA homologous to transposable element *copia* in *Drosophila melanogaster*. Nature *302*, 119-124.

Sogin, M. L., Igold, A., Karlok, M., Nielsen, H. and Engberg, J. (1986). Phylogenetic evidence for the acquisition of ribosomal RNA introns subsequent to the divergence of some of the major Tetrahymena groups. EMBO J. *5*, 3625-3630.

Spangler, E. A., Ryan, T. and Blackburn, E. H. (1988). Developmentally regulated telomere addition in *Tetrahyeman thermophila*. Nucl. Acids Res. *16*, 5569-5585.

Stein-Gavens, S., Wells, W. M. and Karrer, K. M. (1987). A germ line specific DNA sequence is transcribed in Tetrahymena. Dev. Biol. *120*, 259-269.

Streck, R. D., MacGaffey, J. E. and Beckendorf, S. K. (1986). The structure of *hobo* transposable elements and their insertion sites. EMBO J. *5*, 3615-3623.

Sudaresan, V. and Freeling, M. (1987). An extrachromosomal form of the Mu transposons of maize. Proc. Natl. Acad. Sci. USA *84*, 4924-4928.

Tausta, S. L. and Klobutcher, L. A. (1989). Detection of circular forms of eliminated DNA during macronuclear development in *E. crassus*. Cell *59*, 1019-1026.

Varmus, H. and Brown, P. (1989). Retroviruses. in Mobile DNA. Ed. D. E. Berg and M. M. Howe. Washington, D.C.: American Society for Microbiology. 53-108.

Warmington, J. R., Anwar, R., Newlon, C. S., Waring, R. B., Davies, R. W., Indge, K. J. and Oliver, S. G. (1986). A "hot-spot" for Ty transposition on the left arm of chromosome III. Nucl. Acids Res. *14*, 3475-3485.

Weindruch, R. H. and Doerder, F. P. (1975). Age-dependent micronuclear deterioration in *Tetrahymena pyriformis*, syngen 1. Mechanisms of Ageing Development *4*, 263-279.

Xiong, Y. and Eickbush, T. H. (1988). The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol. Cell. Biol. *8*, 114-123.

Yao, M.-C. (1982). Elimination of specific DNA sequences from the somatic nucleus of the ciliate Tetrahymena. J. Cell Biol. *92*, 783-789.

Yao, M.-C., Chio, J., Yokoyama, S., Austerberry, C. F. and Yao, C.-H. (1984). DNA elimination in Tetrahymena: A developmental process involving extensive breakage and rejoining of DNA at defined sites. Cell *36*, 433-440.

" I trust that posterity will read these statements with a feeling of proud and justified superiority."

1939 Worlds Fair time-capsuel