Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 11983–11987, December 1994
Genetics

Evolution of mobile group I introns: Recognition of intron sequences by an intron-encoded endonuclease

NICK LOIZOS*†, ELISABETH R. M. TILLIER*, AND MARLENE BELFORT*‡

*Molecular Genetics Program, Wadsworth Center, New York State Department of Health, Albany, NY 12201-0509; †Department of Microbiology, Immunology and Molecular Genetics, Albany Medical College, Albany, NY 12208; and ‡School of Public Health, State University of New York, Albany, NY 12201-0509

Communicated by Sidney Altman, August 18, 1994 (received for review July 7, 1994)

ABSTRACT **Mobile group I introns are hypothesized to have arisen after invasion by endonuclease-encoding open reading frames (ORFs), which mediate their mobility. Consistent with an endonuclease-ORF invasion event, we report similarity between exon junction sequences (the recognition site for the mobility endonuclease) and intron sequences flanking the endonuclease ORF in the *sunY* gene of phage T4. Furthermore, we have demonstrated the ability of the intron-encoded endonuclease to recognize and cleave these intron sequences when present in fused form in synthetic constructs. These observations and accompanying splicing data are consistent with models in which the invading endonuclease ORF is provided safe haven within a splicing element. In turn the intron is afforded immunity to the endonuclease product, which imparts mobility to the intron.**

Mobile group I introns share the remarkable property of "homing" from intron-containing genes to cognate intronless alleles in a process that is dependent upon an intron-encoded DNA endonuclease. The site-specific endonuclease creates a double-strand break at the exon junction of the intronless allele, thereby facilitating intron acquisition via a break-induced gene conversion event (1–6).

Mobile group I introns can be divided into two functionally and structurally distinct domains. The intron core sequences form the characteristic secondary and tertiary RNA structure necessary for catalyzing splicing (7, 8). Emanating from peripheral loops in this structure are the endonuclease-encoding open reading frames (ORFs) that impart mobility (see Fig. 1*A*). Circumstantial evidence from various systems supports the idea that the core and ORF sequences evolved independently. First, there are several instances of closely related group I introns containing disparate ORFs at different locations, as for example in phage T4 (9) and *Neurospora* (10). Second, the discovery of a homing-type endonuclease encoded by an archaeal intron, which is structurally distinct and splices by a different pathway than the group I introns, provides compelling evidence for the independent origin of the mobility and splicing functions (11). Finally, ORFs that encode homing-type endonucleases that also possess protein splicing activity have been found directly in protein coding sequences, independent of an RNA splicing apparatus (reviewed in refs. 6, 12, and 13).

Previous work has shown that intron mobility depends on the ORF but is independent of core sequences (14), implicating the endonuclease ORF as the primary agent of mobility. This result is consistent with the hypothesis that mobile group I introns arose by playing host to invasive endonuclease ORFs that conferred mobile properties on the introns (2–5). Colonization of a group I intron by an endonuclease ORF could benefit both entities. The endonuclease ORF would be provided safe haven by settling in an intron that is removed by RNA splicing during gene expression, rather than in an essential gene whose function would be disrupted. The endonuclease-encoding group I intron could in turn propagate in a genome that might pressure for its removal (5, 15).

The endonuclease ORF is posited to have colonized the intron by a cleavage event similar to that used to mobilize the intron (14). Consistent with the endonuclease-ORF invasion hypothesis, we report that for the mobile *sunY* intron of phage T4 (16, 17), intron sequences flanking the endonuclease ORF (IL and IR, for *i*ntron *l*eft and *i*ntron *r*ight) bear striking similarity to exon junction sequences (EL:ER, for *e*xon *l*eft:*e*xon right; the recognition site for the mobility endonuclease). Further, the *sunY* intron-encoded endonuclease, I-*Tev* II, is capable of binding and cleaving a synthetic fused IL:IR sequence, but cannot cleave IL or IR alone. These data strongly support the endonuclease-ORF invasion hypothesis and provide a logical basis for how RNA splicing elements have developed an endonuclease-promoted, DNA-based means of propagation.

MATERIALS AND METHODS

Plasmids, Bacteria, and Phage. For I-*Tev* II overproduction, the PCR-derived gene was cloned into plasmid pAII17 (18), to generate pT7-I-*Tev* II, and expressed in BL21(DE3)/pLysS (19). The ORF deletion corresponding to the IL:IR junction, ΔORF1, was constructed by site-directed mutagenesis (this work). Transcription templates contained the wild-type *sunY* intron or deletion variants (ΔORF1 and ΔORF-C1 to -C4; see Fig. 4*A*) in pBSM13 or pNC85 (20, 21) (this work).

I-*Tev* II Preparation. BL21(DE3)/pLysS/pT7-I-*Tev* II, grown to an OD_{650} of 0.2, was induced with 2 mM isopropyl β-D-thiogalactopyranoside and aerated for 30 min at 37°C. Cells were harvested, suspended in lysis buffer [50 mM Tris·HCl, pH 8.0/2 mM EDTA pH 8.0/0.1 mM phenylmethanesulfonyl fluoride/20 mM 2-mercaptoethanol with leupeptin (2 μg/ml) and aprotinin (5 μg/ml)], and lysed by freeze–thawing and sonication. After centrifugation at 10,000 × *g* for 15 min at 4°C, the soluble fraction was used in all assays. One unit of activity is defined as the amount of enzyme required to cleave 250 ng (112 fmol) of linearized target (3.38 kb) to 50% completion in 1 min at 23°C.

DNA Binding Assays. For the experiment in Fig. 2, one of the oligonucleotides (20 pmol) in each duplex was end-labeled with ^{32}P by T4 kinase (BRL). DNAs were annealed by heating complementary oligonucleotides at 90°C in 25 μl of annealing buffer (25 mM Tris·HCl, pH 8.0/50 mM NaCl) for 7 min and cooling to 50°C for 15 min and to 23°C for 15 min. For the experiment in Fig. 3, degenerate, randomized, and control oligonucleotides were annealed to a labeled primer complementary to a sequence outside the region of degeneracy/randomization and rendered double-stranded with Klenow

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: ORF, open reading frame.

polymerase (BRL) to generate a fragment of 65 bp, which was gel purified. For gel mobility-shift experiments, I-*Tev* II (2 units) was incubated for 2 min at 23°C with duplex DNAs (≈8 fmol, 10,000 cpm) in 20 μl of binding buffer [50 mM Tris·HCl, pH 8.0/100 mM NaCl/5% (vol/vol) glycerol/10 mM EDTA with poly(dI-dC) (100 μg/ml)]. Reaction mixtures were loaded onto an 8% polyacrylamide (29:1 acrylamide/*N,N*-methylene bisacrylamide weight ratio) gel in Tris borate buffer; electrophoresis was at 20 mA for 1 hr at 4°C.

DNA Cleavage Assays. Labeled duplexes (≈8 fmol, 10,000 cpm), generated as described above, were incubated in cleavage buffer [50 mM Tris·HCl, pH 8.0/100 mM NaCl/10 mM $MgCl_2$ with poly(dI-dC) (25 μg/ml)] with I-*Tev* II (amounts indicated) for two min at 23°C. An equal volume of stop solution (95% formamide/20 mM EDTA/0.05% bromophenol blue/0.05% xylene cyanol FF) was added and the mixtures were placed on dry ice. Products were separated in 10% polyacrylamide (19:1 acrylamide/bisacrylamide)/8 M urea gels.

RNA Splicing Assays. Precursor RNAs transcribed from linearized templates by T3 RNA polymerase (BRL) were phenol extracted, ethanol precipitated, and incubated for 75 min at 37°C in splicing buffer [40 mM Tris·HCl, pH 7.5/30 mM $MgCl_2$/0.4 mM spermidine/50 mM NH_4Cl/4 mM dithiothreitol with 16 units of RNasin (Promega)] with 3.3 pmol of [α-^{32}P]GTP (3000 Ci/mmol; 1 Ci = 37 GBq). Splice products were separated in a 5% polyacrylamide (39:1 acrylamide/bisacrylamide)/8 M urea gel.

RESULTS

Similarity Between Sequences Flanking the Intron and Those Flanking the Intron ORF. A high degree of sequence identity has been observed between *sunY* exon junction sequence EL:ER, the natural substrate of I-*Tev* II, and IL sequences immediately 5′ to the start of the intron ORF (14). Upon closer inspection, we found sequence identities between IR sequences 3′ of the ORF and exon 2 (Fig. 1*A*). By connecting the intron sequences flanking the ORF, we generated an ORF-less junction we term IL:IR, which is identical in 19 of 23 positions to EL:ER (Fig. 1*B*). The similarity spans precisely the exon recognition sequence of I-*Tev* II determined by DNase I footprinting (23) (Fig. 1*B*).

Functional Similarity Between Intron and Exon Sequences. To test whether I-*Tev* II could recognize the IL:IR sequence, duplex DNAs representing IL:IR, EL:ER, and 5′ and 3′ intron/ORF junctions (IL:OL and OR:IR, respectively) of the *sunY* gene were prepared (Figs. 1*A* and 2*A*). The duplexes were incubated with I-*Tev* II and formation of I-*Tev* II/DNA complexes was monitored by a gel mobility-shift assay (Fig. 2*B*). Although I-*Tev* II was unable to form a complex with OR:IR (Fig. 2*B*, lanes 17–20) the enzyme was able to induce a mobility shift with IL:IR (lanes 5–8) and IL:OL (lanes 13–16) similar to that with its normal substrate, EL:ER (lanes 1–4). These results suggest that the IL sequences 5′ to the ORF may be sufficient to direct complex formation with I-*Tev* II (see also Fig. 3*B*, IL:R6 data).

To determine whether these duplexes could act as authentic substrates for the enzyme, cleavage assays were performed with these DNAs, end-labeled on either the top or the bottom strand (Fig. 2*C*). Neither IL:OL nor OR:IR was able to act as a cleavage substrate for I-*Tev* II (Fig. 2*C*, lanes 10–15). In sharp contrast, the fused intron sequence IL:IR was cleaved by I-*Tev* II to a similar extent as the natural cleavage substrate EL:ER. Moreover, cleavages occurred at the predicted location on both strands of the duplex (Fig. 2*C*, compare lanes 5 and 6 with lanes 2 and 3; for cleavage locations, see Figs. 1*B* and 2*A*). These results suggest that the intron deleted for the endonuclease ORF at the IL:IR junction is a cleavage substrate for I-*Tev* II, whereas the ORF-containing intron is not. Similar experiments performed on the full-length intron containing the ORF support these findings (data not shown).

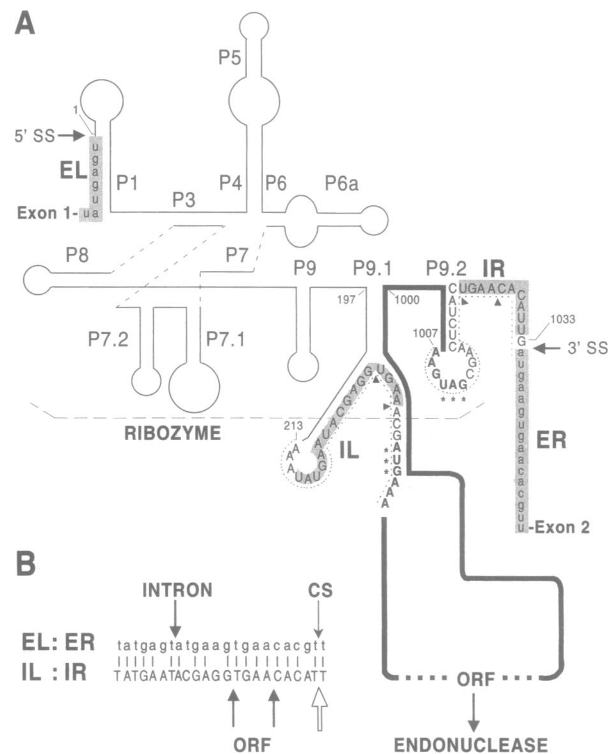

FIG. 1. Similarity between exon sequences flanking an intron and intron sequences flanking an endonuclease ORF. (*A*) Secondary structure of the *sunY* intron. Base-paired regions are designated P1 through P9.2 (22). Specific exon and intron nucleotides are in lowercase and uppercase, respectively. Shading indicates stretches of similarity between exon nucleotides (EL and ER) and intron sequences flanking the ORF (IL and IR). The bold line and nucleotides represent I-*Tev* II ORF sequences, with start and stop codons marked by stars. 5′ SS and 3′ SS indicate 5′ and 3′ splice sites, respectively. The intron is numbered from 1 at the 5′ end to 1033 at the 3′ end. Arrowheads 5′ and 3′ to the ORF correspond to 4-nt stretches over which the IL:IR junction can be generated. The dotted lines demarcate sequences tested in binding and cleavage studies. (*B*) Alignment of exon and intron junctions. EL:ER, exon junction of an intronless *sunY* allele; IL:IR, potential intron junction for an ORF-less *sunY* allele; vertical lines, identical nucleotides; CS, cleavage site of I-*Tev* II; open arrow, predicted cleavage site for IL:IR. The putative endonuclease-ORF insertion site is ambiguous (two upward-directed arrows) because the tetranucleotide sequence T(U)GAA occurs both 5′ and 3′ to the ORF (arrowheads in *A*).

The significance of IL:IR constituting a cleavage substrate for I-*Tev* II was further addressed by examining sequence variability tolerated by this endonuclease. Thus, a pool of duplexes was synthesized with degenerate 23-bp EL:ER sequences containing an average of 4-bp changes per duplex, corresponding to the number of mismatches of the 23-nt alignment of EL:ER with IL:IR (Fig. 1*B*). The degenerate duplex (d[EL:ER]) was bound and cleaved by I-*Tev* II to <20% relative to wild-type EL:ER or IL:IR duplexes (Fig. 3*B*, compare lanes 1 and 3 with lane 5; Fig. 3*C*, compare lanes 1–3 and 4–6 with lanes 7–9). Thus, 4-bp changes per 23-bp EL:ER sequence are consistent with only a low level of binding and cleavage, most of which might be attributed to the 21% of duplexes that, according to the binomial distribution, are expected to have two or fewer changes. These results reflect a high degree of specificity of I-*Tev* II under these experimental conditions and underscore the significance of the functionality of the IL:IR duplex.

Binding to IL and Constraints on IR Sequences. Interestingly, IL:OL contains all of the information required for

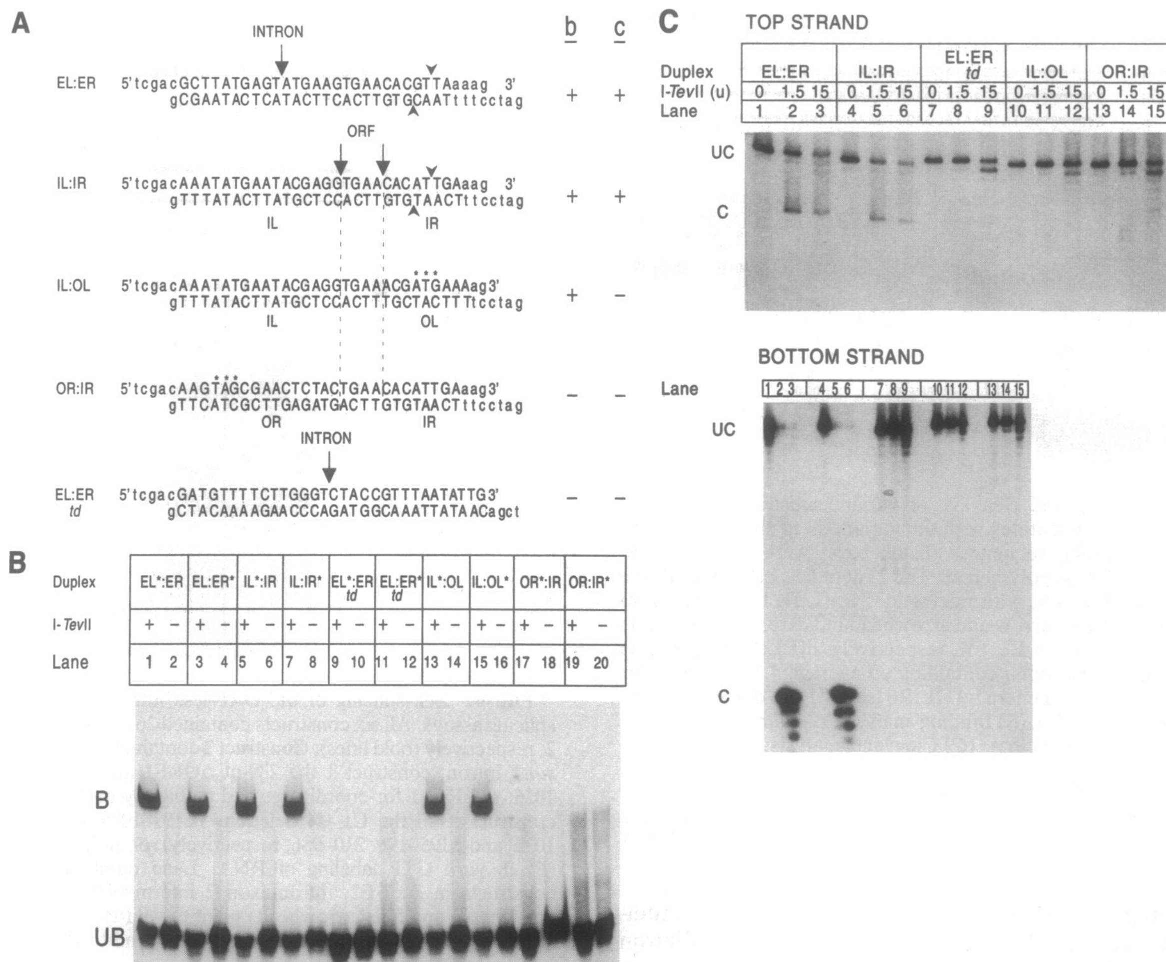

FIG. 2. **Binding and cleavage of intron sequences flanking the ORF by I-*Tev* II.** (*A*) Summary of DNA duplexes and results. EL:ER, exon junction of the intronless *sunY* gene (i.e., I-*Tev* II binding site); IL:IR, postulated intron junction of an ORF-less sunY intron; IL:OL, 5′ intron/ORF junction, with asterisks indicating endonuclease ORF start codon; OR:IR, 3′ ORF/intron junction, with asterisks indicating endonuclease ORF stop codon. The intron sequences comprising these duplexes are marked by dotted lines in Fig. 1*A*. EL:ER *td*, exon junction of the intronless phage T4 *td* gene (negative control). Lowercase letters at the ends of duplexes represent cloning linkers. Arrows are as in Fig. 1*B*. Arrowheads show cleavage sites. The rightmost columns summarize results of binding (b) and cleavage (c) assays from *B* and *C*, respectively. (*B*) Gel mobility-shift assay. Binding reaction mixtures contained the indicated duplexes incubated with (+) or without (−) I-*Tev* II. The radioactively labeled strand of each duplex is marked (*). B, bound complex; UB, unbound DNA. (*C*) Cleavage analysis. Cleavage assay mixtures contained the duplexes shown, incubated with the indicated amount of I-*Tev* II. u, units of I-*Tev* II. UC, uncleaved DNA; C, cleavage product. Lanes 1–15 for the bottom strand correspond to those for the top strand. Extra bands in lanes 3, 6, 9, 12, and 15 reflect nonspecific degradation at the highest enzyme concentration.

binding but not for cleavage by I-*Tev* II (Fig. 2). Therefore the 6 nt at the 3′ end of IL:IR, which differ from those in IL:OL, must bear the cleavage determinants for I-*Tev* II. When these 6 nt were randomized to form IL:6R (Fig. 3*A*), corresponding duplexes were bound by I-*Tev* II to an extent approaching that of EL:ER (95.7%; Fig. 3*B*, compare lanes 1 and 7) consistent with the binding properties of IL:OL. However, the IL:6R duplexes were cleaved to about 1/10th of EL:ER (9.3%; Fig. 3*C*, compare lanes 2 and 3 with lanes 11 and 12). With ≈10% cleavage, we estimate that 1 in 10 random 6-mer sequences 3′ to IL would provide a cut site. These results bear on both the cleavage properties of the enzyme and consideration of the ancestral intron (see *Discussion*).

Splicing of ORF-Less Introns. It was of interest to determine whether an ORF-less intron that contains the IL:IR junction (Figs. 1*B* and 2*A*) can form a splicing-competent structure. We therefore compared 3′ intron deletion derivatives known to be impaired in splicing (21, 24) with ΔORF1, a deletion construct with the precise IL:IR junction (Fig. 4*A*). We used an assay (25) that measures intron excision by labeling the 5′ end of the intron with [α-^{32}P]GTP during the first step of splicing (Fig. 4*B*). The major bands labeled in such a reaction are intron-exon 2 (I-E2), the product of cleavage at the 5′ splice site only, and intron (I), the product of cleavage at both splice sites. In this assay, the ΔORF1 intron containing the IL:IR junction, like the wild type, was capable of complete excision (Fig. 4*B*, I band, lanes 1 and 2). In contrast, precursors from the four control alleles underwent reduced cleavage at the 5′ splice site (I-E2 band of reduced intensity) and were not capable of cleavage at the 3′ splice site (absence of I band) (Fig. 4*B*, lanes 3–6). While the extent of intron excision was reduced in the ΔORF1 RNA relative to the wild type (Fig. 4*B*, lanes 1 and 2, compare I and I-E2 levels) and further compromised under more restrictive splicing conditions (e.g., 3 mM rather than 30 mM $MgCl_2$), exon ligation was shown to be accurate with ΔORF1 RNA (N.L., data not shown). These results indicate that although intron function is abolished by some 3′ deletions, the ΔORF1 intron retains accurate splicing function *in vitro*.

DISCUSSION

Similarities have been identified between intron sequences flanking the *sunY* endonuclease ORF (IL and IR) and the I-*Tev*

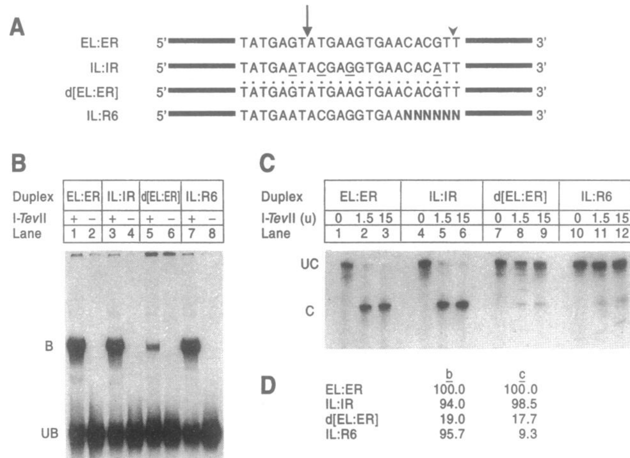

FIG. 3. Binding and cleavage of variant substrates. (*A*) DNA substrates. Oligonucleotides with the variant 23 nt shown and 21 nt of invariant flanking sequences (black bars) were made double-stranded with Klenow polymerase. The 23 nt of EL:ER and IL:IR are as indicated in Fig. 1*B*, with nucleotides in IL:IR that differ from EL:ER underscored. Arrow and arrowhead indicate intron insertion site and cleavage site on EL:ER, respectively. d[EL:ER] represents degenerate duplexes (dots) containing on average 4 bp changes per 23 bp (i.e., ≈17.4% degeneracy). IL:R6 is the IL:IR duplex with the last 6 nt randomized (N). (*B*) Binding analysis. Labeling as in Fig. 2*B*, with duplexes described in *A*. (*C*) Cleavage analysis. The top strand (shown in *A*) was radioactively tagged, with labeling as in Fig. 2*C*. (*D*) Summary of data. The extent of binding (b, from *B*) and cleavage (c, from *C*) is expressed relative to that of the EL:ER duplex, which was assigned a value of 100 (corresponding to ≈50% binding and 90% cleavage, respectively).

II recognition sequence in the exons (EL:ER) (Fig. 1). Additionally, the ability of fused IL:IR sequence of the *sunY* intron to act as a substrate for its cognate endonuclease has been demonstrated (Fig. 2). The significance of finding an I-*Tev* II cleavage site within the intron is supported by the specificity data provided by using degenerate oligonucleotides resembling EL:ER (Fig. 3). Together these results support the hypothesis that endonuclease ORFs invaded introns (2–5) by a similar cleavage event to that used to mobilize the introns (14). Further, although I-*Tev* II has binding specificity for IL sequences, the enzyme can cleave the DNA only when these are fused to appropriate IR sequences (Figs. 2 and 3). Thus the presence of the ORF prevents cleavage by I-*Tev* II, rendering the intron immune to its own gene product, thereby ensuring donor viability in subsequent intron mobility events where the endonuclease is expressed.

A recent report notes similarities between sequences flanking the endonuclease ORFs of the *Chlamydomonas eugametos* and *C. reinhardtii* chloroplast large-subunit rRNA introns and their cognate exon junctions (26). In a survey of alignments of exon junction sequences with ORF-flanking sequences for the five other known group I introns with freestanding endonuclease ORFs (all eight compiled in ref. 6), none displayed the sequence similarities observed in the *Chlamydomonas* and phage T4 systems (ref. 26; this study; E.R.M.T., data not shown). The absence of sequence identities in some cases is not surprising when one considers that intron sequences are under pressure to evolve to optimize splicing while accommodating the newly acquired insert. Therefore, retention of the IL binding site in the *sunY* intron suggests either a recent invasion event or constraints on these sequences for purposes of regulating endonuclease ORF expression (27).

Endonuclease-ORF Invasion and Consideration of the Ancestral Intron. Although the mode of entry of the endonuclease ORF is unknown, invasion by an illegitimate double-

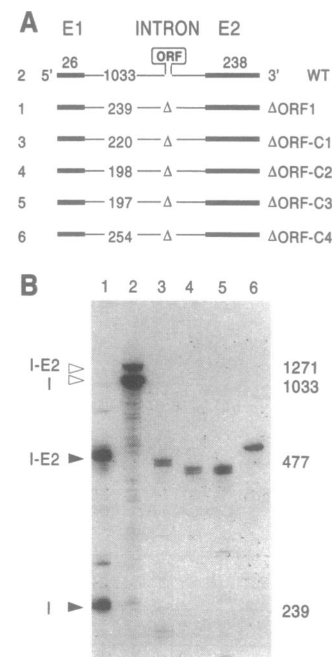

FIG. 4. Self-splicing of the ORF-less intron. (*A*) Transcripts for splicing assays. All six constructs contained 26 and 238 nt of exon 1 and 2, respectively (bold lines). Construct 2 contained the wild-type 1033-nt *sunY* intron, construct 1 the 239-nt ΔORF1 intron (Δ229/233–1022/1026; see Fig. 1 for coordinates and ambiguity in deletion joint), and constructs 3–6 the C1–C4 deletions (Δ197–1009; Δ197–1031; Δ197–1032; and Δ195–196, 210–986, respectively), of indicated intron length. (*B*) *In vitro* GTP labeling of RNA. Lane numbers correspond to constructs in *A*. I-E2, intron-exon 2 intermediate. I, intron. Open arrowheads and filled arrowheads point to wild-type and mutant products, respectively. Numbers at right correspond to RNA length (nt).

strand-break-mediated gene conversion event is possible, given that DNA-breakage enzymes can promote illegitimate recombination (28). Whereas any piece of DNA might in principle be captured within a double-strand break in the intron, endonuclease coding sequences would be selected, given their ability to propagate. In contrast, other sequences would tend to be lost in the streamlining process (15).

Endonuclease-ORF invasion, which would have been tolerated only if some degree of splicing activity were maintained, is proposed to have occurred into functional introns. While the ΔORF1 *sunY* intron bearing the IL:IR junction fulfills the splicing proficiency criteria (Fig. 4), one cannot state with certainty that exactly this sequence existed in the ancestral intron. Although this construct splices accurately, albeit at reduced efficiency (Fig. 4 and data not shown), it lacks the peripheral P9.2 element (Fig. 1*A*), which is involved in 3′ splice site selection and is conserved among all three phage T4 introns (9, 24). This raises the question of whether P9.2 was present in the ORF-less ancestral intron or whether it developed in response to endonuclease-ORF invasion of the *sunY* intron to ensure efficient 3′ splice site selection and then transferred to the other phage introns by recombination (29). Regardless, the finding that ≈10% of 6-mer sequences constituting IR can impart cleavage specificity (Fig. 3) implies a reasonable likelihood that the 6-nt sequence flanking IL in the ancestral intron could have been different from the designated IR sequence, while being consistent with both cleavage properties of the enzyme and splicing of the intron. Conversely, ≈90% of invasion events would have rendered the intron immune to self-cleavage by the endonuclease.

Models for the Evolution of Mobile Introns. A group I intron that had been colonized by an endonuclease ORF (Fig. 5, top) would lack the requisite homing properties except in extraor-

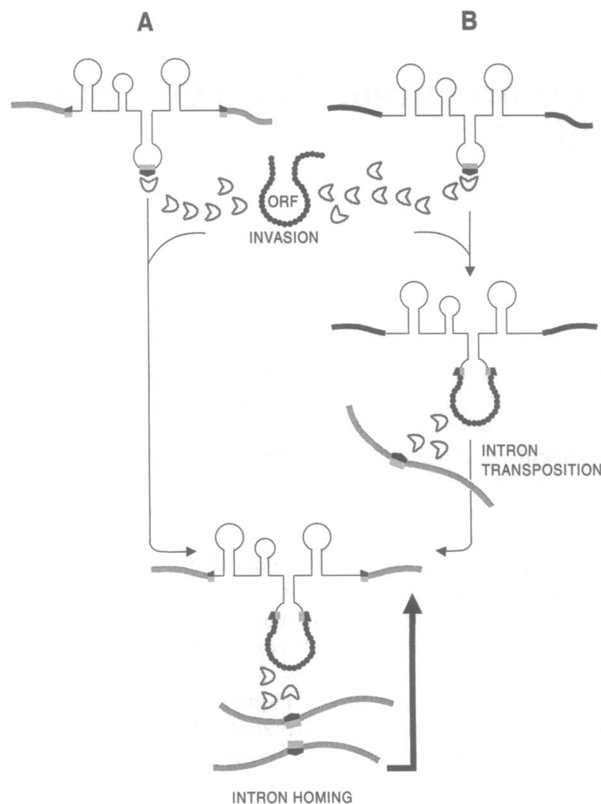

FIG. 5. Evolution of mobile group I introns. The two models (A and B), both of which begin with endonuclease-ORF invasion and culminate in intron homing, are described in the text. Thick lines, exons, solid or shaded to represent different alleles; thin lines, intron core; string of beads, endonuclease ORF; Pac-man symbols, endonuclease; bilayered symbols (intact or split), endonuclease recognition elements. Endonuclease-ORF invasion and intron transposition are considered to be very rare events (thin arrows), whereas intron homing is an extremely efficient process (thick arrow).

dinary circumstances. For homing to be established, cognate intronless alleles would need to be available. Even if intronless alleles were prevalent, which is not necessarily the case, the endonuclease would need to have specificity for the exon junction of these intronless variants (Fig. 5, bottom). The unlikely scenario of the endonuclease evolving this specificity is rendered even less credible by the functional similarity observed between the intron IL:IR and the exon EL:ER sequences (Figs. 1–3). However, one might argue that the bipartite intron would be maintained only on those infrequent occasions when the endonuclease ORFs are localized to introns for which intronless alleles that are also endonuclease targets exist (Fig. 5, model A). The composite intron's persistence in a population would derive from its ability to engage in the extremely efficient homing process. Alternatively, the necessity for intronless alleles that are also endonuclease substrates is bypassed if the intron, containing a newly acquired endonuclease ORF, could transpose at low frequency to a nonallelic locus, predicated on the existence of an endonuclease recognition sequence at that site (Fig. 5, model B) (4, 14). After translocation, the intron would be located in a homing situation, with a pool of intronless alleles containing endonuclease cleavage sites available as recipients for the homing event.

Functional similarity between intron sequences flanking the endonuclease ORF and exon sequences flanking the intron (Figs. 1 and 2) is consistent with the two models in Fig. 5. However, both models presuppose extremely rare events. Model A not only infers the existence of cognate intronless alleles but also assumes sequence relatedness between intron and exon recognition-site sequences, shown to be highly unlikely to exist by chance alone (Fig. 3*B*, d[EL:ER] experiments, and unpublished statistical data). On the other hand, the intron transposition event which is posited in model B remains unproven, although evidence is beginning to mount in its favor (reviewed in refs. 1, 4, and 14). Indeed, double-strand-break-dependent intron mobility has been observed when exon homology has been limited to ≤10 bp on the two sides of the intron (M. Parker and M.B., unpublished data). Regardless, the common feature of the two pathways is that the efficient homing process would tend to stabilize the presence of the mobile intron in a population. Thus the intron, rather than falling victim to invasion by a foreign element, becomes the beneficiary of a mobility apparatus that imparts the ability to propagate. These characteristics would tend to ensure the intron's persistence in its host organism during vertical transmission, while affording the possibility to transfer horizontally to other genomes.

We thank Mary Bryk, Keith Derbyshire, Vicky Derbyshire, Alan Lambowitz, Chip Lawrence, Randy Morse, John Mueller, Mary Reaban, and David Shub for useful discussions; David Shub for *sunY* constructs; Monica Parker for unpublished data; and Maryellen Carl for manuscript preparation. This work was supported by National Institutes of Health Grants GM39422 and GM44844.

1. Dujon, B. (1989) *Gene* **82,** 91–114.
2. Perlman, P. S. & Butow, R. A. (1989) *Science* **246,** 1106–1109.
3. Lambowitz, A. M. (1989) *Cell* **56,** 323–326.
4. Belfort, M. (1990) *Annu. Rev. Genet.* **24,** 363–385.
5. Lambowitz, A. M. & Belfort, M. (1993) *Annu. Rev. Biochem.* **62,** 587–622.
6. Mueller, J. E., Bryk, M., Loizos, N. & Belfort, M. (1993) in *Nucleases*, eds. Linn, S. M., Lloyd, R. S. & Roberts, R. J. (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 111–143.
7. Cech, T. R. (1988) *Gene* **73,** 259–271.
8. Michel, F. & Westhof, E. (1990) *J. Mol. Biol.* **216,** 585–610.
9. Shub, D. A., Gott, J. M., Xu, M.-Q., Lang, B. F., Michel, F., Tomaschewski, J., Pedersen-Lane, J. & Belfort, M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1151–1155.
10. Mota, E. M. & Collins, R. A. (1988) *Nature (London)* **332,** 654–656.
11. Dalgaard, J. Z., Garrett, R. A. & Belfort, M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 5414–5417.
12. Shub, D. A. & Goodrich-Blair, H. (1992) *Cell* **71,** 183–186.
13. Colston, M. J. & Davis, E. O. (1994) *Mol. Microbiol.* **12,** 359–363.
14. Bell-Pedersen, D., Quirk, S., Clyman, J. & Belfort, M. (1990) *Nucleic Acids Res.* **18,** 3763–3770.
15. Darnell, J. E. & Doolittle, W. F. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1271–1275.
16. Shub, D. A., Xu, M.-Q., Gott, J. M., Zech, A. & Wilson, L. D. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 193–200.
17. Quirk, S. M., Bell-Pedersen, D. & Belfort, M. (1989) *Cell* **56,** 455–465.
18. Kong, H., Kucera, R. B. & Jack, W. E. (1993) *J. Biol. Chem.* **268,** 1965–1975.
19. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185,** 60–89.
20. Xu, M.-Q. & Shub, D. A. (1989) *Gene* **82,** 77–82.
21. Xu, M.-Q. (1989) Dissertation (State Univ. of New York, Albany).
22. Burke, J. M., Belfort, M., Cech, T. R., Davies, R. W., Schweyen, R. J., Shub, D. A., Szostak, J. W. & Tabak, H. F. (1987) *Nucleic Acids Res.* **15,** 7217–7221.
23. Loizos, N. (1994) Dissertation (Albany Medical College, Albany, NY).
24. Michel, F., Jaeger, L., Westhof, E., Kuras, R., Tihy, F., Xu, M.-Q. & Shub, D. A. (1992) *Genes Dev.* **6,** 1373–1385.
25. Garriga, G. & Lambowitz, A. M. (1984) *Cell* **38,** 631–641.
26. Marshall, P., Davis, T. B. & Lemieux, C. (1994) *Eur. J. Biochem.* **220,** 855–859.
27. Gott, J. M., Zeeh, A., Bell-Pedersen, D., Ehrenman, K., Belfort, M. & Shub, D. A. (1988) *Genes Dev.* **2,** 1791–1799.
28. Ehrlich, S. D. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. D. (Am. Soc. Microbiol., Washington, DC), pp. 799–832.
29. Bryk, M. & Belfort, M. (1990) *Nature (London)* **346,** 394–396.