The Highly Conserved Family of *Tetrahymena thermophila* Chromosome Breakage Elements Contains an Invariant 10-Base-Pair Core†

Eileen P. Hamilton,* Sondra Williamson, Sandra Dunn,‡ Virginia Merriam,# Cindy Lin,§ Linh Vong,¶ Jessica Russell-Colantonio,‖ and Eduardo Orias

**Department of Molecular, Cellular and Developmental Biology, University of California at Santa Barbara, Santa Barbara, California 93106**

Received 28 October 2005/Accepted 12 January 2006

As a typical ciliate, *Tetrahymena thermophila* is a unicellular eukaryote that exhibits nuclear dimorphism: each cell contains a diploid, germ line micronucleus (MICN) and a polyploid, somatic macronucleus (MACN). During conjugation, when a new MACN differentiates from a mitotic descendant of the diploid fertilization nucleus, the five MICN chromosomes are site-specifically fragmented into 250 to 300 MACN chromosomes. The classic chromosome breakage sequence (CBS) is a 15-bp element (TAAACCAACCTCTTT) reported to be necessary and sufficient for chromosome breakage. To determine whether a CBS is present at every site of chromosome fragmentation and to assess the range of sequence variation tolerated, 31 CBSs were isolated without preconception as to the sequence of the chromosome breakage element. Additional CBS-related sequences were identified in the whole-genome sequence by their similarities to the classic CBS. Forty CBS elements behaved as authentic chromosome breakage sites. The CBS nucleotide sequence is more diverse than previously thought: nearly half of the CBS elements identified by unbiased methods have a variant of the classic CBS. Only an internal 10-bp core is completely conserved, but the entire 15-bp chromosome breakage sequence shows significant sequence conservation. Our results suggest that any one member of the CBS family provides a necessary and sufficient *cis* element for chromosome breakage. No chromosome breakage element totally unrelated to the classic CBS element was found; such elements, if they exist at all, must be rare.

In many eukaryotes, chromatin or chromosomes are eliminated or rearranged during the differentiation of somatic cells while remaining intact in germ line cells (see reference 22 for a review). Well-known examples include the programmed rearrangement of the immunoglobulin genes of the vertebrate immune system and chromatin diminution and chromosome fragmentation in the nematode *Parascaris*. Analogous processes universally occur during the differentiation of the somatic nucleus of the ciliates, which are unicellular eukaryotes. As is typical of ciliates, *Tetrahymena thermophila* has two very different but genetically related nuclei (reviewed in reference 30). The diploid micronucleus (MICN) contains the cell’s germ line, i.e., the source of the genetic information for the sexual progeny. It has five pairs of chromosomes, divides by mitosis or meiosis, and is transcriptionally silent during the vegetative stage of the cell’s life cycle. The polyploid macronucleus (MACN) is the somatic nucleus, responsible for all known vegetative gene expression and thus determining the cell’s phenotype. During sexual reorganization, new MACNs are differentiated from mitotic products of the MICN-derived fertilization nucleus. This differentiation involves site-specific fragmentation of the MICN chromosomes, de novo telomere addition to the newly created ends, and amplification of the resulting MACN chromosomes. In addition, about 6,000 MICN-specific DNA elements (known as internally eliminated sequences [IES]) are removed with ligation of the flanking sequences (9, 18, 21). IES deletion eliminates about 15% of the MICN sequences, much of it repetitive DNA. The MACN contains an estimated 280 chromosomes (8), ranging in size from 21 kb to over 3 Mb (1, 8). Their ends are capped by telomere sequences, which are ∼200- to 400-bp-long tracts of tandem repeats of the hexanucleotide GGGGT T/C CCCCAA (4, 38, 39). The 21-kb rRNA gene chromosome encodes the major ribosomal RNAs (28S and 18S) and is maintained at about 9,000 copies per MACN. All or most of the other MACN chromosomes are estimated to be maintained at about 45 copies per MACN (12).

*T. thermophila* is an excellent model organism for the study of developmentally programmed chromosome breakage because the process is very precise and uniform and has high site specificity. Fragmentation occurs at a 15-bp MICN-specific sequence, the chromosome breakage sequence (CBS). The first few identified CBSs have the sequence TAAACCAACCTCTTT (41), called here the classic CBS. Yao et al. (40) concluded that this *cis*-acting sequence is necessary and sufficient for site-specific chromosome fragmentation during ma-
cromonuclear differentiation. During fragmentation, the CBS and up to ~60 bp of flanking DNA (breakage-eliminated sequences [BES]) are deleted from the MACN genome (14, 16, 42). Previous mutational studies led to the conclusion that the CBS has a stringent sequence requirement (15). Only two CBS variants, each with a single base pair substitution, were previously reported (15, 42), but a recent study (16) found one new variant (and five junctions with variant CBSs) which suggested that CBS element variation might be greater than previously accepted.

The objective of this work was to ascertain the range of sequence variation tolerated in the CBS element and to determine whether it is the only cis-acting element capable of causing chromosome fragmentation. An impartial survey of sequences associated with chromosome breakage was undertaken by walking, from a randomly selected telomere-adjacent sequence, along the MICN DNA and across the ~75-bp BES. Thirty-one CBSs were isolated in this way, which did not rely on any knowledge or assumptions about what a chromosome breakage sequence should or would look like. All of them turned out to be related to the classic CBS. Nine additional CBSs were mined from the whole-genome sequence (available at the TIGR website [13]) based on sequence similarity to the classic CBS element. Unexpectedly, CBS variants accounted for nearly half of the breakage sites isolated, while the rest had the classic chromosome breakage sequence found by Yao et al. (41). A 10-bp core sequence within the CBS is completely conserved. Despite significant variability at its margins, the entire 15-bp segment corresponding to the chromosome breakage sequence is a sharply delimited island of sequence conservation.

**MATERIALS AND METHODS**

**Strains.** The DNAs used were from inbred strains B and C3 of *T. thermophila*, which have been previously used in making genetic maps and in characterizing CBS junctions. Strain SB210 is the inbred strain B derivative whose MACN genome was recently sequenced at TIGR (http://www.tigr.org/db/e2k1/ttg/). Nullisomes are strains missing both copies of a particular MICN chromosome or chromosome arms. All strains have been previously described (16, 17).

**PCR amplification.** Unless otherwise specified, each PCR contained 5 µl of DNA (at 5 ng/µl), 2.5 µl of 10× PCR buffer (500 mM KCl, 100 mM Tris-HCl [pH 8.3], 15 mM MgCl2, 0.01% gelatin), 2.5 µl of 10 mM MgCl2, 4 µl of deoxynucleoside triphosphates (dissolved in water at 1.25 mM each), 1.25 µl of each primer (at 4 µM), 0.125 µl of AmpliTaq DNA polymerase (catalog no. N801-0060; Perkin-Elmer) at 5 U/µl, 4 µl of 10 mM Tris-HCl [pH 8.3], 15 mM MgCl2, 0.01% gelatin), 2.5 µl of 10 mM MgCl2, 4 µl of deoxynucleoside triphosphates (dissolved in water at 1.25 mM each), 1.25 µl of each primer (at 4 µM), 0.125 µl of AmpliTaq DNA polymerase (catalog no. N801-0060; Perkin-Elmer) at 5 U/µl, and H2O as required to bring the reaction volume to 25 µl. The temperature cycling conditions varied according to the particular experiment. Primer design, analysis of PCR products, optimization of primer-anneling temperature, and search for DNA polymorphisms between inbred strains B and C3 were carried out as described in our earlier CBS studies (16).

**Construction of a telomere library using magnetic-capture PCR.** A new and efficient magnetic-capture method was used to make a library enriched for MACN telomere-adjacent sequences as follows.

(i) DNA preparation and adaptor ligation. Cell-ghost DNA was isolated from *T. thermophila* SB210 (17). The DNA was digested to completion with Sau3A, and the enzyme was heat killed (65°C, 20 min). Adaptors with a Sau3A-compatible overhang, constructed by annealing two primers (CPI [CTG ACTACTGAGCTGACGAC] and 5-phosphorylated AC1B [GATCGTCTG AGTCTAGTCTAG]) were ligated overnight at 8°C to the digested DNA. The ligated DNA was used as a template in a PCR with the CPI primer and the Tel-X primer (XCCCAACCCCAACC, where X is TEG biotin [Operon Technologies]). PCR products were as described above, except that 5 µl of DNA (at 100 ng/µl), 2 µl of 10 µM CPI primer, 2 µl of 10 µM Tel-X primer, and 0.25 µl of AmpliTaq DNA polymerase were used. Temperature cycling conditions were 5 min at 94°C, followed by seven cycles of 1 min at 94°C, 1 min at 53°C, and 6 min at 72°C, followed by a terminal extension period of 7 min at 72°C.

(ii) Purification of fragments containing a telomere sequence. PCR products containing the Tel-X primer were captured using streptavidin-paramagnetic particles. The streptavidin-coated magnetic beads (Streptavidin MagSphere paramagnetic particles; Promega) were prepared by transferring 200 µl of beads to a 1.5-ml Microtuf tube, washing them three times with equal volumes of Tris-EDTA (TE) buffer (10 mM Tris HCl [pH 8.0], 1 mM EDTA), and resuspending them in 200 µl of TE. Ten microliters of the PCR mixture was added to the beads. The beads were kept in suspension for 30 min by shaking at 500 rpm on a microcentrifuge tube shaker at room temperature. The beads were collected with a magnet (MagSphere Technology magnetic separation stand; Promega) and washed two times with 0.5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 37°C, two times with 0.5× SSC at 65°C, and once with 0.5× TE at 80°C. The beads were resuspended at each step by repeated pipetting. Fifteen seconds was allowed for bead capture by the magnet. To thermally release captured Tel-X-containing PCR products from the beads, the beads were resuspended in 10 µl of 0.5× TE and incubated for 3 min at 95°C. The beads were immobilized on the magnet for a few seconds, and the supernatant was transferred to a fresh tube. The eluted single-stranded DNA was next PCR amplified. Five microliters of the 95°C eluate was used as a template in a 25-µl typical PCR mixture where CPI and Tel14 (unbiotinylated) were the primers. Temperature cycling conditions were 5 min at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 53°C, and 2 min at 72°C, followed by a terminal extension period of 7 min at 72°C.

(iii) Cloning and sequencing of amplified inserts. PCR products were cloned into plasmid pCR2.1-TOPO (Invitrogen) and transformed into chemically competent TOP10 cells according to the supplier’s instructions. Plasmid DNA was isolated from 40 out of 48 white colonies using a QIAGEN kit. Insert DNA was PCR amplified using the CPI and Tel14 primers. Each PCR mixture contained 5 µl of plasmid DNA (at 5 ng/µl) and used CPI and Tel14 as primers. The temperature cycling conditions were 5 min at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 53°C, and 2 min at 72°C, followed by a terminal extension period of 7 min at 72°C. Plasmid DNA was sequenced using the BigDye Terminator cycle sequencing ready reaction kit (PE Applied Biosystems). Nucleotide sequences were determined using an ABI 310 gel analyzer. The great majority (90%) of the cloned inserts had a telomere sequence at one end and a CPI adaptor sequence at the other, including the Sau3A1 site. Only those inserts, referred to as “Tel clones”, were used for subsequent work.

Tel clones obtained from the above-mentioned library were serially numbered (e.g., Tel 16). A second, smaller telomere library was constructed in a similar manner for an unrelated purpose (J. Russell-Colantonio, E. P. Hamilton, and E. Orias, unpublished observations). This library used size-selected MACN chromosomes in the 1.0- to 1.2-Mb range as the starting material. Tel clones from this second library are named with a letter followed by a number (e.g., Tel B2). A random sample of 32 Tel clones, whose inserts appeared to differ in size, were selected as the unbiased CBS search set, in the absence of any previous knowledge about their associated CBSs.

**Isolation of CBS junctions by inverse PCR.** Most of the CBSs found by an unbiased search in this study were located by the strategy of walking along an MICN DNA sequence, starting from a telomere-adjacent sequence and using inverse PCR (IPCR) (42). The template for the IPCR was MICN DNA that had been digested with a restriction enzyme and then allowed to circularize and ligate in vitro. Unique divergent primers for IPCR were designed using a telomere-adjacent sequence from the Tel clones (Fig. 1A; primer sequences are in Table S1 in the supplemental material). Successful IPCR products are those that bridge the circularization junction and include a usable amount of the BES-adjacent sequence.

The IPCR was carried out as follows. Purified micronuclear DNA (16) was separately digested to completion with several restriction enzymes (see Table S1 in the supplemental material for the enzyme successfully used in each case). Extracted DNA was digested with ligase buffer to 1 ng/µl to favor intramolecular ligation and ligated with 800 units of T4 DNA ligase (New England Biolabs) overnight at 16°C. Ligated DNA was concentrated with a Microcon YM100 centrifugal filter device (Amicon) and used as a template in an IPCR, a typical PCR that used 50 to 100 ng of circularized DNA as a template and the two divergent primers. Cycling conditions were 5 min at 94°C, followed by 40 cycles of 1 min at 94°C, 1 min at 53°C, and 2 min at 68°C, followed by a terminal extension period of 7 min at 68°C; PCR products were either sequenced directly or first cloned in plasmid pCR2.1-TOPO (Invitrogen). In every case, the presence of a Tel clone sequence at one end of the IPCR insert was confirmed.

**Hybridization of Tel clones to existing CBS libraries.** The availability of classic-CBS-enriched libraries (16) provided a shortcut to obtaining the CBS junction sequence for some members of the Tel clone search set. Dot blots representing clones from those libraries were screened with a probe from each of
FIG. 1. Locations and nomenclature of PCR primers used for IPCR walks and CBS PCR tests. Open bar, MACN-destined DNA; vertically hatched bar, telomere sequence; solid bar, CBS element; stippled (including solid) bar, SES; arrows, primer locations and their 5′-to-3′ orientations. (A) Divergent specific primers for IPCR walks, designed using a Tel clone sequence (top line). Divergent proximal primers (DVP) and divergent distal primers (DVD) are relative to the telomere (in MACN DNA) or the CBS (in MICN DNA). RE, restriction enzyme sites. Note that a PCR product is obtained only when the RE-bounded MICN segment (bottom line) has been circularized. (B) Convergent specific PCR primers to test chromosome breakage function, designed using a BES-adjacent sequence. For CBS junctions isolated by IPCR, Tel CVD (convergent distal) is derived from Tel clone sequence and Tel CBS is derived from sequence on the opposite side of the BES, obtained after the IPCR walk. For CBS junctions available as a previously sequenced CBS clone, CBS C and CBS G were designed on opposite sides of the junction; see the work of Hamilton et al. (16) for more details. Note that the use of the two convergent specific primers generates products exclusively templated from MACN DNA (top line); thus, no product is seen when the chromosome arm containing the CBS is missing in the MICN of a nulli- somic strain. Using a convergent specific primer with the Tel primer generates the MACN-templated PCR product only when a telomere is added nearby after complete (or conceivably partial) chromosome breakage.

the selected Tel clones. Labeling, hybridization, and wash conditions were as described previously (16). Eight positive CBS plasmid inserts were identified (Table 1). Plasmid inserts that had not been previously characterized were sequenced.

Identifying CBS-related sequences in the Tetrahymena genome sequence read database. The whole-genome shotgun sequence read database (~1.2 million reads) was downloaded from the TIGR website (http://www.tigr.org/tdb/e2k1/H11032). Sequence reads containing a CBS-related sequence were computationally identified using a Perl script. The 15-bp segment had to satisfy one or both of the following criteria: (i) it differed from the classic CBS (TAAACCAACCTCTTT) by at most a single nucleotide (at any position) and (ii) it contained the 10-bp so-called “completely conserved core” in the correct location (i.e., TAAACCAAC-C-C-C-C, where each hyphen represents any nucleotide). BLAT searches and/or examination of the original traces through the TETRAHYMENA Genome Database allowed the exclusion of reads with putative base-calling errors (less than 5%) and three other reads having sequence traces of poor quality.

Six hundred eight valid reads gave a positive match. From each one, a 115-bp segment was extracted (a CBS tag), containing the CBS-related sequence and up to 50 nucleotides, when available, on either side. Where necessary, CBS tags were converted to the reverse complement so that every tag displayed the C strand of the CBS-related element. Independent sequence reads from the same genome location, identified by their identical CBS tags (or tag fragments), were grouped into 79 nonredundant sets. The nonredundant sets were then tested for alignment to an assembled genome sequence by BLAT tests. They were classified depending on the number of reads per set, on whether or not they were incorporated in the genome assembly, and on the sequence of the CBS-related element, as follows. (i) Group A consisted of 63 multiple-read sets. Every set represents a unique sequence assembled internally in an MACN genome scaffold. None included a CBS variant previously identified as functional. (ii) Group B consisted of six single-read sets. None were found in the MACN genome sequence assemblies or included a CBS variant previously known to be functional. (iii) Group C consisted of 10 single-read sets, each containing a 15-bp sequence already identified as being either the classic CBS or a variant of it compatible with the chromosome breakage function.

All 16 single-read sets came from authentic genome sequences, as confirmed by PCR amplification. Regions of the genome represented by reads of special interest (selected reads in group A and every single read) were further tested for compatibility with the chromosome breakage function, as described below.

| CBS name | Variant | Origin |
|----------|---------|--------|
| 1L-5     | Classic | Tel 44 plasmid |
| 1L-6     | 1A      | Tel 2 plasmid |
| 1L-7     | Classic | Tel 16 |
| 1L-8     | 1A      | Tel 3 |
| 1L-9     | Classic | Tel 30 |
| 1R-7     | Classic | Tel 1 plasmid |
| 1R-8     | Classic | Tel 36 plasmid |
| 1R-9     | 1A, 15A | Tel B2/Tel D1b |
| 1R-10    | Classic | Genome |
| 1R-11    | Classic | Genome |
| 1R-12    | Classic | Genome |
| 1R-13    | 1A, 13A | Tel A2 |
| 1R-14    | 1A      | Tel 20 |
| 1R-15    | Classic | Tel 31 plasmid |
| 1R-16    | Classic | Tel 48 |
| 1R-17    | 1A      | Tel E2 |
| 1R-18    | Classic | Tel 37 |
| 1R-19    | 1A      | Tel 15 |
| 1R-20    | Classic | Tel 22 |
| 1R-21    | 11C, 15A| Tel 29 |
| 1R-22    | 11C, 14C| Tel 45 |
| 1R-23    | Classic | Tel 28 |
| 1R-24    | 11C, 13G| Tel 25 |
| 1R-25    | 11C, 13A| Tel 41 |
| 1R-26    | Classic | Tel 39 plasmid |
| 1R-27    | 15A     | Genome |
| 1R-28    | 1A      | Tel 38 |

* Origin, source of the CBS clone; Tel n alone, IPCR; Tel n plasmid, by hybridization of the Tel clone to a CBS library clone; Genome, whole-genome shotgun sequence read.

b Tel B2 and Tel D1 represent the opposite sides of the same CBS junction, 1R-9, which is attributed to a coincidence.
Testing whether CBS-related elements are compatible with the chromosome breakage function. To further characterize CBS-related elements obtained by any method, two complementary PCR approaches were used: the nullisomic PCR mapping test (6) and the telomere-primed PCR test (16). These tests and their rationale are explained in the legend to Fig. 1B. They aim to answer several questions. (i) Does the CBS-related element reside on an MICN-limited sequence (as expected if it is functional), and if so, in what MICN chromosome arm? (ii) Do telomere addition sites flank the CBS-related element (as expected if it is functional), and if so, at what approximate distances from it on either side? (iii) Is there evidence of a partial function in chromosome breakage? (iv) Does the IPCR product correspond to a valid MICN DNA segment (as expected if no spurious products were generated during restriction and circularization of the MICN DNA used for IPCR)?

The location and guide to nomenclature of convergent specific PCR primers flanking each putative CBS used in these tests are illustrated in Fig. 1B, while primer sequences and PCR conditions are listed in Table S2 in the supplemental material. An invariably directed generic telomere primer (CCCCAAACCCCA ACC) was used for the telomere-primed PCR test. Primer melting temperatures were targeted for 56°C, the melting temperature of the Tel primer. Each (otherwise typical) PCR contained whole-cell genomic DNA as a template.

Measuring sequence conservation at CBS and flanking regions. An assumption-free search for conserved DNA motifs in the regions of chromosome breakage was made using the MEME (Multiple Expectation-maximization Motif Elicitation) motif discovery tool (2). A training set, consisting of a multFASTA file with the entire (unaligned) available sequences for the 40 CBS regions obtained by an unbiased search (Table 1), was submitted to the MEME website (http://meme.sdsc.edu/meme/meme.html). The lengths of the individual sequences ranged from 159 to 1,908 bp, totaled 29.2 kb, and averaged 729 bp. The motifs were allowed to have any sequence and length (between 6 and 50 bp), occur any number of times (0 included) in any sequence, and be present anywhere in the sequence. The entire set of published chromosome breakage regions was then searched for motifs discovered by MEME using the Motif Alignment Search Tool (MAST; http://meme.sdsc.edu/meme/mast.html) (3). For every match, the program reports its occurrence probability in a random sequence having the same background base pair composition. As a control, the same motifs were searched for in 20 randomized 29.2-kb sequences with the same overall nucleotide frequencies.

To measure and display nucleotide conservation, we used the “sequence logo” method (31), based on classic logarithmic measures of diversity (information content or entropy) (32). The level of conservation, as well as the contribution of every nucleotide to that measure, is displayed at every position of the sequence alignment. Sequence logo plots were generated by pasting aligned CBSs at the WebLogo site (http://weblogo.berkeley.edu/logo.cgi) (11). The source of the sequences was as follows. For functional CBSs, 62 published CBS junctions were used (16; this paper). To measure conservation at CBS-flanking positions, all 62 sequences were used. At CBS element positions, only the subset of 40 CBS elements obtained in an unbiased way (see above) was used. (Inclusion of CBS elements identified solely by hybridization to the classic CBS would have resulted in a spuriously high conservation measure at variable CBS positions.) For control sequences, a nonredundant set of 44 CBS-related elements obtained from the genome sequence database was used. These sequences were chosen because they contained the completely conserved CBS core sequence and were represented by multiple sequence reads. All these segments were found to be MACN destined (see Results) and thus represent nonfunctional CBS-related sequences.

Nucleotide sequence accession numbers. The accession numbers for those sequences which were previously unpublished are DQ395098 to DQ395112.

RESULTS

Chromosome breakage elements in *Tetrahymena thermophila* belong to a conserved family. To assess the functional range of sequence variation within the CBS and to determine whether a CBS is present at every site of chromosome fragmentation, a search for chromosome breakage sites in the *T. thermophila* genome was conducted in a manner independent of any previous knowledge about the chromosome breakage sequence. The approach was to walk along an MICN chromosome by IPCR, starting from randomly selected MACN-destined, telomere-adjacent sequences (see Materials and Methods). Thirty-one junctions, representing about 10% of all CBS junctions, were identified in this way. All of the chromosome breakage sites contained a recognizable member of the CBS family.

Ten additional CBS junctions were identified by a computational search of the whole-genome sequence (see Materials and Methods), based on sequence similarity to the classic CBS. All but one (CBS 3R-4 [Table 1]) represent CBS junctions not previously identified or identified exclusively by biased searches. Although not found by means of a totally unbiased search, these nine functional CBS elements do contribute unbiased information regarding variability within the CBS element.

To test whether (or reconfirm that) chromosome breakage occurs in the MICN genome near each newly identified CBS element and to determine whether breakage is all or none, two complementary tests were performed using genomic DNA: the nullisomic PCR mapping test and the telomere-primed PCR test. Their rationale is illustrated in Fig. 1B and is explained in the legend to Fig. 1 (see also Materials and Methods and references 6 and 16). In combination, these sensitive PCR tests ought to allow detection of incomplete chromosome breakage and telomere addition, which could result if a CBS element has only a partial chromosome breakage function. Every CBS element listed in Table 1 had all the features of a fully functional CBS. All 40 gave PCR bands with every nullisomic strain tested except those missing the MICN chromosome that the CBS element resides on, showing that the CBS-related element resides in an MICN-limited sequence and that chromosome breakage is complete. They also gave telomere-primed specific PCR products of the expected size, verifying the presence of correctly oriented flanking-telomere addition sites, the hallmark of chromosome breakage (data not shown). The nullisomic test also generated an assignment of each CBS to an MICN chromosome arm, used as a prefix in the final name of the CBS. The 40 CBS junctions are distributed among all the MICN chromosomes (Table 1).

Surprisingly, only slightly more than half of the sites (22 out of 40) had the classic CBS element; the remaining sites had a variant of it. Table 2 shows the full range of variation present in functional CBSs. Variation was observed at five positions (1, 11, 13, 14, and 15) of the CBS; no variation was found at a conserved core, AAACCAACC-C (where the hyphen represents position 11). These results show that the CBS element tolerates much more variability than previously suspected, but only at restricted positions.

The *Tetrahymena* genome sequence as a source of nonfunctional CBS-related sequences. The *Tetrahymena thermophila* macronuclear genome sequence provided us with an untapped, assumption-free source of nonfunctional CBS variants. The 1.2 million sequence reads in the whole MACN genome sequence database were computationally “mined” for CBS-related elements, as described in Materials and Methods.

CBS-related sequences were found at 69 distinct genomic locations (see Table S3 in the supplemental material). Sixty-three of these (group A) are represented by multiple sequence reads and are embedded at internal locations in MACN chromosome scaffolds. Since every test substantiated their macronuclear origin, we conclude that these substitutions render the CBS-related elements incompatible with the chromosome breakage function at each genome location where they occur, regardless of sequence context.

At least one representative of each single-base-pair substit-
TABLE 2. Functional sequence variation among naturally occurring T. thermophila chromosome breakage site elements

| CBS element sequence | Type or variant | No. of sequences identified by: | Unbiased searches | WGS | Prev | Total | All searches |
|----------------------|----------------|-------------------------------|-------------------|-----|------|-------|--------------|
| TAACCACTCTTTTT       | Classic        | 17                            | 5                 | 4   | 26   | 53    |              |
| A                    | 1A             | 5                             | 2                 | 7   | 12   |       |              |
| A                    | 1A, 13A        | 1                             | 1                 | 1   |      |       |              |
| A                    | 1A, 15A        | 1                             | 1                 | 1   |      |       |              |
| A                    | 11C, 13A       | 1                             | 1                 | 1   |      |       |              |
| A                    | 11C, 13G       | 1                             | 1                 | 1   |      |       |              |
| A                    | 11C, 14C       | 1                             | 1                 |     |      |       |              |
| A                    | 11C, 15A       | 1                             | 1                 | 1   |      |       |              |
| A                    | 13A            | 2                             | 2                 | 2   |      |       |              |
| A                    | 13C            | 1                             |                   |     |      |       |              |
| A                    | 14A            | 2                             | 2                 | 2   |      |       |              |
| A                    | 14A, 15A       | 1                             | 1                 | 1   |      |       |              |
| A                    | 14C            | 1                             |                   |     |      |       |              |
| A                    | 15A            | 2                             | 2                 | 2   |      |       |              |

Summary: 31 9 6 46 82

Notes:
- Dash, no nucleotide difference from the classic CBS element; underscore, completely conserved core; W, A, or T; Y, C, or T; N, any nucleotide; H, A, C, or T.
- IPCR, inverse PCR walk or Tel clone hybridized to a CBS library clone, this work; WGS, T. thermophila genome sequence, this work; Prev, previous published work.
- Isolated by any method, including biased searches and more recent unpublished observations.
- A 1A variant was found by IPCR and in the genome sequence, so it is listed only under IPCR.

TABLE 3. Selected nonfunctional CBS–related segments found by genome sequence mining

| Read set group | Substitution(s) | No. of reads | Tel-primed products | Nullisomics giving PCR products | Conclusion |
|----------------|-----------------|--------------|---------------------|--------------------------------|------------|
| A              | 2T              | 12           | None                | All                             | MAC destined |
| A              | 5A              | 11           | None                | All                             | MAC destined |
| A              | 5G              | 11           | None                | All                             | MAC destined |
| A              | 5T              | 8            | None                | All                             | MAC destined |
| A              | 6A              | 4            | None                | All                             | MAC destined |
| A              | 6A              | 10           | None                | All                             | MAC destined |
| A              | 6G              | 5            | None                | All                             | MAC destined |
| A              | 6T              | 8            | None                | All                             | MAC destined |
| A              | 9G              | 8            | None                | All                             | MAC destined |
| A              | 10T             | 5            | None                | All                             | MAC destined |
| A              | 11A             | 6            | None                | All                             | MAC destined |
| A              | 11G             | 7            | None                | All                             | MAC destined |
| A              | 13A, 14C, 15A   | 6            | None                | All                             | MAC destined |
| B              | 10A             | 1            | None                | All but MIC 2L                  | IES on MIC 2L |
| B              | 1A, 11A, 13C, 14C, 15C | 1 | None | All | IESd |
| B              | 1A, 11A, 15C    | 1            | None                | All                             | IESd |
| B              | 1A, 13A, 15C    | 2            | None                | All                             | IESd |
| B              | 1A, 13A, 15C    | 1            | None                | All                             | IESd |
| B              | 1G, 1H, 13C, 14A | 1   | None                | All but MIC 1L                  | IES on MIC 1L |

Notes:
- Underscore, allowed by itself; boldface, forbidden by itself; lightface, untested function by itself.
- Specific primer flanking each CBS-like element produced specific PCR products of the expected size when paired with the telomere primer.
- PCR products of the expected size were obtained with all nullisomics or all but MIC X (all except those missing MIC chromosome X).
- These CBS-like elements are nonfunctional in chromosome breakage, as shown by the absence of telomere-primed product. These reads are almost certainly MIC derived and represent repetitive IESs present on multiple MIC chromosomes, since each gave a PCR product with all the nullisomics tested and each has significant homology to a known IES or is likely to contain an IES (see Table S3 in the supplemental material for full explanations).
stitution in the conserved core) and a probability of chance occurrence that was at least 3 to 4 orders of magnitude higher than that found for functional CBSs in the genome.

These findings are illustrated in Fig. 2A by plotting sequence conservation in the neighborhood of the CBS element for 62 distinct, aligned functional CBS junctions (see Materials and Methods). Clearly, the 15-bp segment corresponding to the CBS is a sharply delimited island of sequence conservation relative to the rest of the surrounding region.

As a control, conservation for 44 genomic segments (represented by sequence read sets in group A [Table 3]) that had the correctly positioned conserved CBS core, AAACCAACC-C, but were nonfunctional in chromosome breakage was measured. Unlike in functional CBS elements, conservation at those positions that were free to vary within the 15-bp CBS-related segment (positions 1, 11, and 13 to 15) is as low as in flanking sequences (Fig. 2B). This indicates that the sequence of the completely conserved core per se causes no intrinsic limitation of nucleotide diversity at the variable positions; thus, the conservation seen in functional CBS elements is likely function related.

The combined results rigorously show that, in spite of sig-

FIG. 2. Nucleotide conservation measurements across the 15-bp CBS-related and flanking sequences. (A) Conservation measurements in 62 functional CBSs. (B) Conservation measurements in 44 nonfunctional CBS-related elements (controls). (C) Nucleotide frequencies in the 62 functional CBSs shown in panel A. The CBS element occupies positions 21 to 35 in all panels, emphasized by the vertical lines in panels B and C. Nucleotide conservation was calculated at every position, as described in Materials and Methods, using aligned C-strand nucleotide sequences. At any given position in the logo plot, 2 bits represent maximum conservation (i.e., the same nucleotide occupies that position in every CBS junction), and 0 bits corresponds to no conservation (i.e., maximum diversity; all 4 nucleotides are equally frequent at that position). Note that (i) in panel B, positions corresponding to the completely conserved CBS core (22 to 30 and 32) have been marked with dots because variation at those positions was forbidden by the method used for the identification of nonfunctional CBS and (ii) the average conservation of flanking region positions for nonfunctional CBS-like elements (B) is significantly lower than that observed for regions flanking functional CBS (A). This likely reflects a bias for higher GC content in the flanking sequence introduced by the selection for the GC-rich (50%) CBS conserved core sequence. For example, 20 of the 44 CBS-like elements and surrounding regions match genes in the predicted proteome, which statistically is GC richer than the intron or intergenic region (36).
significant variability at its margins, the conserved CBS element is indeed 15 bp long, its boundaries are sharply delimited, and the entire 15-bp sequence is likely maintained by purifying selection.

The distribution of functional CBS variants around the MICN genome is not entirely random. Examination of the MICN chromosome arm locations of functional CBS variants (Table 1) shows three potential “jackpots” (1A and 15A in MICN chromosome 1R and 11C in MICN chromosome 4R). The numbers are small, but the 11C substitutions, which occur exclusively in chromosome 4R, rise to the level of high statistical significance, determined as follows. The T. thermophila MICN chromosomes are roughly metacentric. The mean number of 11C substitutions per chromosome arm is 0.4 (four 11C substitutions in 10 chromosome arms). If we assume that every chromosome arm has the same length, the Poisson probability that a chromosome arm will have four or more 11C substitutions is 0.00077. If we further consider that chromosome 4 is the second-smallest chromosome, the probability is even lower. Thus, 11C substitutions cluster in the left arm of chromosome 4 significantly more than expected by chance alone.

DISCUSSION

New view of the family of conserved chromosome breakage elements of T. thermophila. A previous experimental study (15) led to the conclusion that the CBS has a remarkably long and rigorous sequence requirement. The findings reported here reveal the other side of the coin: a family of closely related elements that includes a surprising degree of variation (Table 2). Ten novel sequence variants were found, involving changes at 5 of the 15 positions of the CBS. Added to the three previously identified variants (15, 16, 41, 42), there are currently 13 known functional CBS variants, found at 29 genomic locations (Table 2). Since functional variants are found at nearly half of the impartially characterized CBS junctions, we now estimate that more than 100 chromosome breakage sites around the genome are occupied by variant CBS elements. This degree of variability was not seen earlier because the most extensive previous searches for CBS elements (16, 41) were biased in favor of the classic CBS element.

Variability across the CBS element is neither randomly nor symmetrically distributed. All functional CBS-related elements identified so far share a precisely placed and completely conserved 10-bp core (-AAACCAACC-C-, where the hyphens represent positions allowed to vary) (Table 2). Every single-base-pair substitution in this conserved core (encountered and tested in this study, 11 in total) produced a nonfunctional CBS. Furthermore, no substitution is allowed at a minimum of four of these positions (5, 6, 9, and 10) (Table 4).

These findings change our view of the CBS element. While it might still be regarded as a long, stringent sequence, it now must be looked upon more precisely as having an invariant core embedded within a significantly variable recognition sequence. In this light, the CBS element resembles a regulatory protein binding site more than it does a restriction enzyme site. This finding is puzzling because the variability built into existing regulatory elements brings about the modulation of their activity. In contrast, and despite an earlier report of partial function in certain variant CBS elements (15) (Table 4), our tests show that chromosome breakage behaves as an all-or-none phenomenon (see further discussion below).

One clue to reconciling such sequence variability with an all-or-none function comes from the findings that all known functional CBS variants have no more than two substitutions and that a CBS triple variant, containing three individually allowed substitutions at marginal positions (13A, 14C, and 15A) (Table 3), is nonfunctional. Thus, there are limits to the amount of CBS variability compatible with the chromosome breakage function. It seems plausible that the macromolecule that interacts with the CBS element must have a minimum threshold of binding affinity in order to trigger chromosome breakage. One or two changes at variable positions could allow the CBS to stay above the threshold affinity and would be tolerated, while additional changes, even if individually allowed, could reduce the affinity, and thus breakage efficiency, below the threshold and would be selected against.

This view also provides a consistent explanation for the striking T-A frequency asymmetry found at marginal positions (1 and 13 to 15) (Fig. 2C). This asymmetry is puzzling, considering that both nucleotides are compatible with breakage function and that the positions immediately adjacent to the CBS are at mutational equilibrium (equal A and T frequencies). If natural selection against below-threshold affinity is sufficiently strong, then the nucleotides occupying the four marginal positions are interdependent, and the number of T-to-A substitutions in the group as a whole cannot exceed a limit well short of mutational equilibrium.

Additional observations on CBS element variability. Nearly 15% of the estimated number of breakage sites in the genome (37 sites, all but 6 from this study) (Table 2) have now been isolated by starting with sequences adjacent to randomly selected telomeres and walking along MICN-limited DNA. None
was found to lack a recognizable member of the known CBS family. Thus, this family is the major, if not the only, type of chromosome breakage cis element in the *T. thermophila* genome.

Nine additional functional CBS-related elements were identified in the whole-genome sequence. Because they were found by searching for sequence similarities to the classic CBS, they contribute unbiased information regarding variability within the CBS family. Further observations about this wider set of 46 functional CBS-related elements can be summarized as follows. (i) The invariant core contains exclusively C and A nucleotides and resembles the telomere repeat unit (CCCCCA [common bases are underlined]). This similarity is extended in the 11C functional variant (TAAACCAACCTCTTT). The invariant core includes a continuous stretch of 9 base pairs, which represents nearly one full turn of a B DNA double helix.

(ii) Most of the functional variation is observed at marginal positions (positions numbered 1 and 13 to 15) and consists mainly of T-to-A substitutions (Table 2 and Fig. 2). This may merely reflect the general high A + T bias seen in the vicinity of the CBS. However, 15C is a nonfunctional single-base-pair substitution at a marginal position (Table 4).

(iii) In spite of the variation observed at the margins, the results of the nucleotide diversity analysis (Fig. 2A) clearly show significant sequence conservation at every position classically attributed to the chromosome breakage element. Thus, it seems justified to refer to the entire 15-bp segment as the "CBS box".

(iv) Previous studies have contributed information about variant CBSs (10, 15, 16, 20, 42). A comparison of all published observations on the chromosome breakage function of CBS single-bp substitutions in *T. thermophila* is shown in Table 4. Examination of this table shows that the studies agree for 9 (3 functional, 6 nonfunctional) of 12 substitutions of the consensus CBS. The discrepancies are discussed further below.

Any one member of the CBS family is necessary and sufficient for chromosome breakage. The classic CBS element was shown by hybridization to be MICN limited (41). Our work has found no exceptions and has extended that characterization to every functional variant so far identified. Our impartial MICN genome-wide computational search found no case where the classic CBS element, or any functional variant of it, exists as a nonfunctional element anywhere in the MICN genome. Yao et al. (40) showed that the chromosome breakage function of the classic CBS element is context independent and concluded that it is a necessary and sufficient cis-acting element for chromosome breakage. Our work provides abundant evidence that the classic CBS is not necessary, as a range of variants can supply the necessary function. Instead, we suggest that one (any one) functional member of the CBS element family is a necessary and sufficient cis element for chromosome breakage.

**Does partial chromosome breakage occur at natural CBS-related sequences?** As indicated above, our work shows three quantitative discrepancies with results from the plasmid DNA microinjection chromosome breakage assay (15). The discrepancies relate to substitutions that the earlier assay classified as partially functional but that the current study found to be nonfunctional (6A) or fully functional (13C and 14C). The 6A substitution received special scrutiny because it represents a change in the completely conserved CBS core. PCR tests, independently performed at each of the four locations in the genome where this CBS occurs, showed no evidence of even partial chromosome breakage or telomere addition near these sites.

Two types of analyses of natural junctions have now failed to reveal cases of partial chromosome breakage. First, PCR tests of 78 natural junctions (mainly from reference 16, Table 2 in this work, and our unpublished observations) have not revealed any cases of partial chromosome breakage (i.e., no BES sequences remain in the MACN, as sensitively determined by nullisomic PCR mapping). Second, the sizes of more than 100 MACN chromosomes have been determined by pulsed-field electrophoresis (5; E. P. Hamilton, S. Williamson, H. Ryder, and E. Orias, unpublished results). Additional bands, which could have been attributed to partial chromosome breakage function, were not observed, including the case of two MACN chromosomes carrying 6A-substituted, CBS-related elements. If we assume that chromosome fragments are generated by partial breakage during MACN differentiation, we must conclude that they are quickly lost by natural selection during the vegetative growth of the clone. This could happen only if partial breakage caused significantly decreased cell fitness, which, in the long term, would result in an all-or-none function being favored by natural selection.

The considerations described above cast strong doubts on the occurrence of partial chromosome breakage in natural *T. thermophila* junctions. We suggest that the microinjected plasmid DNA assay may sometimes fail to adequately reproduce the function of the (still largely undefined) chromosome breakage pathway in its natural context.

**Evidence for CBS junction duplication.** The 11C group of CBS elements shows statistically significant concentration in a particular MICN chromosome arm (4L) (Table 2). A plausible explanation for this clustering is that the four 11C-containing CBS junctions are members of a clone, derived from an 11C progenitor, which arose by successive duplications with a preference for restricted distribution in the MICN genome. If so, two observations suggest that these duplications would be relatively ancient: (i) all four 11C variants have accumulated a second substitution, encompassing three different positions; and (ii) when the 11C junctions are aligned by BLASTn analysis in all pairwise combinations, no significant regions of sequence similarity that include the CBS junction are found (data not shown). In contrast, instances of putatively more-recent junction duplications, as indicated by significant sequence identity, have been reported (5; E. P. Hamilton and E. Orias, unpublished observations). In all but one of these cases, groups are also restricted to the same MICN chromosome arm. The combined observations suggest a mechanism of junction duplication with a preference for short-range transposition, possibly via a transposable element. A number of eukaryotic DNA transposable elements show such preference (for examples, see references 23, 24, and 34).

Although the classic CBS, TAAACCAACCTCTTT, comprises only a slight majority of the CBS elements found by unbiased methods (26 out of 46 junctions) (Table 2), the nucleotide occupying each position of the classic CBS is by far the most frequent nucleotide (78 to 90% in variable positions) among the entire set. Thus, the classic sequence represents a unique consensus sequence of all the functional variants. If the
CBS element indeed spreads through the Tetrahymena genome by junction duplication, then it would be simplest to hypothesize that the ancestral CBS had the classic sequence.

**Implications of this work regarding the CBS endonuclease.** Knowledge about the mechanism of programmed chromosome fragmentation is limited. The cleavage reaction is presumably carried out by a specific “CBS endonuclease”, not yet identified. Telomere additions to newly created chromosome ends appear to be mechanistically coupled to chromosome breakage (14). There are indications (but no rigorous evidence) that RNAs may be involved in the process, as chromosome breakage is severely inhibited by the loss of Twi1p (25), a PAZ/Piwi domain protein that binds and stabilizes a 28-nucleotide class of double-stranded Tetrahymena RNAs. Members of this class, “scan RNAs”, are proposed to guide IES DNA elimination by a sequence-specific mechanism analogous to those used for RNA silencing and heterochromatin formation (26, 27).

Although the identity of the putative CBS endonuclease is not known, this analysis suggests that the molecule that binds to the CBS element should recognize a nonpalindromic 15-bp box containing a 10-bp invariant core and five other positions that tolerate different degrees of variability. A homing endonuclease homolog has been discussed as a candidate for the CBS endonuclease (15, 35) because such enzymes recognize long (14- to 30-bp) pseudopalindromic sites and tolerate some sequence flexibility (19). Homing endonuclease homologs are encoded by putative mobile genetic elements within the *T. thermophila* MICN genome (37).

Based on a search of autoannotated *T. thermophila* gene predictions, we propose that a DDE endonuclease related to the CBS is conserved in *T. thermophila* CBS superfamily of DNA-mediated transposons, p. 565–610. Annu. Rev. Genet. 689–699.

**ACKNOWLEDGMENTS**

This work was supported by National Institutes of Health grant RR-02391 from the National Center for Research Resources. Sequences and gene predictions were obtained from The Institute for Genomic Research website at http://www.tigr.org and through the Tetrahymena Genome Database at http://www.ciliate.org/. We thank Teisha Rowland and all the anonymous reviewers for helpful suggestions that improved the manuscript.

**REFERENCES**

1. Altshuler, M. L., and M. C. Yao. 1985. Macromolecular DNA of *Tetrahymena thermophila* exists as defined subchromosomal-sized molecules. Nucleic Acids Res. 13:5817–5831.

2. Bailey, D. L., and C. Elkan. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc. Int. Conf. Intell. Syst. Mol. Biol. 4:22–36.

3. Bailey, T. L., and M. Gribskov. 1998. Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14:48–54.

4. Blackburn, E. H., and J. C. Gall. 1978. A tandemly repeated sequence at the terminus of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. J. Mol. Biol. 120:33–53.

5. Cassidy-Hanley, D., Y. Bisharyan, V. Fridman, J. Gerber, C. Lin, E. Orias, J. D. Orias, H. Ryder, L. Vong, and E. P. Hamilton. 2005. Genome-wide characterization of *Tetrahymena thermophila* chromosome breakage sites. 2. Physical and genetic mapping. Genetics 170:1623–1631.

6. Cassidy-Hanley, D., M. C. Yao, and P. J. Bruns. 1994. A method for mapping germ line sequences in *Tetrahymena thermophila* using the polymerase chain Bruns, Genetics 133:489–502.

7. Chalker, D. L., and M. C. Yao. 2001. Nongenic, bidirectional transcription precedes and may promote developmental DNA deletions in *Tetrahymena thermophila*. Genes Dev. 15:1287–1298.

8. Conover, R. K., and C. F. Weil. 1986. Macromolecular DNA molecules of *Tetrahymena thermophila*. Mol. Cell. Biol. 6:900–905.

9. Coyne, R. S., D. L. Chalker, and M.-C. Yao. 1996. Genomic downsizing during ciliate development: nuclear division of labor through chromosome restructuring. Annu. Rev. Genet. 30:557–578.

10. Coyne, R. S., and M.-C. Yao. 1996. Evolutionary conservation of sequences directing chromosome breakage and rDNA palindromic formation in tetrahymenine ciliates. Genetics 144:1479–1487.

11. Croteau, G. E., G. Hon, J. M. Chandoura, and S. E. Brenner. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188–1190.

12. Doerder, F. P., J. C. Deak, and J. H. Lief. 1992. Rate of phenotypic assortiment in *Tetrahymena thermophila*. Dev. Genet. 13:126–132.

13. Eisen, J. A. 3 November 2005, date last modified. *Tetrahymena thermophila* genome project. [Online.] http://www.tigr.org/tdb/e2k1/ttg/.

14. Fan, Q., and M.-C. Yao. 1996. New telomere formation coupled with site-specific chromosome breakage in *Tetrahymena thermophila*. Mol. Cell. Biol. 16:1267–1274.

15. Fan, Q., and M. C. Yao. 2000. A long stringent sequence signal for programmed chromosome breakage in *Tetrahymena thermophila*. Nucleic Acids Res. 28:895–900.

16. Hamilton, E., P. Bruns, C. Lin, V. Merriam, E. Orias, L. Vong, and D. Cassidy-Hanley. 2005. Genome-wide characterization of *Tetrahymena thermophila* chromosome breakage sites. I. Cloning and identification of functional sites. Genetics 165:1611–1621.

17. Hamilton, E. F., and E. Orias. 2000. Genetically mapping new mutants and cloned genes. Methods Cell Biol. 62:260–280.

18. Jahn, C. L., and L. A. Klobutcher. 2002. Genome remodeling in ciliated protozoa. Annu. Rev. Microbiol. 56:489–520.

19. Jurica, M. S., and B. L. Stoddard. 1999. Homing endonucleases: structure, function and evolution. Curr. Opin. Life Sci. 5:1304–1320.

20. Kapler, G. M., and E. H. Blackburn. 1994. A weak germ-line excision mutation blocks developmentally controlled amplification of the rDNA minichromosome of *Tetrahymena thermophila*. Genes Dev. 8:54–95.

21. Karrer, K. 2000. Tetrahymena genetics: two nuclei are better than one. Methods Cell Biol. 62:127–186.

22. Kloc, M., and B. Zagrodzinska. 2001. Chromatin elimination—an oddity or a common mechanism in differentiation and development? Differentiation 68:84–91.

23. Kunze, R., and C. F. Weil. 2002. The h4T and CACTA superfamilies of plant transposons, p. 565–610. In N. L. Craig, R. Craige, M. Gellert, and A. M. Lambowitz (ed.), Mobile DNA II. ASM Press, Washington, D.C.

24. Machida, C., H. Onouchi, J. Koizumi, S. Hamada, E. Semiarti, S. Torikai, and Y. Machida. 1997. Characterization of the transposition pattern of the Ac element in *Arabidopsis thaliana* using endonuclease I-SceI. Proc. Natl. Acad. Sci. USA 94:8675–8680.

25. Mochizuki, K., A. Fine, T. Fujisawa, and M. A. Gorovsky. 2002. Analysis of a psi-related gene implicates small RNAs in genome rearrangement in Tetrahymena. Cell 110:689–699.

26. Mochizuki, K., and M. A. Gorovsky. 2004. Small RNAs in genome rearrangement in *Tetrahymena*.Curr. Opin. Genet. Dev. 14:181–187.

27. Mochizuki, K., and M. A. Gorovsky. 2005. A Dicer-like protein in *Tetrahymena* has distinct functions in genome rearrangement, chromosome segregation, and meiotic prophase. Genes Dev. 19:77–89.

28. Nakagawa, H., J. K. Lee, J. Hurwitz, R. C. Allaire, and S. Nakamura. 2002. Fission yeast CEN-PB homologs nucleate centromeric heterochromatin by promoting heterochromatintin-specific histone tail modifications. Genes Dev. 16:1766–1778.
29. Ohzeki, J., M. Nakano, T. Okada, and H. Masumoto. 2002. CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J. Cell Biol. 159:765–775.

30. Prescott, D. M. 1994. The DNA of ciliated protozoa. Microbiol. Rev. 58:233–267.

31. Schneider, T. D., and R. M. Stephens. 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18:6097–6100.

32. Shannon, C. E. 1948. A mathematical theory of communication. Bell System Tech. J. 27:379–423 and 623–656.

33. Smit, A. F., and A. D. Riggs. 1996. Tiggers and other DNA transposon fossils in the human genome. Proc. Natl. Acad. Sci. USA 93:1443–1448.

34. Tower, J., G. K. Karpen, N. Craig, and A. C. Spradling. 1993. Preferential transposition of Drosophila P elements to nearby chromosomal sites. Genetics 133:347–359.

35. Ward, J. G., P. Blomberg, N. Hoffman, and M. C. Yao. 1997. The intranuclear organization of normal, hemizygous and excision-deficient rRNA genes during developmental amplification in Tetrahymena thermophila. Chromosoma 106:233–242.

36. Wuitschick, J. D., and K. M. Karrer. 1999. Analysis of genomic G + C content, codon usage, initiator codon context and translation termination sites in Tetrahymena thermophila. J. Eukaryot. Microbiol. 46:239–247.

37. Wuitschick, J. D., P. R. Lindstrom, A. E. Meyer, and K. M. Karrer. 2004. Homing endonucleases encoded by germ line-limited genes in Tetrahymena thermophila have APETELA2 DNA binding domains. Eukaryot. Cell 3:685–694.

38. Yao, M.-C. 1989. Site-specific chromosome breakage and DNA deletion in ciliates, p. 715–734. In D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.

39. Yao, M.-C., E. H. Blackburn, and J. Gall. 1981. Tandemly repeated C-C-C-A-A hexanucleotide of Tetrahymena rDNA is present elsewhere in the genome and may be related to the alteration of the somatic genome. J. Cell Biol. 90:513–520.

40. Yao, M.-C., C. H. Yao, and B. Monks. 1990. The controlling sequence for site-specific chromosome breakage in Tetrahymena. Cell 63:763–772.

41. Yao, M.-C., K. Zheng, and C. H. Yao. 1987. A conserved nucleotide sequence at the sites of developmentally regulated chromosomal breakage in Tetrahymena. Cell 48:779–788.

42. Yu, G. L., and E. H. Blackburn. 1991. Developmentally programmed healing of chromosomes by telomerase in Tetrahymena. Cell 67:823–832.