Nucleic Acids Research, 1997, Vol. 25, No. 10 **1883–1889**

Evolution of internal eliminated segments and scrambling in the micronuclear gene encoding DNA polymerase α in two *Oxytricha* species

David C. Hoffman and David M. Prescott*

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309-0347, USA

Received February 24, 1997; Accepted March 27, 1997

DDBJ/EMBL/GenBank accession nos U89697, U89698

ABSTRACT

To learn about the evolution of internal eliminated segments (IESs) and gene scrambling in hypotrichous ciliates we determined the structure of the micronuclear (germline) gene encoding DNA polymerase α (DNA pol α) in *Oxytricha trifallax* and compared it to the previously published structure of the germline DNA pol α gene in *Oxytricha nova*. The DNA pol α gene of *O.trifallax* contains 51 macronuclear-destined segments (MDSs) separated by 50 IESs, compared to 45 MDSs and 44 IESs in the *O.nova* gene. This means that IESs and MDSs have been gained and/or lost during evolutionary divergence of the two species. Most of the MDSs are highly scrambled in a similar non-random pattern in the two species. We present a model to explain how IESs, non-scrambled MDSs and scrambled MDSs may be added and/or eliminated during evolution. Corresponding IESs in the two species differ totally in sequence, and junctions between MDSs and IESs are shifted by 1–18 bp in *O.trifallax* compared to the *O.nova* gene. In both species a short region of the gene is distantly separated from the main part of the gene. Comparison of the gene in the two species shows that IESs and scrambling are highly malleable over evolutionary time.

INTRODUCTION

Two profound alterations have occurred in the micronuclear (germline) genome of hypotrichous ciliates during evolution. First, internal eliminated segments (IESs) have been introduced into most micronuclear genes. IESs are short, AT-rich, unique sequences bounded by short repeats (reviewed in 1). The micronuclear genome of *Oxytricha nova* is estimated to contain ~200 000 IESs (2). Second, in some micronuclear genes macronuclear-destined segments (MDSs), which are separated by IESs, have been scrambled (reviewed in 2). IESs are excised and MDSs are re-ordered and precisely spliced during macronuclear development after cell mating (conjugation) to produce gene-sized molecules competent for transcription and translation.

To gain an insight into the possible significance of IESs and of MDS scrambling, we are investigating how these alterations in micronuclear genes may change during evolution. In this paper we compare the structure of the micronuclear DNA pol α gene in *Oxytricha trifallax* with that of the gene in the closely related species *O.nova* (3). We present a model to explain how IESs and scrambled MDSs may be added to and/or subtracted from the micronuclear genome during evolution.

MATERIALS AND METHODS

Cell culture, isolation of nuclei and preparation of genomic DNA

Culturing of *O.nova* and *O.trifallax*, isolation of macro- and micronuclei and preparation of macro- and micronuclear DNAs were performed as described earlier (4).

Nuclease BAL31 treatment of micronuclear DNA preparations and PCR

Micronuclear DNA preparations were treated with nuclease BAL31 (Boehringer-Mannheim, Indianapolis, IN) to remove trace macronuclear DNA contamination as follows: ~300 ng DNA was digested with 1.5 U BAL31 in 1× BAL31 buffer (20 mM Tris–HCl, pH 7.2, 600 mM NaCl, 12.5 mM $MgCl_2$, 12.5 mM $CaCl_2$ and 1 mM EDTA) in a total volume of 10 µl at 30°C for 15 min. For PCR, treated DNA was added to a 100 µl PCR reaction mixture. The mixture was 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris–HCl, pH 8.8 at 25°C, 2–4 mM $MgSO_4$ and 0.1% Triton X-100 containing the four dNTPs (125 mM each; Pharmacia) and 10–20 pmol each primer. Either macronuclear DNA (10 ng) or untreated (~150 ng) or BAL31-treated (~300 ng) micronuclear DNA was used as template. After a hot start at 94°C for 10 min, 1 U Vent (exo^-) DNA polymerase (NEB) was added and thermal cycling initiated. PCR was performed in either a Hybaid thermal reactor (National Labnet, Woodbridge, NJ) or an Ericomp thermal reactor (Ericomp, San Diego, CA) for 36–40 cycles of: 94°C for 30 s, 58–65°C for 45 s and 72°C for 40–60 s.

Subcloning

PCR products were purified with the Qiaquick PCR Cleanup Kit (Qiagen). PCR products with restriction sites incorporated into the primers were digested with restriction enzymes and repurified. PCR products were cloned into either pUC19 (NEB) or pGEM-7zf+ (Promega). Ligations were performed using T4 DNA ligase (NEB). Competent *Escherichia coli* strain DH5α

*To whom correspondence should be addressed. Tel: +1 303 492 8381; Fax: +1 303 492 7744; Email: prescotd@spot.colorado.edu

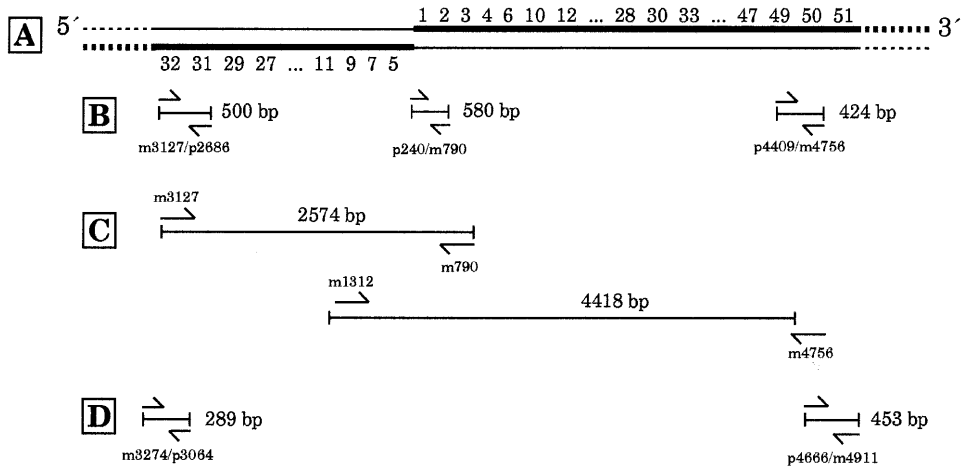

Figure 1. PCR strategy for analyzing the micronuclear DNA pol α gene of *O.trifallax*. (**A**) Linear representation of the structure ultimately determined for the micronuclear DNA pol α gene in *O.trifallax*. The thin line is the template for transcription. An inversion is present between MDSs 1 and 5. Some MDS numbers have been omitted (...) for space reasons. IESs (not depicted) are present between successive MDSs, beginning with IES 1 separating MDSs 32 and 31, followed by IES 2 between MDSs 31 and 29, etc. Dashed lines depict sequences flanking the gene in the micronucleus. (**B**) PCR products obtained from non-scrambled segments of the *O.trifallax* micronuclear DNA pol α gene using primers designed from the macronuclear gene sequence. Primers are labeled below the PCR products. (**C**) Long micronucleus-specific PCR products obtained with primers spanning the inversion point in the micronuclear gene (the inversion point was initially predicted from the *O.nova* micronuclear gene). (**D**) PCR products obtained from *O.trifallax* micronuclear DNA that include the ends of MDSs 32 and 51. PCR products are not drawn to scale. Primers designed from the parts of the macronuclear gene sequence corresponding to MDSs 34, 36, 38, 40, 42, 44, 46 or 48 gave no PCR products when paired with primers p2686 or p3064 above. The sequences have been submitted to GenBank, accession numbers U89697 (main portion of micronuclear gene) and U89698 (PCR product containing MDSs 34-36-38-40-42).

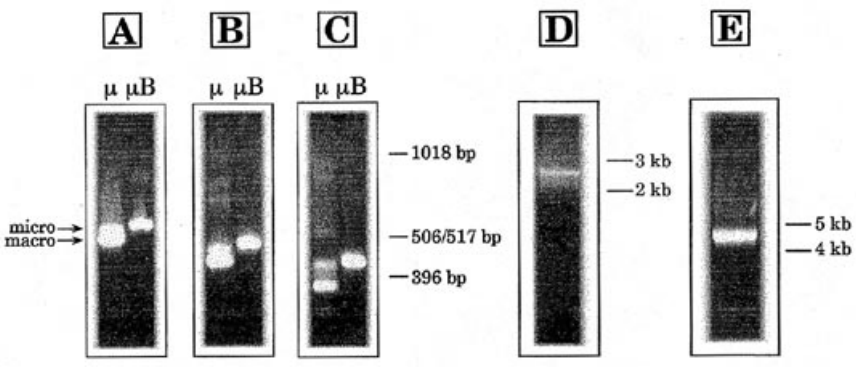

Figure 2. PCR products from the DNA pol α gene amplified from untreated micronuclear DNA (μ) or micronuclear DNA treated with the nuclease BAL31 (μB). (**A**) Primer pair p240/m790 yielded a micronuclear product of 580 bp (including parts of MDSs 1 and 2 and the intervening IES), and a macronuclear product of 635 bp. (**B**) Primer pair p3127/m2686 yielded a micronuclear product of 500 bp (including parts of MDSs 31 and 32 and the intervening IES), and a macronuclear product of 449 bp. (**C**) Primer pair p4409/m4756 yielded a micronuclear product of 424 bp (including parts of MDSs 49 and 50 and the intervening IES), and a macronuclear product of 355 bp. (**D**) Primer pair m3127/m790 yielded a micronucleus-specific product of 2574 bp extending from MDS 32 across the inversion point and into MDS 2. (**E**) Primer pair m1312/m4756 yielded a micronucleus-specific product of 4418 bp, extending from MDS 9 across the inversion point and into MDS 50. Primer pairs used in (**D**) and (**E**) amplify the micronuclear DNA pol α gene but not the macronuclear gene; both primers in each pair anneal to the non-coding strand of the gene and thus give no product from macronuclear DNA.

was transformed by the standard heat shock method as described elsewhere (5) or by a high efficiency method without heat shock (6).

Plasmid DNA preparation and sequencing

Small scale preparations of plasmid DNA were obtained by the alkaline lysis method using the Qiaprep Spin Plasmid Miniprep Kit (Qiagen). Plasmid inserts were sequenced by the dideoxynucleotide chain termination method using Sequenase DNA polymerase (Amersham Life Sciences) and [α-^{35}S]dATP (Dupont/New England Nuclear) incorporation. Some PCR products were sequenced without cloning by the same method using the Sequenase PCR Product Sequencing Kit (Amersham Life Sciences). Sequencing primers were obtained from either DNA International, Integrated DNA Technologies Inc., Gibco Life Sciences (Grand Island, NY) or Only DNA (Midland, TX). Plasmid sequencing gels were 6% Long Ranger acrylamide (FMC Bioproducts) in 7 M urea, 1× TBE (89 mM Tris–borate, 89 mM boric acid and 2 mM EDTA, pH 8.0). PCR product sequencing gels were 6% Long Ranger, 7 M urea, 1× GTB (89 mM Tris–HCl, 29 mM taurine and 0.54 mM EDTA, pH 8.0). Sequence alignments and analysis were performed on an Apple Macintosh computer with the AssemblyLIGN/MacVector sequence analysis package (IBI-Kodak, Rochester, NY).

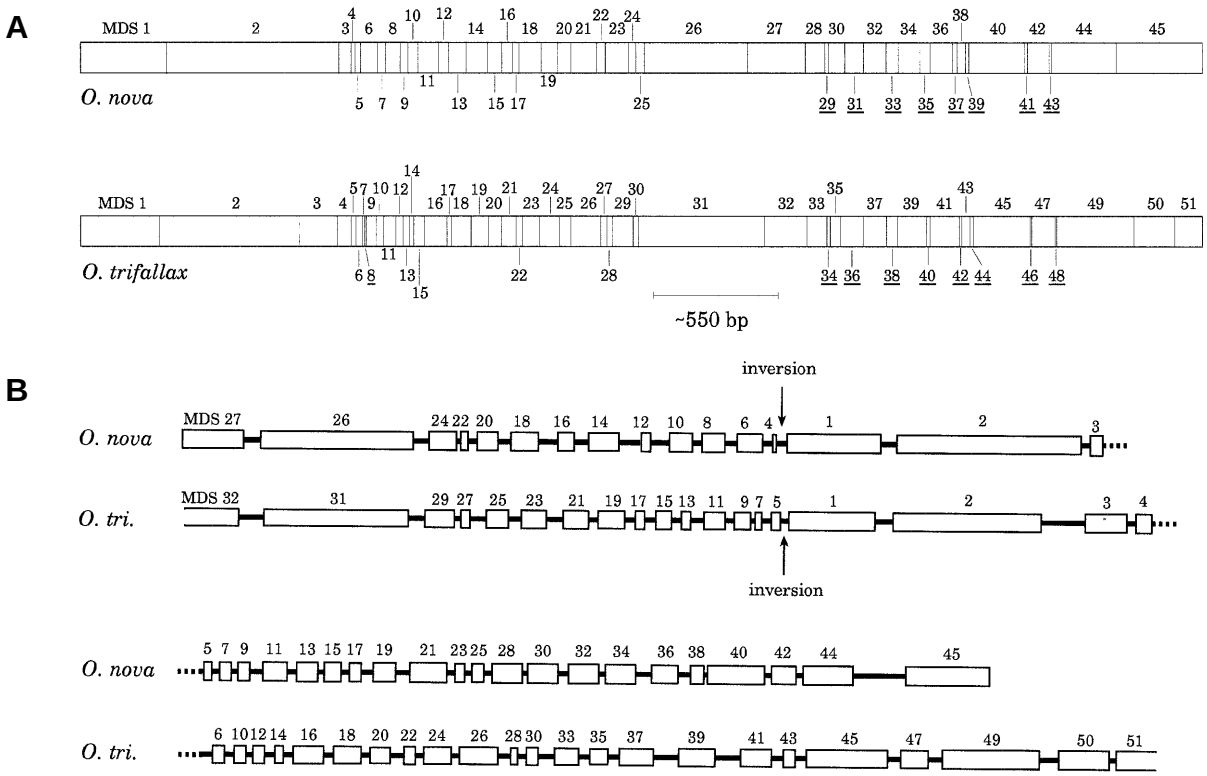

Figure 3. (**A**) Comparison of the number and relative size of MDSs in the macronuclear DNA pol α genes of *O.nova* and *O.trifallax*. The numbers of MDSs that are missing from the main body of the micronuclear gene in each species are underlined. (**B**) The scrambled arrangement of the main body of the DNA pol α genes in *O.nova* and *O.trifallax*. MDSs are open boxes; IESs are thick black lines. MDSs 32 and 51 of *O.trifallax* are drawn as open-ended boxes because the sequence at their 5′ or 3′ ends, respectively, have not been determined.

RESULTS

Sequence of the macronuclear DNA pol α gene of *O.trifallax*

The *O.trifallax* DNA pol α clone was isolated from a macronuclear genomic library using the *O.nova* gene as probe. A single λ clone was purified by standard procedures and shown to contain an insert of 4952 bp, not including 20 bp of telomeric repeats at the 5′-end and 28 bp at the 3′-end. The presence of telomeres indicates that the λ clone contained an intact, macronuclear gene-sized molecule. The complete macronuclear sequence has been submitted to the GenBank sequence database, accession no. U59426, and is described elsewhere (7).

Sequence of the micronuclear DNA pol α gene of *O.trifallax*

PCR products from the micronuclear DNA pol α gene of *O.trifallax* were obtained by assuming that the *O.trifallax* gene was scrambled in a similar pattern to the *O.nova* gene (Fig. 1A). The micronuclear genes encoding actin I and the α telomere-binding protein in *O.nova* are similarly scrambled in *O.trifallax* (8; J.D.Prescott, M.L.DuBois and D.M.Prescott, unpublished results). First, three PCR primer pairs designed from the *O.trifallax* macronuclear DNA pol α gene were used to amplify portions of the gene predicted to be non-scrambled based on the structure of the *O.nova* micronuclear gene (Fig. 1B). The three primer pairs were applied to total macronuclear DNA (control), total micronuclear DNA and total micronuclear DNA treated with nuclease BAL31. The BAL31 treatment was done because preparations of micronuclei are usually contaminated with a few fragments of

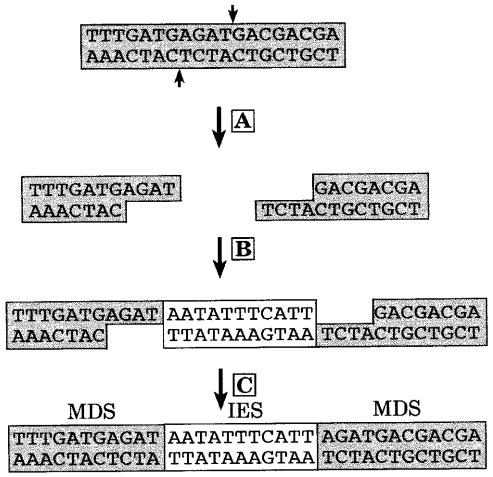

Figure 4. A model for IES insertion into micronuclear genes (1). (**A**) A double-stranded break is made in chromosomal DNA, leaving single-stranded ends. (**B**) A short, AT-rich piece of DNA is inserted into the break. (**C**) The gaps are filled in and ligated. This model results in the introduction of an IES and the creation of the repeat sequences that are presumed to guide IES removal during macronuclear development. MDSs are grey boxes; IESs are open boxes. Repeat sequences are underlined. Cut sites are indicated with short arrows.

macronuclei per ~10^4 micronuclei. The contaminating macronuclear DNA in untreated micronuclear DNA preparations generates a non-scrambled but shorter PCR product. Brief BAL31 nuclease treatment removes the short macronuclear DNA molecules

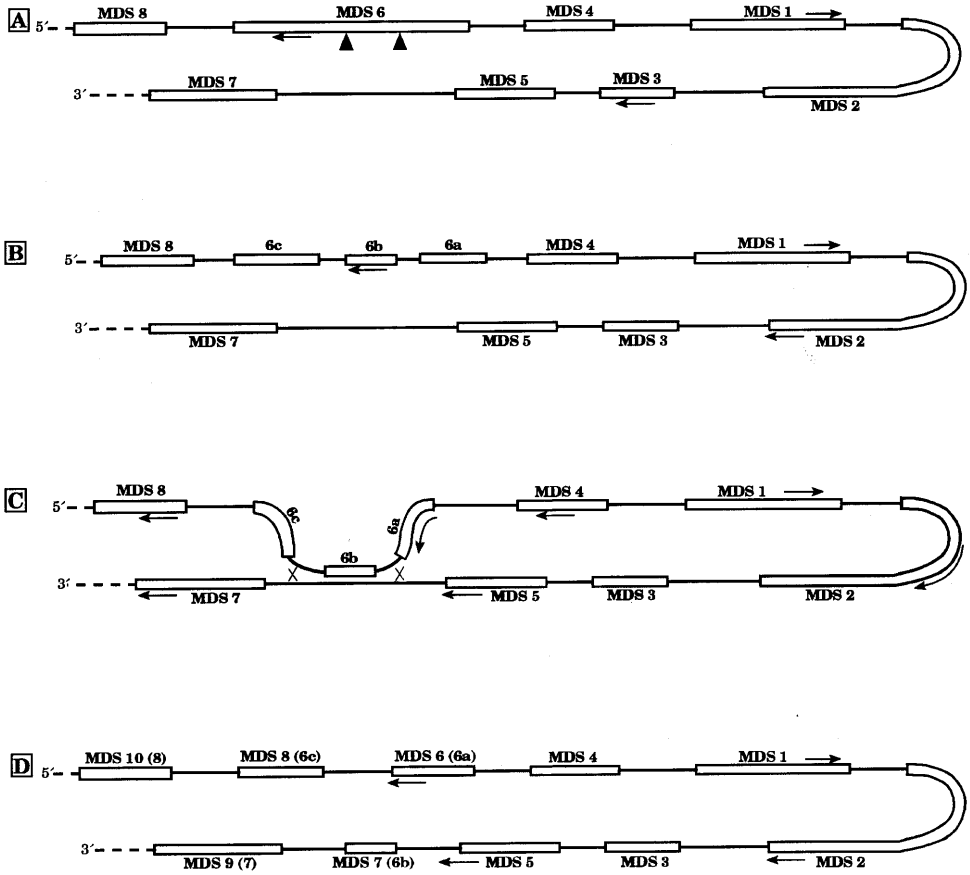

Figure 5. A model for the introduction of additional scrambled MDSs into the micronuclear DNA pol α gene in *O.trifallax* (M. DuBois, personal communication). (**A**) Two new IESs are inserted into MDS 6 at the black wedges, by the mechanism described in Figure 4B. The micronuclear gene with two new, non-scrambled MDSs (MDSs 6a, 6b and 6c have replaced MDS 6) folds into a hairpin similar to that thought to be required for unscrambling during macronuclear development (3). (**C**) Two intramolecular recombination events (indicated by crosses) occur between the two IESs flanking MDS 6b and the IES separating MDSs 6 and 7, thereby moving MDSs 6c, 6a and 6b into non-randomly scrambled positions as shown in (**D**). MDSs in (D) have been renumbered to reflect the addition of two MDSs to the gene; old MDS numbers are included in parentheses. MDSs are open boxes; IESs are thin black lines. Arrows indicate the direction of transcription.

(~5 kb for the DNA pol α gene), but leaves the very long (>500 kb) micronuclear DNA molecules intact (9,10).

Figure 2A–C shows the results of PCR with untreated and BAL31-treated micronuclear DNA using primers designed from the macronuclear sequence. PCR with untreated micronuclear DNA as template yielded two products, one from contaminating macronuclear DNA and one from micronuclear DNA (Fig. 2A–C, lanes marked μ). PCR with a BAL31-treated micronuclear template yielded only the micronucleus-specific product (Fig. 2A–C, lanes marked μB). The micronucleus-specific PCR products generated by each of these three primer pairs were cloned into the plasmid pUC19. Sequencing revealed the presence of portions of two non-scrambled MDSs separated by an IES in each case.

Next, a PCR product was generated with the primer pair m3127/m790, extending from MDS 32 to MDS 2 (Figs 1C and 2D). Most of the rest of the gene was amplified with the primer pair m1312/m4756 (Figs 1C and 2E). This yielded a single PCR product extending from MDS 11 to MDS 50 (Fig. 1C). Because of the inversion, both PCR primers in each pair used in these experiments (m3127/m790 and m1312/m4756) anneal to the same strand, obviating the need to treat the genomic micronuclear DNA template with nuclease BAL31 (see Fig. 2D and E). Last, primers designed from the macronuclear gene sequence were used to generate short PCR products from BAL31-treated micronuclear DNA to obtain the sequence at the ends of MDSs 32 and 51 (Fig. 1D). From the sequences of these overlapping PCR products a map of the micronuclear DNA pol α gene in *O.trifallax* was derived (Fig. 1A).

The macronuclear DNA pol α gene-sized molecule of *O.trifallax* is 4952 bp long, excluding telomeres, with an open reading frame (ORF) of 4542 bp (7). Comparison of the macronuclear gene-sized molecule with the ~7.2 kb of micronuclear sequence obtained by PCR revealed that the micronuclear DNA pol α gene is divided into at least 51 MDSs separated by at least 50 IESs. Numbers were assigned to MDSs according to the orthodox order in which they appear in the translationally competent macronuclear gene (Fig. 3A). Forty-two MDSs are present in the cloned PCR products of the *O.trifallax* gene, but nine scrambled MDSs containing 193 bp out of the 4542 bp ORF (~4% of the ORF) are missing. The 42 MDSs contained in the micronuclear PCR clones are in the order 32-31-29-27-25-23-21-19-17-15-13-11-9-7-5-1-2-3-4-6-10-12-14-16-18-20-22-24-26-28-30-33-35-37-39-41-43-45-47-49-50-51 (Fig. 3B). Fifteen MDSs at one end of the gene (MDS 32 and odd numbered MDSs 31–5) are inverted with respect to the remaining 27 (MDSs 1–4, 6, even numbered MDSs 10–30, odd numbered MDSs 33–47 and MDSs 49–51). The nine

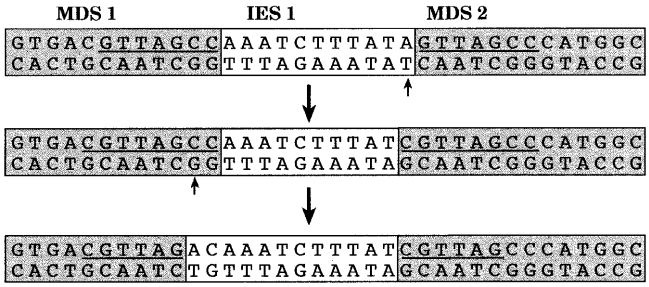

Figure 6. A model for MDS–IES junction shifting (1). A single base pair change in the IES from an AT to a CG lengthens the repeat sequence by 1 bp. A second single base pair mutation from a CG to an AT shortens the repeat to 6 bp and creates an MIJ shift of 2 bp towards the 5′ end of the gene. MDS sequences are in grey boxes; IES sequences are in open boxes. Repeats are underlined, and mutations are indicated with short arrows. Modified from Prescott and DuBois (1).

MDSs that were not present in PCR clones are 8 and 34-36-38-40-42-44-46-48 (compare the *O.trifallax* macronuclear and micronuclear configurations in Fig. 3A and B).

The missing MDSs 34–48 interdigitate with MDSs 33-35-37-39-41-43-45-47-49 in the micronuclear sequence determined from cloned PCR products. Assuming the minimum of eight MDSs (the missing 187 bp could theoretically be divided into more than eight MDSs), these missing MDS sequences range in length from 5 to 94 bp, excluding repeats at the ends of MDSs because their lengths are not known (see below for data on repeats). The ninth missing MDS, MDS 8 (6 bp without repeats) is presumably present separately somewhere else in the genome. Because the cloned PCR sequences were obtained using primers designed from the macronuclear sequence, flanking sequences could not be determined.

The previously characterized *O.nova* micronuclear gene, selected from a micronuclear DNA library in bacteriophage λ (3), lacks eight MDSs (~230 bp) that correspond approximately to the same region missing from the *O.trifallax* gene. The MDSs missing from the *O.nova* gene are not present in the first several kilobases of flanking DNA on either end of the cloned micronuclear gene. PCR with BAL31-treated *O.trifallax* micronuclear DNA as template and primers designed from the macronuclear sequence identified a product of ~700 bp (data not shown) that contains MDSs 36-38-40-42-44 (163 bp of the 193 bp missing from the ORF). The remaining missing MDSs, 34, 46 and 48, are presumably present in the DNA adjacent to the segment that yielded the 700 bp PCR product. MDS 8 could not be identified in micronuclear DNA by PCR because of its short length (6 bp without repeats). The location of the segment containing MDSs 36-38-40-42-44 (and presumably MDSs 34, 46 and 48) relative to the bulk of the gene in the *O.trifallax* genome remains unknown.

Repeat sequences at the ends of MDSs

Repeats (6–15 bp) are present at the ends of those MDSs that must be re-ordered during macronuclear development and repeats of 2–8 bp are present where MDSs are already in the correct order with respect to one another and only require IES excision (Table 1). These repeat pairs are thought to guide recombination between MDSs, splicing them in the orthodox order, with accompanying elimination of one copy in each repeat pair and elimination of the intervening IESs (3,4,10).

Table 1. Repeat pairs of the *O.trifallax B* micronuclear DNA pol α gene

5′ MDS/IES junction	MDS no.	3′ MDS/IES junction
5′ telomere addition site	1	AGATA
AGATA	2	ATT
ATT	3	ATA
ATA	4	ATGATGAGTGGAAT
ATGATGAGTGGAAT	5	AACAGAAC
AACAGAAC	6	AGAAATATG
AGAAATATG	7	n.d.
n.d.	9	TTATCATT
TTATCATT	10	AAAATAAT
AAAATAAT	11	GTTTCTTG
GTTTCTTG	12	ATGCAAA
ATGCAAA	13	TAAAATGA
TAAAATGA	14	AGAGGAG
AGAGGAG	15	TAATGATGG
TAATGATGG	16	ATGGTGAG
ATGGTGAG	17	AAAATCAA
AAAATCAA	18	AAAGCATGCTTG
AAAGCATGCTTG	19	GATTTCAAGAAAA
GATTTTAAGAAAA	20	GTTACTCTTG
GTTACTCTTG	21	GCTCAATAAAA
GCTCAATAAAA	22	ATCTTG
ATCATG	23	AAAACTT
AAAACTT	24	GAGAGATAGA
GAGAGATAGA	25	TAGTTGCTC
TAGTTGCTC	26	AAGCTAGATTTT
AAGCTAGATTTT	27	GGAGGATC
GGAGGATC	28	CAAGATAA
CAAGATAA	29	GTTCAACT
GTTCAACT	30	ATAAGACTTTGATGA
ATAAGACTTTGATGA	31	CTAATGAA
CTAATGAA	32	n.d.
n.d.	36	CTTGAGAT
CTTGAGAT	37	AAAGTAGTTTAG
AAAGTAGTTTAG	38	CACTTTCAA
CACTTTCAA	39	ATGAAAAATAA
ATGAAAAATAA	40	CCTTGGATCA
CCTTGGATCA	41	AAGAGTGAAT
AAGAGTGAAT	42	TGAACAACTTT
TGAACAACTTT	43	GTGCTTAG
GTGCTTAG	44	n.d.
n.d.	49	ATAAAA
ATAAAA	50	AT
AT	51	3′ telomere addition site

Mismatched nucleotides in two repeat pairs are underlined.
n.d., not determined.

DISCUSSION

The scrambled pattern of the DNA pol α gene

The DNA pol α gene is the seventh gene in *O.nova* for which a comparison has been made between micronuclear and macronuclear versions. Four of these genes, R1 (11), C2 (12), the gene encoding β telomere-binding protein (13) and the genes encoding heat shock protein 70 (14; Lindauer *et al.*, in preparation) and histone H4 (Arkind *et al.*, in preparation) contain IESs but are non-scrambled. Three genes, encoding actin I (4,14), α telomere-binding protein (11) and DNA pol α (3), are scrambled. The micronuclear actin I and α telomere-binding protein genes in *O.trifallax* are scrambled in essentially the same pattern as in *O.nova*, but with

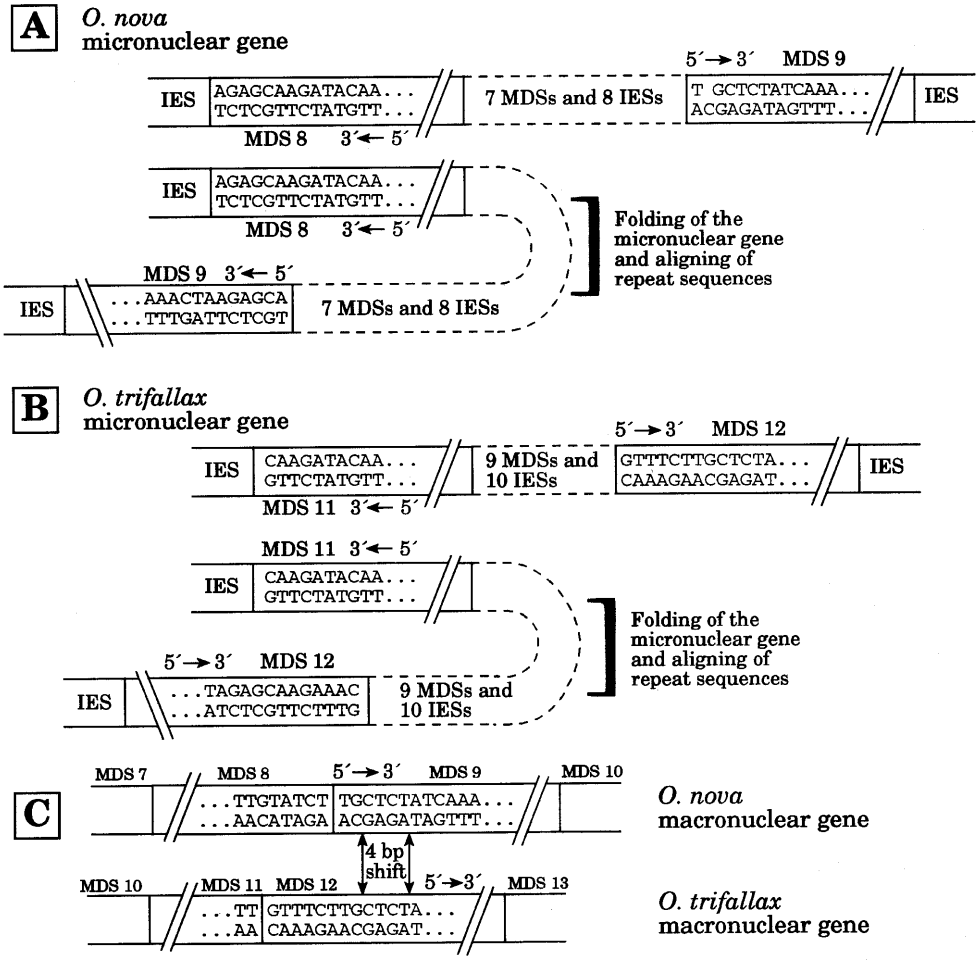

Figure 7. An example of MDS–IES junction shifting. (**A**) Folding of micronuclear DNA of *O.nova* to align two copies of the repeat sequence TGCTCT at the junction of MDS 8 with an adjacent IES and at the junction of MDS 9 with an adjacent IES. After alignment of repeat pairs, a site-specific recombination event is postulated to occur that joins the two MDSs (see panel C) and removes the intervening IES along with one copy of the repeat sequence. (**B**) Folding of micronuclear DNA of *O.trifallax* to align two copies of the repeat sequence GTTTCTTG at the junction of MDS 11 with an adjacent IES and at the junction of MDS 12 with an adjacent IES. (**C**) Partial macronuclear sequence of MDSs 8 and 9 of *O.nova* and MDSs 11 and 12 of *O.trifallax* after unscrambling. MDSs 8 and 9 of *O.nova* contain the same nucleotide sequence as MDSs 11 and 12 in *O.trifallax*. The MDS/IES junction shift for the repeat guiding the unscrambling of MDSs 8 and 9 in *O.nova is* 4 bp to the 5′ end of the gene compared to the same region of *O.trifallax* (MDSs 11 and 12), yielding the same macronuclear sequence after unscrambling.

additional IESs and MDSs (8; J.D.Prescott, M.L.DuBois and D.M.Prescott, unpublished results).

The scrambled patterns of the DNA pol α micronuclear genes in *O.nova* and *O.trifallax* are non-random. Like the αTP gene (14 MDSs in *O.nova*), most of its MDSs fall into odd and even numbered series. However, unlike the αTP gene, ~1.9 kb of the DNA pol α gene are inverted relative to the rest of the gene (Fig. 3B). The actin I micronuclear gene in *O.nova* and *O.trifallax* contains a single, inverted MDS (MDS 2), but the scrambling pattern of the actin I gene appears random, unlike the DNA pol α gene. Thus, the three scrambled micronuclear genes characterized so far show three different combinations of inversions and random/non-random scrambled patterns of MDSs. Most of the MDSs in all three genes are scrambled, but a few are not: two in the actin I gene, three in the αTP gene, seven in the *O.nova* DNA pol α gene and nine in the *O.trifallax* DNA pol α gene. Although the scrambled micronuclear DNA pol α gene shares structural features with both previously characterized scrambled genes, it represents a new and different scrambling pattern.

Like the αTP and actin I genes, the scrambled pattern of DNA pol α has no discernible relationship to conserved amino acid domains within the predicted polypeptide. For example, the macronuclear DNA pol α gene in both *O.nova* and *O.trifallax* contains 12 domains of amino acid sequence conserved among a wide variety of other organisms from bacteriophage to humans and yeast (7,16), but these domains do not correspond in any detectable way to individual MDSs or groups of MDSs. In some cases, conserved domains are divided among two or three scrambled MDSs.

A comparison of the micronuclear DNA pol α gene of *O.nova* and *O.trifallax*

The DNA pol α gene in *O.trifallax* is scrambled in essentially the same way as in *O.nova*, but the *O.trifallax* gene is subdivided into 51 MDSs by 50 IESs, compared to 45 MDSs and 44 IESs in *O.nova* (Fig. 3B). PCR experiments failed to detect any non-scrambled or alternatively scrambled copies of the gene in

the micronuclear genome. The first 1234 bp in the *O.trifallax* gene are subdivided into four non-scrambled MDSs (MDSs 1–4; Fig. 3A), and the first 1233 bp in *O.nova* are subdivided into three non-scrambled MDSs (MDSs 1–3; Fig. 3A). In *O.nova* the 3′-end of the gene is divided into MDSs 44 and 45 by a single long IES of 223 bp, but in *O.trifallax* this region is divided into MDSs 49–51 by two short IESs of 69 and 10 bp (Fig. 3A). The other four additional MDSs in *O.trifallax* compared to *O.nova* are scrambled. The eight MDSs that are missing from the main body of the micronuclear gene of *O.trifallax* correspond closely in position in the ORF to the eight MDSs that are missing from the cloned micronuclear gene of *O.nova*. One scrambled MDS (MDS 8) is not present in the cloned micronuclear PCR product from *O.trifallax*. There is no corresponding MDS in *O.nova*.

The micronuclear DNA pol α gene in *O.trifallax* (Fig. 3B) contains an inversion in the same position as in *O.nova*. This strongly suggests that the DNA pol α gene became scrambled before *O.trifallax* and *O.nova* diverged from their common ancestor.

Evolution of the scrambled pattern in the micronuclear DNA pol α gene of two *Oxytricha* species

Comparison of the micronuclear DNA pol α gene in *O.nova* and *O.trifallax* showed that multiple IESs have either been introduced into the *O.trifallax* gene or lost from the *O.nova* gene. A model for the gain or loss of IESs and MDSs in a non-scrambled configuration has been proposed in which IESs are introduced with concomitant creation of short repeats, as explained in Figure 4.

Scrambled MDSs could be introduced in the micronuclear DNA pol α gene of *O.trifallax* while preserving its complex, non-random scrambled pattern as shown in Figure 5. In this example, two new IESs are inserted into MDS 6 of the *O.nova* micronuclear gene (Fig. 5A). Two intramolecular recombination events between the new IESs flanking MDS 6b and the IES separating MDSs 5 and 7 (indicated by crosses in Fig. 5C) remove MDS 6b from its position between MDSs 6a and 6c and splice it into the IES present between MDSs 5 and 7, thereby adding two IESs to the total and increasing the number of non-randomly scrambled MDSs in the *O.trifallax* micronuclear gene by two (Fig. 5D). Similar subdivision of MDS 10 in the *O.nova* gene could account for the two additional non-randomly scrambled MDSs present in the *O.trifallax* micronuclear gene.

Shifting of MDS/IES junctions

MDS/IES junctions can shift either in the 5′ or 3′ direction along micronuclear DNA (1,8). The IESs separating scrambled MDSs in the micronuclear actin I gene of *O.trifallax* are shifted 1–14 bp in the 5′ or 3′ direction relative to the corresponding IES in *O.nova*, and the IES separating the two non-scrambled MDSs has apparently shifted 138 bp in the 5′ direction compared to the *O.nova* actin I gene (8). Junctions are shifted in the DNA pol α gene by 1–18 bp. A proposed mutational model of junction shifting (Fig. 6) is supported by the high rate at which mutations accumulate in IES sequences. Two examples of junction shifting identified by comparison of the *Oxytricha* DNA pol α genes are shown in Figure 7. The largest shift in the DNA pol α gene is 18 bp (data not shown), which completely changes the sequence of the repeat at the ends of the two MDSs that are to be spliced during macronuclear development.

Overall, comparison of the micronuclear DNA pol α genes in *O.nova* and *O.trifallax* shows that IES insertion and/or loss, with modification of the scrambling patterns of MDSs, occur continuously during evolution. These changes illustrate the extraordinary malleability of the germline genome in *Oxytricha* species. Such malleability is due in part to DNA sequence manipulations that are unlike other currently known phenomena.

ACKNOWLEDGEMENTS

The authors wish to thank Michelle DuBois for suggesting the model described in Figure 5, DeWight Williams for critical reading of the manuscript and Gayle Prescott for assistance with manuscript preparation. This work was supported by NSF grant MCB9506287 to D.M.P.

REFERENCES

1 Prescott,D.M. and DuBois,M.L. (1996) *J. Eukaryotic Microbiol.*, **43**, 432–441.
2 Prescott,D.M. (1994) *Microbiol. Rev.*, **58**, 233–267.
3 Hoffman,D.C. and Prescott,D.M. (1996) *Nucleic Acids Res.*, **24**, 3337–3340.
4 Prescott,D.M. and Greslin,A.F. (1992) *Dev. Genet.*, **13**, 66–74.
5 Hanahan,D. (1983) *J. Mol. Biol.*, **166**, 557–580.
6 Pope,B. and Kent,H.M. (1996) *Nucleic Acids Res.*, **24**, 536–537.
7 Hoffman,D.C. and Prescott,D.M. (1997) *J. Mol. Evol.*, in press.
8 DuBois,M. and Prescott,D.M. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 3888–3892.
9 Jahn,C.L. (1988) *Exp. Cell Res.*, **177**, 162–175.
10 Mitcham,J.L., Lynn,A.J. and Prescott,D.M. (1992) *Genes Dev.*, **6**, 788–800.
11 Ribas-Aparacio,R.M., Sparkowski,J.J., Proulx,A.E., Mitchell,J.D. and Klobutcher,L.A. (1987) *Genes Dev.*, **1**, 323–336.
12 Klobutcher,L.A., Jahn,C.L. and Prescott,D.M. (1984) *Cell*, **36**, 1045–1055.
13 Mitcham,J.L., Prescott,D.M. and Miller,M.K. (1994) *J. Eukaryotic Microbiol.*, **41**, 478–480.
14 Anderson,R.C., Lindauer,K.R. and Prescott,D.M. (1996) *Gene*, **168**, 103–107.
15 Greslin,A.F., Prescott,D.M., Oka,Y., Loukin,S.H. and Chappell,J.C. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6264–6268.
16 Wang,T.S.-F., Wong,S.W. and Korn,D. (1989) *FASEB J.*, **3**, 14–21.