GENOME GYMNASTICS: UNIQUE MODES OF DNA EVOLUTION AND PROCESSING IN CILIATES

David M. Prescott

In some ciliates, the DNA sequences of the germline genomes have been profoundly modified during evolution, providing unprecedented examples of germline DNA malleability. Although the significance of the modifications and malleability is unclear, they may reflect the evolution of mechanisms that facilitate evolution. Because of the modifications, these ciliates must perform remarkable feats of cutting, splicing, rearrangement and elimination of DNA sequences to convert the chromosomal DNA in the germline genome (micronuclear genome) into gene-sized DNA molecules in the somatic genome (macronuclear genome). How these manipulations of DNA are guided and carried out is largely unknown. However, the organization and manipulation of ciliate DNA sequences are new phenomena that expand a general appreciation for the flexibility of DNA in evolution and development.

CILIATES
Single-celled organisms containing a micronucleus (germline nucleus), a macronucleus (somatic nucleus) and cilia for swimming and food capture.

RNA SELF-SPLICING
Removal of the intron sequence from precursor RNA and splicing of the mature RNA by catalytic action of the intron RNA, showing the enzyme-like activity of RNA.

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Colorado 80309-0347, USA. e-mail: prescotd@spot.Colorado.edu

CILIATES make up the phylum Ciliophora, an ancient group of organisms that originated perhaps as long as 2×10^9 years ago[1]. These organisms have evolved into many widely divergent lineages consisting of at least 8,000 species, with undoubtedly thousands more remaining to be discovered. In recent years, ciliates have become increasingly important in molecular genetic research. Their general mode of development, which generates many thousands of individual DNA molecules, and therefore telomeres, per cell has made them particularly useful for studying telomeres. The repeat structure of telomeric DNA[2] and the enzyme telomerase[3] were first discovered in the ciliate *Tetrahymena thermophila*. Ciliates of the hypotrichous group (class Polyhymenophora, subclass Spirotricha, order Hypotrichida) were the first organisms in which the protein–DNA structure of telomeres was extensively analysed[4,5]. The solution to the 5′ primer problem — how to replicate the end of a linear DNA molecule — was also first proposed from the study of ciliate telomeres[6]. Another important genetic mechanism discovered first in ciliates (*Tetrahymena*) was RNA SELF-SPLICING[7].

Hypotrichous ciliates also possess extraordinary organizational features in their micronuclear and macronuclear genomes and carry out extensive developmental manipulations of their genomic DNA. These DNA phenomena, which considerably expand our view of genetic versatility, are the subject of this review.

Germline and somatic nuclear DNA

A single ciliate organism has two kinds of nuclei: a germline diploid nucleus (micronucleus) and a somatic nucleus (macronucleus). The number of nuclei per cell varies in different ciliates — *Oxytricha trifallax* has two micronuclei and two macronuclei, whereas *Oxytricha nova* has four micronuclei and two macronuclei (FIG. 1). The micronucleus undergoes meiosis during cell mating, haploid micronuclei are exchanged between the two cells in a mating pair, and a new diploid micronucleus is formed in each cell (FIG. 1). The mating cells separate, and a macronucleus develops from a copy of the new diploid micronucleus. Simultaneously, unused haploid micronuclei and the old macronucleus are destroyed.

Micronuclear DNA consists of extremely long molecules typical of eukaryotic chromosomes. Genes are scattered along this DNA and are separated by large

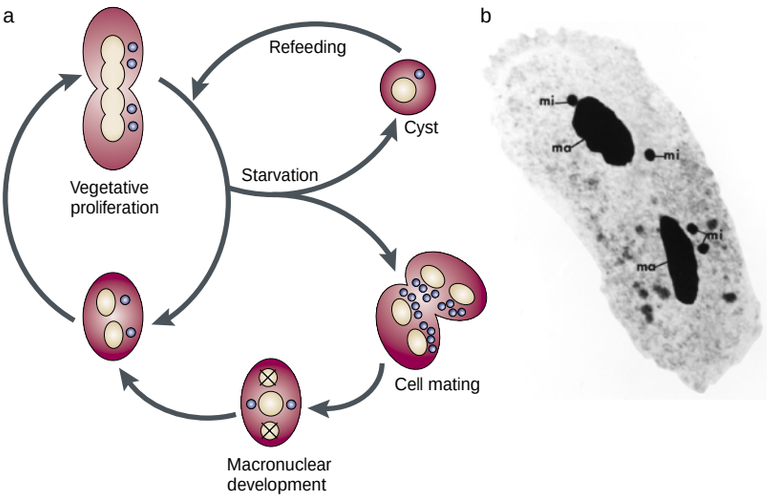

Figure 1 | **The behaviour of ciliate nuclei. a** | The life cycle of *Oxytricha trifallax* (two macronuclei (cream) and two micronuclei (blue)). When food organisms (usually the alga *Chlorogonium*) are available, ciliates proliferate continuously. Starvation induces cyst formation or cell mating. Only one macronucleus and one micronucleus are retained in a cyst. Dried, frozen cysts remain viable for decades; cysts hatch in water containing food organisms and resume proliferation. In mating, two cells join, undergo meiosis of micronuclei (eight haploid micronuclei are shown in each cell in a mating pair) and exchange haploid micronuclei. A diploid micronucleus is formed by the fusion of two haploid micronuclei. The new diploid micronucleus divides mitotically and one of them develops into a new macronucleus. The unused haploid micronuclei (not shown) and old macronuclei (X's) degenerate. The remaining new diploid micronucleus divides again to give two micronuclei. At the end of development, the macronucleus divides so that the nuclear complement of two macronuclei and two micronuclei is reconstituted. The cells resume proliferation if food is present or otherwise form cysts. **b** | Light micrograph of *Oxytricha nova*, stained to show the four micronuclei (mi) and the two macronuclei (ma).

stretches of spacer DNA (FIG. 2). The telomeres of micronuclear DNA are made up of hundreds of repeats of the sequence 5′-CCCCAAAA-3′ and end with a t-loop[8,9]. The t-loop is formed by a foldback of a single-stranded 3′ tail that invades a short stretch of the double-stranded telomere region.

Macronuclear DNA, in sharp contrast to micronuclear DNA, exists as short molecules that range in size from a few hundred base pairs (bp) to around 15,000 bp, with an average of about 2,000 bp. With few exceptions, each of the thousands of macronuclear molecules comprises a single gene, and has a consistent structure — a 5′ non-coding leader, an open reading frame and a 3′ non-coding trailer. The ends consist of a repetitive 20-bp telomere sequence (5′-$C_4A_4C_4A_4C_4$-3′) and a single-stranded 3′ tail. A heterodimeric protein binds tightly to the single-stranded tail, forming a protective telomere cap. So micronuclear and macronuclear DNAs use two very different mechanisms to form the telomere: a t-loop for the long micronuclear DNA molecules and a protein cap for the short macronuclear DNA molecules.

From chromosomal DNA to gene-sized DNA

The gene-sized molecules in the macronucleus are derived from micronuclear DNA when a copy of the new diploid micronucleus formed during cell mating develops a new macronucleus. A micronucleus begins this development by forming POLYTENE CHROMOSOMES. The polytene chromosomes are subsequently destroyed to yield gene-sized macronuclear DNA molecules to which telomeric repeats are added by telomerase. This release of macronuclear molecules is accompanied by elimination of the long spacers between genes in the micronuclear DNA. Spacer DNA accounts for around 95% of the SEQUENCE COMPLEXITY of micronuclear DNA (in *O. nova*); the gene-sized macronuclear molecules, which encode all nuclear RNA, have around 5% of the sequence complexity present in micronuclear DNA. In the final stage of macronuclear development, each gene-sized molecule is amplified to give one to several thousand copies. The mature macronucleus contains about 25×10^6 short DNA molecules (in *O. nova*); these are the smallest DNA molecules known to occur in nature. The 25×10^6 molecules provide 50×10^6 telomeres per macronucleus, which makes hypotrichs particularly useful for studying telomere structure.

In hypotrichous species from the Euplotes genus, excision of macronuclear molecules from micronuclear DNA seems to be guided by a 10-bp sequence that is 17 bp upstream or, more usually, 17 or 18 bp downstream of the cutting points for excision[10,11]. No such consensus sequence is present in other hypotrichs (such as Oxytricha, Stylonychia, Gastrostyla, Pleurotricha, Paraurostyla and Uroleptus), which are distantly related to Euplotes. This implies that the specificity of gene excision is achieved by different molecular mechanisms in the two groups of organism.

In the Oxytricha/Stylonychia group (but not in Euplotes), the ends of macronuclear molecules just inside the telomeres have an anomalous base composition. The 50 bp at the two ends (excluding the telomere repeats) violate 'PARITY RULE 2' (see also PARITY RULE 1). This rule, which is borne out by extensive observations on DNA base composition[13], states that A = T and G = C within a strand of DNA[12], and follows logically if the substitution rate for each base is equal in the two strands. Violation of parity rule 2 occurs, for example, when codon usage is biased towards codons rich in a particular base — commonly A or G. In the 50 bases at the two 5′ ends of macronuclear DNA molecules, which are part of the non-coding 5′ leaders and 3′ trailers, A > T and G > C (D.M.P. and S.J. Dizick, unpublished observations). This anomaly is not observed in the rest of the 5′ leaders and 3′ trailers, or in DNA immediately flanking a macronuclear sequence, as it resides in a micronuclear chromosome. So the ends of macronuclear molecules in micronuclear DNA are characterized by 50-bp islands of anomalous base composition and may function to direct excision of macronuclear genes from micronuclear chromosomes.

These DNA-processing events raise fundamental questions. How is the excision of the thousands of gene-sized macronuclear sequences from micronuclear DNA achieved? How is the spacer DNA between the excised genes marked for elimination? Furthermore, micronuclear versions of genes are generally transcriptionally silent and become activated during macronuclear development. How is this achieved? How is the amplification of gene-sized molecules regulated? What controls and directs destruction of superfluous haploid micronuclei

POLYTENE CHROMOSOME
A giant chromosome formed by many replications of the DNA. The replicated DNA molecules tightly align side-by-side in parallel register, creating a non-mitotic chromosome visible by light microscopy.

SEQUENCE COMPLEXITY
The number of different DNA sequences in a genome, originally measured by the rate of reassociation of heat-denatured DNA.

PARITY RULE 2
This rule derives from parity rule 1. In the absence of strand bias for mutation or selection, A = T and C = G within a strand of the double helix.

PARITY RULE 1
In the absence of strand bias for mutation or selection, the 12 substitution rates between all four bases reduce to six rates, that is A→T = T→A, G→C = C→G, A→G = T→C, G→A = C→T, C→A = G→T and A→C = T→G.

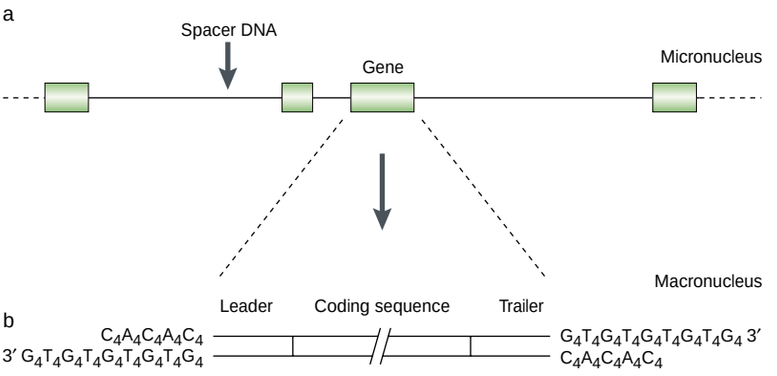

Figure 2 | **Ciliate genome organization. a** | Arrangement of genes in micronuclear DNA. **b** | Excision of a gene and telomere addition during macronuclear development. (Adapted from REF. 29.)

and the old macronuclei after cell mating? A micronucleus divides by mitosis but a macronucleus divides amitotically. How is this switchover accomplished during macronuclear development? Even without answers to these questions, the strategy seems clear. DNA processing eliminates the burden of non-coding DNA, producing a greatly simplified, stripped-down genome with a high copy number of each gene, which can support rapid vegetative cell proliferation.

Enigmatic interruptions in micronuclear genes

DNA processing in hypotrichs is considerably more complex than elimination of spacer DNA, release of gene-sized molecules and addition of telomere repeats. A comparison of the sequences of macronuclear molecules with their micronuclear precursors reveals that micronuclear versions are interrupted by multiple, short, (A+T)-rich, non-coding sequences called internal eliminated segments (IESs), which are eliminated during macronuclear development[14]. A typical example of a micronuclear precursor gene is the *O. trifallax* gene that encodes β-telomere-binding protein (βTP)[15] (FIG. 3). The first IES interrupts the micronuclear precursor in the 5′ non-coding leader just before the ATG start codon. Five more IESs interrupt the protein-coding region. Extrapolation from 12 micronuclear genes that have been sequenced in various hypotrichs (excluding Euplotes) gives an estimate of 100,000 to 200,000 IESs per haploid genome. Some species, such as *O. trifallax*, are particularly rich in IESs. All of these IESs are spliced out of all genes by an extraordinary amount of recombinational activity in a few hours during macronuclear development, primarily during the polytene chromosome stage[16]. The excision of IESs from a particular gene seems to follow a defined temporal programme. For example, of the nine IESs in the *actin I* gene in *O. trifallax* (FIG. 4), IES6 is always removed first, followed a few hours later by the removal of the remaining eight IESs in rapid succession (D.M.P. and E.A. Hewitt, unpublished observations).

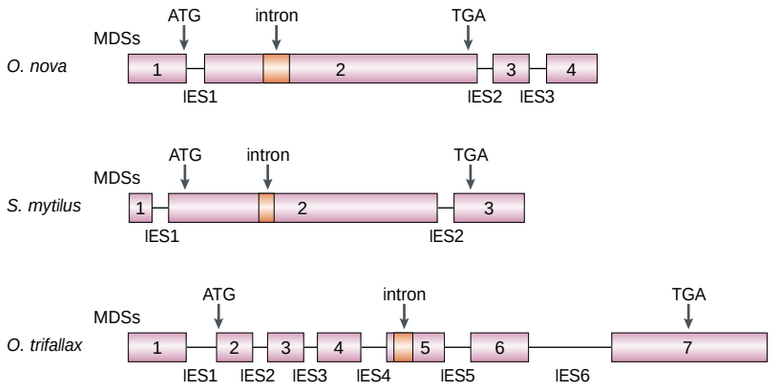

Figure 3 | **Interruptions in micronuclear genes.** Structure of the non-scrambled gene that encodes β-telomere-binding protein in *Oxytricha nova*, *Stylonychia mytilus* and *Oxytricha trifallax*. Internal eliminated segments (IESs) are lines between gene segments called macronuclear-destined segments (MDSs, purple blocks). The intron (red) interrupts the coding sequence at the same codon in all three species. (Adapted from REF. 18.)

Much is known about the structure and behaviour of IESs in evolution and development, although their significance remains a mystery. Most IESs in hypotrichs are <100 bp long, but a few are as large as 600 bp. IESs consist of unique sequences, containing 70–100% A+T, with no consensus sequence(s) or other shared sequence patterns. The number and positions of IESs in a particular gene have changed, sometimes greatly, during evolution. The micronuclear copy of the *βTP* gene contains two IESs in *Stylonychia mytilus*, three in *O. nova*, and six in *O. trifallax* (FIG. 3). None of the IESs interrupts the *βTP* gene in the same place in the three species, and none shares any length or sequence identity from species to species. Clearly, IESs are gained or lost during speciation. Comparison of the gene that encodes DNA polymerase α (pol α) in different ciliate species provides strong evidence that new IESs may be inserted into a gene during evolution[17], but there is no clear evidence that they may be lost. In keeping with their non-coding property, IESs accumulate mutations at a high rate and, oddly, at a higher rate than introns. Corresponding IESs in a particular gene in two ciliate species are almost always totally non-identical in sequence, in contrast to the coding regions of the gene, which are highly conserved in nucleotide and predicted protein sequences. Even the two alleles of a micronuclear gene in a cloned organism may differ substantially in their corresponding IES sequences, whereas the coding regions in the alleles are almost identical in nucleotide sequence.

It is difficult to imagine how IESs could originate *de novo* within a gene. It seems more likely that IESs are derived from DNA elsewhere in the genome and are inserted into genes in random events. Spacer DNA between micronuclear genes has the same (A+T)-rich composition as IESs, so it is possible that IESs originate from spacer DNA. It is unclear what drives the insertion of pieces of spacer DNA into genes to form IESs. However, it is clear that insertion of an IES creates a pair of repeats that flanks the ends of the new IES. This might occur through the introduction of a staggered cut into the target gene and ligation of a blunt-ended piece of (A+T)-rich DNA to the resultant single-stranded overhangs, followed by filling in of the single-stranded

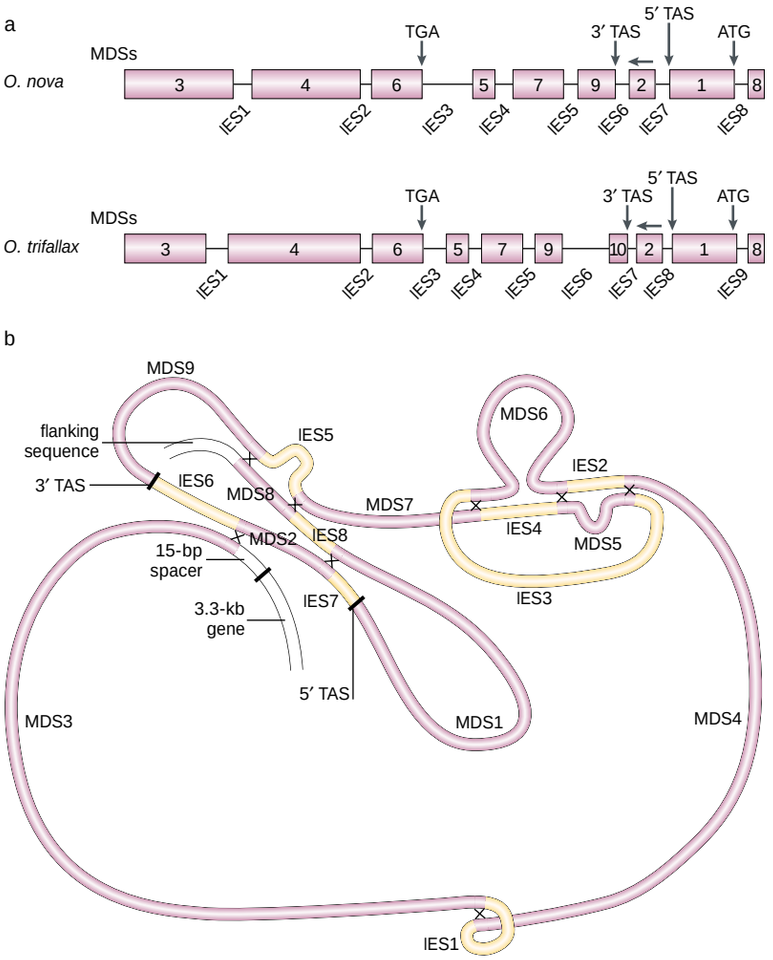

Figure 4 | **Scrambled genes. a** | Structure of the scrambled *actin I* gene in *Oxytricha nova* and *Oxytricha trifallax*. Macronuclear-destined segments (MDSs) are blocks; internal eliminated segments (IESs) are lines between blocks. MDS2 is inverted (indicated by the horizontal arrow) in both organisms. (Adapted from REF. 27.) **b** | Recombination model of unscrambling of the *actin I* gene in *O. nova*. Folding of micronuclear DNA aligns pairs of repeat sequences at the ends of MDSs. Recombinations between pairs of repeats are indicated by 'X's. The *actin I* gene is linked by a 15-bp spacer to a 3.3-kb gene. (TAS, telomere addition site.) (Adapted from REF. 21.)

gaps[18]. Whatever the origin of IESs, the ability to remove them precisely must have pre-existed in the cell when the first IESs were formed. This suggests that IES excision during macronuclear development is mediated by enzymes or factors that evolved earlier for other manipulations of DNA.

How are IESs recognized and then excised? The pairs of repeats that flank IESs may have a function in these activities. Recombination between the repeats would account for the observed retention of a single copy of the repeat in the macronuclear gene. However, the repeat sequences are too short to function as the sole indicator of IES ends. For example, the average length of repeats is only 4 bp in the *βTP* gene of *O. trifallax*. In some cases, the sequence of the repeat occurs several times within the IES and also a short distance upstream or downstream of the IES.Other sequence information must therefore be required to mark the ends of IESs. Perhaps the repeat sequences have no role in identifying and excising IESs but only function in the ligation of the ends of gene segments left by IES excision.

In the ciliate *Paramecium tetraurelia*, DNA in the old macronucleus seems to direct correct excision of an IES in the new micronucleus[19,20]. In a key experiment, a DNA construct consisting of part of a non-essential gene that contained an IES was injected into a vegetative macronucleus, where that IES is normally absent[19]. The injected *P. tetraurelia* was allowed to proliferate and then induced to enter the sexual cycle. In the formation of the new macronucleus, the particular IES was not removed during conversion of the micronuclear version of the gene into the macronuclear version. Removal of other IESs was unaffected. This result indicates that the old macronucleus may influence IES excision in the new micronucleus. When a restriction site was added to the IES before injection of the partial gene construct, the IES that was present in the gene after development of the new macronucleus lacked the restriction site. So the injected construct had not itself become part of the endogenous gene, but seems to have instructed the developing macronucleus to retain the IES in the endogenous gene. A plausible interpretation is that DNA (or RNA) is transferred from the old macronucleus to the new, developing macronucleus to act as a template to guide IES excision, and then the template is eliminated. This experimental design could be applied to hypotrichs to test the generality of the hypothesis that IES removal (and possibly excision of genes from micronuclear chromosomes) is guided by the old macronucleus before it is destroyed.

IESs might be mere curiosities that are tolerated because micronuclear genes are not transcribed. It does not matter that IESs disable transcription because the organisms can remove and destroy IESs efficiently during macronuclear development. But perhaps the significance of IESs lies in the division of genes into multiple blocks, as discussed in the next section.

Gene segments created by IES insertions

Gene segments created by the insertion of IESs in a gene are called macronuclear-destined segments, or MDSs. The six IESs in the micronuclear *βTP* gene of *O. trifallax* divide the gene into seven MDSs (FIG. 3). These MDSs do not correspond to domains within a protein but are random divisions of the gene. Perhaps then, recombination among IESs might facilitate evolution by creating new combinations and arrangements of MDSs. However, IESs interrupt individual codons at any position with equal frequency. So recombination between two IESs, to link two MDSs in a new combination, would result in a shift in the reading frame two-thirds of the time and change the amino-acid sequence encoded by the downstream MDS. This does not necessarily obviate the idea of MDS shuffling as an important evolutionary mechanism, but perhaps makes it less plausible. In addition, the MDSs that flank an IES contain repeats of 2–7 bp at their junctions with the IES. For example, MDS1 in the *βTP* gene of *O. trifallax* has the sequence 5′-CAGTA-3′ at its 3′ end and MDS2 has the

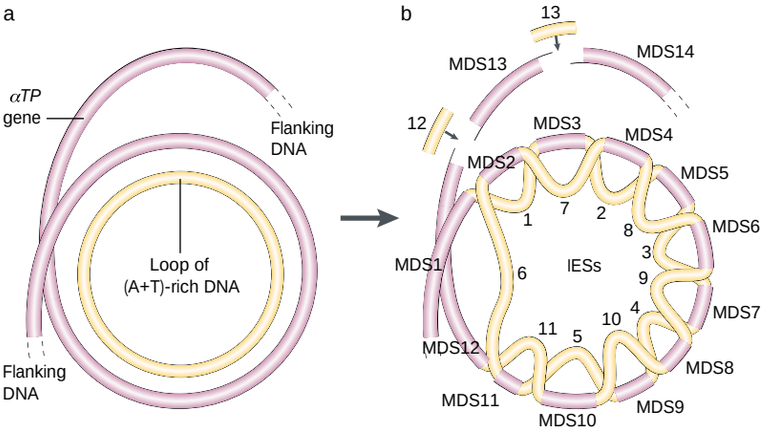

Figure 5 | **A model for nonrandom scrambling.** The model illustrates how simultaneous internal eliminated segment (IES) insertions lead to nonrandom scrambling of the micronuclear gene that encodes α-telomere-binding protein (αTP) in *Oxytricha nova*. **a** | IESs are derived from a loop of (A+T)-rich DNA (spacer DNA) and **b** | are inserted into the *αTP* gene at staggered cuts. Insertion of IESs 12 and 13 are separate events. **c** | Final order of the scrambled macronuclear-destined segments (MDSs). (Adapted from REF. 24.)

same sequence at its 5′ end. The repeat sequences that flank each IES are very variable (all pairs of repeats in a gene consist of different sequences). Therefore, recombination between two IESs would create an IES with non-matching flanking sequences at its ends. If the pairs of repeats are important in guiding IES excision or MDS ligation during macronuclear development, then the absence of a repeat pair might make IES excision and MDS ligation impossible. But this is not the case, as shown by the behaviour of repeat pairs whose positional relationships have been disrupted by recombination, which scrambles MDSs.

Scrambled MDSs

In some micronuclear genes, the MDSs are out of order. In the gene encoding actin I in *O. nova*, eight IESs divide the micronuclear gene into nine MDSs in a largely random, scrambled order: 3–4–6–5–7–9–2–1–8 (REF. 21) (FIG. 4). MDSs 3 and 4 are separated by the first IES; they are not scrambled relative to each other and share a pair of direct repeats. MDS3 has the sequence 5′-AATC-3′ at its 3′ end, where it joins IES1. The downstream end of IES1 joins with MDS4, which begins with 5′-AATC-3′. The other seven MDSs are scrambled, but pairs of repeats still define their orthodox order. For example, MDSs 4 and 5 are scrambled because they are separated by the segment IES2–MDS6–IES3, but they still share a pair of repeats: 5′-CTCCCAAGTCCAT-3′ at the 3′ end of MDS4 is repeated at the 5′ end of MDS5. Moreover, MDS2 is in the inverted orientation; its upstream end contains the sequence 5′-CTTGACGACTCC-3′, which is an inverted repeat of the sequence, 5′-GGAGTCGTCAAG-3′, at the upstream end of MDS3, to which it must ligate after reinversion to become unscrambled during macronuclear development. Similarly, the other end of MDS2 shares an inverted-repeat pair with the downstream end of MDS1. Remarkably, the ligation of MDS1 with reinverted MDS2 brings together an AT sequence at the 3′ end of MDS1 with a G at the 5′ end of MDS2 to form the ATG start codon for *actin I*.

The pairs of repeats for scrambled MDSs in the *actin I* gene range in length from 9–13 bp (average of 11 bp), in contrast to repeat lengths of 2–7 bp (average of 4 bp) for non-scrambled MDSs in several genes. This suggests that longer repeats are required for unscrambling of MDSs than for ligation of non-scrambled MDSs, but whether the longer repeats are adequate to guide unscrambling specifically is questionable. Further guidance may be required, as is the case for the short repeats of non-scrambled MDSs.

The presence of pairs of repeats suggests a recombination model of MDS unscrambling in which members of a pair align in register by folding of micronuclear DNA (FIG. 4b). Recombination between the aligned repeats would join all MDSs in the orthodox order and would remove one copy of each repeat with its adjacent IES. A problem with the folding/recombination model is the insufficiency of information in the short repeats to guide correct folding; multiple copies of a repeat sequence sometimes occur within an IES and in both upstream and downstream MDSs.

How did the *actin I* gene become scrambled? A simple hypothesis is to assume that in a common ancestor to *O. nova* and *O. trifallax*, eight IESs were inserted into the *actin I* gene, creating nine MDSs in the orthodox order. Subsequently, intraduplex recombination between IESs, including recombination that inverted MDS2, created the current scrambled pattern[22]. Such recombination might be tolerated because the mechanism already present for excising IESs and ligating MDSs could join MDSs in the orthodox order when guided by the repeat pairs and, possibly, by sequence information provided by the DNA molecules in the old macronucleus. It seems likely that IES6, which creates non-scrambled MDSs 9 and 10 in *O. trifallax* (FIG. 4), was added more recently, after divergence from *O. nova*. (Coincidentally, IES6 is the first to be removed during macronuclear development.) However, the IES recombination hypothesis does not account for lengthening of the repeat sequence (from an average of 4 bp to 11 bp) when non-scrambled MDSs are converted to scrambled MDSs.

Nonrandomly scrambled genes

The micronuclear gene that encodes α-telomere-binding protein (αTP) in *O. nova* contains 13 IESs that divide it into 14 MDSs, most of which are scrambled [23]. Unlike the pattern in the *actin I* gene, the *αTP* gene is scrambled in the nonrandom pattern, 1–3–5–7–9–11–2–4–6–8–10–12–13–14. No inversions are present, and MDSs 12, 13 and 14 are not scrambled relative to each other.

The nonrandom pattern of MDS scrambling suggests a recombination model for the origin of scrambling that is different from simple recombination between IESs[24]. IESs 1–11 are hypothesized to have originated simultaneously from a loop of (A+T)-rich DNA

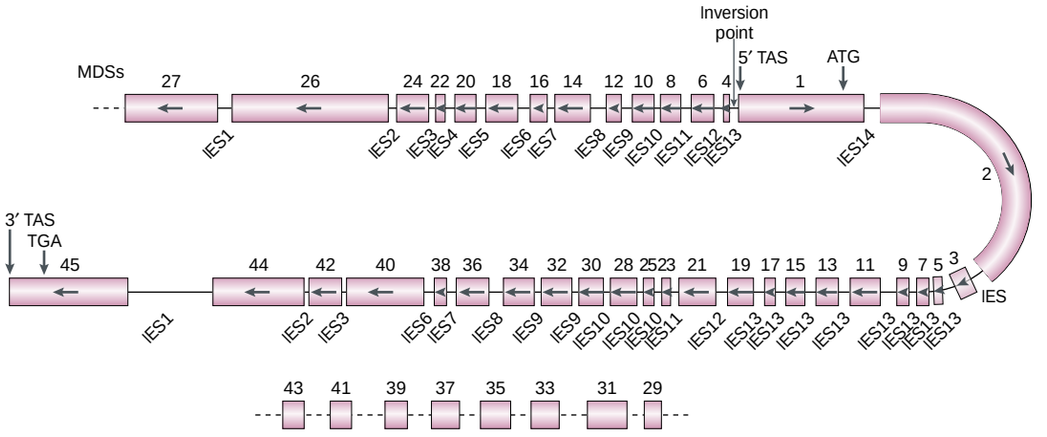

Figure 6 | **An extremely scrambled gene.** Arrangement of macronuclear-destined segments (MDSs) in the gene that encodes DNA polymerase α in *Oxytricha nova*. MDSs are boxes; internal eliminated segments (IESs) are lines between boxes. Arrows in boxes indicate the 5′→3′ polarity of the coding strand. The MDSs are present at two separate chromosomal locations in the micronucleus. The distances between the MDSs at the second (lower) location are unknown and so are represented as dotted lines. (TAS, telomere addition site.) (Adapted from REF. 18.)

(spacer DNA) aligned with the *αTP* gene, followed by a series of recombinations between the two (FIG. 5). The recombinations occur at staggered cuts in the gene DNA to create pairs of repeats. Non-scrambled MDSs 12, 13 and 14, formed by insertion of IESs 12 and 13, are separate events in time in this model.

The *pol α* gene in *O. nova* is extremely scrambled[25]. It contains 44 IESs and 45 MDSs. Most of the MDSs are arranged in a nonrandom odd/even pattern as in the *αTP* gene, but with an inversion of about one-third of the MDSs that places MDS1 near the middle of the gene (FIG. 6). The model for scrambling of the *αTP* gene can be applied to the *pol α* gene as well, with the inversion occurring after scrambling of the MDSs. Occurrence of the inversion before scrambling in the odd/even pattern of MDSs is difficult to imagine. The structure of the micronuclear *pol α* gene is further complicated by the separation of the scrambled MDSs into two loci. Eight MDSs occur at a separate locus whose positional relationship to the main locus is unknown. Unscrambling requires that the two loci be brought together and aligned to permit the appropriate series of recombinations between odd- and even-numbered MDSs.

The *pol α* gene is scrambled in a very similar pattern in two other hypotrichous ciliates — the same inversion is present, and the gene is split into two loci — but there are some significant differences. In *Stylonychia lemnae*, the gene includes three more IESs and three more MDSs than in *O. nova*[26]. In *O. trifallax*, the *pol α* gene has 50 IESs and 51 MDSs[17], six more than in *O. nova*. Two IESs are inserted into two MDSs that are non-scrambled in *O. nova* to create two further non-scrambled MDSs. The other four extra MDSs are integrated into the odd/even scrambled pattern. Finally, one very short MDS is not present in either of the two loci in *O. trifallax* or *S. lemnae* and remains unaccounted for.

It is apparent that the *pol α* gene became scrambled in an evolutionary progenitor of *O. nova*, *S. lemnae* and *O. trifallax*. As in the *actin I* and *αTP* genes, IESs and MDSs continued to be added to the *pol α* gene during evolution, but in this case some of the new MDSs are in scrambled configurations that indicate scrambling can occur by recombination between IESs.

The unscrambling of genes, particularly the *pol α* gene, implies an extraordinary molecular mechanism of DNA cutting and splicing, about which little is known. Cell clones in which these germline DNA modifications are present can undergo intraclonal mating, develop new macronuclei, and yield new vegetative clones with 95 to 100% efficiency, proving that the modifications are reversed with remarkable precision during macronuclear development.

Shifting of MDS–IES junctions

Comparison of the *actin I*, *αTP* and *pol α* genes among species reveals an important property of MDSs and IESs: they change in length, location and composition during speciation. For example, the lengths of MDS1 in the *αTP* gene of *O. trifallax*, *S. mytilus* and *O. nova* are 208, 230 and 239 bp, respectively, and the lengths of IES1 in the three species are 43, 111 and 48 bp, respectively, reflecting insertions or deletions within the IES. IES1 is also totally different in sequence in the three organisms, indicating a high rate of accumulation of TRANSITIONS and TRANSVERSIONS in this non-coding DNA. MDSs generally contain coding sequences and are, by comparison, strongly conserved.

The lengths, locations and compositions of MDSs and IESs can also change by shifting of MDS–IES junctions[27]. This is illustrated in FIG. 7 for MDSs 11 and 12 in the *αTP* gene of *O. trifallax* and the corresponding MDSs in *O. nova*. The MDS–IES junctions are different and the repeat sequences are different in the two organisms. Nevertheless, when the MDSs are ligated during macronuclear development, they yield the same macronuclear sequence in the two organisms. This is most readily explained by a series of mutations that shorten or lengthen MDSs at their ends and simultane-

TRANSITION
A mutation in which a purine is replaced by a purine, or a pyrimidine is replaced by a pyrimidine.

TRANSVERSION
A mutation in which a pyrimidine is replaced by a purine or vice versa.

Figure 7 | **A model for MDS–IES junction shifting.** Sequence of the 3′ end of scrambled macronuclear-destined segment (MDS) 11, the beginning of internal eliminated segment (IES) 12, the end of IES5 and the 5′ end of MDS12 in the scrambled genes that encode α-telomere-binding protein (αTP) in *Oxytricha trifallax* (top line) and *Oxytricha nova* (bottom line). In the scrambled genes, six IESs and five MDSs separate MDS11 from MDS12 (see FIG. 5). Repeat sequence pairs are underlined. A model is shown for shifting of MDS–IES junctions in the *αTP* gene of *O. trifallax* by a series of mutations creating hypothetical intermediates, ending with the gene as it is in *O. nova*. Arrows indicate mutations.

ously lengthen or shorten the adjacent IES during evolution. This is illustrated by the series of mutations in FIG. 7 that could have converted the *O. trifallax* sequence structure into the *O. nova* sequence structure. As a result of junction shifting, MDS11 in *O. trifallax* is longer by 11 bp than in *O. nova*, and MDS12 in *O. trifallax* is shorter by 11 bp than in *O. nova*. IES12 is shorter by 11 bp and IES5 is longer by 11 bp in *O. trifallax* relative to *O. nova*. Some single mutations result in two or three base shifts at junctions, for example, lines 7 and 8 in FIG. 7. The sequence of the repeat has changed completely from *O. trifallax* to *O. nova*, and is 1 bp shorter in *O. nova*. Appropriate mutations must occur in loose coordination at the two junctions to preserve a repeat pair, although the sequence of the repeat changes.

The junction shifts between species for scrambled MDSs are modest; the longest observed so far is 19 bp in the *pol α* gene between *O. nova* and *O. trifallax*. For non-scrambled MDSs, shifts can extend over at least a few hundred bp. The reason for such a difference in junction mobility is obscure. The consequence of junction shifts is that MDS lengths can change substantially, with sequence being switched from one MDS to another, but with no change in sequence. So the macronuclear sequence in FIG. 7 is unaltered by junction shifting.

Conclusion

The insertion of IESs into micronuclear genes, the concomitant creation of MDSs and the scrambling of MDSs in some genes in ciliates represent new evolutionary phenomena in molecular genetics. Superficially, these ciliate phenomena bear some similarity to DNA manipulations in other organisms. For example, the unscrambling of MDSs seems similar to V(D)J recombination in vertebrates, which generates immunoglobulin genes, but the two processes are fundamentally different in mechanism, design and consequence. Also, in some respects IESs resemble introns. Both interrupt coding regions of genes, but their mechanisms of excision are completely different; IESs are excised from DNA, introns from RNA. IESs also accumulate mutations at a significantly higher rate than introns. IESs shift along DNA molecules, but there is no convincing evidence that introns can shift along DNA. Introns create the opportunity for alternative splicing of exons to generate different gene products, but IESs lack that function because they are absent from transcribed (macronuclear) genes. Nevertheless, borrowing the concept of exon shuffling suggests the possibility that MDSs may recombine in the germ line to evolve new genes. If such MDS shuffling has occurred during evolution, it might be detected in contemporary organisms by searching for MDSs that are common to two or more micronuclear DNA genes, although this test may be complicated by the ability of MDS boundaries to change by IES shifting. At present, insight into the significance of the MDS/IES phenomena is being sought by long-term mutagenesis, which might alter MDS/IES patterns in informative ways. We can be confident that the spectacular contortions shown by ciliate genomes have played an important part in their evolution — a phenomenon that has been summed up incisively as the 'evolution of evolvability'[28].

Links

DATABASE LINKS *Tetrahymena thermophila* | *Oxytricha trifallax* | *Oxytricha nova* | Euplotes | Oxytricha | Stylonychia | Gastrostyla | Pleurotricha | Paraurostyla | Uroleptus | *Stylonychia mytilus* | *Paramecium tetraurelia* | *Stylonychia lemnae*

FURTHER INFORMATION David Prescott's lab page

ENCYCLOPEDIA OF LIFE SCIENCES Ciliophora | Euplotes | Tetrahymena | Developmentally programmed DNA rearrangements

1. Wright, A.-D. G. & Lynn, D. H. Maximum ages of ciliate lineages estimated using a small subunit rRNA molecular clock: Crown eukaryotes date back to the Paleoproterozoic. *Arch. Protistenkd.* **148**, 329–341 (1997).
2. Blackburn, E. H. & Gall, J. G. A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *J. Mol. Biol.* **120**, 33–53 (1978).
3. Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43**, 405–413 (1985).
4. Price, C. M. Telomere-binding proteins of ciliated protozoa. *Nucleic Acids Mol. Biol.* **9**, 299–307 (1995).
5. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A. & Schultz, S. C. Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA. *Cell* **95**, 963–974 (1998).
6. Zahler, A. M. and Prescott, D. M. Telomere terminal transferase activity in the hypotrichous ciliate *Oxytricha nova* and a model for replication of the ends of linear DNA molecules. *Nucleic Acids Res.* **16**, 6953–6972 (1988).
7. Brehm, S. L. & Cech, T. R. Fate of an intervening sequence ribonucleic acid: Excision and cyclization of the *Tetrahymena* ribosomal RNA intervening sequence *in vivo*. *Biochemistry* **22**, 2390–2397 (1983).
8. Griffith, J. D. *et al*. Mammalian telomeres end in a large duplex loop. *Cell* **97**, 503–514 (1999).
9. Murti, K. G. & Prescott, D. M. Telomeres of polytene chromosomes in a ciliated protozoan terminate in duplex DNA loops. *Proc. Natl Acad. Sci. USA* **96**, 14436–14439 (1999).
10. Klobutcher, L. A. *et al.* Conserved DNA sequences adjacent to chromosome fragmentation sites in *Euplotes crassus*. *Nucleic Acids Res.* **26**, 4230–4240 (1998).
11. Klobutcher, L. A. Characterization of *in vivo* developmental chromosome fragmentation intermediates in *Euplotes*

crassus. *Mol. Cell* **4**, 695–704 (1999).
12. Sueoka, N. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. *J. Mol. Evol*. **40**, 318–325 (1995); erratum **42**, 323 (1996).
13. Frank, A. C. & Lobry, J. R. Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* **238**, 65–77 (1999).
14. Klobutcher, L. A., Jahn, C. L. & Prescott, D. M. Internal sequences are eliminated from genes during macronuclear development in the ciliated protozoan *Oxytricha nova*. *Cell* **36**, 1045–1055 (1984).
The original discovery of internal eliminated segments in a germline gene of a hypotrich.
15. DuBois, M. L. & Prescott, D. M. Volatility of IESs in germline genes of hypotrichous ciliates. *Mol. Cell. Biol*. **17**, 326–337 (1997).
The discovery of internal eliminated segment shifting in the *actin I* gene of hypotrichs.
16. Tausta, S. L. & Klobutcher, L. A. Internal eliminated sequences are removed prior to chromosome fragmentation during development in *Euplotes crassus*. *Nucleic Acids Res*. **18**, 845–853 (1990).
This work proved that excision of internal eliminated segments leaves intact the polytene chromosomes formed during macronuclear development, and is apparently unrelated mechanistically to the eventual breakup of the polytene chromosomes into gene-sized DNA molecules.
17. Hoffman, D. C. & Prescott, D. M. Evolution of internal eliminated segments and scrambling in the micronuclear gene encoding DNA polymerase α in two *Oxytricha* species. *Nucleic Acids Res*. **25**, 1883–1889 (1997).
The complex scrambled patterns in the gene that encodes DNA polymerase α in two hypotrich species.
18. Prescott, D. M. & DuBois, M. L. Internal eliminated segments (IESs) of *Oxytrichidae*. *J. Euk. Microbiol*. **43**, 432–441 (1996).
19. Duharcourt, S., Keller, A.-M. & Meyer, E. Homology-dependent maternal inhibition of developmental excision of internal eliminated sequences in *Paramecium tetraurelia*. *Mol. Cell. Biol*. **18**, 7075–7085 (1998).
These experiments in *Paramecium* implicate the old macronucleus in directing the excision of internal eliminated segments in the developing macronucleus.
20. Forney, J. D., Yantiri, F. & Mikami, K. Developmentally controlled rearrangement of surface protein genes in *Paramecium tetraurelia*. *J. Euk. Microbiol*. **43**, 462–467 (1996).
21. Prescott, D. M. & Greslin, A. F. Scrambled *actin I* gene in the micronucleus of *Oxytricha nova*. *Dev. Genet*. **13**, 66–74 (1992).
22. Prescott, D. M. The evolutionary scrambling and developmental unscrambling of germline genes in hypotrichous ciliates. *Nucleic Acids Res*. **27**, 1243–1250 (1999).
23. Mitcham, J. L., Lynn, A. J. & Prescott, D. M. Analysis of a scrambled gene: The gene encoding α-telomere-binding protein in *Oxytricha nova*. *Genes Dev*. **6**, 788–800 (1992).
24. Prescott, J. D., DuBois, M. L. & Prescott, D. M. Evolution of the scrambled germline gene encoding α-telomere binding protein in three hypotrichous ciliates. *Chromosoma* **107**, 293–303 (1998).
A model of the origin of scrambling of the gene that encodes α-telomere-binding protein is presented, based on the structure of the germline gene in three hypotrichs.
25. Hoffman, D. C. & Prescott, D. M. The germline gene encoding DNA polymerase α in the hypotrichous ciliate *Oxytricha nova* is extremely scrambled. *Nucleic Acids Res*. **24**, 3337–3340 (1996).
26. Landweber, L. F., Kuo, T.-C. & Curtis, E. A. Evolution and assembly of an extremely scrambled gene. *Proc. Natl Acad. Sci. USA* **97**, 3298–3303 (2000).
27. DuBois, M. & Prescott, D. M. Scrambling of the *actin I* gene in two *Oxytricha* species. *Proc*. *Natl Acad. Sci. USA* **92**, 3888–3892 (1995).
28. Radman, M., Matic, I. & Taddei, F. Evolution of evolvability. *Ann. NY Acad. Sci*. **870**, 146–155 (1999).
29. Prescott, D. M. Invention and mystery in hypotrich DNA. *J. Euk. Microbiol*. **45**, 575–581 (1998).

Acknowledgements
This work is supported by the NIGMS and the NSF.