

# Characterization of multigene families in the micronuclear genome of *Paramecium tetraurelia* reveals a germline specific sequence in an intron of a centrin gene

Laurence Vayssié, Linda Sperling and Luisa Madeddu\*

Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, 91198 Gif-sur-Yvette Cedex, France

Received November 1, 1996; Revised and Accepted January 14, 1997

DDBJ/EMBL/GenBank accession nos U76539, U76540, U76541

## ABSTRACT

In *Paramecium*, as in other ciliates, the transcriptionally active macronucleus is derived from the germline micronucleus by programmed DNA rearrangements, which include the precise excision of thousands of germline-specific sequences (internal eliminated sequences, IESs). We report the characterization of micronuclear versions of genes encoding *Paramecium* secretory granule proteins (trichocyst matrix proteins, TMPs) and *Paramecium* centrans. TMP and centrin multigene families, previously studied in the macronuclear genome, consist of genes that are co-expressed to provide mixtures of related polypeptides that co-assemble to form respectively the crystalline trichocyst matrix and the infraciliary lattice, a contractile cytoskeletal network. We present evidence that TMP and centrin genes identified in the macronucleus are also present in the micronucleus, ruling out the possibility that these novel multigene families are generated by somatic rearrangements during macronuclear development. No IESs were found in TMP genes, however, four IESs in or near germline centrin genes were characterized. The only intragenic IES is 75 bp in size, interrupts a 29 bp intron and is absent from at least one other closely related centrin gene. This is the first report of an IES in an intron in *Paramecium*.

## INTRODUCTION

The crystalline contents of *Paramecium* secretory granules and a number of elements of the *Paramecium* cortical cytoskeleton have been studied at the protein level and are known to assemble from heterogeneous families of immunologically related polypeptides (1–4). For both the trichocyst matrix proteins (TMPs) and the major polypeptides of one of the cortical cytoskeletal arrays, the infraciliary lattice (ICL), we have recently shown that most or all of this heterogeneity is situated at the level of primary structure: both TMPs and ICL polypeptides are encoded by multigene families (5,6). TMPs are *Paramecium*-specific proteins (7), while the ICL polypeptides are centrans, EF-hand calcium binding proteins that have been highly conserved throughout eukaryotic evolution (8).

The TMP and centrin multigene families, estimated to contain ~100 and ~20 members respectively, share several characteristics. First, they are organized into subfamilies. Each subfamily codes for a distinct protein, but within each subfamily, several genes code for nearly identical proteins. All members of these families so far characterized (12 TMP genes or gene fragments from three different subfamilies; three ICL genes from the same subfamily) contain short introns (between 23 and 29 bp) (5–7), the small size being a characteristic of all known *Paramecium* introns (9–11). The most unusual feature of these multigene families, however, is that all members seem to be constitutively co-expressed, thus producing mixtures of related polypeptides that co-assemble. We have suggested that the use of multigene families to assure a microheterogeneity of structural proteins may be a general morphogenetic strategy in *Paramecium* (5,6).

*Paramecium*, like all ciliated protozoa, presents nuclear dimorphism; each cell possesses a somatic nucleus (macronucleus) that is responsible for gene expression during vegetative growth and a transcriptionally inactive germline nucleus (micronucleus) that is involved in sexual processes (conjugation and autofertilization). The germline micronucleus is diploid and contains conventional chromosomes of an average size of 2000 kb. The somatic macronucleus has a DNA content ~800 times greater than the micronuclear haploid value, divides amitotically during vegetative growth and consists of small acentric chromosomes (50–800 kb). During sexual processes, the micronucleus undergoes meiosis followed by fertilization to form the zygotic nucleus. The old macronucleus is degraded and a new one is formed by extensive, programmed rearrangements of the DNA of a mitotic copy of the zygotic nucleus (12), involving chromosome fragmentation, amplification and *de novo* telomere addition, as well as the precise excision of thousands of germline-specific elements known as internal eliminated sequences (IESs; 13–15).

Before tackling evolutionary questions raised by the TMP and ICL families of co-expressed genes, it is important to see whether all of them are present in the micronucleus or are somehow produced by somatic rearrangements during macronuclear development. As the only micronuclear sequences that have been characterized so far in *Paramecium* are the coding and flanking regions of surface antigen genes (16–19), it is also important to characterize other micronuclear genes and, in particular, genes that contain introns. By screening a *Paramecium tetraurelia* micronuclear library (16), we have obtained evidence that the

\*To whom correspondence should be addressed. Tel: +33 1 69 82 43 92; Fax: +33 1 69 82 31 50; Email: madeddu@cgm.cnrs-gif.fr

previously characterized macronuclear TMP and ICL genes are indeed present in the micronucleus. No germline-specific sequences were found in the coding regions of TMP genes. However, an IES was found in an intron of a germline centrin gene. Furthermore, this IES is absent from at least one other closely related micronuclear centrin gene.

## MATERIALS AND METHODS

### Isolation and characterization of phage clones

A library constructed in  $\lambda$ GEM-11 with *Paramecium tetraurelia* (stock 51) micronuclear genomic DNA, partially digested with *Sau3A*, was kindly provided by John R. Preer (Institute of Molecular Biology, Indiana University, Bloomington, IN) (16); a library constructed in  $\lambda$ EMBL3 with *Sau3A*-partially digested *Paramecium tetraurelia* (strain d4-2) macronuclear genomic DNA was a gift of Eric Meyer (Laboratoire de Gntique Molculaire, Ecole Normale Suprieure, Paris, France). The libraries were screened using T1 and ICL1 subfamily-specific  $^{32}$ P-labelled probes (272 and 597 bp respectively), obtained by PCR amplification in the presence of [ $\alpha$ - $^{32}$ P]dATP as described previously (5). Selected clones were isolated according to standard techniques (20). The positive phage plaques were further characterized by using the first round phage stocks as substrate for PCR amplification of T1- or ICL1-selected gene regions. For each phage plaque stock, a small aliquot of the storage medium (SM; 0.1 M NaCl, 10 mM MgSO<sub>4</sub>, 0.01% gelatin, 20 mM Tris-HCl, pH 7.5) was heated at 75°C for 20 min. Aliquots of 0.5  $\mu$ l (about 1/1000 of the total volume of the stock) were then used for each PCR sample. The oligonucleotides used to prime these reactions and their positions with respect to the gene maps, as well as the templates and primers used to generate the  $^{32}$ P-labeled probes, are given in Figures 1A and 2A and their legends. The amplification products were then analyzed by Southern blot hybridization, using as  $^{32}$ P-labeled probes either the subfamily-specific gene fragments used for library screening or (for the T1 products) the gene-specific oligonucleotide probes given below. Selected PCR products were recovered using the QIAquick spin kit (Qiagen), then cloned into the *Sma*I site of plasmid pUC18. Macronuclear and micronuclear inserts containing the ICL1-d gene were identified and characterized by restriction digestion and Southern blot analysis using the ICL1 subfamily-specific probe; *Eco*RI phage restriction fragments (~4.5, 1.5 and 0.28 kb) were subcloned into plasmid pUC18. DNA sequences were determined using the T7 sequencing kit (Pharmacia).

### PCR amplification

PCR reactions (50  $\mu$ l) contained 200 pmol each primer, 0.1 mM dNTPs and 2 U Taq DNA polymerase (Boehringer). Reactions were overlaid with 50  $\mu$ l mineral oil and carried out for 30 cycles of denaturation at 90°C for 30 s, annealing at 48°C for 45 s and extension at 72°C for 90 s, using an OmniGene Temperature Cycler (Hybaid).

### Southern blots

DNA was fractionated by electrophoresis through agarose gels and transferred to Hybond-N<sup>+</sup> filters (Amersham) in 0.4 M NaOH. Hybridizations were carried out according to Church and Gilbert (21) at 60°C (PCR-generated probes). The sequences of

the T1 gene-specific probes, corresponding to the third intron of each of six members of the T1 subfamily, are as follows (5):

T1-a, GTATGTATCCCTTGTTAACCCCTTTCATAG;  
 T1-b, GTAACCAATCCTTATTAACCTCGCCCTAG;  
 T1-c, GTAATGTCTAATTGATATATCCTCTAG;  
 T1-d, GTAATTCTAACCTAATATCCTATAG;  
 T1-e, GTAATTCTAACAACTTAAATGATAG;  
 T1-f, GTATTCTATTTTCTCATTCTAG.

For these oligonucleotide probes hybridization temperatures were between 43 and 52°C, according to the estimated  $T_m$  of the hybrids. The membranes were then washed at the same temperatures chosen for hybridization with decreasing concentrations of SSC (1 $\times$  SSC = 150 mM NaCl, 15 mM sodium citrate, pH 7.2) (20) in the presence of 0.1% SDS, 2 $\times$  SSC for 30 min, followed by 0.2 $\times$  SSC for 20 min. Autoradiograms were obtained by exposing the filters to Hyperfilm-MP films (Amersham).

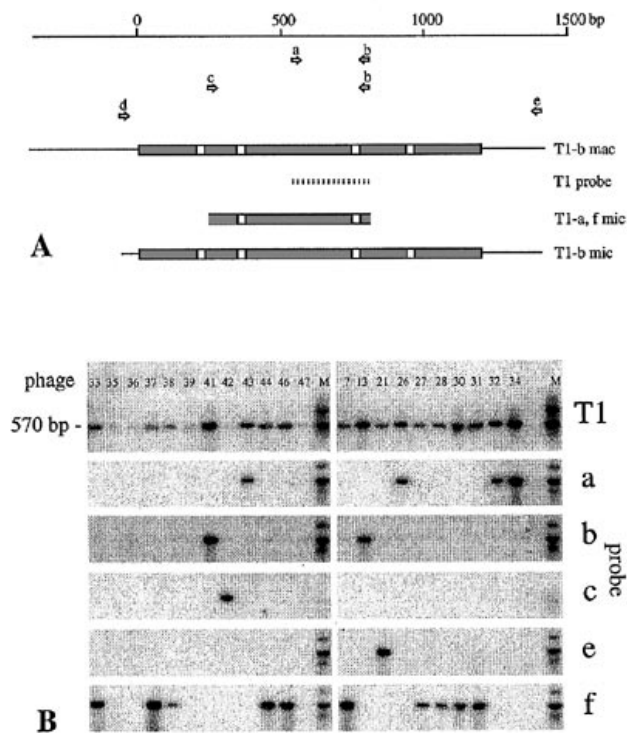
## RESULTS

### Micronuclear TMP genes

Our previous characterization of macronuclear TMP genes involved the determination of complete sequences for three genes (T1-b, T2-c and T4-a) cloned from a macronuclear library (7). These three genes, representative of three different subfamilies, code for proteins that present a common organization but share only ~25% sequence identity. For several other members of each subfamily, the sequences of gene fragments generated by PCR were established (5). Southern blot experiments using exon sequences as probes allowed us to show that there are between four and eight different genes in each subfamily, consistent with the number of different sequences found among the PCR products. The introns in these genes, and in particular six paralogous introns characterized in the T1 genes, were shown to be unique elements in the *Paramecium* genome and to constitute gene-specific probes (5; L. Vayssi, unpublished data).

In order to characterize micronuclear TMP genes, a micronuclear library (16) was screened by hybridization to a T1 subfamily-specific exon probe (Fig. 1). Twenty-two positive phage plaques were chosen for further analysis. The first round stocks, containing the partially purified phage particles, served as substrates for PCR using partially degenerate primers that amplify ~570 bp of the T1 genes. The 22 PCR products were then hybridized with gene-specific probes consisting of synthetic oligonucleotides that exactly correspond to the third intron of each of six different T1 genes (5). Each gene-specific probe hybridized to at least one amplification product and each amplification product hybridized to no more than one gene-specific probe (Fig. 1B). This region of the T1 genes does not appear to contain germline-specific sequences, as the amplification products were of the same size as those obtained using macronuclear DNA templates.

As the library we screened was estimated to be contaminated with ~20% macronuclear inserts (16), we cannot prove, in the absence of germline-specific sequences, that any given insert is of micronuclear origin. However, the consistency of the results obtained for 22 clones strongly suggests that all T1 genes are present in the micronucleus and that the different T1 genes found in the macronuclear genome are not produced by reorganization of the sequences during macronuclear development. Further support for this conclusion was obtained by cloning and

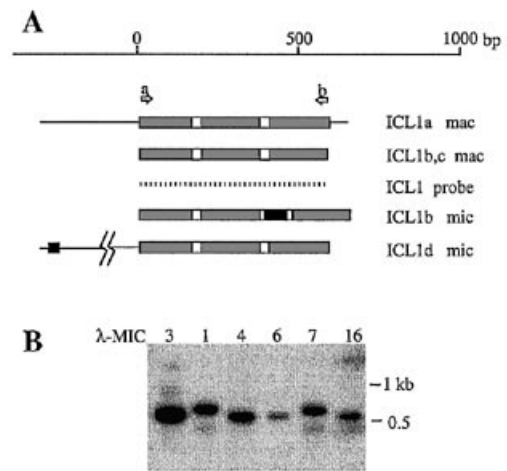


**Figure 1.** Characterization of T1 genes. (A) Map of T1 genes showing macronuclear and putative micronuclear DNA regions that have been sequenced. The T1-b macronuclear gene previously characterized (7) is shown for reference. Coding sequences are represented by gray boxes, introns by white boxes and non-coding sequences flanking the ATG initiation and the TGA STOP codons by a thin solid line. On the upper part of the figure, arrows indicate the positions of the PCR primers used to generate the subfamily-specific probe for library screening (shown in the lower part of the figure as a double dotted line) and to amplify phage DNA. A subclone of the macronuclear T1-b gene (pmgh9; 7) was used as template to generate the T1 subfamily-specific probe. T1 gene regions that were amplified, cloned and sequenced are indicated in the lower part of the figure. The sequences of the oligonucleotides used are: a, 5'-GAACACAAYGAWGCTATYGG-3'; b, 5'-CAACTCTCAARTTATTGAAGGC-3'; c, 5'-GGATGAGGAATTAGC-3'; d, 5'-TGGAA-TTTTTAACAC-3'; e, 5'-CAGTAGAGTTCCATT-3'. (B) Hybridization of phage PCR products with gene-specific probes. T1 amplification products of 570 bp were generated for each of 22 first round phage stocks, separated by 1.8% agarose gel electrophoresis and analyzed by Southern blotting. The same filter was sequentially hybridized with the T1 subfamily-specific probe and different gene-specific intron probes (see Materials and Methods). In addition to the results presented, the ICL1-d probe hybridizes with the PCR product corresponding to phage 39 (data not shown). Three PCR products hybridized with none of the gene-specific probes; they may correspond to two other T1 subfamily members found in RT-PCR experiments (5) but whose gene (and therefore intron) sequences have not been determined.

sequencing amplification products covering part (T1-a and T1-f) or all (T1-b) of three different T1 genes (see Fig. 1A): the sequences were identical to those previously obtained for macronuclear T1 genes (5,7; L.Madeddu, unpublished results).

#### A micronuclear centrin gene contains an IES within an intron

The *Paramecium* ICL, a contractile network that constitutes the innermost element of the cortex, is composed of six immunologically related  $\text{Ca}^{2+}$  binding proteins (2). Distinct N-terminal microsequences were found for four of the ICL polypeptides and for one of them, ICL1, macronuclear DNA sequences were obtained (6; Fig. 2A). These sequences, along with data from



**Figure 2.** Characterization of ICL1 genes. (A) Map of ICL1 genes showing macronuclear and micronuclear DNA regions that have been sequenced. The ICL1 macronuclear genes previously characterized (6) are shown for reference. Coding and non-coding sequences, as well as PCR primers used for probe generation and characterization of micronuclear library inserts, are represented as in Figure 1A. Germline-specific sequences (IESs) are represented by a black box. To generate the ICL1 subfamily-specific probe (double dotted line), we used as template a cloned PCR-generated DNA fragment corresponding to the entire coding region of the ICL1-b gene (6). The oligonucleotides used are: a, 5'-GGCACGAAGAGGATAGTAACCACCACCC-3'; b, 5'-GCAAAGGTC-TTTTTTGTGTCATAATGTTGTAG 3'. (B) Analysis of phage PCR products. ICL1 amplification products were generated for each of seven first round phage stocks, separated by 1.3% agarose gel electrophoresis and analyzed by Southern blotting using the ICL1 subfamily-specific probe.

Southern blot and reverse transcription-polymerase chain reaction (RT-PCR) experiments, revealed that at least three different co-expressed genes, ICL1-a, ICL1-b and ICL1-c, code for nearly identical polypeptides and that these ICL1 polypeptides are *Paramecium* centrins.

The micronuclear library was screened with an ICL1 subfamily-specific probe (Fig. 2A) and seven positive phage plaques were chosen for further study. As for the characterization of T1 genes, the first round phage stocks were used as substrates for PCR. Four of the amplification products obtained using primers that amplify the ICL1 coding regions are the same size as the product of macronuclear DNA amplification (~600 bp; Fig. 2B). Three other amplification products appeared to be larger (Fig. 2B and data not shown).

The three PCR products of larger size were cloned and sequenced. Each one is identical to the macronuclear ICL1-b gene, but with a 75 bp insertion in the second intron (Fig. 3). The inserted sequence presents all of the characteristics of a *Paramecium* IES. It is bounded by the dinucleotide 5'-TA-3' and only one 5'-TA-3' remains in the macronuclear sequence. The AT content is very high (92%). Finally, the ends of the element fit quite well (6/7 nt match) the IES inverted terminal repeat consensus established by Klobutcher and Herrick (22) on the basis of 20 IESs from *Paramecium* surface antigen genes (Table 1).

The phages whose amplification products were the same size as the product of macronuclear DNA amplification,  $\lambda\text{MIC}3$ ,  $\lambda\text{MIC}4$  and  $\lambda\text{MIC}16$ , were purified to homogeneity by further rounds of plaque hybridization. Restriction profiles of the phage DNA indicated that  $\lambda\text{MIC}3$  and  $\lambda\text{MIC}4$  contain the same insert, different from that of  $\lambda\text{MIC}16$ . In order to obtain the correspon-

```

1  GGCACGAAGAGGATAGTAACCACCACCCCTAACACAATAAGCCCCACCAC 50
   A R R G Q Q P P P Q Q Q Q A P P Q
51  AAAAAAATAAGCTGGCAAATTAACCCAGCCGAATTTGTTAAGCCTGGA 100
   K N Q A G K F N P A E F V K P G
101 TTGACCGAAGAAGAGTCTTAGAATAAAGAAGCATTCGACTTATTGTA 150
   L T E E E V L E I K E A F D L F D
151 TACAGATGGgtataatatcatacttattatttagTACTCAATCCATTGAT 200
   T D G T Q S I D
201 CCAAAGGAATTAAGAGCTGCAATGACCTCATTAGGATTTGAAGCCAAAAA 250
   P K E L K A A M T S L G F E A K N
251 CTAACAATTTATCAAATGATAGTATTGGATACTGATGGAAGTGGAC 300
   Q T I Y Q M I S D L D T D G S G Q
301 AAATTGATTTTGCTGAGTTTGTAAATFGATGACTGCTAGAAATCAGCGAG 350
   I D F A E F L K L M T A R I S E
351 AGAGATTCCAAAGCTGACATTTAAgtacataaatacattagatcctaactc 400
   R D S K A D I Q
401 atttttatgattttcttataattattataattttcttaataataatttttattaa 450
451 tttgattatatcactcaattcattctagAAAGTCTTCAACCTTTTCGATT 500
   K V F N L F D S
501 CAGAAAGAGCTGGAGTTATCACTTTGAAGGATCAGAAAAGTAGCTAAA 550
   E R A G V I T L K D L R K V A K
551 GAATTAGGAGAAACCATGGATGATCTGAATGTGAAGAGATGATTGATAG 600
   E L G E T M D D S E L Q E M I D R
601 AGCTGATTTCTGATGGAGATGCTTAAAGTGACCTTTGAAGATTTCTACAACA 650
   A D S D G D A Q V T F E D F Y N I
651 TTATGACAAAAAAGACCTTTTGC 672
   M T K K T F A

```

**Figure 3.** Sequence of the micronuclear ICL1-b gene. Coding nucleotides are in upper case with the corresponding amino acids in single letter code beneath them. The two introns are in lower case and the IES which interrupts the second intron is in underlined lower case. The 5'-TA-3' terminal repeats of the IES are in bold. The nucleotides corresponding to the PCR primers used to amplify this sequence are in italics (see Fig. 2). DDBJ/EMBL/GenBank accession no. U76539.

ding macronuclear chromosome fragments, we screened a macronuclear library for ICL1 genes and, after purification of several phages, identified one whose restriction profile was very similar to that of  $\lambda$ MIC3 and  $\lambda$ MIC4. Subcloning of  $\lambda$ MIC4 and of the corresponding macronuclear phage ( $\lambda$ MAC10) inserts and sequencing of the ICL1 gene contained in 1.5 and 0.28 kb *EcoRI* fragments (see Fig. 4) revealed 100% identity of the micronuclear and macronuclear sequences and identify a fourth member of the ICL1 subfamily, ICL1-d. The ICL1-d sequence (DDBJ/EMBL/GenBank accession no. U76540) is 95% identical to that of the ICL1-a gene (6). Since the same sequence had previously been found among RT-PCR products (L.Madeddu, unpublished observation), we can conclude that, like the other ICL1 genes, ICL1-d is expressed.

As the germline ICL1-b gene contains an IES, while the ICL1-d sequence obtained from the micronuclear library does not, we looked for IESs elsewhere on the 16 kb insert in order to verify the micronuclear nature of  $\lambda$ MIC4. We sequenced the entire 1.5 kb *EcoRI* fragment containing part of the ICL1-d gene, as well as the ends of a 4 kb *EcoRI* fragment found to be larger in the micronuclear phage digest than in the macronuclear one. Comparison of the sequences obtained for the corresponding macronuclear and micronuclear fragments allowed us to identify three germline-specific sequences, presented in Figure 4 along with their positions on a restriction map of  $\lambda$ MIC4. These IESs

appear to be in non-coding sequences. All three are of small size, two being to our knowledge the shortest IESs found so far in *Paramecium* (26 and 27 bp respectively). The terminal repeats of these elements are given in Table 1. The larger micronuclear *EcoRI* fragment whose ends we sequenced must contain another 0.4 kb of IESs, in addition to the two that we characterized (26 and 45 bp); the ends of the corresponding micronuclear and macronuclear fragments present the same sequences while their size differs by ~0.5 kb. We thus have identified a 5 kb region of the micronuclear genome which contains a minimum of four IESs, and perhaps many more. In the context of the present analysis, they provide confirmation of the micronuclear origin of the  $\lambda$ MIC4 insert and allow us to conclude that at least one germline ICL1 gene (ICL1-b) contains an IES and at least one (ICL1-d) does not.

**Table 1.** IES inverted terminal repeats

ICL1-b IES	Left	5'-TACATTAG-3'	6/7
	Right	5'-TATAATCA-3'	6/7
$\lambda$ MIC4 IES-1	Left	5'-TATGGATG-3'	5/7
	Right	5'-TATAGATG-3'	6/7
$\lambda$ MIC4 IES-2	Left	5'-TATTGTTG-3'	6/7
	Right	5'-TATACATA-3'	5/7
$\lambda$ MIC4 IES-3	Left	5'-TATAGTCG-3'	7/7
	Right	5'-TATAGTGA-3'	7/7
IES consensus		5'-TAYAGYNR-3'	

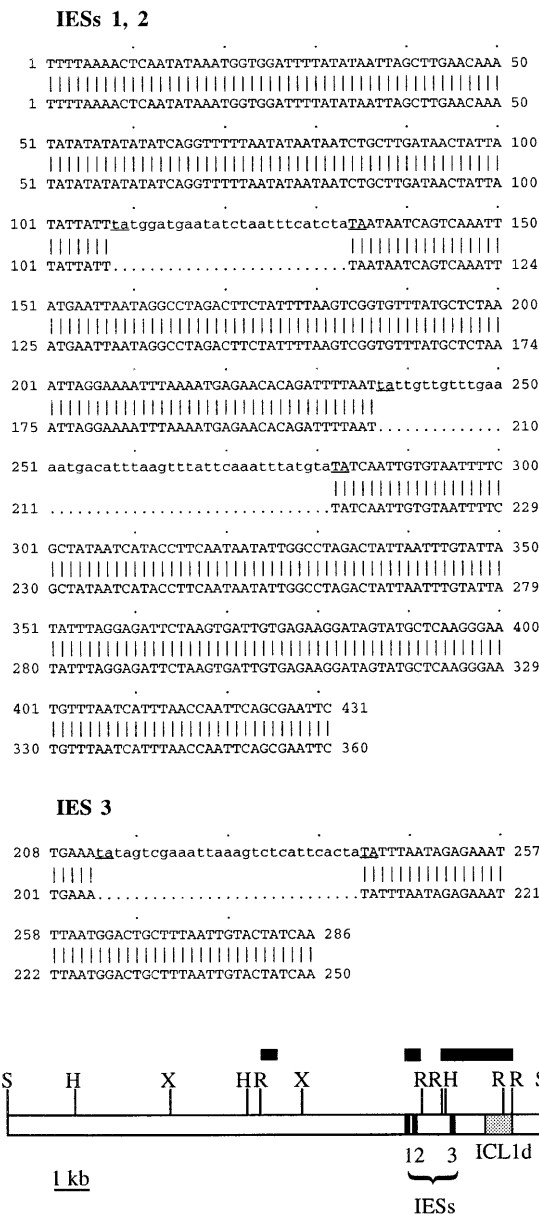
The inverted terminal repeats of the IESs characterized in the present study are compared with the inverted terminal repeat consensus established by Klo-bucher and Herrick (21) on the basis of a statistical analysis of 20 IESs from *Paramecium* surface antigen genes. The number of nucleotides that match the consensus is given to the right of each sequence.

## DISCUSSION

We have screened a *Paramecium* micronuclear library in search of germline genes corresponding to members of two multigene families previously characterized in the macronuclear genome (5-7). The ensemble of the analysis provides direct evidence that TMP and ICL genes identified in the macronucleus are also present in the micronucleus and that the macronuclear genes are co-linear with the micronuclear ones. Evidence in favour of a direct relationship between micronuclear and macronuclear members of these families of co-expressed genes constitutes a necessary prelude to consideration of the evolutionary issue they raise, namely how and why a unicellular organism has generated and maintains in its genome so many genes coding for very similar proteins. The only intragenic IES we found, in the ICL1-b gene, interrupts an intron. Moreover, this IES is absent from other ICL1 subfamily members.

### IESs in multigene families

The four ICL1 genes that have been characterized share 85-95% nucleotide identity and most likely arose through duplication of an ancestral gene. This is all the more probable as the positions of two introns are strictly conserved among the ICL1 genes. One of two micronuclear ICL1 genes characterized contains an IES (in the second intron), while the other does not, indicating that an IES has been either gained or lost since duplication of the ancestral ICL1 gene. Another possibility is that the duplication(s) occurred



**Figure 4.** Characterization of  $\lambda$ mic4. The lower part of the figure shows a restriction map of the phage insert. S, *Sst*I; X, *Xba*I; H, *Hind*III; R, *Eco*RI. The shaded box represents the ICL1-d gene (whose complete sequence and 5' upstream flanking region have been deposited in the DDBJ/EMBL/GenBank database, accession no. U76540) and the black boxes represent the IESs. The thick black lines over the map indicate the regions that were sequenced. Comparison of the micronuclear sequences with the sequences of the corresponding macronuclear regions (from phage  $\lambda$ MAC10) allowed identification of three IESs, as shown in the upper part of the figure. Sequences retained in the macronucleus are in upper case and micronuclear-specific sequences are in lower case. The 5'-TA-3' direct repeats that bound each IES, one of which is retained in the macronucleus, are underlined. The sequence of the region containing IES 1 and IES 2 has been deposited in the DDBJ/EMBL/GenBank database, accession no. U76541.

after developmental excision of the IES, thus fixing a macronuclear version of the gene in the micronucleus, much as reverse transcription of spliced RNA molecules can create intron-free genes (23).

This situation is similar to that found for *Paramecium* A and B surface antigen genes, which share 70% nucleotide identity and belong to a family of ~10 genes characterized by mutually exclusive expression (24). The 8 kb A gene contains eight IESs within coding and immediate upstream sequences (16), while the B gene contains only four IESs, three and probably all four of which are in conserved positions with respect to the IESs of the A gene (18). It thus appears likely that some or all of these IESs were present in the ancestral surface antigen gene and were maintained, while the others were lost or gained subsequent to gene duplication. Although IES position seems to be conserved, the sequence and size of paralogous IESs are highly variable, in favour of the hypothesis that only the inverted terminal repeats, which resemble those of Tc1-related transposons, and a minimal size (see below) are necessary for developmental excision of *Paramecium* IESs (22).

**Small IESs and tiny introns**

We report here the first example in *Paramecium* of an IES in an intron. In *Tetrahymena*, all known IESs are located in non-transcribed regions of the genome with one exception: a germline-specific sequence (named mse2.9) was found in an intron of a gene of unknown function (25). Since IES excision is imprecise and generates some junction heterogeneity in *Tetrahymena* (26), it was suggested that IESs would not be tolerated in *Tetrahymena* coding sequences (25). Subsequent analysis of caryonidal clones confirmed that excision of mse2.9 does generate considerable junction microheterogeneity (27).

In hypotrich ciliates and in *Paramecium*, IESs are found in both coding and non-coding regions (13,15). Even if *Paramecium* IES excision is not always precise (an example of boundary microheterogeneity in excision of a *P.primaurelia* IES from a non-coding region has been observed; A.Le Mouel, K.Dubrana and L.Amar, personal communication), it clearly can be when IESs are situated in coding sequences. Evaluation of more micronuclear genes is needed in order to better evaluate IES distribution in coding and non-coding regions of the genome.

The finding of an IES in an intron may provide a clue as to the small size (19–33 bp) of *Paramecium* introns (9–11). Not only are these introns among the smallest characterized in any organism, but their size is remarkably homogeneous: no introns of larger size have yet been found. It is expected that these introns are spliced by a classic nuclear spliceosome machinery (28), since a preliminary statistical analysis of available intron sequences indicates that they harbour at least some of the signals found in yeast and vertebrate nuclear introns (C.Thermes and Y.Daubenton-Carafa, personal communication). The complexity of the spliceosome and the number of genes involved in RNA splicing rule out independent evolution of this process in *Paramecium*. It is, however, striking that the largest introns are ~30 bp in size (and most frequently bordered by GTA...TAG), while IESs (always bounded by TA...TA) tend towards a size of 28 bp, the IES size most frequently found so far in *Paramecium*, beyond which their developmental excision has been postulated to be inefficient (15,22). Although highly speculative, it seems worth considering the possibility that IESs which have become so small that they escape DNA excision can turn into introns and be removed by splicing. Such an adaptation to the removal of defectively small IESs could have driven *Paramecium* introns to their present small size, by exerting selective pressure for optimization and specialization of the splicing reaction on very small substrates. It will be

interesting to perform somatic transformation experiments with the micronuclear ICL1-b gene in order to determine whether the presence of the 75 bp IES within the second intron will inhibit RNA splicing.

## ACKNOWLEDGEMENTS

We are particularly grateful to Claire Bertrand for her participation in the present study during the preparation of the Magistre de Biotechnologie. We also wish to thank Laurence Amar, Mireille Betermier, Eric Meyer and Claude Thermes for useful discussions, Eric Meyer and John Preer for kindly providing us with phage libraries and Janine Beisson and Jean Cohen for critical reading of the manuscript. This work was financed by the GREG Genome Project (grant no. 94/70) and the CNRS. L.V. was supported by a pre-doctoral fellowship from the Ministre de l'Enseignement Suprieur et de la Recherche and L.M. by a 'poste rouge' from the CNRS.

## REFERENCES

- Tindall,S.H., DeVito,L.D. and Nelson,D.L. (1989)*J. Cell Sci.*, **92**, 441–447.
- Garreau de Loubresse,N., Klotz,C., Vignes,B., Rutin,J. and Beisson,J. (1991)*Biol. Cell*, **71**, 217–225.
- Sperling,L., Keryer,G., Ruiz,F. and Beisson,J. (1991)*Dev. Biol.*, **148**, 205–218.
- Coffe,G., Le Caer,J.-P., Lima,O. and Adoutte,A. (1996)*Cell Motility Cytoskeleton*, **34**, 137–151.
- Madeddu,L., Gautier,M.-C., Vayssi,L., Houari,A. and Sperling,L. (1995)*Mol. Biol. Cell*, **6**, 649–659.
- Madeddu,L., Klotz,C., Le Caer,J.-P. and Beisson,J. (1996)*Eur. J. Biochem.*, **238**, 121–128.
- Gautier,M.-C., Sperling,L. and Madeddu,L. (1996)*J. Biol. Chem.*, **271**, 10247–10255.
- Salisbury,J.L. (1995)*Curr. Opin. Cell Biol.*, **7**, 39–45.
- Dupuis,P. (1992)*EMBO J.*, **11**, 3713–3719.
- Russell,C.B., Fraga,D. and Hinrichsen,R.D. (1994)*Nucleic Acids Res.*, **22**, 1221–1225.
- Fraga,D. and Hinrichsen,R.D. (1994)*Gene*, **147**, 145–148.
- Sonneborn,T.M. (1974) In King,R. (ed.), *Handbook of Genetics*. Plenum Publishing Corp., New York, NY, pp. 469–594.
- Prescott,D.M. (1994)*Microbiol. Rev.*, **58**, 233–267.
- Yao,M.C. (1996)*Trends Genet.*, **12**, 26–30.
- Klobutcher,L.A. and Herrick,G. (1996)*Prog. Nucleic Acids Res. Mol. Biol.*, **56**, in press.
- Steele,C.J., Barkocy-Gallagher,G., Preer,L.B. and Preer,J.R. (1994)*Proc. Natl. Acad. Sci. USA*, **91**, 2255–2259.
- Amar,L. (1994)*J. Mol. Biol.*, **236**, 421–426.
- Scott,J., Leeck,C. and Forney,J. (1994)*Nucleic Acids Res.*, **22**, 5079–5084.
- Meyer,E. and Keller,A.-M. (1996)*Genetics*, **143**, 191–202.
- Sambrook,J., Fritsch,E. and Maniatis,T. (1989)*Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Church,G.M. and Gilbert,W. (1984)*Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
- Klobutcher,L.A. and Herrick,G. (1995)*Nucleic Acids Res.*, **23**, 2006–2013.
- Maestre,J., Tchenio,T., Dhelin,O. and Heidmann,T. (1995)*EMBO J.*, **14**, 6333–6338.
- Caron,F. and Meyer,E. (1989)*Annu. Rev. Microbiol.*, **43**, 185–188.
- Heinonen,T.Y.K. and Pearlman,R.E. (1994)*J. Biol. Chem.*, **269**, 17428–17433.
- Austerberry,C.F. and Yao,M.C. (1988)*Mol. Cell. Biol.*, **8**, 3947–3950.
- Li,J. and Pearlman,R.E. (1996)*Nucleic Acids Res.*, **24**, 1943–1949.
- Moore,M.J., Query,C.C. and Sharp,P.A. (1993) In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.