Genomic Organization and Developmental Fate of Adjacent Repeated Sequences in a Foldback DNA Clone of *Tetrahymena thermophila*

Almuth H. Tschunko,[1] Robert H. Loechel, Ning C. McLaren and Sally Lyman Allen[2]

Department of Biology, University of Michigan, Ann Arbor, Michigan 48109

Manuscript received September 2, 1986
Revised copy accepted July 20, 1987

ABSTRACT

DNA sequence elimination and rearrangement occurs during the development of somatic cell lineages of eukaryotes and was first discovered over a century ago. However, the significance and mechanism of chromatin elimination are not understood. DNA elimination also occurs during the development of the somatic macronucleus from the germinal micronucleus in unicellular ciliated protozoa such as *Tetrahymena thermophila*. In this study foldback DNA from the micronucleus was used as a probe to isolate ten clones. All of those tested (4/4) contained sequences that were repetitive in the micronucleus and rearranged in the macronucleus. The presence of inverted repeated sequences was clearly demonstrated in one of them by electron microscopy. DNA sequence analysis showed that the left portion of this clone contains three tandem, directly repeated copies of a 340-bp sequence, a 120-bp portion of which appears in inverted orientation at a 1.6-kb distance. This clone, pTtFB1, was subjected to a detailed analysis of its developmental fate. Subregions were subcloned and used as probes against Southern blots of micronuclear and macronuclear DNA. We found that all subregions defined repeated sequence families in the micronuclear genome. A minimum of four different families was defined, two of which are retained in the macronucleus and two of which are completely eliminated. The inverted repeat family is retained with little rearrangement. Two of the families, defined by subregions that do not contain parts of the inverted repeat, one in the "loop" and one in the "right flanking region," are totally eliminated during macronuclear development—and contain open reading frames. A fourth family occurs in the "loop" region and is rearranged extensively during development. The two gene families that are eliminated are stable in the micronuclear genome but are not clustered together as evidenced by experiments in which DNAs from nullisomic strains are used to map family members to specific micronuclear chromosomes. The inverted repeat family is also stable in the micronuclear genome and is dispersed among several chromosomes. The significance of retained inverted repeats to the process of elimination is discussed.

SOME 100 years ago—first NUSSBAUM—and then AUGUST WEISMANN proposed the germ line theory which recognized the separate roles of the germ cells from somatic cells (reviewed by TOBLER 1986). The germ cells are those which maintain the genealogy of the species. The somatic cells are those which build up the organism. Accompanying the germ line/soma differentiation in some organisms are the phenomena of chromatin diminution and chromosome elimination. This was first observed by BOVERI (1887) in a nematode. Subsequent research by many cytologists showed that chromosome/chromatin elimination occurs not only in nematodes, but in six orders of insects, arachnids, copepods, in some vertebrates, and in plants (TOBLER 1986). More recently, molecular analysis has shown that somatic cell variations may arise during normal development as a result of DNA loss and rearrangements. Rearrangements have been found in such diverse systems as immunoglobin switching in mammals (TONEGAWA 1983), antigen variation in Trypanosomes (BORST and CROSS 1982), and mating type interconversion in yeast (NASMYTH 1982). Each of these rearrangements may involve some sort of DNA elimination.

Chromatin elimination also occurs in the ciliated protozoa during formation of the somatic macronucleus from the germinal micronucleus. In certain groups of ciliates (holotrichs) DNA elimination is modest. For example, 10–20% of the micronuclear genome is eliminated from the macronucleus of *Tetrahymena thermophila*. In other groups of ciliates (hypotrichs) 90–95% of the micronuclear genome is absent from the mature macronucleus, since elimination of DNA sequences occurs on a grand scale during macronuclear development. The ciliate macronucleus can therefore be used as a convenient model system for studying the process of DNA elimination.

In *T. thermophila* other genomic reorganizations occur with development. The retained 80–90% is rearranged, fragmented into 300–500 subchromosomal molecules to which are added tandem repeats of

[1] Present address: Department of Biology, Marietta College, Marietta, Ohio 45750.
[2] To whom correspondence should be addressed.
The sequence data presented in this article have been submitted to the EMBL/GenBank Data Libraries under the Accession No. Y00605.

the C_4A_2 hexanucleotide the new telomeres, and then amplified to produce a mature macronucleus containing 45 genomic equivalents of DNA (reviewed in BRUNK 1986).

With the advent of DNA cloning and Southern analysis, elimination of micronuclear DNA sequences could be examined directly. Many different DNA clones of *T. thermophila* have been investigated as to their fate during macronuclear development. Analysis of randomly selected micronuclear clones suggests that 20–25% contain sequences that are lost from the mature macronucleus (KARRER 1983; YAO 1982). Some cloned micronuclear sequences are completely eliminated, especially those that contain C_4A_2 repeats (BRUNK *et al.* 1982; CHERRY and BLACKBURN 1985; YAO 1982; YOKOYAMA and YAO 1984). Others are retained but rearranged (HOWARD and BLACKBURN 1985; WHITE and ALLEN 1985, 1986; WHITE, EL-GEWELY and ALLEN 1985; YAO *et al.* 1984). Although both repetitive and single copy sequences are represented among the cloned sequences that are eliminated, families of repeated sequences are especially abundant. Thus most of the sequences that are retained in the macronucleus are present in single copy, with only a few exceptions to the rule so far reported (ALLEN *et al.* 1985; ALLITTO and KARRER 1986; KIMMEL and GOROVSKY 1976; MARTINDALE, MARTINDALE and BRUNS 1986).

Elimination of sequences occurs at two different types of sites (reviewed in BRUNK 1986). It occurs at or near fragmentation sites, and it occurs at sites internal to the subchromosomal macronuclear DNA molecules. At the sites internal to the macronuclear DNA molecules deletion of DNA is accomplished by breakage and reunion of the retained flanks (YAO *et al.* 1984). An analysis of 20 randomly selected macronuclear DNA clones suggested that there are about 5000 of these deletion sites in the genome, each separated on the average by about 33 kb of DNA (YAO *et al.* 1984).

In the vicinity of each elimination site a "signal" sequence would be expected to be present that would facilitate the deletion process. Especially important might be a sequence capable of folding up to produce a particular secondary structure. One such type of sequence would be inverted repeated sequences. Indeed, short inverted repeats have been reported at sites where DNA is eliminated (KLOBUTCHER, JAHN and PRESCOTT 1984; YAO, ZHU and YAO 1985). Much longer inverted repeats were identified in the foldback DNA fraction from the amicronucleate strain of *T. pyriformis* GL (BORCHSENIUS; BORKHSENIUS *et al.* 1978) but have not been studied in either the micronuclear or macronuclear genomes of micronucleate strains of the *T. pyriformis* complex. The foldback DNA fraction studied by BORCHSENIUS and his colleagues had interesting properties. It was present in 4%, or more, of the macronuclear genome and its average percentage of G + C content was similar to unfractionated DNA. Each repeat unit was at least 230 bp in length and present in low copy number in the genome. Some of the units appeared to be organized as large blocks into which single copy DNA was imbedded. These segments appeared to be discrete in length and of different lengths since sharp bands were seen in 5% polyacrylamide gels (BORCHSENIUS and MERKULOVA 1980).

The purpose of this study was to isolate clones of micronuclear DNA which contain long inverted repeated sequences and then investigate whether they played a role during macronuclear development and differentiation. Ten foldback DNA clones were isolated. One clone (pTtFB1) includes both halves of a 120-bp inverted repeat separated by a central loop sequence. The "left" half is part of a larger 340-bp sequence which is itself tandemly repeated. A segment of the loop and the right flanking sequences are eliminated from the mature macronucleus. However, the inverted repeated sequences define a family of sequences that is retained in the macronucleus, with little rearrangement, and is present in significant copy number.

MATERIALS AND METHODS

Strains: The following cell lines of *T. thermophila* were used in this study: DI21481 (inbred strain D × congenic strain D/1), B2079X6 (inbred strain B), CU362 (nulli-3), CU383 (nulli-4), CU354 (nulli-5), CU359 (nulli-2,3,5), CU371 (nulli-2R,1R), CU414 (nulli-2,3,4,5) and CU373 (nulli-1R). The origin of the inbred strains and congenic strains has been previously outlined (ALLEN and GIBSON 1973; ALLEN *et al.* 1984). The nullisomic strains were derived from B strain clones in P. J. BRUN's laboratory (Cornell University), and have been described in detail (BRUNS and BRUSSARD 1981).

Growth of cultures and DNA isolation: Cultures were maintained in 1% proteose-peptone at 16° and transferred monthly. When fungal contamination was present, as it was in the cultures of the nullisomic strains, the culture medium was amended with the two fungicides Fungizone (0.25 mg/ml) present in Antibiotic-Antimycotic (Gibco Laboratories) and nystatin (5 units/ml) sold as Mycostatin (Squibb). Growth of larger cultures for nuclear isolation was described by ALLEN *et al.* (1983). Micronuclei and macronuclei of high purity were isolated by the Percoll procedure as detailed by ALLEN *et al.* (1983). Micronuclear preparations routinely had only 0.2–0.5% macronuclear DNA contamination, while macronuclear preparations were contaminated with only 0.02–0.05% micronuclear DNA. After lysis of the nuclei, the DNA was purified on a cesium chloride gradient.

Isolation of a foldback fraction from micronuclear DNA: Micronuclear DNA in 0.12 M potassium phosphate buffer (KPB), pH 6.8, was broken by sonication to give fragments 200–2000 bp in length. The sonicated DNA was denatured by boiling for 5 min, cooled on ice, and passed through a 50° hydroxylapatite column. Up to 0.5 hr elapsed between boiling and completing the passage through the column; thus a C_0t value of 3.3 was theoretically attained. Some repeated DNA would have renatured by then, but not unique sequences (ALLEN and LI 1974; BORCHSENIUS *et al.* 1978). The column was rinsed with 0.12 M KPB to elute

the single-stranded DNA and then with 0.5 M KPB to elute the double-stranded DNA. The double-stranded DNA fractions were pooled, dialyzed against 0.1 × SSC (SSC = 0.15 M NaCl, 0.015 M sodium citrate), and concentrated by means of butanol extraction.

The concentrated fraction of foldback DNA was ethanol precipitated and resuspended in S1 buffer (30 mM acetate buffer, pH 4.6, 300 mM NaCl, 4.5 mM $ZnCl_2$), using proportions of 27 μg foldback DNA/100 ml volume/9 units of S1 nuclease (Bethesda Research Laboratories). Digestion was carried out at room temperature for 30 min. The digestion mixture was deproteinated by adding half a volume of 20% diethyl pyrocarbonate in chloroform/isoamyl alcohol (24:1), shaken for 2.5 hr in the cold, centrifuged to separate the phases, and the aqueous layer was reextracted once more with 1 hr shaking. To the aqueous phase ⅕ volume of 4 M ammonium acetate was added before ethanol precipitation.

Cloning, screening, subcloning, and isolation of plasmid DNA: Micronuclear DNA from strain DI21481 was partially digested with *Mbo*I and cloned into the *Bam*HI site of pKH47, a pBR322 derivative. The clones were transformed into *Escherichia coli* strain RH202 (ADAMS *et al.* 1979) and the transformed colonies were stored at −80°. The clone bank was screened by colony hybridization using S1 nuclease digested, foldback fraction from micronuclear DNA that had been labeled by nick-translation with [α-^{32}P]-dATP (RIGBY *et al.* 1977; MANIATIS, JEFFREY and KLEID 1975).

Subclones were also made. After separating fragments by electrophoresis through low melting point agarose, the fragments to be used as vector and insert were cut out. The DNA was extracted by adding four volumes of buffer [0.3 M NaCl, 10 mM Tris-HCl, pH 7.4, 2 mM disodium ethylenediamine tetraacetate (EDTA), 1% (w/v) SDS], melting at 65°, mixing and then serially extracting the protein first with buffer-saturated phenol, then with (25:24:1) phenol/chloroform/isoamyl alcohol, and last with (24:1) chloroform/isoamyl alcohol. The DNA was precipitated by addition of ethanol. In the latest experiments, fragments were separated on regular agarose, cut out and extracted by means of the IBI Analytical Electroeluter (International Biotechnologies, Inc.) according to their instructions. Ligations to the plasmid vector pBR322 were carried out according to MANIATIS, FRITSCH and SAMBROOK (1982). The subclones were transformed into *E. coli* strain HB101 and the transformed colonies were stored at −80°.

Small amounts of plasmid DNA were isolated according to BIRNBOIM and DOLY (1979). For greater purity, phenol and chloroform extractions were done before the ethanol precipitation step. When 1–2-mg quantities were needed, a modified SDS/high salt cleared lysate procedure was used (GUNSALUS, ZURAWSKI and YANOFSKY 1979), followed by CsCl/ethidium bromide gradient centrifugation in a vertical rotor.

Heteroduplexes and electron microscopy: Denaturation solutions contained 12–15 μg/ml *Pst*I-cut pBR322, 12–15 μg/ml *Pst*I-cut clone, 0.02 M EDTA, pH 7.4, and 0.1 M NaOH. Denaturation was carried out for 10 min at room temperature. The denaturation solution was adjusted to about 0.1 M Tris, pH 7.1, and 50% formamide, and incubation was continued for 30 min at room temperature to allow renaturation. For mounting the DNA onto grids, the hypophase solution was 5% formamide in water or in 0.01 × TE (0.01 M Tris, 0.001 M EDTA), pH 8.5. Because of the low percentage of G+C content of *T. thermophila* (25% G+C) a 35% hyperphase spreading solution was used: 35% formamide, 0.1 M Tris, pH 8.5, 0.01 M EDTA, 0.5 μg/ml of heteroduplexed DNA and of each molecular standard (pBR322 for double-stranded DNA, M13 phage for single-stranded DNA), and 0.04 mg/ml cytochrome *c*, type VI (Sigma C-7752). This hyperphase was spread immediately and the DNA was mounted onto grids. The grids were then stained with 5 × 10^{-5} molar uranyl acetate, destained with 90% ethanol, and shadowed, with 80:20 platinum-palladium in a Kinney SC-3 High Vacuum Evaporator (The New York Air Brake Company). A Phillips 300 electron microscope and Kodak electron microscope film 4489 were used.

Restriction mapping, gel electrophoresis and blotting: Digestions using tenfold excess of enzyme were carried out by the method of MANIATIS, FRITSCH and SAMBROOK (1982) using restriction endonucleases purchased from Bethesda Research Laboratories, Amersham, Boehringer-Mannheim, and New England Biolabs. Several different DNA preparations for each nuclear source were used in replicate experiments. The use of excess enzyme and different DNA preparations helped to control the potential problem of partial digestion. Genomic DNA was allowed to digest overnight. Plasmid DNA was digested for a few hours. For cloning purposes *Pvu*II was allowed to digest only 15 min to avoid damage to the DNA ends by contaminating exonucleases in the *Pvu*II enzyme preparation. When DNA was to be digested with two different enzymes which required two different buffer conditions, the enzyme requiring the lower salt concentration was allowed to digest first. Then the solution was adjusted to the second buffering conditions.

Submarine gels consisted of 0.6–2.4% agarose with ethidium bromide (50 μg/100 ml) dissolved in Tris borate buffer (89 mM Tris, 89 mM boric acid, 2.5 mM EDTA, pH 8.3). One-fifth volume of 30% glycerol saturated with bromphenol blue was added to the DNA sample, and the mixture was incubated at 65° for 10 min before loading. Electrophoresis of genomic DNA fragments was usually carried out at 35 mA for 2–3 days. Electrophoresis of digested plasmid DNA was carried out at higher amperage for a shorter period of time.

The method of SOUTHERN (1975) was employed to transfer DNA from the gel to a nitrocellulose filter, with slight modifications. The agarose gels were irradiated for 5 min on a 360 nm transilluminator to nick the DNA and thus improve the transfer to nitrocellulose. The gel was then soaked in: 0.5 M NaOH, 1.5 M NaCl for 30 min, water for 5 min, 0.5 M Tris-HCl (pH 7), 3 M NaCl for 30 min and last 20 × SSC for 5 min. The nitrocellulose (0.45 μm pore size) was equilibrated in 2 × SSC for 5 min. The blot was left overnight. Then the nitrocellulose filter was soaked, DNA-side up, in 2 × SSC for 5 min, air-dried for 30 min, and baked for 1.5–2 hr in an 80° vacuum oven.

Radioactive DNA labeling, hybridization and autoradiography: The nick-translation kit from Bethesda Research Laboratories was used according to their instructions. DNA was labeled with [α-^{32}P]dATP and routinely gave 10^8 cpm/μg of DNA. Unincorporated nucleotides were separated from the DNA by passing the mixture through a 10 ml Sephadex G-50 column, rinsing with TE buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA).

Prehybridization, hybridization and all of the rinses were done at 62° because of the low (25%) G+C content of Tetrahymena DNA. In a heat sealable bag, the nitrocellulose filter was soaked in prewarmed 3 × SSC for 30 min at 62°. This was replaced with the prehybridization solution: 0.2% bovine serum albumin, 0.2% Ficoll, 0.2% polyvinyl pyrrolidone, 0.1% SDS, 3 × SSC, 50–100 μg/ml sheared calf thymus DNA, which had been boiled 10 min and cooled on ice. After 6 hr at 62° the solution was replaced with the hybridization solution: for a full-sized 350 cm^2 filter, 30 ml of prehybridization solution plus 4 × 10^7 cpm of probe, boiled 10 min and quick-cooled on ice. Hybridization at 62°

proceeded for at least 36 hr. The filters were then rinsed at 62° six times in 2 × SSC, 1% SDS for 20 min each, once in 0.2 × SSC, 1% SDS for 30 min, and once in 2 × SSC for 10 min. The filters were briefly air-dried, wrapped in Saran wrap, and exposed to Kodak XAR-5 film with an intensifying screen at −80°C. In some experiments the radioactive probe was stripped from the filter using several washes at 90° in probe-stripping buffer (modified slightly after S. HOROWITZ, University of Rochester, personal communication). The buffer contained: 0.02% w/v bovine albumin, 0.02% Ficoll, 0.02% polyvinyl pyrrolidone, 0.02% SDS, 0.4 mM sodium pyrophosphate, 0.4 mM EDTA (free acid), 0.2 mM disodium EDTA, 5 mM Tris (pH 8.0) in 0.3 × SSC. The filter was then reprobed with a different labeled sequence.

Biotin labeling of DNA: The DNA was labeled by nick-translation with two different biotin nucleotides present simultaneously (biotin-11-dUTP and biotin-7-dATP). The prehybridization and hybridization procedure was the same as for ^{32}P-labeled probes except that incubation was done at 54° due to the lower affinity of biotin-labeled probes for DNA. The protocols followed for the post hybridization rinses and the color detection assay were modifications of the procedure provided by Bethesda Research Laboratories with their biotin detection kit.

DNA sequencing: The ^{35}S-dideoxy (Sanger) sequencing method was used. The directions provided with the sequencing kit from Bethesda Research Laboratories were followed with few modifications. The DNA fragments to be sequenced were cloned into the RF (replicative form) of M13mp10 or M13mp11. *E. coli* strains JM101 or Tg1 were transformed with these clones. Using the minipreparation procedure of BIRNBOIM and DOLY (1979), RF preparations were made from numerous putative positive clones (clear phages on X-gal overlay). RF preparations were digested and the resulting fragments were separated by electrophoresis to determine the size of the inserts. The sequence of clones with inserts of the correct size was then determined. Valid sequences were those that (a) were from M13 clones having inserts of correct sizes, (b) had the correct cloning junction, and (c) had all of the restriction sites known from the detailed restriction map. The sequences were analyzed by several of the BIONET programs (IntelliGenetics, Inc., Palo Alto, California).

RESULTS

Isolation and preliminary characterization of foldback DNA clones: The foldback fraction of micronuclear DNA makes up only 1–2% of the micronuclear genome as determined by hydroxylapatite chromatography. A similar percentage was found for macronuclear DNA. The fraction from micronuclear DNA was isolated, purified, digested with S1-nuclease, radioactively labeled by nick-translation, and used as a probe in colony hybridization. A library of micronuclear DNA clones in pKH47 was probed in duplicate. Ten of the colonies showing intense hybridization were restreaked and reprobed with another preparation of the foldback fraction of micronuclear DNA. All ten hybridized with the probe. All ten putative clones were retransformed into *E. coli* strain HB101.

The ten clones were tested by a dozen restriction endonucleases to find enzymes that (a) cut singly into the vector (needed for linearization), or (b) cut the plasmid in the insert (needed for restriction mapping of the clone). Crude maps were made of all ten clones. When tested for cross homology, the clones could be arranged into about five groups (1, 3, 4, 8; 7, 9, 10; 2; 5; 6). Four of the clones that showed no cross-reactions were then tested for their repetitiveness in genomic DNA and for rearrangement during macronuclear development. *Hind*III or *Eco*RI digests of micro and macronuclear DNA were blotted and hybridized in duplicate with each of the clones. In each case the cloned sequence appeared to be repeated in the micronucleus, and repeated but rearranged in the macronucleus (data not shown).

Presence of an inverted repeated sequence (IRS) in the clones: The foldback DNA clones (pTtFB clones) were originally isolated by probing the micronuclear DNA library with S1-nuclease digested foldback fraction of micronuclear DNA and choosing the colonies which hybridized most strongly. This screening procedure may have yielded clones containing (a) a full IRS, (b) half of an IRS only, or (c) one or more copies of a repeated non-inverted sequence, which was isolated because enough undigested copies of tail sequences were left in the probe, or because some of the most highly repeated sequences of the genome could renature quickly enough to be isolated on hydroxylapatite along with the foldback DNA fraction. Therefore, the presence of an IRS in the pTtFB clones cannot be assumed but must be demonstrated.

Heteroduplexes of each of the ten clones with pBR322 were examined electron microscopically for the presence of a full IRS in the insert. The homoduplex structure of a very short IRS would look the same as two DNA strands lying across each other. Thus this procedure selects for inverts that are at a minimum 50–100 bp in length. Only clone pTtFB1 clearly showed a homoduplex structure in the insert. Under 50% formamide hyperphase conditions, half of the molecules showed an apparent IRS bounding a large loop sequence (*i.e.*, transposon-like structure). Under less stringent conditions—35% formamide, which is the correct renaturing condition for Tetrahymena's 25% G+C content—all heteroduplexes show this homoduplexing IRS. In addition, most also show an apparent point of contact within the loop near the IRS (Figure 1).

There was noticeable heterogeneity in the distance of the major homoduplexed region from the two ends of the insert as indicated by the *arrows* in Figure 1. The ratio of the lengths of two flanking regions is about 1:1 in Figure 1a, about 1:1.5 in Figure 1b, and about 1:4 in Figure 1c. The DNA sequencing data explain this heterogeneity. Clone pTtFB1 does indeed contain a large 120 bp IRS; however, the "left" copy of the IRS is part of a 340 bp sequence which is tandemly repeated three times at the "left" end of the

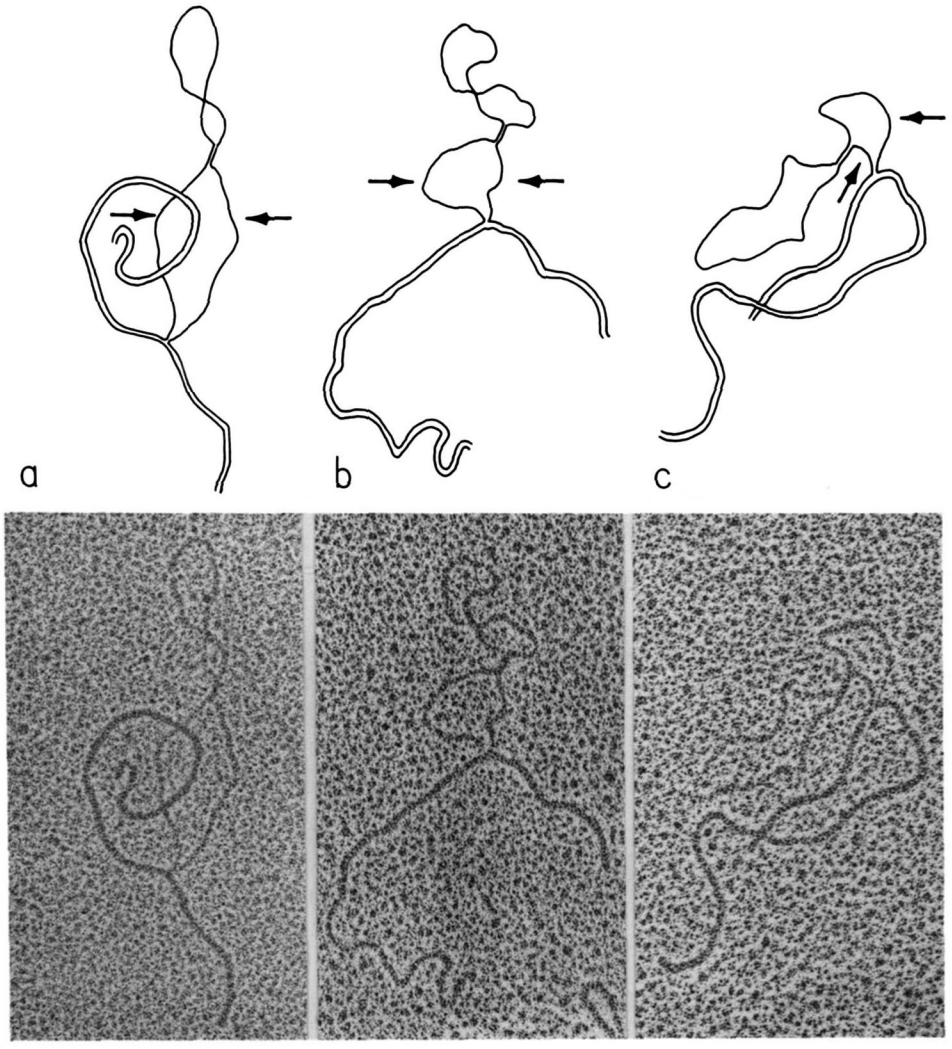

FIGURE 1.—Electron micrographs of heteroduplexes of pBR322 and pTtFB1 (35% formamide). Note the heterogeneous location of the homoduplexed region within the loop. The ratio of the lengths of the two flanking regions is (a) 1:1, (b) about 1:1.5, (c) about 1:4. Also, (a) and (b) show a second point of contact within the loop.

clone (Figure 2). Clone pTtFB1 therefore has four copies of the repeated 120 bp sequence—from left to right the directly repeated 3L, 2L, and 1L copies and the inverted R copy. The R copy is separated from the 1L copy by the 1.6-kb "loop" region. When the "R" copy of the IRS family bonds to the "1L" copy of the left tandem repeat, the flank length ratio is 900 bp:900 bp, or the observed 1:1 ratio. When the "R" copy bonds to the "2L" copy, the flank length ratio is 600 bp:900 bp, or the observed 1:1.5. When the "R" copy bonds to the "3L" copy, the ratio is 250 bp:900 bp, or the observed 1:4.

Computer analysis of the sequence data revealed the following features: (a) The R copy of the 120 bp sequence is approximately 83% homologous to 2L, the middle member of the three left end representatives (see *top* section of Figure 3 for their comparison). (b) Within each copy of the 120 bp sequence occur pairs of short direct repeats (Figure 3). (c) Homology between two copies of the inverted repeat (2L and R, for example) deteriorates gradually at the ends. (d) Because the ends of the inverted repeated sequence are not clearly delineated, it is difficult to look for transposon-like target site duplications. The repetition of the left part of the IRS argues against this IRS being a typical transposable element. (e) The two tandem repeats of the 340 bp region for which the sequence is completely known are approximately 93% homologous to each other (Figure 4). (f) There are several possible short inverted sequences within the A+T-rich loop region which could correspond to the minor contact point seen electron microscopically (data not shown).

Restriction map, terminology and subclones: Figure 2 also shows the restriction map of pTtFB1. The sites containing six base pair sequences recognized by restriction enzymes are indicated above the top map. Sites with 4 or 5 bp recognition sequences are indicated below the maps. Note that the 6 bp sequences are clustered in the right flanking region and that the

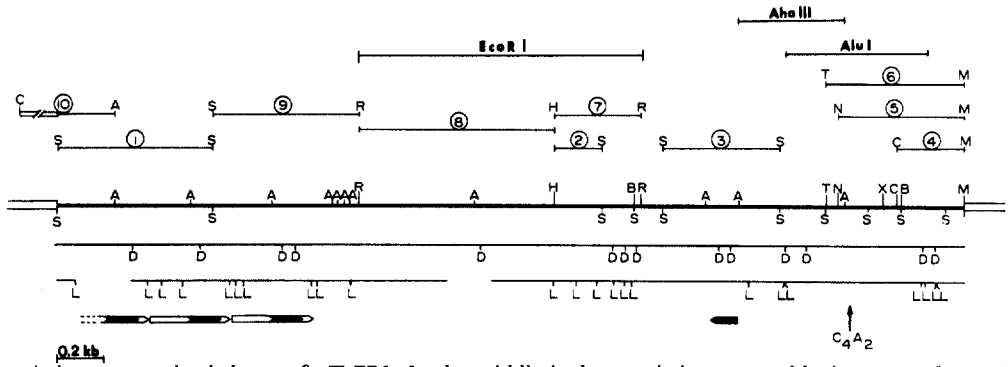

FIGURE 2.—Restriction map and subclones of pTtFB1. In the middle is the restriction map, with six-cutters shown above and a four-cutter shown below the map. Above the map are indicated the ten subclones generated, ① through ⑩, as well as the *Alu*I fragment, *Aha*III fragment and *Eco*RI fragment used as probes in the experiments. Below the map is the *Dde*I (D) map and the *Alu*I (L) map. Horizontal arrows below these maps indicate the locations of the 120-bp inverted repeated sequences (*black arrows*) and the 340-bp direct repeats (*open arrows*). The location of a single copy of the C_4A_2 hexanucleotide is indicated. The *Bam*HI site was recreated only at the cloning junction on the right. A = *Aha*III, B = *Bgl*II, C = *Cla*I, D = *Dde*I, H = *Hind*III, L = *Alu*I, M = *Bam*HI, N = *Hin*cII, R = *Eco*RI, S = *Sau*3A, T = *Hae*III, X = *Xba*I. The sequence of 3.3 kb of the 3.7-kb insert is known.

```
          220        230        239        249        259
211  CAGTOCTOTA GGTTATGGTT TAGCAA-ATT GCATTAATCT CTCAAATTTG   2L
          **  *          *          *  *  *
21   CAG-GCT-TA GGTTATGCTT TAGCAATATT TCACTAATCT CTCAAATTTG   R
            28         38         48         58         68

          269        279        289        299        309
     AAACTTGATC TTAAGTAAAA ACAGTTTATT TGTTTTGGAT TGTGTTATTT   2L
          *  *       *          *             *  *
     ACACTTGGTC TTAGGTAAAA ACAGTTTTTT TGTTTTGGAT TGTTTTTTTT   R
            78         88         98        108        118

          319        329
     TTAATTATAA AAACTAATTA A      330                       2L
          **         **  •
     TTCTTTATAA AAATGAATTT A      139                       R
            128        138
```

```
          10         19         29         39         49
1    AATTGGTGCA -ACAGGTGCA TCAGGCTTAG GTTATGCTTT AGCAATATTT   R
               ****                   *  *       **  **
188  AATTGGTGCA TTTGGGTGCA TCAGGCTTAG GTTCTGGTTT AGCAAATT8G   1L
          197        207        217        227        237

          59         69         79         89         99
     CACTAATCTC TCAAATTTGA CACTTGGTCT TAGGTAAAAA CAGTTTTTTT   R
          *          *          **                    *
     CATTAATCTC TCAAATT4GA CACTTTATCT TAGGTAAAAA CAGTTTATTT   1L
          247        257        267        277        287

          109        119        129        139
     GTTTTGGATT GTTTTTTTTT TCTTTATAAA AATGAATTTA AAA      142   R
                  **  •      **         **   *
     GTTTTGGATT GTGATATTTT TAATTATAAA AACTAATTAA AAA      330   1L
          297        307        317        327
```

FIGURE 3.—Homology between the "2L," "1L" and "R" copies of the inverted repeated sequence family. (Refer to Figure 2.) *Top*: alignment of the "2L" sequence (*upper line*) and the "R" sequence which is shown inverted (*lower line*). *Bottom*: Alignment of the "R" sequence which is shown inverted (*upper line*) and the "1L" sequence (*lower line*). Pairs of short, direct repeats are indicated. BIONET's ambiguity code: 0 = may be G, 8 = may be T.

```
          11         21         31         41         51
TTTCTTTTTC TTTTTATTGA GCTTGCTTTT TTTTGTTCTT TCTTTGTTAA   1L
                     *
TTTCTTTTTC TTTTTATTGT GCTTGCTTTT TTTTGTTCTT TCTTTGTTAA   2L
          10         20         30         40         50

          61         71         81         91        101
TTTAGCTCAT CATTTTATTT TCAACTTTTG CTTTTATTAT TAATTAATCT   1L
                     *
TTTAGCTCAT CATTTTATTT TCTACTTTTG CTTTTATTAT TAATTAATCT   2L
          60         70         80         90        100

         111        121        131        141        151
TATTATTTCA TATTTTTTCC TCATTTTTTC CTTCTAATAA ATTATTTAAT   1L
                                             **
TATTATTTCA TATTTTTTCC TCATTTTTTC CTTCTAATAA GCTATTTAAT   2L
         110        120        130        140        150

         160        170        180        189        199
CATATA-TTT TTCTTCAATA TTTAAAAAGT AACAATTA-A ATTGGTGCAT   1L
*    *                          * *  *  *        *
GATATATTTT TTCTTCAATA TTTAAAAAGT TAGAAOTAGA ATTGGTG-AT   2L
         160        170        180        190        199

         209        216        226        236        246
TTGGGTGCAT -CAG-GCT-T AGGTTCTGGT TTAGCAAATT 8GCATTAATC   1L
*           *  **  *       •                 *
ATGGGTGCAT OCAGTOCTOT AGGTTATGGT TTAGCAAATT -GCATTAATC   2L
         209        219        229        239        248

         256        266        276        286        296
TCTCAAATT4 GACACTTTAT CTTAGGTAAA AACAGTTTAT TTGTTTTGGA   1L
           *  *  *       *
TCTCAAATTT GAAACTTGAT CTTAAGTAAA AACAGTTTAT TTGTTTTGGA   2L
         258        268        278        288        298

         306        316        326        336
TTGTGATATT TTTAATTATA AAAACTAATT AAAAACAATA AAGCTCTTT    1L
    *
TTGTGTTATT TTTAATTATA AAAACTAATT AAAAACAATA AAGCTCTTT    2L
         308        318        328        338
```

FIGURE 4.—Alignment of the middle 340-bp sequence (2L) and the right 340-bp sequence (1L). BIONET's ambiguity code: 0 = may be G, 8 = may be T, 4 = probably T.

restriction enzyme *Aha*III has eleven sites in pTtFB1. This abundance of *Aha*III sites is not surprising because Tetrahymena DNA has only 25% G+C content and the recognition site for *Aha*III is the hexanucleotide TTTAAA.

The "left" portion of the clone is defined as that portion which contains the tandem, directly repeated sequences and is the end of the insert closest to the *Eco*RI site of the vector DNA in clone pTtFB1. The "right" portion of the clone contains the cluster of *Xba*I-*Cla*I-*Bgl*II restriction sites and is the end of the insert closest to the *Sal*I site of the vector DNA in the clone.

Various subregions of clone pTtFB1 were cloned

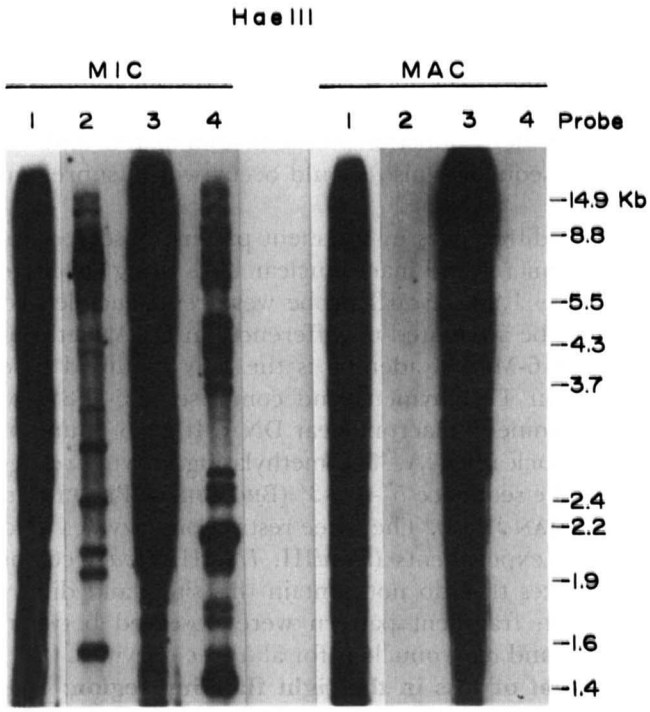

FIGURE 5.—Genomic repetitiveness and developmental rearrangement of subregions of pTtFB1. *Hae*III-digested micronuclear (MIC) and macronuclear (MAC) DNA was probed with subclones 1, 2, 3 and 4. Size markers are indicated (lambda DNA digested with *Ava*I plus *Bgl*II).

into pBR322 to obtain the ten subclones indicated in Figure 2. "Region 2" or "sequence 2" of pTtFB1 is cloned to give "subclone 2" which is labeled to give "probe 2." When "probe 2" is hybridized to Southern blots of digested genomic DNA, the resulting banding pattern defines "family 2."

pTtFB1 is a valid micronuclear DNA clone: To check for cloning artifacts, micronuclear DNA was digested with various combinations of enzymes whose sites bracketed region 2. Southern blots of the gels were probed with subclone 2. In each digestion, a genomic fragment of the expected size (based on the restricted map) did hybridize to the probe (data not shown). These data argue against pTtFB1's being a cloning artifact.

Clone pTtFB1 is not a contaminating macronuclear DNA clone from the clone bank because region 2 (in the loop) and region 4 (right flank) are absent from the macronuclear genome (see below).

Genomic organization: Four of the subclones of pTtFB1 were used as probes in hybridization experiments against genomic blots of micronuclear and macronuclear DNA. Figure 5 shows the results of an experiment in which micronuclear DNA and macronuclear DNA were separately digested with *Hae*III. The fragments were separated by electrophoresis, transferred onto nitrocellulose by Southern blotting, and probed with subclones 1 (left flank), 2 (portion of the loop), 3 (right copy of the inverted repeated sequence), and 4 (right flank).

In the micronuclear genome, all four subclones define families of repeated sequences (Figure 5). Families 1 and 3 (defined by subclones 1 and 3) are present in higher copy number than the loop and right flank families (2 and 4, respectively). Subclone 10 (which contains the left-most 240 bp *Sau*3a-*Aha*III subregion of subclone 1) and subclone 3 were labeled with biotin in order to get clearer bands in the hybridization patterns. These two subclones defined micronuclear families having similar banding patterns (data not shown). The similarity of the two patterns may be due to (a) the homology of the partial 3L copy of the inverted repeated sequence family in subclone 10 and the R copy in subclone 3, or (b) false cross homology due to the low %G+C content of the DNA flanking these inverted repeated sequence copies in the subclones.

In the macronuclear genome, the inverted repeated sequence family (families 1 and 3) remains abundant (Figure 5). Using biotin-labeled probe 1, an array of thin, tightly spaced bands was observed in micronuclear and macronuclear DNA in contrast to the black smudges seen when the probe was labeled with ^{32}P (compare Figures 5 and 6). This allowed comparison of the banding patterns in micronuclear and macronuclear DNA. The patterns were very similar especially if the filter was held at a slant in examining the lanes. (The pattern of thin, closely spaced bands was stronger on the filter than in the photograph). The similarity in banding pattern is also evident in autoradiograms using short exposure times (see Figure 8a). Thus, the majority of the bands are not rearranged in the macronucleus.

In contrast to the behavior of the inverted repeated sequence family, all members of family 2 (a 220-bp portion of the loop) and family 4 (in the right flanking region) appear to be eliminated during macronuclear development (Figure 5). The procedure was sensitive enough to detect plasmid DNA at concentration levels equivalent to single copy DNA (data not shown).

Extent of loss in the loop region: Although family 2 is eliminated, the whole loop region bounded by the inverted repeated sequence is not eliminated during macronuclear development. This retention was determined initially by using a 1.2-kb *Eco*RI fragment from the loop region as a probe against micronuclear as well as macronuclear DNA which had been singly digested with *Hind*III, *Hae*III, or *Hpa*I (data not shown). This 1.2-kb loop probe hybridized to a repeated micronuclear family as expected. However, this family was retained and extensively rearranged during macronuclear development. Thus the sequences present in the entire loop region do not act as a unit during macronuclear development.

The 1.2-kb *Eco*RI fragment encompasses the subregions cloned in subclones 7 and 8 (see Figure 2). Subclone 7 includes the region cloned as subclone 2

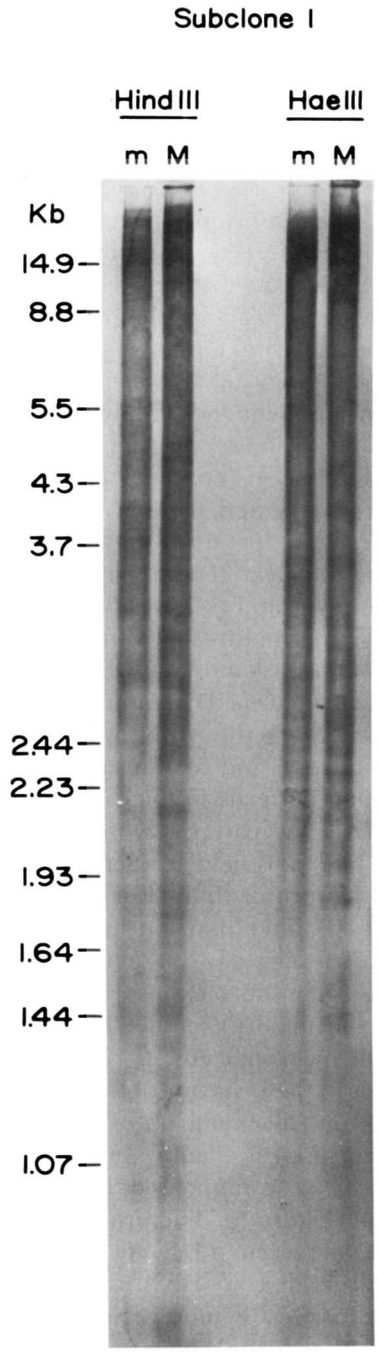

FIGURE 6.—Double biotin-labeled probe hybridized against micronuclear and macronuclear DNA. Either *Hind*III or *Hae*III was used to digest 10 μg of micronuclear DNA (m) and of macronuclear DNA (M). A 1300-bp *Eco*RI-*Sal*I fragment which contains the insert was isolated from subclone 1 doubly-labeled with biotin-11-dUTP and biotin-7-dATP, and used as a probe. Size markers are indicated (lambda DNA digested with *Ava*I plus *Bgl*II).

as well as the *Sau*3A-*Eco*RI region immediately to the right. When subclone 7 was used as a probe against micronuclear and macronuclear DNA which had been singly digested with *Hind*III or *Hae*III, it hybridized to a repeated micronuclear family but it did not hybridize to macronuclear DNA (Figure 7b). Thus all of subregion 7 is eliminated. When subclone 8 was used as a probe against micro- and macronuclear blots, it hybridized to a family of sequences that was retained and rearranged in the macronucleus (Figure 7a). Thus the region of the loop retained in the macronucleus occurs within the subregion represented by subclone 8. The "left" junction between eliminated and retained sequences also should occur within subregion 8.

The differences in fragment pattern observed between micro- and macronuclear blots using subclone 8 or the 1.2-kb *Eco*RI probe were reproducible and cannot be attributed to differences in DNA methylation. *N*-6-Methyl adenine is the only modified base found in Tetrahymena and comprises 0.6–0.8% of the adenine in macronuclear DNA. It is not found in micronuclear DNA. The methylating enzyme recognizes the sequence 5′-AT-3′ (BROMBERG, PRATT and HATTMAN 1982). The three restriction enzymes used in our experiments (*Hind*III, *Hae*III, *Hpa*I) cut at sequences that do not contain this site—and differences in fragment pattern were observed between micro- and macronucleus for all three enzymes.

Extent of loss in the right flanking region: The eliminated family in the right flanking region was defined by probing genomic DNA with subregion 4 which starts 200 bp to the right of the sequence 5′—CCCCAA—3′. An open reading frame begins immediately 3′ of this sequence (A. H. TSCHUNKO, K. H. BLAKLEY, R. H. LOECHEL, N. C. MCLAREN, and S. L. ALLEN, manuscript in preparation). The location on the map of the C_4A_2 hexanucleotide is shown in Figure 2. We wished to locate the junction between sequences that are eliminated and those that are retained in the right flanking region in the macronucleus. To this end, several probes were used that extended for various distances to the left: subclone 4; subclone 5, which contains the segment found in subclone 4 as well as an additional 240 bp and extends 40 bp to the left of the C_4A_2 hexanucleotide; subclone 6, which contains the segment found in subclone 5 as well as an additional 50 bp and extends 90 bp to the left of the C_4A_2 hexanucleotide; an *Alu*I fragment, which overlaps subclone 6 and extends 260 bp to the left of the C_4A_2; a 450 bp *Aha*III fragment which overlaps the *Alu*I fragment and lies totally to the left of the C_4A_2; and subclone 3, which overlaps the *Aha*III fragment and contains a copy of the inverted repeat. These probes were used against Southern blots of micronuclear and macronuclear DNA digested with *Hind*III, *Bgl*II, or *Eco*RI. In the *Hind*III experiment the autoradiograms were exposed for different time spans—from 1 hr to 120 hr. Some of the results are shown in Figure 8. The micronuclear pattern of bands was similar in the blots probed with subclones 4, 5, 6, the *Alu*I fragment, and the *Aha*III fragment, but differed from the pattern in the blot probed with subclone 3. The latter represents the family of sequences expected for the inverted repeat family.

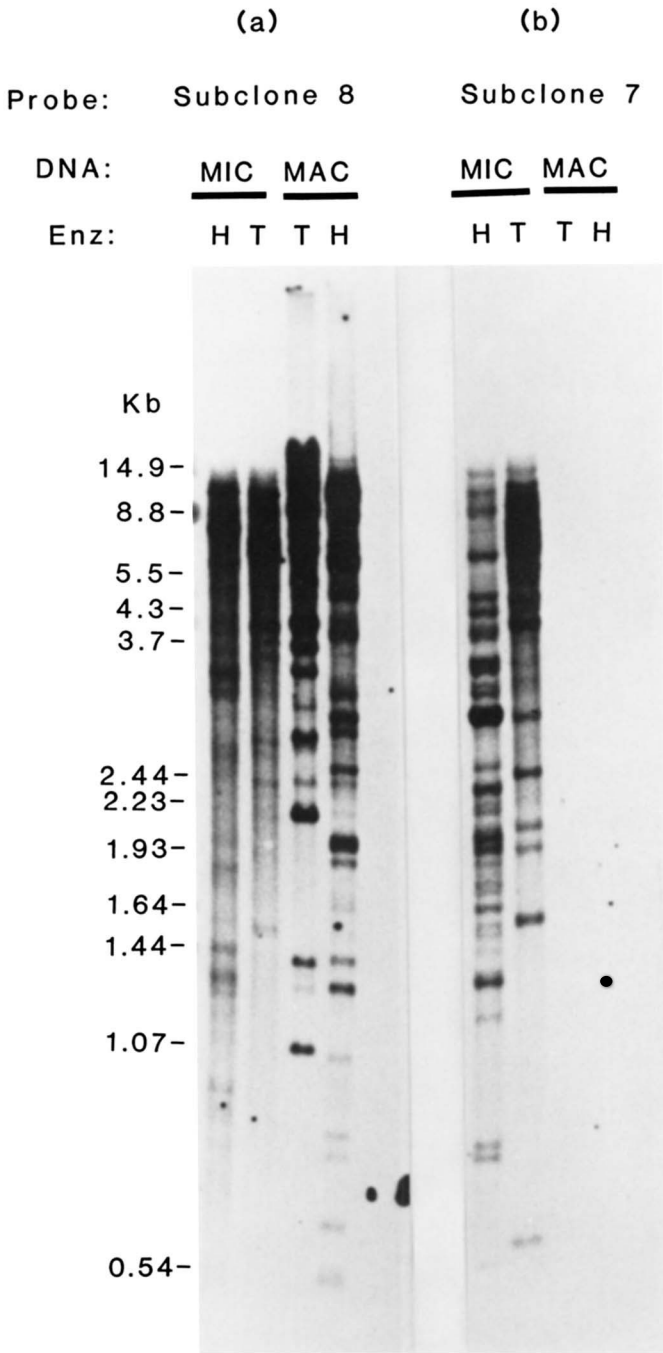

FIGURE 7.—Two families in the loop region behave differently during macronuclear development. Micronuclear (m) or macronuclear (M) DNA was digested with *Hind*III (H) or *Hae*III (T), and 5 μg of digested DNA was loaded in each lane (In this experiment *Hae*III may not have cut completely.) The blots were probed with (a) subclone 8 or (b) subclone 7. Size markers are indicated (lambda DNA digested with *Ava*I plus *Bgl*II). The autoradiograms shown were exposed for 30 hours. In a 14 day exposure (not shown) a few very faint bands were seen in the macronuclear DNA blots probed with subclone 7. These lined up with prominent micronuclear DNA bands, suggesting that the macronuclear signal was due entirely to contaminating micronuclear DNA.

There are a couple of bands that differ in the blots probed with the *Alu*I fragment compared to those probed with subclones 4, 5, and 6, and a few more differences in the blots probed with the *Aha*III fragment. Since the patterns are basically the same, the same family of sequences is being revealed. Thus, the 1 kb segment in the right flanking region appears to hybridize to a single family.

Retention of sequences in the macronucleus was seen for the inverted repeat family, as expected. Note that the pattern and the intensity of the bands in the macronuclear blot were quite similar to that of the micronuclear blot probed with subclone 3 (Figure 8a); for all other probes the signal given off by the macronuclear blots was reduced. It was considerably reduced in the case of subclones 4, 5 and 6. Note that the intensity and pattern of bands seen in the macronuclear blots at 120 hr for these probes was similar to that observed for a *1-hr* exposure of the micronuclear blots (Figure 8, d, e and f). The purity of the macronuclear DNA used in these experiments was known with respect to the degree of micronuclear DNA contamination. This figure was 0.0167%, or on an equivalent weight basis—1 μg micronuclear DNA/6000 μg macronuclear DNA—for single copy sequences. If there were 100 copies of the sequence per genome, the signal expected from micronuclear contamination in the macronuclear DNA would be 1/60. This is roughly similar to the observed signal at 120 hr. For these probes (subclones 4, 5, 6) we conclude that the signal we are seeing in the macronuclear DNA blots is due to contaminating micronuclear DNA. When the *Alu*I and *Aha*III fragments were used as probes, a reduced but slightly stronger signal was given off by the macronuclear blots (Figure 8, b and c). Moreover, the macronuclear pattern is not similar to the micronuclear pattern; instead, it is more like the macronuclear pattern seen for probe 3 (Figure 8a). Thus the sequences that are retained are shared by those hybridizing to probe 3. The most likely interpretation of these results is that the region between the inverted repeat and the C_4A_2 hexamer contains or has homology to the junction between eliminated and retained sequences.

Nullisomic mapping of families 2, 3 and 4: In nullisomic strains of *T. thermophila* the micronucleus lacks both copies of one or more of its five chromosomes. These strains are viable because the macronucleus, normal in these strains, is the nucleus responsible for gene expression. Chromosomal mapping of the members of family 2 (loop), family 3 (inverted repeat), and family 4 (right flank) was accomplished by hybridizing the subclone of interest to the *Hind*III-digested micronuclear DNA of various nullisomic strains and of the normal B strain. For example, a 2.3-kb fragment is present in the micronuclear DNA of the normal B strain (panel *b* of Figure 9). This fragment was assigned to chromosome *3* because it is absent in the micronuclear DNA of the nulli-3 strain (CU362, missing chromosome *3*) but is present in the micronuclear DNA of the nulli-4, nulli-5, and nulli-

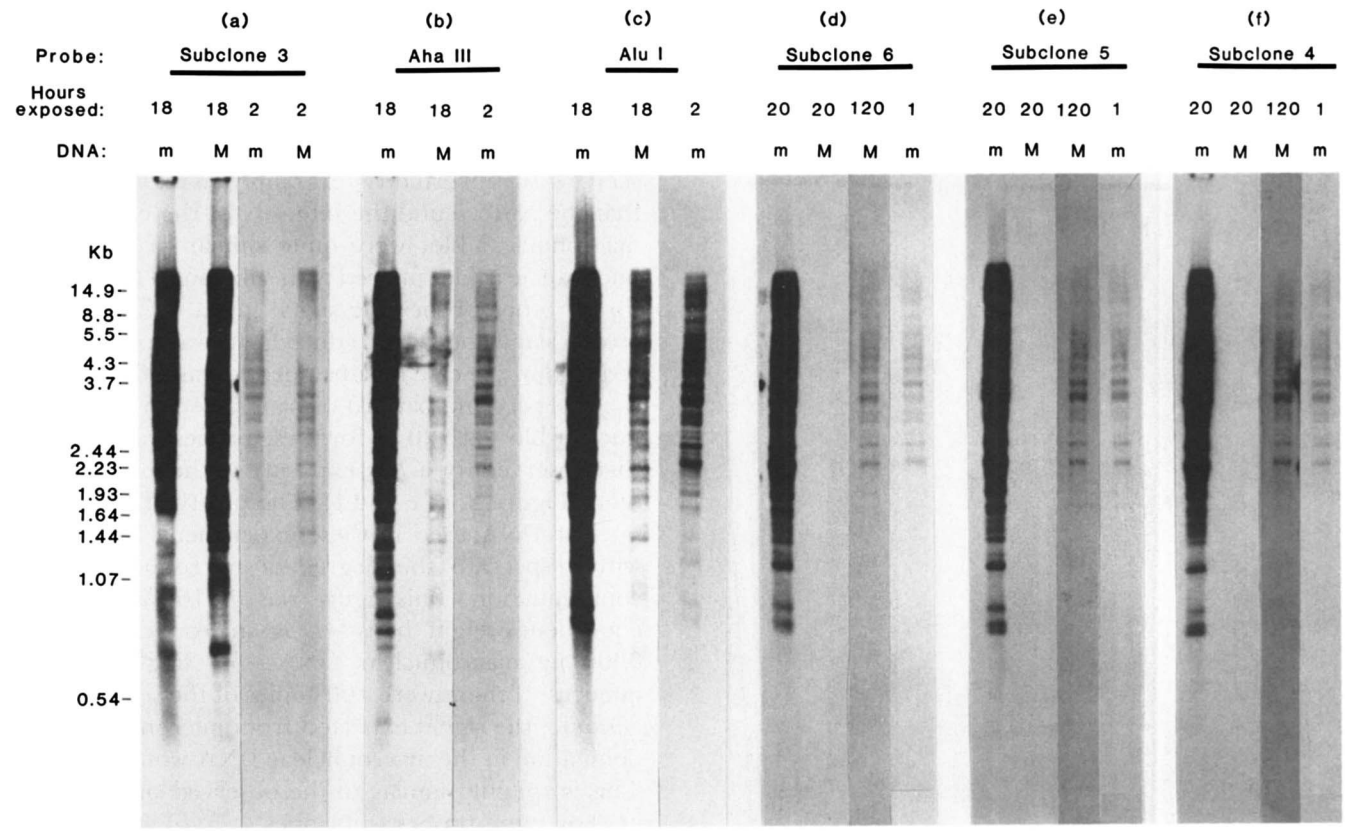

FIGURE 8.—Developmental alterations in the right flanking region of pTtFB1. Micronuclear (m) and macronuclear (M) DNA was digested with *Hind*III and probed with (a) subclone 3, (b) the *Aha*III fragment, (c) the *Alu*I fragment, (d) subclone 6, (e) subclone 5, and (f) subclone 4. Each lane contained 5 μg of digested DNA. Three sets of digestions were loaded into alternating lanes of a single gel, electrophoresis was carried out, the DNA was transferred to nitrocellulose filters, and the blots were probed with subclone 4, 5 or 6, using a total of 2×10^7 cpm for each probe. The probes were stripped off the blots and then the blots were probed with subclone 3 (1×10^7 cpm), the *Aha*III fragment (5×10^6 cpm), or the *Alu*I fragment (5×10^6 cpm). For (d), (e) and (f) the autoradiograms were exposed for 1, 20 and 120 hr at −80°. For (a), (b), and (c) the autoradiograms were exposed for 2 and 18 hours. Size markers are indicated (lambda DNA digested with *Ava*I plus *Bgl*II).

1R,2R strains. Some fragments cannot be mapped because the chromosomal deficiencies of some nullisomic strains are incompletely characterized. For example, the 1.2-kb fragment in panel *b* of Figure 10 is absent from nulli-5 and nulli-2,3,5 (mapping it to chromosome *5*) but it is present in nulli-2,3,4,5 and absent in nulli-1R (mapping it to chromosome *1*). In spite of the discrepancies between some nullisomic strains, it was determined that members of family 2 are on all five chromosomes but are nonrandomly distributed (Figure 9). About half (18) of this family's 39 *Hind*III bands are missing from the nulli-3 lane and mapped to chromosome *3* while only three to at most nine bands mapped to chromosome *1*.

The members of family 4 are also nonrandomly distributed among the five chromosomes (Figure 10). Of this family's 65 *Hind*III bands, about half (at least 27 bands) mapped to chromosome *1* and only four bands mapped to chromosome *3*. Thus the relative arrangement of the loop and the right flank as seen in pTtFB1 is not tightly conserved throughout the genome.

Only three of the nullisomic strains were used in experiments with family 3, the inverted repeat family. The strains included were CU362 (nulli-3), CU359 (nulli-2,3,5) and CU414 (nulli-2,3,4,5). Some of the fragments mapped to chromosome *3*, some to chromosome *1*, a few to chromosome *4*, and the remainder mapped to chromosome *2* or *5* (data not shown). Thus members of the inverted repeat family are distributed on at least four of the five chromosomes.

Stability of families 2, 3 and 4 in the genome: All three families appear to be quite stable in the micronuclear genome. When *Hind*III-digested micronuclear DNA from the distantly related DI and B strains were probed with subclone 2, about four of the 39 bands were different between these strains (Figure 11a). Thus the loop sequence, flanked by an inverted repeated sequence in pTtFB1, is probably not a member of a family of highly mobile transposable elements. Similarly, when the *Hae*III-digested micronuclear DNA from the DI and B strains were probed with subclone 6, about three or four of the 35 bands were different between these strains (Figure 11b). Finally, as far as we can tell, the pattern of bands seen in blots of *Hind*III-digested micronuclear DNA from the DI and B strains was similar, if not identical, when probed with subclone 3 (data not shown). However, the in-

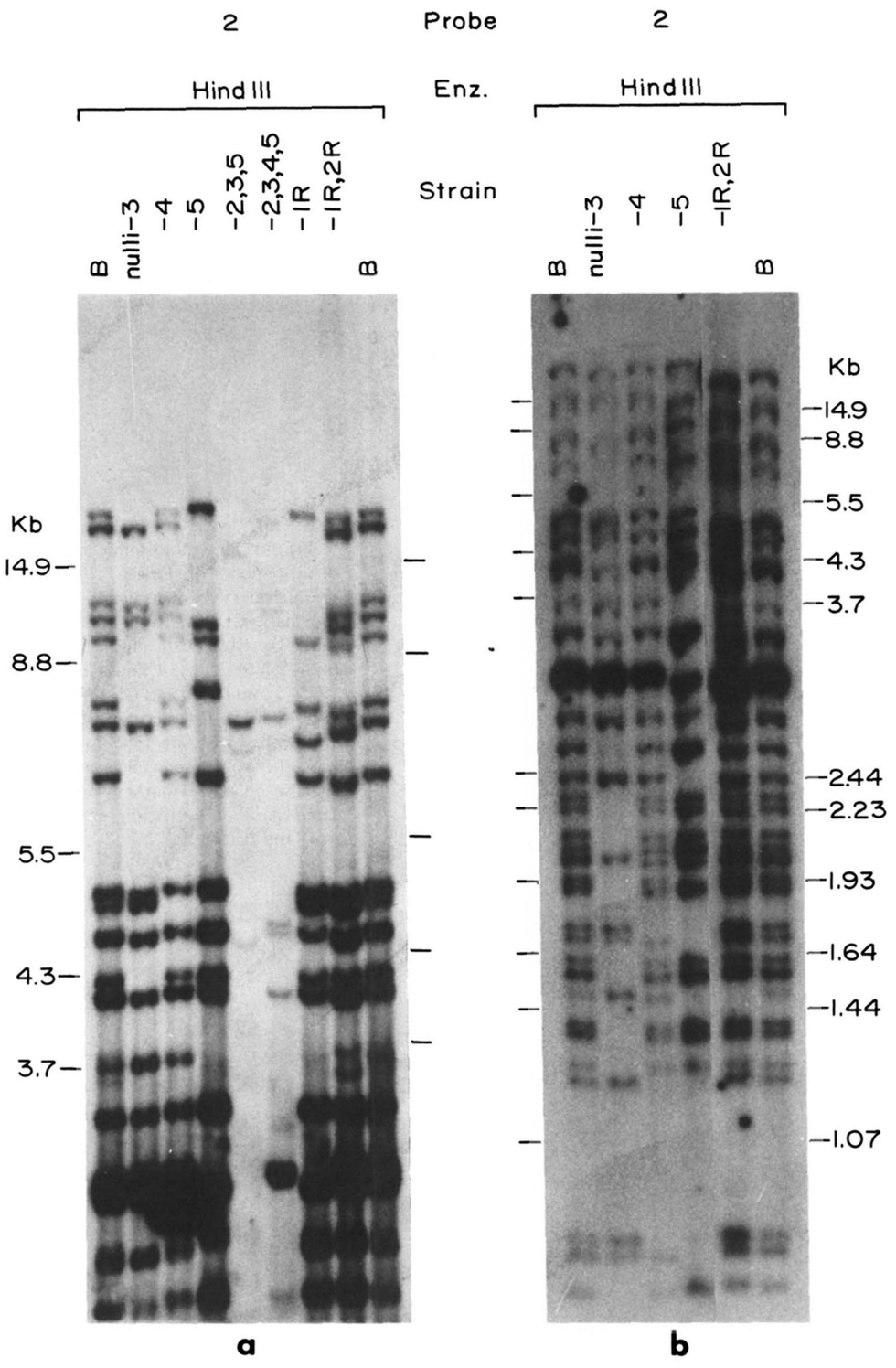

FIGURE 9.—Nullisomic mapping of subregion 2 of clone pTtFB1. Micronuclear DNA of strains B2079X6 (normal), CU362 (nulli-3), CU383 (nulli-4), CU354 (nulli-5), CU359 (nulli-2,3,5), CU414 (nulli-2,3,4,5), CU373 (nulli-1R), and CU371 (nulli-1R,2R) were digested with *Hind*III and probed with subclone 2. Panel *a* was a 0.7% agarose gel; panel *b* was a 1% agarose gel. Size markers are indicated (lambda DNA digested with *Ava*I and *Bgl*II).

verted repeat family is abundant, and differences would not be seen easily.

DISCUSSION

Foldback DNA occurs generally in eukaryotes but varies in its content and complexity from short palindromes to long interrupted inverted repeats, such as the FB elements of Drosophila, which range in length from 300 to 1500 bp (POTTER 1982). These sequences are thought to play various regulatory and structural roles because of their potential for assuming alternate structures as the DNA breathes. The foldback DNA of *T. thermophila* may also be special. We found that it comprises a similar but small percentage of the micro- and macronuclear genomes. This percentage was less than half of that reported by BORCHSENIUS *et al.* (1978) for the amicronucleate strain GL of *T. pyriformis*.

One of the ten FB micronuclear DNA clones was

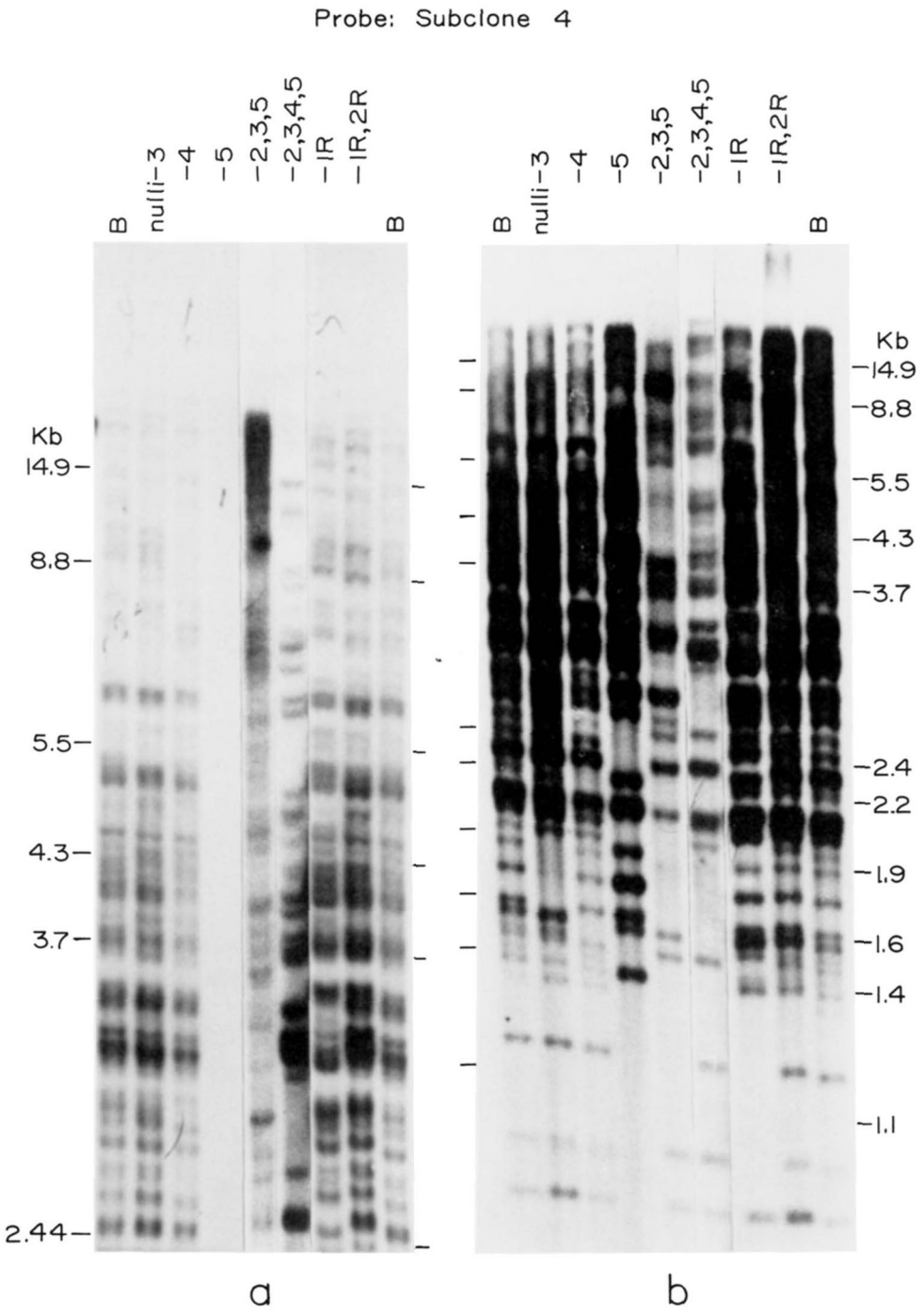

FIGURE 10.—Nullisomic mapping of subregion 4 of clone pTtFB1. Micronuclear DNA of strains B2079X6 (normal), CU362 (nulli-3), CU383 (nulli-4), CU354 (nulli-5), CU359 (nulli-2,3,5), CU414 (nulli-2,3,4,5), CU373 (nulli-1R), and CU371 (nulli-1R,2R), were digested with *Hind*III and probed with subclone 4. Panel *a* was a 0.7% agarose; panel *b* was a 1% agarose gel. Size markers are indicated (lambda DNA digested with *Ava*I and *Bgl*II).

shown by heteroduplex analysis to contain both halves of a 120-bp inverted repeat. DNA sequencing showed that this was not a typical inverted repeat, since the left copy is part of a 340-bp sequence which is tandemly repeated at least three times. This more complex structure may explain BORCHSENIUS observations that individual inverts were organized into larger units. The G+C content of the 120 bp invert is 25%, the same percentage as the G+C content of unfractionated DNA. A similar finding was reported by BORCHSENIUS and his colleagues for their foldback DNA fraction (BORKHSENIUS *et al.* 1978). However, the G+C content of the direct repeat portion of the 340-bp sequence is much lower (18–19%). This suggests an organization on the left in which the inverts are imbedded in low G+C DNA.

The uncharacteristic presence of not just one but three tandem copies in inverted orientation on the left argues against pTtFB1 being a typical transposable element. So does the lack of any obvious direct duplication of the host sequence on either side of the inverted repeat. Moreover, the inverted repeat rep-

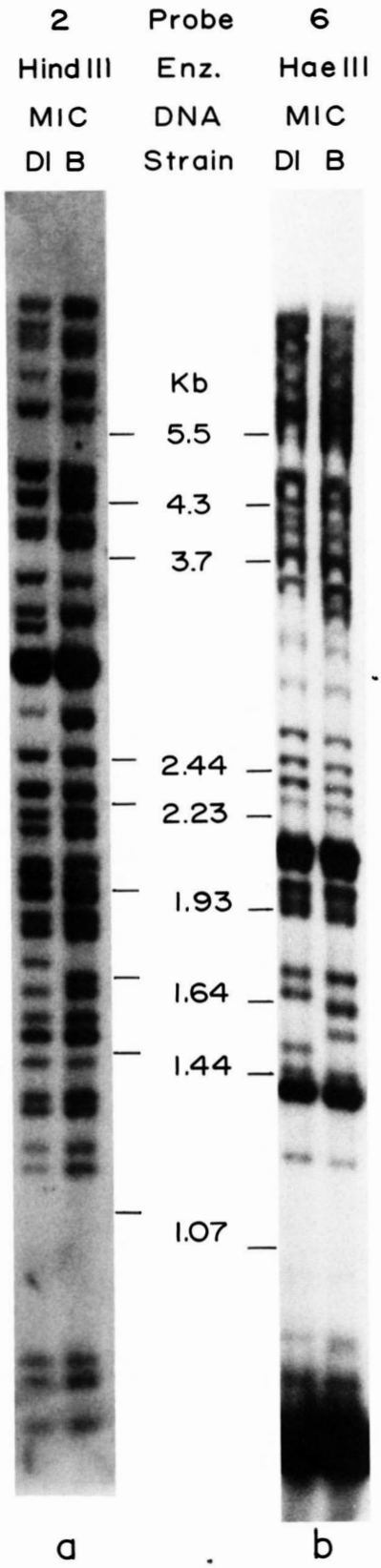

FIGURE 11.—Stability of families 2 and 4 (6) in genome. *Hind*III-digested micronuclear DNA of strains DI21481 and B2079X6 were probed with subclone 2 (panel *a*) and *Hae*III-digested micronuclear DNA of strains DI21481 and B2079X6 were probed with subclone 6 (panel *b*). Size markers are between the two panels (lambda DNA digested with *Ava*I and *Bgl*II).

resents a family that has some 100, or more, members in the micronucleus, is stable in the micronuclear genome, and is retained in the macronucleus with little rearrangement. The only reported transposable element of *T. thermophila* is the Tel-1 family of elements which is completely eliminated from the macronucleus (CHERRY and BLACKBURN 1985).

The inverted repeat family has 100, or more members which are retained in the macronucleus with only about one-third rearranged. The size of this family is much larger than that of the inverted repeats described in the macronucleus of *T. pyriformis* GL by BORCHSENIUS and his colleagues. There, 80% of the zero-time renaturation fraction included single copy sequences and 20% included sequences repeated "tens of times" (BORKHSENIUS *et al.* 1978). The size of the inverted repeat family reported here is considerably larger than the small families described by ALLITTO and KARRER (1986) or MARTINDALE, MARTINDALE and BRUNS (1986) but is similar in magnitude to the 5S RNA or tRNA gene families which number in the hundreds in the macronucleus (ALLEN *et al.* 1985; KIMMEL and GOROVSKY 1976; PEDERSEN *et al.* 1984).

A single C_4A_2 hexanucleotide is observed 5′ of the start of the open reading frame in the right flanking region. Direct repeats of this hexanucleotide make up the macronuclear telomeres, with variations of this sequence occurring in multiple copy internally in micronuclear limited DNA (reviewed in BRUNK 1986). Since a single hexanucleotide would be expected on a random basis every 33 kb of micronuclear DNA, it is not clear what biological role, if any, is played by the single copy of this sequence in the region we have analyzed.

The different developmental fates of the regions that lie between or surround the inverted repeats of clone pTtFB1 are summarized in Figure 12. Within the loop region that lies between the left and right copies of the inverted repeat are two types of families. The right portion of the loop defines a repeated, eliminated family while the rest of the loop region defines a repeated, retained and rearranged family. The right flanking region located to the right of the "R" copy of the invert contains a repeated family that appears to be completely eliminated. Thus the inverted repeat sequences are retained in the macronucleus but are surrounded by sequences that are eliminated or highly rearranged during development. Is this arrangement of sequences coincidental or do the retained inverted repeat sequences play a role in the elimination process?

On the one hand we know that the inverted repeat family and the two eliminated families appear to be conserved in the genome judging from the few changes in restriction patterns seen between distantly related strains. On the other hand, the exact organi-

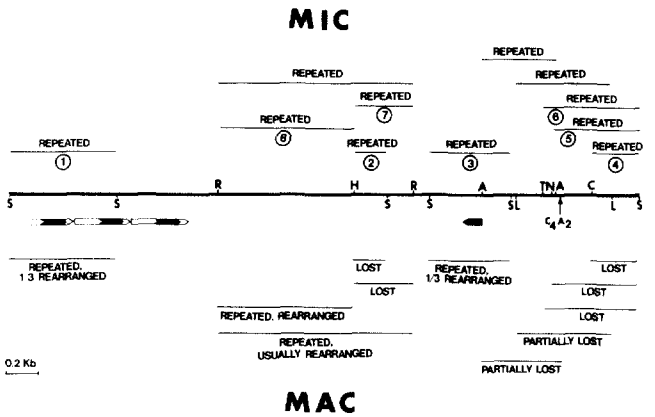

FIGURE 12.—Summary of the repetition and developmental alterations of the subregions of pTtFB1. Above the restriction map of pTtFB1 is indicated the repetition in the micronucleus (MIC) of the various subregions. Below the restriction map are indicated the repetition and developmental fate in the macronucleus (MAC) of the various subregions. The restriction map in the center has above it the six-cutters and two of the 11 *Aha*II sites, and below it only a few pertinent *Sau*3A and *Alu*I sites. (A = *Aha*III, C = *Cla*I, H = *Hind*III, L = *Alu*I, N = *Hinc*II, R = *Eco*RI, S = *Sau*3A, T = *Hae*III). Of the *wide arrows* below the map, the *black ones* represent the 120-bp inverted repeated sequences, the *open ones* represent the 340-bp direct repeats. The location of the C_4A_2 hexanucleotide is indicated.

zation of sequences seen in clone pTtFB1 does not appear to be preserved at numerous positions within the genome judging from the results of nullisomic mapping. We know that the inverted repeat family is dispersed to several micronuclear chromosomes; thus each of the eliminated families could be located near an inverted repeat. However, the two eliminated families are not located together as a rule. This conclusion is reached by the following reasoning. On chromosome *3* at least 18 members of family 2 (within the loop region) occur, but only four members of family 4 (within the right flanking region). On chromosome *1* at least 27 members of family 4 occur, but only three to nine members of family 2. If each of the bands seen in the autoradiograms (Figures 9 and 10) are single-copy, then there are at most 7–13 cases in which members of family 2 and 4 are potentially near each other, although not necessarily in the arrangement seen in pTtFB1.

Elimination of micronuclear sequences occurs at fragmentation sites and at deletion sites internal to the subchromosomal macronuclear DNA molecules (reviewed in BRUNK 1986). The number of fragmentation sites should be equal to the number of different subchromosomal DNA molecules in the macronucleus, *i.e.*, 300–500. The number of deletion sites is estimated at about 5000, each site separated by about 33 kb of DNA, provided the sites are evenly distributed in the micronuclear genome (YAO *et al.* 1984). Thus the number of internal deletion sites is ten to twenty times the number of fragmentation sites. The average amount of DNA eliminated at each site is about 8 kb (BRUNK 1986). This value was estimated using the following considerations. The total amount of DNA eliminated from the micronuclear genome (220,000 kb) is about 20%, or 44,000 kb. The total number of elimination sites is 5000 (from deletion) and 500 (from fragmentation). Dividing 44,000 kb by 5,500 sites gives the value of 8 kb. This is, however, an average value. Many of the well-characterized sites contain small deletions although clones with longer deletions are known. Those with repetitive sequences or with C_4A_2 repeats tend to be the longest (BRUNK *et al.* 1982; HOWARD and BLACKBURN 1985; WHITE and ALLEN 1986).

If elimination is completely random, we would expect deletion sites spaced 33 kb apart in the micronuclear genome. However, complete randomness in the distribution of sites may not occur. We know, for example, that the spacing may depend, in part, on the type of sequence eliminated. Unique sequences are less likely to be eliminated than are repetitive sequences, and sites involving repetitive sequences appear to be spaced more closely together (HOWARD and BLACKBURN 1985). MARTINDALE, MARTINDALE and BRUNS (1986) estimated that a total span of 50–75 kb surrounded their nonrepetitive genes that were found to be unrearranged. Their results suggested that elimination may not be random with respect to the genes near which it occurs. Out of eight cloned genes, seven of the genes were unique and none were associated with rearranged sequences, whereas all copies of the one repetitive cloned gene were associated with rearranged DNA. Nonrandomness of eliminated sequences also occurs in the hypotrichs. For example, BOSWELL *et al.* (1983) found that micronuclear-destined genes were clustered in the micronuclear genome while large blocks of repetitive DNA appeared to be eliminated. Nonrandomness of eliminated sequences also occurs in other eukaryotic organisms (reviewed in TOBLER 1986). Heterochromatic regions and particular chromosomal parts, such as telomeres and centromeres, tend to be eliminated.

What about long inverted repeats and elimination? What would the expected frequency be of random association of the retained inverted repeats with eliminated sequences? The foldback fraction represents 1–2% of the *T. thermophila* DNA, or 2200–4400 kb of the 220,000-kb micronuclear genome. On a random basis—with deletion sites spaced 33 kb apart—we would expect a total of 67–133 sites of elimination in this DNA fraction. Four of the ten foldback DNA clones were tested with respect to their developmental fate. All four contained repetitive DNA that was rearranged in the macronucleus. Altogether the cloned segments make up a 20-kb length of DNA. With a minimum of one elimination site per clone, the sites are spaced about 5 kb apart. In the case of pTtFB1 there are two segments eliminated that are about 500 bp apart, provided that the "R" copy of the inverted

repeat is retained during the processing of this region during macronuclear development. Thus sites of elimination do occur more frequently in the foldback DNA fraction than expected. It is, however, repetitive DNA that is eliminated, and repetitive DNA appears to be more frequently eliminated.

Short inverted repeats appear to be associated with DNA breakage during gene conversion of several different eukaryotic genes (KRAWINKEL, ZOEBELEIN and BOTHWELL 1986). In this case it was suggested that these sequences form stem-loop structures that promote recombination. In Tetrahymena short inverted repeats have been discovered at or near the fragmentation sites associated with the processing of ribosomal RNA, and a stem-loop structure was proposed as an intermediate in the excision of this region from the micronuclear chromosome (YAO, ZHU and YAO 1985). Here, the inverted repeats are also eliminated.

Long inverted repeats have been found at the ends of certain transposons in Drosophila, sea urchins, and Dictyostelium (POTTER 1982; LIEBERMANN *et al.* 1983; CAPPELLO, HANDELSMAN and LODISH 1985) but for the reasons given earlier we do not believe the structure of the inverts found in pTtFB1 are typical of transposons and therefore suggest that they have another role. In their case they are not eliminated but sequences that lie between or flank them may be eliminated. Their structure is also more complex in that the 120-bp invert on the left is imbedded in a larger (340 bp) direct repeat. We propose that these regions participate in the elimination process by generating structures that are recognized by specific DNA binding proteins that are involved in processing the DNA during macronuclear development. Perhaps these proteins protect the inverted repeats from being eliminated while they or other proteins cause surrounding regions to be cleaved.

Of particular interest and the subject of a subsequent communication is the nature of the germ line specific sequences that are eliminated. It was suggested by HOWARD and BLACKBURN (1985) that a primary function of eliminated sequences is in the DNA rearrangements that occur concomitantly with their elimination rather than the possible micronuclear-specific products which such sequences might encode. This may indeed be true for the eliminated sequences previously studied. However, the eliminated sequence families described here differ in being associated with inverted repeats that are retained in the macronucleus. Both the loop family and the family in the right flanking region are relatively stable in the micronuclear genome, since distantly related strains show similar restriction fragment patterns. Both families are totally eliminated from the macronucleus. In both regions in pTtFB1 there occur open reading frames, and both families may have members which are functional (A. H. TSCHUNKO, K. H. BLAKLEY, R. H. LOECHEL, N. C. MCLAREN, and S. L. ALLEN, manuscript in preparation). Perhaps one or both of these sequences encodes proteins that bind to the inverted repeat sequence under certain structural conditions to facilitate the processing of DNA during macronuclear development. An important outcome is their own elimination.

We would like to thank KAREN BLAKLEY, MIKE AVIDANO, SUE HASEGAWA, MATT HOBERG, JOHN MCAREE, THOMAS MITCHELL and BOB PRITCHARD for their help on various aspects of the research, and SUSAN L. ALLEN for her help with the BIONET analysis. This work was supported by grants from the National Science Foundation (PCM-830052) and the National Institutes of Health (GM-33112). A. H. T. was supported in part by the NIH National Research Service Award No. 1-T32-GM-07544 from the National Institute of General Medical Sciences.

LITERATURE CITED

ADAMS, J., T. KINNEY, S. THOMPSON, L. RUBIN and R. B. HELLING, 1979 Frequency-dependent selection for plasmid containing cells of *Escherichia coli*. Genetics **91:** 627–637.

ALLEN, S. L., and I. GIBSON, 1973 Genetics of Tetrahymena. pp. 307–373. In: *Biology of Tetrahymena*, Edited by A. M. ELLIOTT. Dowden, Hutchinson & Ross, Stroudsburg, Penna.

ALLEN, S. L., and C. I. LI, 1974 Nucleotide sequence divergence among DNA fractions of different syngens of *Tetrahymena pyriformis*. Biochem. Genet. **12:** 213–233.

ALLEN, S. L., T. C. WHITE, J. P. LANGMORE and M. A. SWANCUTT, 1983 Highly purified micro- and macronuclei from *Tetrahymena thermophila* isolated by percoll gradients. J. Protozool. **30:** 21–30.

ALLEN, S. L., P. R. ERVIN, N. C. MCLAREN and R. E. BRAND, 1984 The 5S ribosomal RNA gene clusters in *Tetrahymena thermophila*: strain differences, chromosomal localization, and loss during micronuclear ageing. Mol. Gen. Genet. **197:** 244–253.

ALLEN, S. L., P. R. ERVIN, T. C. WHITE and N. C. MCLAREN, 1985 Rearrangement of the 5S ribosomal RNA gene clusters during the development and replication of the macronucleus in *Tetrahymena thermophila*. Dev. Genet. **5:** 181–200.

ALLITTO, B. A., and K. M. KARRER, 1986 A family of DNA sequences is reproducibly rearranged in the somatic nucleus of *Tetrahymena*. Nucleic Acids Res. **14:** 8007–8025.

BIRNBOIM, H. C., and J. DOLY, 1979 A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7:** 1513–1523.

BORCHSENIUS, S. N., and N. A. MERKULOVA, 1980 Discrete length of palindromic segments from *Tetrahymena* DNA. FEBS Lett. **119:** 90–92.

BORCHSENIUS, S. N., N. A. BELOZERSKAYA, N. A. MERKULOVA, V. G. WOLFSON and V. I. VOROB'EV, 1978 Genome structure of *Tetrahymena pyriformis*. Chromosoma **69:** 275–289.

BORKHSENIUS, S. N., N. A. BELOZERSKAYA, N. A. MERKULOVA and V. I. VOROB'EV, 1978 Palindromic, repetitive, and unique sequences in the DNA of the macronucleus of the infusorian *Tetrahymena pyriformis* GL. Mol. Biol. **12:** 676–688.

BORST, P., and G. A. M. CROSS, 1982 Molecular basis for trypanosome antigenic variation. Cell **29:** 291–303.

BOSWELL, R. E., C. L. JAHN, A. F. GRESLIN and D. M. PRESCOTT, 1983 Organization of gene and non-gene sequences in micronuclear DNA of *Oxytricha nova*. Nucleic Acids Res. **11:** 3651–3663.

BOVERI, T., 1987 Über differenzierung der zellkerne während der furchung des eies von *Ascaris megalocephala*. Anat. Anz. **2:** 688–693.

BROMBERG, S., K. PRATT and S. HATTMAN, 1982 Sequence specificity of DNA adenine methylase in the protozoan *Tetrahymena thermophila*. Nucleic Acids Res. **11:** 5131–5145.

BRUNK, C. F., 1986 Genome reorganization in *Tetrahymena*. Int. Rev. Cytol. **99:** 49–83.

BRUNK, C. F., G. S. TSAO, C. H. DIAMOND, P. S. OHASI, N. N. G. TSAO and R. E. PEARLMAN, 1982 Reorganization of unique and repetitive sequences during nuclear development in *Tetrahymena thermophila*. Can. J. Biochem. **60:** 847–853.

BRUNS, P. J., and T. B. BRUSSARD, 1981 Nullisomic *Tetrahymena*: eliminating germinal chromosomes. Science **213:** 549–551.

CAPPELLO, J., K. HANDELSMAN and H. F. LODISH, 1985 Sequence of Dictyostelium DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. Cell **43:** 105–115.

CHERRY, J. M., and E. H. BLACKBURN, 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.

GUNSALUS, R. P., G. ZURAWSKI and C. YANOFSKY, 1979 Structural and functional analysis of cloned deoxyribonucleic acid containing the *trpR-thr* regions of the *Escherichia coli* chromosome. J. Bacteriol. **140:** 106–113.

HOWARD, E. A., and E. H. BLACKBURN, 1985 Reproducible and variable genomic rearrangements occur in the developing somatic nucleus of the ciliate *Tetrahymena thermophila*. Mol. Cell Biol. **5:** 2039–2050.

KARRER, K. M., 1983 Germ-line specific DNA sequences are present on all five micronuclear chromosomes in *Tetrahymena thermophila*. Mol. Cell Biol. **3:** 1909–1919.

KIMMEL, A. R., and M. A. GOROVSKY, 1976 Numbers of 5S and tRNA genes in macro- and micronuclei of *Tetrahymena pyriformis*. Chromosoma **54:** 327–337.

KLOBUTCHER, L. A., C. L. JAHN and D. M. PRESCOTT, 1984 Internal sequences are eliminated from genes during macronuclear development in the ciliated protozoan *Oxytricha nova*. Cell **36:** 1045–1055.

KRAWINKEL, U., G. ZOEBELEIN and A. L. BOTHWELL, 1986 Palindromic sequences are associated with sites of DNA breakage during gene conversion. Nucleic Acids Res. **14:** 3871–3882.

LIEBERMANN, D., B. HOFFMAN-LIEBERMANN, J. WEINTHAL, G. CHILDS, R. MAXSON, A. MAURON, S. N. COHEN and L. KEDES, 1983 An unusual transposon with long terminal inverted repeats in the sea urchin *Strongylocentrotus purpuratus*. Nature **306:** 342–347.

MANIATIS, T., E. F. FRITSCH and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

MANIATIS, T., A. JEFFREY and A. G. KLEID, 1975 Nucleotide sequence of the rightward operator of phage λ. Proc. Natl. Acad. Sci. USA **72:** 1184–1188.

MARTINDALE, D. W., H. M. MARTINDALE and P. J. BRUNS, 1986 Tetrahymena conjugation-induced genes: structure and organization in macro- and micronuclei. Nucleic Acids Res. **14:** 1341–1354.

NASMYTH, K. A., 1982 Molecular genetics of yeast mating type. Annu. Rev. Genet. **16:** 439–500.

PEDERSEN, D. S., M.-C. YAO, A. R. KIMMEL and M. A. GOROVSKY, 1984 Sequence organization within and flanking clusters of 5S ribosomal RNA genes in *Tetrahymena*. Nucleic Acids Res. **12:** 3003–3021.

POTTER, S. S., 1982 DNA sequence of a foldback transposable element in *Drosophila*. Nature **297:** 201–204.

RIGBY, P. W. J., M. DIECKMAN, C. RHODES and P. BERG, 1977 Labeling deoxyribonucleic acid to higher specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113:** 237–251.

SOUTHERN, E., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517.

TOBLER, H., 1986 The differentiation of germ and somatic cell lines in nematodes. pp. 1–69. In: *Results and Problems in Cell Differentiation, Vol. 13, Germ Line-Soma Differentiation*, Edited by W. HENNIG. Springer-Verlag, Berlin.

TONEGAWA, S., 1983 Somatic generation of antibody diversity. Nature **302:** 575–581.

WHITE, T. C., and S. L. ALLEN, 1985 Macronuclear persistence of sequences normally eliminated during development in *Tetrahymena thermophila*. Dev. Genet. **6:** 113–132.

WHITE, T. C., and S. L. ALLEN, 1986 Alternative processing of sequences during macronuclear development in *Tetrahymena thermophila*. J. Protozool. **33:** 30–38.

WHITE, T. C., M. R. EL-GEWELY and S. L. ALLEN, 1985 Eliminated sequences with different copy numbers clustered in the micronuclear genome of *Tetrahymena thermophila*. Mol. Gen. Genet. **201:** 65–75.

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., S.-G. ZHU and C.-H. YAO, 1985 Gene amplification in *Tetrahymena thermophila*: formation of extrachromosomal palindromic genes coding for rRNA. Mol. Cell. Biol. **5:** 1260–1267.

YAO, M.-C., J. CHOI, S. YOKOYAMA, C. F. AUSTERBERRY and C.-H. YAO, 1984 DNA elimination in Tetrahymena: a developmental process involving extensive breakage and rejoining of DNA at defined sites. Cell **36:** 433–440.

YOKOYAMA, R., and M.-C. YAO, 1984 Internal micronuclear DNA regions which include sequences homologous to macronuclear telomeres are deleted during development in *Tetrahymena*. Nucleic Acids Res. **12:** 6013–6116.

Communicating editor: J. E. BOYNTON