Spliceosomal introns in the deep-branching eukaryote

*Trichomonas vaginalis*

Štěpánka Vaháčová*, Weihong Yan*, Jane M. Carlton*, and Patricia J. Johnson*†

*Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, CA 90095; and †The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850

Edited by W. Ford Doolittle, Dalhousie University, Halifax, NS, Canada, and approved February 2, 2005 (received for review October 11, 2004)

Eukaryotes have evolved elaborate splicing mechanisms to remove introns that would otherwise destroy the protein-coding capacity of genes. Nuclear premRNA splicing requires sequence motifs in the intron and is mediated by a ribonucleoprotein complex, the spliceosome. Here we demonstrate the presence of a splicing apparatus in the protist *Trichomonas vaginalis* and show that RNA motifs found in yeast and metazoan introns are required for splicing. We also describe the first introns in this deep-branching lineage. The positions of these introns are often conserved in orthologous genes, indicating they were present in a common ancestor of trichomonads, yeast, and metazoans. All examined *T. vaginalis* introns have a highly conserved 12-nt 3′-splice-site motif that encompasses the branch point and is necessary for splicing. This motif is also found in the only described intron in a gene from another deep-branching eukaryote, *Giardia intestinalis*. These studies demonstrate the conservation of intron splicing signals across large evolutionary distances, reveal unexpected motif conservation in deep-branching lineages that suggest a simplified mechanism of splicing in primitive unicellular eukaryotes, and support the presence of introns in the earliest eukaryote.

S

splicing of nuclear premRNA plays a central role in eukaryotic gene expression and contributes to gene regulation and protein diversity (1, 2). Splice-site (SS) motifs, recognized by small nuclear RNAs and protein components of the spliceosome, coordinate multiple steps in splicing (1). Precision is required to generate mRNAs that encode functional proteins, yet the SS motifs that define exon/intron junctions are short and often weakly conserved. Two types of spliceosomal introns, U2- and U12-dependent, have been described in eukaryotes (reviewed in ref. 3). Groups I and II introns, capable of self splicing in vitro and present in some eubacterial genomes, are also found in organellar genomes of unicellular eukaryotes and plants (reviewed in refs. 4 and 5). Most spliceosomal introns are U2-dependent and are bordered at the 5′ end by a consensus of 6 nt, encompassing the invariant 5′ GT. The 5′ splice-site consensus in yeast (*Saccharomyces cerevisiae*) introns is generally highly conserved, whereas the consensus in mammalian introns is often limited to the 5′ GT. The 3′ SS motif of mammalian and yeast U2-dependent introns is shorter and typically limited to the last 3 nt (T/CAG3') (1, 3, 6, 7). In addition to the 5′ and 3′ SS motifs, yeast introns also require a conserved branch-point sequence (TACTAAC), which is only loosely conserved in mammalian U2-dependent introns.

Spliceosomal introns are common in animal and plant genes, absent in eubacterial and archaeal genes, and found in several unicellular protists. The prevalence of splicing in protists is unknown, and the properties of protist introns vary considerably. In euglenoids and trypanosomes, both transsplicing and cis-splicing U2-dependent introns are present (8). Although virtually all trypanosome premRNAs undergo transsplicing, only one cis-splicing intron has been identified (9). In euglenoids, U2-dependent cis-splicing introns are common; however, a subset of nuclear genes are also present that contain unconventional introns thought to be excised by a spliceosome-independent mechanism (10, 11). In contrast, apicomplexan parasites *Plasmodium* and *Toxoplasma* and entamoebids (12) appear to have only conventional cis-splicing introns and no transsplicing.

The two deepest-branching protist lineages are the diplomonad and parabasalid lineages (13–15). Only three diplomonad genes are known to contain introns (16, 17), and no intron-containing genes have been reported in the parabasalid lineage. However, the presence of a highly conserved essential spliceosomal protein, PRP8, in the parabasalid, *Trichomonas vaginalis* (18), provides evidence that splicing is likely to occur. The spliceosomal machinery and the RNA motifs required for splicing have not been examined for either of these lineages. Such analyses are critical, because the only known intron from the diplomonad, *Giardia*, lacks the canonical 5′ GT essential for splicing in yeast and metazoans (16), suggesting that splicing mechanisms in deep-branching protists may differ (19).

Here we examine the motifs required for the parabasalid *T. vaginalis* to splice heterologous introns and characterize the first introns found in this lineage. These introns share an extremely conserved extended 3′ SS motif that encompasses the branch point and is necessary for splicing with the only known intron in *Giardia*. Their positions within respective genes are also often conserved in orthologous genes in distal lineages, indicating the presence of introns in this deep-branching lineage before its divergence from its common ancestor with plants and animals.

Materials and Methods

**Cultures.** *T. vaginalis* strain T1 was maintained in Diamond’s complete media supplemented with iron (20).

**Generation of Chloramphenicol Acetyltransferase (CAT) Constructs Containing Introns.** The CAT gene was inserted into the Asp-718 and *Bam*HI restriction sites of a pBluescript vector that had previously been subjected to site-specific mutagenesis to disrupt the two *Pvu*II restriction sites. Forward and reverse oligomers corresponding to the sequence of each specific intron tested were annealed and blunt-end ligated into the *Pvu*II site of the CAT gene within this construct. The CAT genes containing these introns were subsequently inserted at the Asp-718 and *Bam*HI restriction sites to the *T. vaginalis* expression vector Master-Neo (21) that contains a neomycin phosphotransferase selectable marker. The sequence of each construct was verified by sequencing.

**Transfections.** Transfections were conducted as described (20), and cultures were selected with the drug Gentamicin (GIBCO) (G418) for 5 days before harvesting to assess CAT activity.

This paper was submitted directly (Track II) to the PNAS office. Freely available online through the PNAS open access option.

Abbreviations: SS, splice site; CAT, chloramphenicol acetyltransferase; STK, serine-threonine protein kinase.

†Present address: Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

© 2005 by The National Academy of Sciences of the USA
CAT Assay. CAT assays were performed on cell lysates by using radioactively labeled chloramphenicol C^{14}, as described by Promega.

DNA and RNA Isolation. Genomic DNA from *T. vaginalis* was isolated by using the guanidium hydrochloride method as described (22). Total RNA was isolated by using a lithium chloride procedure and enriched in mRNA by using a Promega mRNA isolation kit.

RT-PCR Analyses. The first-strand synthesis for RT-PCR was generated with SuperScript RTIII (Invitrogen) on RNaseH-treated mRNA with a CAT-specific reverse primer, AAGCGCCTGATTAAAACCTTGCT. Four microliters of the resulting cDNA was then used for PCR with a combination of CAT-specific primers (CATFor2, TCAGTGTGCTCAATTGACTCTATAACCAG; CATR2, TCATCAGGCCGGCAGAAATGT). The final PCR products were resolved on a 2% agarose gel.

Confirmation of *Trichomonas* Endogenous Introns. RNA was isolated by TRizol preparation (Invitrogen). The first-strand cDNA was generated on total RNA with the oligo dT primer by using Superscript RT III reverse transcriptase. A list of oligonucleotides can be found in Supporting Text, which is published as supporting information on the PNAS web site.

RNA Blot Analysis. Total RNA or mRNA was separated on a 1.2% agarose gel and blotted to nitrocellulose by using standard procedures. Hybridization probes were labeled by nick translation by using DNA polymerase (Boehringer), and hybridizations were conducted at 65°C in 3× SSC, followed by washing at the same temperature, reducing the salt concentration to 0.1× SSC.

DNA Sequencing. All RT-PCR products were subcloned into the TOPO vector (Invitrogen). DNA sequencing was performed with M13 forward and reverse primers by using the Big Dye version 3 labeling kit (Applied Biosystems). The sequence was obtained by using an Applied Biosystems 3700 Capillary DNA Analyzer.

Algorithm for the Intron Search. A PERL script was written to search the preliminary *T. vaginalis* sequence data, obtained from The Institute for Genomic Research through the web site (www.tigr.org/db/t2k1/tvg), for the presence of the following motif: GYNNNGYN_{n}ACTAACAACACAG, where Y, T or C and n, any nucleotide. As indicated by N_{n0}, no size limit was set for the distance between the putative 5' SS motif (GYNNGY) and the putative 3' SS motif (ACTAACAACACAG).

Results

**Trichomonads Are Capable of Splicing Short Introns and Require Conserved Splicing Motifs.** To address whether the primitive eukaryote *T. vaginalis* contains the splicesomal machinery required for intron removal, we developed a functional assay to test its ability to splice heterologous introns. Introns were inserted into the PvuII site in the CAT reporter gene contained in a *T. vaginalis* expression vector (21), creating either an in-frame stop codon or shifting the ORF. This would require accurate splicing of the intron from the premRNA for the production of CAT activity in cells transfected with these constructs.

Initially, we tested whether *T. vaginalis* could remove introns from two related anaerobic parasitic protists, a 35-nt intron from the *Giardia intestinalis* ferredoxin gene (16) and a 46-nt intron from an *Entamoeba histolytica* serine/threonine kinase gene (23), as well as a 302-nt intron from the *Trypanosoma cruzi* poly(A) polymerase gene (9) and a 46-nt modified version of an adenovirus intron (24). Although the three introns from E. histolytica, *T. cruzi*, and adenovirus-2 contained canonical 5' GT-AG3' SS motif and putative branch site sequences, none of the introns were removed from the CAT gene (data not shown). In contrast, analysis of *T. vaginalis* transfectants harboring a CAT gene containing the *G. intestinalis* intron (Int2) and mutations of this 35-nt sequence (Int1 and Int3–12). Mutations are indicated in bold and underlined. The positive control (at the top) contains a construct with an intron-less CAT gene and the negative control (at the bottom) has a CAT gene with the Int1 intron inserted in the opposite orientation. CAT activity is expressed as cpm of n-butyrylated chloramphenicol per microgram of protein. The number of independent transfections tested are designated in brackets, and standard error bars are shown. (B) RT-PCR products separated on a 2% agarose gel show two populations of CAT DNAs corresponding to spliced and unspliced mRNA in *T. vaginalis* cells transfected with the Int1 construct (GT) and the Int4 (GC) constructs. No spliced CAT cDNA is detected in cells transfected with the Int2 construct containing wild-type *Giardia* intron (CT) or the Int1 construct with the intron inserted in the reverse orientation (neg). Pos, RT-PCR reaction by using cells transfected with the intronless CAT construct. (C) RNA blot analysis of total RNA from transfected cells. The blot was hybridized with a CAT gene probe showing that cells without CAT activity do transcribe the CAT RNA. A ferredoxin probe was used as an internal standard for RNA loading.

**Fig. 1.** *T. vaginalis* can splice a 35-nt intron. (A) CAT activity in *T. vaginalis* cells transfected with plasmid constructs of a CAT gene containing the wild-type *Giardia* intron (Int2) and mutations of this 35-nt sequence (Int1 and Int3–12). These data demonstrate that cells without CAT activity do transcribe the CAT RNA. A ferredoxin probe was used as an internal standard for RNA loading.
abolished CAT activity (Fig. 1A, Ints 2, 3, 5, 6, and 10–12), with the exception of an intron containing 5′ GC-AG3′ sites (Int4). These findings are consistent with reports demonstrating that a minority of introns contain a 5′ GC SS (25). We also observed that changing the third and fourth or the fifth and sixth dinucleotides at the 5′ SS to TG abolished CAT activity (Fig. 1A; Int7 and Int8), indicating a critical role for the first 6 nt at the 5′ SS in premRNA splicing in *T. vaginalis*. To test whether a sequence identical to the consensus branch site (ACTAAC) of yeast introns (26) is required for splicing in *T. vaginalis*, we mutated this sequence in the 35-nt intron (Fig. 1A, Int9) and found that CAT activity was abolished. Combined, these results show that the splicing machinery necessary to remove U2 spliceosomal introns from premRNAs is present in trichomonads, and that the same motifs used by yeast and metazoa are required for splicing in this deep-branching eukaryote.

To confirm that CAT activity reflects splicing and to demonstrate precise intron removal, we used RT-PCR to clone and sequence CAT mRNA. As shown in Fig. 1B, two products of the predicted size for unspliced and spliced mRNAs are found in cells transfected with the CAT gene introns bordered by either a GT (Int1) or a GC (Int4) 5′ SS. In contrast, transfectants that did not express detectable CAT activity (e.g., CT/Int2 and negative control cells/Int13) produce only a single RT-PCR product, corresponding in length to unspliced RNA (Fig. 1B). These data confirm that transfectants without activity produce CAT premRNA but are incapable of splicing it. Sequencing of the RT-PCR products showed accurate splicing by *T. vaginalis* of the canonical introns in Int1 and Int4 transfectants (Fig. 1B) and demonstrated that longer RT-PCR products contain the intron. The presence of unspliced RNA in Int1 and Int4 transfectants indicates inefficiency of splicing that might be attributed to the short length of the intron or the artificial exons that flank it. To demonstrate that the lack of gene transcription does not account for the absence of CAT activity, Northern blot analysis of total RNA was also performed, and all examined *T. vaginalis* transfectants harboring intron-containing CAT genes were shown to contain CAT premRNA (Fig. 1C).

A *T. vaginalis* Gene Contains an Intron with Extended Conserved Motifs Found in the Only Known *Giardia* Intron. Having demonstrated the presence of splicing machinery in *T. vaginalis*, we searched for intron-containing genes on release of sequence data from the *T. vaginalis* Genome Sequencing Project (www.tigr.org/tdb/e2k1/tvg) by examining homologues of intron-containing genes in other protists. An ORF encoding a putative poly(A) polymerase (PAP1) (9) that is interrupted by a 94-nt sequence flanked by canonical 5′ GT-AG3′ SS was identified. Inspection of this putative intron exposed an unexpected identity of the last 12 nt, ACTAACCCACACAG, with those of the 35-nt *Giardia* intron (Fig. 2A). Moreover, 5 of the 6 nt at the putative 5′ SS of the two sequences are also identical (Fig. 2A and C) and similar to that found in yeast introns. RT-PCR with PAP1-specific primers was used to show that the 94-nt sequence is an intron, because it is removed from the premRNA precisely at the predicted positions (Fig. 2B). The conservation at the 5′ SS and the atypical extended strict conservation at the 3′ SSs strongly suggest a functional role for these motifs. To test this, these sequences were mutated in the 35-nt CAT intron (Int1) and assayed for CAT activity. Single- or double-nucleotide mutations of 6 nt at the 5′ SS revealed that several positions could vary and still support splicing (T to C at position 3 and GT to AC or GC at positions 5 and 6; see Fig. 2C, Int13, Int15, and Int16). However, splicing activity was lost when the C at the fourth position was preceded by a G (GTGCCG) (Fig. 2C; Int14). Similarly, CAT activity was observed only when the C at position 6 was preceded by an A or a G (Fig. 2C; Int15 and Int16), because placing a T or C at position 5 abolished splicing (Fig. 2C; Int17 and Int18). Strict sequence requirements for the 12 nt at the 3′ end were found to be necessary for splicing (Fig. 2C, Int19 to Int25). Two single-nucleotide mutations in the putative branch site, the conserved A and T (Int 19, ACCAAC and Int 20, ACTACC) also abolished splicing, consistent with the role for...
SS, GTATGT, and the remaining five differ by only 1 nt motif, 34 of the 39 hits contain the same 6 nt at the putative 5' H11032 any nucleotide to be in positions 3 and 4 at the 5' in length from 59 to 196 nt. Although our search motif allowed and ACTAACACACAG-3 H11032 that T. vaginalis (3STK) gene (stk196), and the transcription factor TAF6 (taf6– H20862 e2k1 for introns is shown (at the top), and the 5' and 3' motifs and length of the 39 sequences identified are listed (at the bottom). (B) RT-PCR products generated by using T. vaginalis genomic DNA (gDNA) or mRNA (cDNA) as the template and primers flanking the putative introns in genes encoding a TATA-binding protein-associated factor (Taf6–81), a small C-terminal domain phosphatase (Scp67), and a STK (STK 196), this sequence as a branch site motif. Similarly, we observed that splicing is inactivated when the branch motif ACTAAC is moved 2 or 7 nt upstream (Int24 and Int25), indicating a spatial requirement for the motif immediately upstream of the 3' SS. Finally, three additional mutants provide evidence that the entire 12-nt motif is required for splicing, because alterations in all positions 3–7 nt upstream of the 3' AG disrupted splicing (Fig. 2C; Int 21, Int 22, and Int 23). These results led to the prediction that T. vaginalis introns are typically bordered by 5'–GYAYGY and ACTAACACACAG–3' conserved motifs.

Large-Scale Search for Intron-Containing Genes in the T. vaginalis Genome. Strict conservation of 5' and 3' SS and placement of the putative branch site adjacent to the 3' SS motif in introns from Giardia and T. vaginalis, as well as the observed requirement for these sequences for splicing, led us to search for other T. vaginalis intron-containing genes by using in silico methods. A database containing 2.7-fold coverage of the T. vaginalis genome (www. tigr.org/tdb/e2k1/tvg) for introns is shown (at the top), and the 5' and 3' motifs and length of the 39 sequences identified are listed (at the bottom). (B) RT-PCR products generated by using T. vaginalis genomic DNA (gDNA) or mRNA (cDNA) as the template and primers flanking the putative introns in genes encoding a TATA-binding protein-associated factor (Taf6–81), a small C-terminal domain phosphatase (Scp67), and a STK (STK 196), positions 3–7 nt upstream of the 3' AG disrupted splicing (Fig. 2C; Int 21, Int 22, and Int 23). These results led to the prediction that T. vaginalis introns are typically bordered by 5'–GYAYGY and ACTAACACACAG–3' conserved motifs.

Intron Positions in Orthologous Genes from Organisms Representing Other Eukaryotic Lineages. Introns found in the same position in homologous genes of phylogenetically diverse species are interpreted to be “old” introns, whereas those in a unique position are referred to as “new” introns (30). To determine whether T. vaginalis introns are old or new, we compared their position to that of introns in orthologous genes from a protist and various yeast and metazoa. Sixty percent of the T. vaginalis introns were found to interrupt the gene in the exact position of an intron in an orthologous gene from at least one other eukaryotic lineage (Tables 1 and 5). Moreover, an intron was found in another eukaryotic orthologue within four amino acids of the position observed for all T. vaginalis introns. Given the extreme divergence of T. vaginalis proteins, this likely reflects amino acid divergence around an old intron as opposed to the addition of a new intron. These data indicate that some T. vaginalis genes contain old introns, present before the divergence of parabasalids from other eukaryotic lineages. Moreover, they reinforce the conclusion that introns were present when unicellular eukaryotes first evolved, and that their paucity in most extant protists and yeasts is the result of subsequent loss (31, 32).

Discussion

We have shown that T. vaginalis, a protist of the parabasalid deep-branching eukaryotic lineage, possesses splicing machinery that functionally depends on standard conserved motifs and has identified the first intron-containing genes in this lineage. Our studies illustrate the conservation of functional splicing signals across large evolutionary expanses, show that T. vaginalis contain
old introns, and imply that spliceosomal introns were present in genes of the common ancestor of all extant eukaryotes.

In addition to the conservation of typical U2 introns, extended 5’ and 3’ splicing motifs were found in all examined T. vaginalis introns. These extended motifs, which are also found in the only reported intron in Giardia (16, 33), a protist belonging to the sister diplomonad lineage, appear to be required for splicing, because their mutation abolishes splicing. Strikingly, 100% nucleotide identity was observed in five of six positions at the 5’ SS with 92% identity at the remaining position in 41 putative T. vaginalis introns and the Giardia intron. Likewise, 11 or 12 nt at the 3’ SS of these introns are identical, with the remaining position exhibiting 95% identity. Although the first 6 nt at the 5’ SS are often conserved in U2 introns, particularly in yeast introns (3), the extraordinarily high level of conservation of 12 nt at the 3’ SS in T. vaginalis and Giardia introns is unprecedented.

Moreover, the remarkable conservation of this trait between two lineages and their requirement for splicing imply a difference in SS recognition and/or intron excision/ligation by spliceosomes in these unicellular organisms. On the other hand, endogenous genes with introns of variable lengths and exon/intron compositions that allow for a degree of nucleotide flexibility incompatible with our assay may be present that do not require strict conservation of SS sequences.

Remarkably, the putative branch site motif (ACTAAC) required for splicing in T. vaginalis is contained within the highly conserved 12-nt 3’ motif, from −7 to −12 nt. Attempts to exchange this sequence with upstream sequences eliminated splicing, indicating that the distance between the 3’ SS and this element is critical. This spacing requirement is also consistent with the ability of T. vaginalis to splice the 5’ GT-Giardia intron (Int1) and not three others, all of which contain standard 5’ and 3’ SSs but do not have their branch site motifs adjacent to the 3’ SS. This immediate proximity of the branch site and the 3’ end in T. vaginalis introns, an extremely rare trait of yeast and metazoan introns, promotes speculation that their spliceosomes may combine the steps of branch site and 3’ SS recognition. Interestingly, an essential spliceosomal protein, Prp8, that interacts with both 5’ and 3’ canonical motifs and the branch site (34), is the only reported spliceosomal protein in T. vaginalis (18). Thus, the close juxtaposition of the branch site and the 3’ SS may reflect a simplified mechanism for spliceosomal assembly in T. vaginalis, where Prp8 plays an indispensable role. In silico analyses to determine whether genes encoding other critical spliceosomal proteins are present in trichomonads await a complete coverage of the genome sequence. The structural similarity between T. vaginalis and Giardia spliceosomal introns and group II self-splicing introns, a class of introns thought to be evolutionarily related to spliceosomal small nuclear RNA (snRNA) machinery (5, 35), is noteworthy. The distance (7 nt) between the branch site and the 3’ SS of group II introns is the same as that observed for the putative branch site contained with the conserved 12 nt at the 3’ SS of Trichomonas and Giardia introns. The use of analogous splicing pathways by group II introns and nuclear premRNA has been taken to support a common evolutionary origin (35). The possibility that T. vaginalis and Giardia introns represent evolutionary intermediates possessing features of both systems is an exciting one, with important mechanistic implications. Despite strong conservation of canonical 5’ GT-AG3’ and branch site motifs in T. vaginalis introns implying that trichomonad snRNAs would likewise be conserved, these RNAs have yet to be identified.

We have shown that Trichomonas is capable of splicing a 35-nt intron, whereas most eukaryotic introns have a minimal length of ~50 nt (36). This minimal length has been interpreted to reflect physical constraints imposed by the spliceosome. Whether the ability to splice such short introns by T. vaginalis and Giardia reflects significant differences in their splicing components and/or mechanisms awaits further analyses.

The complete sequence and annotation of genomes of numerous eukaryotes have allowed large-scale comparisons of intron position(s) in phylogenetically diverse organisms (31). With rare exceptions, intron-poor unicellular eukaryotes have been found to have a significant 5’ bias in intron position, a property not observed in metazoan genes (31, 37). Our analyses of T. vaginalis introns are consistent with these findings, because introns are typically found within the first one-third of the gene. These data support the argument that the paucity of introns in unicellular eukaryotes results from a secondary loss, driven by homologous recombination between the gene and an intronless cDNA derived by reverse transcription of the corresponding mRNA (31, 37, 38), as opposed to introns simply being less abundant during early eukaryotic evolution.

Two theories have dominated the “origin of introns” debate. The first posited that introns were present in the common ancestor of all cells and were subsequently lost in extant eubacteria, archaea, and certain unicellular eukaryotes (the “intron-early” hypothesis or the exon theory of genes) (28, 39–42), and the second postulated that introns were added to genes in eukaryotes and never existed in orthologous genes in intron-deficient organisms (43, 44) (the “intron-late” hypothesis). At the crux of this debate is the issue of whether introns were present in the last common ancestor of archaea, eubacteria, and eukaryotes. The data presented here support the intron-early hypothesis.

A mixed model of intron evolution incorporating features of both the intron-early and intron-late hypotheses has recently been embraced (27–30, 37, 39, 45–48). Comparative genomics have revealed that numerous introns present in a common ancestor of fungi, nematodes, and arthropods were lost in these lineages but retained in vertebrates and plants (29, 39). When present, such introns are located in the same position in divergent orthologues and are referred to as old introns. Unique positions for introns in the genes of specific groups of vertebrates and plants have also provided evidence that introns can be added to genes (49), so-called new introns (39, 48, 50). Consistent with the hypothesis that the 5’ positional bias of introns in T. vaginalis genes is indicative of secondary loss of introns, we have shown that many of the retained introns in T. vaginalis genes are old. The position of ~60% of

### Table 1. Conservation of intron position in T. vaginalis genes and in orthologous genes from other eukaryotic lineages

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pap1</td>
<td>94</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cent 116</td>
<td>122</td>
<td>127</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Scc 67</td>
<td>70</td>
<td>76</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Taf6 81</td>
<td>91</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Skt 93</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Skt 59</td>
<td>68</td>
<td>105</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Skt 70</td>
<td>78</td>
<td>-</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>Skt 110</td>
<td>114</td>
<td>134</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Skt 99</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Skt 196</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>4</td>
</tr>
</tbody>
</table>

Genes are abbreviated as follows: PAP, poly(A) polymerase; SCP, small C-terminal domain phosphatase; and Taf6, TBP-associated factor TAF6, where numbers correspond to intron length. Organisms are abbreviated as follows: H.s., Homo sapiens; D.m., Drosophila melanogaster; C.e., Caenorhabditis elegans; A.t., Arabidopsis thaliana; S.c., Saccharomyces cerevisiae; S.p., Schizosaccharomyces pombe; and P.f., Plasmodium falciparum. 0 indicates an intron at the same position, and 1, 2, 3, or 4 indicates the presence of an intron shifted in position by one to four amino acids relative to the T. vaginalis intron. − indicates the absence of an intron in a similar position in the orthologous gene. See Table 5 and Supporting Text for detailed analyses.
examined *T. vaginalis* introns are conserved in orthologous genes from at least one other eukaryotic lineage. This observation and the requirement of canonical motifs for splicing in this deep-branching eukaryote indicate that introns and the spliceosomal machinery necessary for their excision were present in the ancestral eukaryote. This places spliceosomes on the roster of essential defining eukaryotic features and establishes the presence of introns in all examined eukaryotes.

We thank Drs. Doug Black, Richard Stefl, and Nancy Sturm, as well as S. N. Sturm and our colleagues in the lab, for helpful discussions and critical comments on the manuscript, and we thank Reviewer no. 1 for insightful comments. This work was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health Grant R01AI050537, and a Co-Operative Agreement Award (U01AI050913) to The Institute for Genomic Research to sequence the genome of *T. vaginalis*. P.J.J. is a Burroughs Wellcome Scholar in Molecular Parasitology.