The low information content of *Neurospora* splicing signals: implications for RNA splicing and intron origin

RICHARD A. COLLINS,[1] JASON E. STAJICH,[2] DEBORAH J. FIELD,[1] JOAN E. OLIVE,[1] and DIANE M. DEABREU[1]

[1]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8
[2]Department of Plant Pathology and Microbiology, University of California, Riverside, California 92521, USA

ABSTRACT

When we expressed a small (0.9 kb) nonprotein-coding transcript derived from the mitochondrial VS plasmid in the nucleus of *Neurospora* we found that it was efficiently spliced at one or more of eight 5′ splice sites and ten 3′ splice sites, which are present apparently by chance in the sequence. Further experimental and bioinformatic analyses of other mitochondrial plasmids, random sequences, and natural nuclear genes in *Neurospora* and other fungi indicate that fungal spliceosomes recognize a wide range of 5′ splice site and branchpoint sequences and predict introns to be present at high frequency in random sequence. In contrast, analysis of intronless fungal nuclear genes indicates that branchpoint, 5′ splice site and 3′ splice site consensus sequences are underrepresented compared with random sequences. This underrepresentation of splicing signals is sufficient to deplete the nuclear genome of splice sites at locations that do not comprise biologically relevant introns. Thus, the splicing machinery can recognize a wide range of splicing signal sequences, but splicing still occurs with great accuracy, not because the splicing machinery distinguishes correct from incorrect introns, but because incorrect introns are substantially depleted from the genome.

Keywords: RNA splicing; splicing specificity; mitochondrial plasmids

INTRODUCTION

In multicellular eukaryotes most nuclear genes encoding mRNAs contain introns, typically multiple introns per gene. Introns are recognized and removed by the spliceosome, a complex molecular machine comprising five small nuclear RNAs (U-snRNAs) and well over a hundred proteins (Wahl et al. 2009). In addition to the integral spliceosomal proteins, many organisms encode *trans*-acting proteins, such as SR and hnRNP proteins, that can bind to sequences in exons or introns to enhance or inhibit splicing at a given site (Chen and Manley 2009). Differential histone modification and chromatin structure have also been implicated in splice site usage efficiency (Zhou et al. 2014).

RNA splicing is widely considered to be a high fidelity process with estimates of incorrect splice site usage being 10^2–10^5-fold lower than that of the correct splice site (Fox-Walsh and Hertel 2009). Consistent with this accuracy, spliceosomal proofreading mechanisms have been identified in some organisms that favor the use of certain splice sites over others in the same transcript (Semlow and Staley 2012). The poly(A) mRNA population, enumerated by sequencing of cDNAs or expressed sequence tags (ESTs) and/or high-throughput RNA-seq, is almost exclusively composed of RNAs that contain an open reading frame potentially capable of being translated into a protein. RNAs spliced at sites that would introduce a premature stop codon or otherwise produce a nonfunctional mRNA have also been detected, but are present at very low abundance (Pickrell et al. 2010), consistent either with a high fidelity of splicing or with quality control pathways efficiently degrading "incorrectly spliced" RNAs.

Some pre-mRNAs are spliced constitutively (or very preferentially) at a single 5′ and 3′ splice site; others are alternatively spliced. The use of high sensitivity RT-PCR and RNA-seq has revealed that most mammalian genes give rise to alternatively spliced mRNAs, although many isoforms are present in very low abundance (Blencowe 2006). The ratios of the different alternatively spliced forms can vary substantially among tissues in a given organism, consistent with *trans*-acting factors regulating the relative efficiency of splice site usage. Some alternatively spliced RNAs encode proteins with different biological functions; others are observed or predicted to produce nonfunctional RNAs. There is debate in the

Corresponding author: rick.collins@utoronto.ca
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.047985.114.

field about how much alternative splicing is biologically relevant versus being a consequence of errors or noise in the splicing process.

The original motivation for the work described in this paper was unrelated to pre-mRNA splicing. We set out to investigate the mechanism by which the *Neurospora* mitochondrial VS plasmid (Collins and Saville 1990; Saville and Collins 1990) enters and spreads through a mitochondrial population. VS is one of many different kinds of circular and linear plasmids found in mitochondria, but not nuclei, of a variety of fungi (Griffiths 1995; Galagan and Kennell 2007; Hausner 2012). The possible mechanism that we intended to examine was that the naturally abundant, circular VS RNA might be imported into mitochondria, perhaps analogous to the import of tRNAs in some other organisms. Instead, we discovered that when we expressed VS RNA in the nucleus it was spliced at multiple sites. Here we describe the implications of these observations for the specificity of the spliceosome, the origin of introns, and the evolution of eukaryotic plasmids.

RESULTS

We established an experimental system in which we intended to produce circular VS RNA in the cytosol of *Neurospora* to see if the RNA might be imported into mitochondria. This system comprises a DNA containing a partial multimer of the VS plasmid under the control of a nuclear RNA polymerase II promoter and terminator integrated into the *Neurospora* nuclear genome (Supplemental Fig. S1). We hypothesized that the VS transgene would be expressed like a typical Pol II-transcribed intronless mRNA, becoming capped, polyadenylated, and exported to the cytosol. We expected a proportion of the transcripts to undergo VS ribozyme-mediated cleavage and ligation to excise and circularize a VS monomer, as we have observed with in vitro transcription and processing of similar partial-multimer constructs, and as occurs naturally in the mitochondria of plasmid-containing strains (Saville and Collins 1991).

We chose a random integration strategy rather than targeting the VS transgene to a specific locus because there was no obvious "right place" to insert a DNA that did not naturally occur in the nuclear genome. To select for transformants we included a hygromycin-resistance gene, also under the control of Pol II signals, adjacent on the same transforming DNA (Supplemental Fig. S1C). We selected hyg^r transformants that grew well, indicating that the transforming DNA integrated in a region of the genome that allowed transcription and did not interfere with essential cellular functions. To avoid potential confusion due to contamination from natural VS-containing strains in the laboratory, we have done most of our work with transformants that contain sequence variants of VS that we have engineered to be distinct from known natural alleles of VS: Some contain point mutations and some resemble deletion mutants observed in natural or laboratory isolates of the VS plasmid (Supplemental Fig. S1D–F). Genomic DNA from each transformant was examined by PCR, cloning and sequencing and was confirmed to contain only the intended VS sequence.

We found that the variant VS RNAs were expressed from these nuclear transgenes, as detected by RT-PCR using total cellular RNA as a template and an arbitrarily chosen pair of primers that amplify a short region of VS RNA. RNA abundance, as inferred from intensity of RT-PCR bands on ethidium bromide stained gels varied among transformants and was typically lower than that of the hyg^r mRNA encoded adjacent on the transforming DNA. More careful examination of the status of VS RNA in the transformants using primers designed to amplify the predicted full-length transcript unexpectedly showed that very little of the expected transcript was present in the transformants. Instead, several shorter RNAs were detected (Fig. 1).

Cloning and sequencing of RT-PCR products showed that the shorter RNAs contained one or more of several internal deletions compared with the genomic sequence from which they were transcribed (Fig. 2; Supplemental Table S1A; Supplemental Fig. S2). Examination of the sequences of the deleted regions showed that the 5′ ends matched the weak consensus sequences for *Neurospora* nuclear spliceosomal 5′ splice sites (5′ss), the 3′ ends matched the nuclear 3′ splice site (3′ss) consensus, and each deletion contained a reasonable match to the predicted branchpoint (bpt) consensus sequence at a distance from the 3′ss that was typical of the distance in annotated *Neurospora* nuclear introns (Fig. 2; Supplemental Table S2; Supplemental Fig. S3). Apparently the nuclear-expressed VS RNAs looked sufficiently like natural Pol II transcripts that the nuclear splicing machinery recognized and spliced them, even though they would not have been expected to contain introns.

We have observed splicing of nuclear-expressed VS transcripts at multiple positions, including alternative splicing using different 5′ or 3′ splice sites. More than one splice was often observed in a single RNA (Supplemental Table S1A). Either of two commonly observed splices (nucleotides 740–807 or 744–807) removes an essential part of the VS ribozyme that would be required for ribozyme-mediated processing of VS RNA. Introducing a VS ribozyme active-site null mutation (Sood and Collins 2002) into the nuclear expression construct still allowed splicing at the same positions observed in wild type, confirming that VS ribozyme activity is not involved in these splicing events (Supplemental Table S1A, mutant NTC). Point mutations designed to abolish or create 5′ or 3′ splice sites had the effects expected for nuclear pre-mRNA splicing. For example, the commonly used 5′ splice sites at positions 740 or 744 are preferentially spliced to nucleotide 807 in the wild-type sequence. Changing that 3′ss from AG to AC abolished splicing at that position; introducing a new 3′ss nearby (T815G, creating a potential alternative AG 3′ss) resulted in splicing of 740 or 744 to nucleotide 816, using the new 3′ss (Supplemental Table S1A,

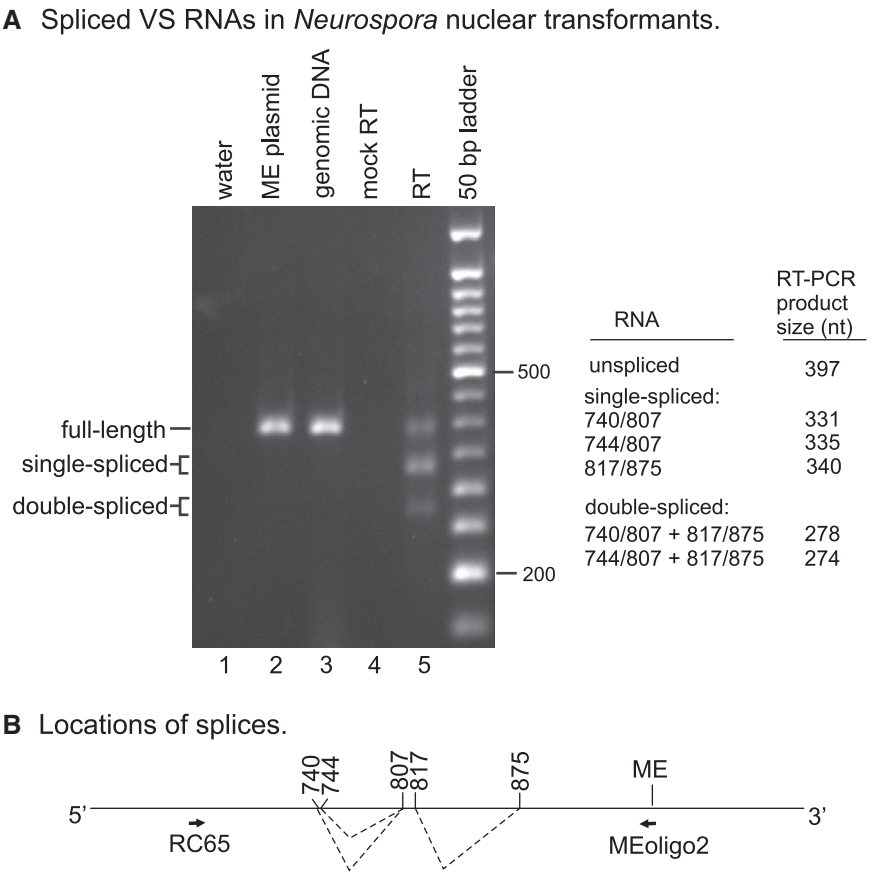


FIGURE 1. Detection of spliced forms of VS RNA in a *Neurospora* nuclear transformant expressing the ME deletion variant. (*A*) The templates indicated *above* lanes *1–5* were amplified by PCR with primers RC65 and MEoligo2 and detected by ethidium bromide staining following agarose gel electrophoresis. ME plasmid (lane *2*) refers to the DNA used to transform *Neurospora*. Genomic DNA (lane *3*), reverse transcription (RT) reaction (lane *5*), or mock RT (lane *4*) from total cellular RNA were from an arbitrarily chosen individual *Neurospora* transformant. Random hexamers were used to prime the RT reaction. The mock RT sample was treated the same as the RT sample except that no reverse transcriptase was added. (*B*) RT-PCR products were cloned and sequenced to determine the locations of splices (indicated by dashed lines). Numbers refer to the standard VS RNA numbering system (see Fig. 2; Supplemental Fig. S1). See Supplemental Table S1 for a summary of all splices detected in all transformants and Supplemental Figure S2 for the sequences of all observed introns.

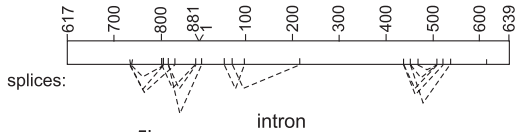
mutants NTD and NTE). Considering that VS RNA is naturally mitochondrial and would not be expected to have undergone positive selection to create or maintain nuclear spliceosomal introns, we found the number and efficiency of splicing events to be surprisingly high.

This unexpected splicing is not unique to the VS plasmid. Using a bioinformatic approach we analyzed other *Neurospora* mitochondrial plasmid sequences for the presence of potential introns and predicted candidate introns in all of them (Supplemental Table S3). We experimentally examined splicing of a 664-nt region of one additional plasmid RNA, called the Varkud (or V) plasmid (Akins et al. 1988), using the same nuclear transgene approach. Within this small region of the V RNA five different splicing events were observed, including splicing at alternative 3′ splice sites (Fig. 2; Supplemental Table S1B). Because two different mt plasmid RNAs are each spliced at multiple sites using many variants of the nuclear 5′ss and branchpoint consensus sequences, including several that are rarely used in natural nuclear introns, we conclude that the *Neurospora* spliceosome is inherently very promiscuous, in the sense that it can recognize a wide range of splice site sequences (analogous to the biochemists' usage of promiscuous to describe proteins that can bind to a number of partners) (Copley 2015).

Analysis of *Neurospora* RNA-seq data also supports the hypothesis that the spliceosome is not always as accurate as generally thought. Supplemental Table S2 lists 401 nuclear genes for which RNA-seq identified alternative 5′ss (222 genes), 3′ss (156 genes), or both (23 genes) in addition to those annotated using current gene models in the Broad Institute database. Many of these alternative splicing events produce mRNAs of low abundance that encode out-of-frame proteins and/or premature termination codons. In none of these cases is there reason to expect that the alternative splicing is biologically relevant. We interpret these to be additional examples of the promiscuity of the *Neurospora* spliceosome.

The observation that functional spliceosomal intron sequences occur frequently in mitochondrial plasmids led us to examine the occurrence of splicing-related sequences and predicted introns in *Neurospora* nuclear genes, intergenic sequences and randomized sequences. This analysis let us explore the hypothesis that potential introns are frequent in random sequence, and that intronless nuclear genes are intronless because they have been depleted of splicing-related sequences that would otherwise be present by chance in random sequence of the same nucleotide composition. We first analyzed a large sample of annotated intron-containing nuclear gene sequences using a range of criteria to define an intron (Fig. 3; Supplemental Fig. S4A). The most relaxed criteria examined (CTRAY branchpoint, HAG 3′ss and one of the most frequently occurring thirty-one 5′ss sequences; where R = A or G, Y = T or C, and H = A, T, or C) include 67% of annotated introns; the most stringent criteria (CTAAC branchpoint, YAG 3′ss, and one of the six most frequently occurring 5′ss sequences) include only 22% of annotated introns. Using the most relaxed criteria, random sequences are predicted to contain introns at approximately the same frequency as real intron-containing genes, slightly

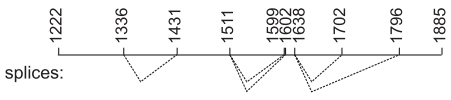
A VS plasmid



splice[a]	5'ss sequence	5'ss rank[b]	intron length (nt)	predicted bpt[c]	bpt to 3'ss (nt)	3'ss
740-807	GTAAGT	1	66	CTAAC	14	AAG
740-833	GTAAGT	1	92	gTAAT	14	TAG
744-807	GTcAGT	10	62	CTAAC	14	AAG
801-875	GTAAGa	7	73	aaAAC	19	AAG
817-875	GTAAtT	33	57	aaAAC	19	AAG
817- 5	GTAAtT	33	68	aTAAC	16	AAG
442-514	GTtTGT	21	71	CTGAT	10	TAG
442-516	GTtTGT	21	73	CTGAT	12	gAG
442-530	GTtTGT	21	73	CTGAT	27	TAG
457-514	GTAGGT	6	56	CTGAT	11	TAG
457-524	GTAGGT	6	66	tTAAT	14	AAG
457-530	GTAGGT	6	72	tTAAT	20	TAG
461-514	GTGTGa	52	52	CTGAT	11	TAG
461-530	GTGTGa	52	68	tTAAT	20	TAG
43-99	GTATGT	3	57	CTAAa	11	TAG
60-231	GTATaT	41	170	tTGAC	13	TAG
60-602[d]	GTATaT	41	65	aaAAC	15	TAG
60-602[d]	GTATaT	41	120	aaAAC	15	TAG
60-565[d]	GTATaT	41	83	CTGAC	7	TAG
518-565	GTAAGC	5	49	tTGAC	7	TAG
nuclear introns[e]	GTAAGT		50 to 300 nt	CTAAC	9 to 25 nt	TAG

5'ss 3'ss

B V plasmid



splice[a]	5'ss sequence	5'ss rank[b]	intron length (nt)	predicted bpt[c]	bpt to 3'ss (nt)	3'ss
1336-1431	GTAtGg	22	94	CTAAT	25	TAG
1511-1599	GTtAGT	9	87	tTGAC	37	TAG
1511-1602	GTtAGT	9	90	tTGAC	40	TAG
1638-1702	GTAgGT	6	63	gTAAC	23	TAG
1638-1796	GTAgGT	6	157	tTAAT	10	TAG

FIGURE 2. Summary of splices observed in mitochondrial plasmid RNA variants expressed in the nucleus. Splices are indicated by dashed lines *below* the diagrams of (*A*) VS RNA and (*B*) V RNA. The tables show the sequences and spacing of splicing-related sequences of each splice observed. Uppercase letters match the nucleotide(s) at the corresponding position in the consensus sequence derived from all nuclear introns. [a]Numbers refer to the last base of the upstream exon and the first base of the downstream exon using the standard sequence numbering (*top*). [b]The position of each 5′ss sequence on the list of all 5′ss observed in natural nuclear introns ranked by decreasing frequency of occurrence; see Supplemental Table S2. [c]The predicted branchpoint (bpt) adenosine (underlined) in the best match to the consensus sequence CTRAY located between 9 and 25 nt upstream of the 3′ss. [d]Splicing of nucleotides 60–565 or 602 removes introns of different sizes using different predicted branchpoints in deletion constructs XE and ME in which the distance between these splice sites has been shortened enough to create an intron of a size more typical of natural nuclear introns (see Supplemental Table S2). [e]Summary of splicing sequences in natural *Neurospora* nuclear introns; in the sequence diagrams letter height is proportional to frequency of occurrence; the stated range of intron lengths and branchpoint to 3′ss distances comprise ~95% of all natural nuclear introns; see Supplemental Table S2 for raw data.

more than one intron per kilobase (Fig. 3A; Supplemental Fig. S4D). Using a series of increasingly stringent prediction criteria, annotated intron-containing genes showed an enrichment of predicted introns relative to randomized sequences (Fig. 3B; Supplemental Fig. S4D), consistent with the idea the "good" intron sequences have been subject to positive selection.

In contrast, predicted introns are under-represented in intronless genes when compared with randomized sequences of the same number, length, and base composition or to sets of intergenic sequences (Fig. 3B,C; Supplemental Fig. S4D). Using the most stringent intron-definition criteria, the entire population of 1064 intronless *Neurospora* nuclear genes is predicted to contain only six potential introns. In contrast, three sets of 1064 randomized sequences are predicted to contain 137 ± 6 such introns, a depletion of >20-fold in the intronless nuclear sequences. Analysis of individual categories of splicing sequences reveals that intronless genes show an underrepresentation of candidate 5′ss, 3′ss, and branchpoint sequences compared with randomized sequences (Fig. 4; Supplemental Fig. S4B). The depletion of predicted introns that use the most frequently observed splicing sequence variants is even slightly greater than expected from simply multiplying the fold-depletion of the individual splicing sequences (expected = 14.3-fold; Supplemental Fig. S4B, column O). This additional depletion is consistent with the few remaining potential (but apparently unused) splicing sequences being located at unfavorable relative spacings or occluded by RNA secondary structures that further decrease their probability of contributing to a functional intron. Using less stringent intron-definition criteria, the depletion of predicted introns in intronless genes becomes progressively less severe (Fig. 3B): that is, intronless genes are strongly depleted of "good" potential introns, but not all potential introns.

Figure 5 shows the frequencies of predicted introns in intronless genes and randomized sequences from three other fungi, using the same prediction criteria as used for *Neurospora* in Figure 3C. Randomized sequences from each of these fungi are predicted to contain introns at about the same frequency as randomized *Neurospora* genes (~0.08 introns per kilobase). In all of these organisms, predicted introns are substantially under-represented (by 11- to 31-fold) in annotated intronless genes. As with *Neurospora* intronless genes, this underrepresentation of predicted introns arises from a combination of depleting 5′ss, branchpoint and 3′ss sequences (Supplemental Fig. S5). We conclude that selection-driven depletion of intron-specifying sequences is a common feature in the evolution of intronless genes in fungi.

DISCUSSION

We report here that sequences that function as nuclear spliceosomal introns occur with high frequency in mitochondrial plasmids of *Neurospora*. In the most thoroughly

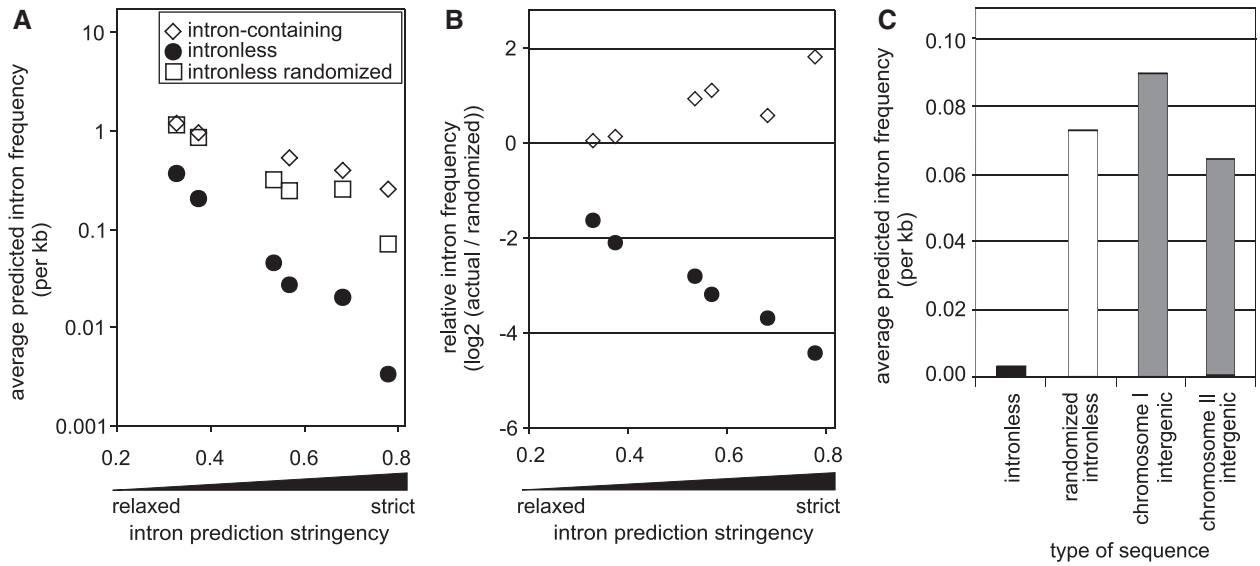
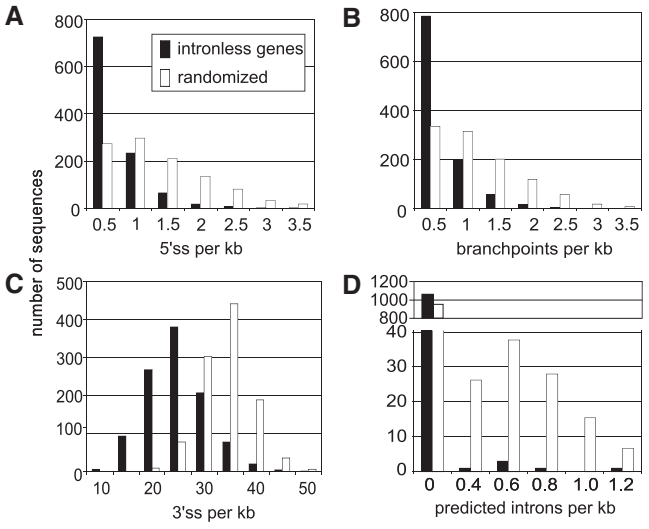


FIGURE 3. Depletion of predicted introns in *Neurospora* intronless genes. (*A*) Introns were predicted in sets of annotated intron-containing (open diamonds), intronless (filled circles), or randomized intronless (open squares) sequences. The randomized sequences retained the lengths and nucleotide compositions of the intronless sequences. Six sets of prediction criteria were used (differing in their definition of 5′ss, branchpoint and 3′ss sequences). Each set of criteria was assigned a stringency score, plotted on the *x*-axis, that is derived from the fraction of annotated introns in intron-containing genes that met each set of criteria (see Materials and Methods; Supplemental Fig. S4A). (*B*) The average frequency of predicted introns in intron-containing (open diamonds) or intronless (filled circles) sequences was divided by the frequency in randomized sequences to show enrichment in intron-containing genes and depletion in intronless genes. (*C*) Average frequencies of predicted introns in the indicated categories of sequences, using the most strict intron prediction criteria (criteria #6 in Supplemental Fig. 4A). See Supplemental Figure S4 for complete frequency distributions and additional analyses.

characterized example, the VS plasmid transcript can be spliced in at least a dozen different ways, often more than once in a single RNA molecule.

It seems unlikely that VS RNA, a small (~0.9 kb), naturally mitochondrial, nonprotein-coding RNA, has been under positive selection to evolve and retain multiple sets of functional nuclear splicing signals. It is possible that the sequences that function as splicing signals arose coincidentally due to selection for some other process that selects for similar sequences. More likely, they are simply present by chance, an interpretation supported by analyses of randomized and intergenic sequences (Fig. 4; Supplemental Fig. S4). It would not be especially surprising to observe splicing of a foreign RNA in the rare situation where that RNA contained appropriately positioned sequences that matched the frequently used 5′ss, branchpoint and 3′ss of the host's endogenous genes. Indeed, sporadic examples of unanticipated splicing of foreign RNAs in other organisms have also been reported, beginning with splicing of jellyfish green fluorescent protein mRNA in a plant host (Haseloff et al. 1997). However, the large number of splicing events observed in *Neurospora* VS and V RNAs using many different 5′ss, 3′ss and branchpoint sequences, some of which are poor matches to the nuclear splicing consensus sequences, indicates that the spliceosome can recognize and splice a wide range of sequences and that it does not require a large amount of information to identify a sequence that will function as an intron.

The conclusion that *Neurospora* splicing signals are short and degenerate is also supported by an analysis of natural *Neurospora* nuclear intron-containing genes (Fig. 2; Supplemental Table S2; Schwartz et al. 2008). Although certain 5′ss sequences occur more frequently than others (consistent with positive selection for these sequences), the spliceosome utilizes >200 variants of a GTnnnn 5′ss consensus sequence. The branchpoint consensus is also degenerate (CTRAY), and 18% of nuclear intron branchpoints deviate from even this consensus. The 3′ss YAG comprises 90% of annotated introns; however, there are an additional several hundred introns that use AAG. Similar functional, but biologically irrelevant, splicing sequences occur very frequently and within an appropriate range of distances even in the 0.9-kb mitochondrial VS RNA: The eight 5′ss sequences used in VS RNA range from the most frequently used nuclear 5′ss, used in 19% of nuclear introns, to the 52nd-ranked 5′ss used in only 0.1% of nuclear introns (highlighted in red in Supplemental Table S2A, column P). Additionally, multiple different splices were observed in a 0.6 kb region of a nuclear-expressed transcript derived from another plasmid (the V plasmid) (Fig. 2B) and bioinformatic analyses of random sequence DNA (Fig. 4; Supplemental Fig. S4) or intergenic sequences (Fig. 3C; Supplemental Fig. S4) predict potential introns at frequencies similar to those found in annotated intron-containing genes. All of these observations indicate that the specificity of the *Neurospora* splicing



E

splicing sequence	mean intronless	mean randomized	fold-depletion	p-value
5'ss	0.34	1.06	3.10	1.3 E-92
bpt	0.27	0.89	3.25	8.4 E-86
3'ss	22.38	31.48	1.41	2.8 E-151
intron	0.003	0.072	21.48	6.1 E-25

FIGURE 4. (*A–D*) Distributions of frequencies of splicing-related sequences and predicted introns in *Neurospora* intronless genes (filled bars) and random sequences (unfilled bars) using the most stringent prediction criteria shown in Figure 3. (*E*) Mean frequency (per kilobase) of each of the indicated splicing sequences. Fold depletion = (mean of randomized sequences/mean of intronless sequences). Wilcoxon rank sum tests were used to compare the distributions of the indicated type of splicing sequence in annotated intronless genes to each of three sets of randomized sequences of the same length, nucleotide composition, and number ($n = 1064$). The average of the *P*-values from each of the three tests is shown.

apparatus is not sufficient to distinguish biologically relevant introns from random sequences that, by chance, contain splicing signals distributed in such a way as to frequently create introns.

Our analysis of *Neurospora* nuclear genes that lack introns shows that they are depleted of (1) candidate branchpoint sequences (especially CTAAC, which is the perfect complement of the *Neurospora* U2 snRNA (Supplemental Fig. S3C) and the most frequently used branchpoint sequence in real *Neurospora* nuclear introns; (2) 5′ss sequences matching those most frequently used in real nuclear genes; and (3) YAG 3′ss sequences, which are also the most frequently used 3′ss in real nuclear introns. We cannot determine if this depletion is due to the role of these sequences in specifying introns, or to selection for or against other sequences with the coincidental result that intron-specifying sequences became depleted. Nonetheless, when presented with the opportunity to splice RNAs that have not been under selection to remove potential splice sites, such as the mitochondrial VS or V RNAs, the *Neurospora* spliceosome reveals its intrinsic ability to recognize a wide range of splicing signals by splicing these foreign RNA at multiple sites. We suggest that splicing of natural *Neurospora* pre-mRNAs only appears to be a high-specificity process, not because of the accuracy of the spliceosome, but because evolution has limited or removed the opportunity for missplicing by weakening or eliminating splicing sequences at locations other than those comprising biologically relevant introns.

Other factors likely also contribute to the appearance, but not the reality, of splicing as a high-specificity process. Quality control mechanisms, such as nonsense-mediated, nonstop, and no-go decay, and possibly other pathways (Lynch and Kewalramani 2003; Dumesic et al. 2013; Kawashima et al. 2014), minimize the proportion of incorrectly spliced RNAs in the cytosolic poly(A) mRNA population. Indeed, VS and V transcripts expressed from nuclear transgenes are present at low abundance, consistent with their being rapidly degraded. Previous studies that inferred high splicing fidelity from sequencing of poly(A) mRNA may have greatly underestimated the amount of missplicing that occurred but was removed by downstream events and therefore absent from the pool of RNAs analyzed.

The ability of contemporary spliceosomes to recognize and cleave specific sequences might provide a genome defense mechanism by degrading RNAs from foreign sources that have not been depleted of potential introns and will, by chance, frequently contain splicing sites. RNAs of apparently foreign origin transcribed from many different plasmids have been described in mitochondria of many filamentous fungi, but plasmids are not present in their nuclei (Griffiths 1995; Klassen and Meinhardt 2007). Our observations suggest that selection to avoid being inactivated by the spliceosome could contribute to lack of plasmids in the nucleus and their frequent presence in mitochondria and cytosol which lack spliceosomes.

In summary, the observation of rampant splicing of nuclear-expressed transcripts of two arbitrarily chosen mitochondrial plasmid RNAs demonstrates that functional spliceosomal intron sequences are frequently present by

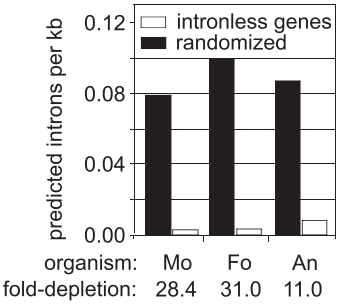


FIGURE 5. Depletion of predicted introns in fungal intronless genes. Introns were predicted using the most stringent criteria (criteria #6 in Supplemental Fig. 4A) in annotated intronless genes and sets of randomized sequences of the same number, length, and nucleotide composition from *Magnaporthe oryzae* (Mo), *Fusarium oxysporum* (Fo), and *Aspergillus nidulans* (An). See Supplemental Figure S5 for depletion of individual splicing signal sequences in these organisms and additional analyses.

chance, and that the contemporary *Neurospora* spliceosome exhibits enough flexibility to splice them. Analyses of natural intronless *Neurospora* nuclear genes suggests that they are intronless at least in part because evolution has depleted them of sequences that would specify introns, not because a high-specificity splicing apparatus or accessory factors distinguish biologically relevant from irrelevant splicing sites.

MATERIALS AND METHODS

Plasmids and *Neurospora* transformation

The family of constructs designed to express variants of the VS and V mitochondrial plasmids in the nucleus of *Neurospora* transformants is described in Supplemental Figure S1. Plasmid DNA linearized with HindIII was transformed via electroporation into conidia from *Neurospora* strain FGSC 2225 (Mauriceville-1c). Transformants were selected on plates that contained enough hygromycin (hyg) to stop the growth of untransformed cells (Colot et al. 2006). From each transformation experiment several of the largest, most healthy-looking colonies were transferred to Vogel's slants containing hygromycin and grown until well conidiated. Conidia were inoculated into liquid Vogel's minimal medium, typically 10 mL, grown overnight with shaking at 25°C and harvested by vacuum filtration. The presence of the expected VS or V transgene was confirmed by PCR, cloning, and sequencing of genomic DNA.

RT-PCR

Harvested mycelium (~0.2 g) was disrupted by grinding with one to two volumes of acid-washed sand in 1.5 mL microcentrifuge tube using a matching plastic pestle. Total RNA was extracted using 1 mL of Tri-Reagent (Chomczynski 1993) according to the manufacturer's instructions (Molecular Research Center). cDNA was synthesized from 3 to 5 µg of RNA in a 20 µL reaction using SuperScript III reverse transcriptase (Invitrogen) and the primers indicated in Supplemental Table S1 (column G). PCR was performed on 1.5 µL of the cDNA reaction in a 25 µL PCR reaction using the primers indicated in the figures and tables to detect designated portions of the VS or V transgene transcripts; parallel control PCR reactions were performed to detect transcripts of the actin and/or hyg^r genes as positive controls for the quality of the RNA preparation.

Bioinformatic analyses

5′ss and 3′ss sequences and intron lengths were tabulated from annotated introns extracted from release 10 of the *Neurospora* genome annotation version 12 (Supplemental Methods S1). Candidate branchpoint (bpt) sequences in annotated introns were predicted by searching for over-represented motifs in the last 40 nt of each intron, as described in Supplemental Figure S3.

Examination of introns based on expressed RNA was performed using two RNA-seq data sets from NCBI Short Read Archive (accessions: SRP018429, SRP005694) and assembled into transcripts with Trinity (version r2012-01-25) (Grabherr et al. 2011). The transcript sequences were aligned to the genome with PASA (Haas et al. 2008) with intron aware aligners to identify 5′ and 3′ss usages.

Randomized sequences matching the lengths and nucleotide frequencies of actual *Neurospora* sequences were generated by random sampling of single nucleotides (without replacement) from actual sequences (Supplemental Methods S2). Genes shorter than 900 nt were omitted because analyses (not shown) revealed that the frequency of predicted introns, even in randomized sequences, decreased with decreasing gene length below ~900 nt for statistical reasons. The lists of intron-containing, intergenic, intronless, and randomized intronless sequences are presented in Supplemental Methods S3.

Prediction of introns is described in detail in Supplemental Methods S1. Six sets of prediction criteria were used, differing in their user-defined lists of sequences comprising 5′ss, branchpoint, and 3′ss (see Supplemental Fig. S4). The range of allowed intron sizes was limited to 50–300 nt, which comprises 95% of annotated introns; the range of branchpoint-to-3′ss distances was 9–25 nt, comprising 96% of annotated introns. For each set of prediction criteria, the fraction of annotated introns that meet those criteria (f_P) was counted, and a stringency score for those criteria was defined as $1 - f_P$. For example, the most relaxed set of criteria (#1) encompasses 67% of annotated introns, thereby having a score of 0.33. The strictest criteria (#6) were met by only 22% of annotated introns, thereby having a score of 0.78. Algorithms using each set of criteria were used to predict introns in intron-containing, intronless, intergenic, and randomized sequences.

Intronless genes from *Fusarium oxysoporum* f. sp. lycopersici 4287, *Magnaporthe oryzae* 70-15 and *Aspergillus nidulans* FGSC A4 were downloaded from FungiDB (Stajich et al. 2012) (http://fungidb.org/fungidb/) by selecting genes annotated as having one exon (i.e., no introns), and analyzed as described above for *Neurospora*.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This study was supported by the Canadian Institutes for Health Research (grant MOP-12837 to R.A.C.) and the Canada Research Chairs Program. J.E.S. was supported by the Alfred P. Sloan Foundation. We thank Nathalie Moati for construction of intermediate plasmids used to make pNTA and Caitlin Collins for performing the rank sum tests in Figure 4.

Received September 8, 2014; accepted February 13, 2015.

REFERENCES

Akins RA, Grant DM, Stohl LL, Bottorff DA, Nargang FE, Lambowitz AM. 1988. Nucleotide sequence of the Varkud mitochondrial plasmid of *Neurospora* and synthesis of a hybrid transcript with a 5′ leader derived from mitochondrial RNA. *J Mol Biol* **204:** 1–25.

Blencowe BJ. 2006. Alternative splicing: new insights from global analyses. *Cell* **126:** 37–47.

Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat Rev Mol Cell Biol* **10:** 741–754.

Chomczynski P. 1993. A reagent for the single-step simultaneous isolation of RNA, DNA and proteins from cell and tissue samples. *Biotechniques* **15:** 532–537.

Collins RA, Saville BJ. 1990. Independent transfer of mitochondrial chromosomes and plasmids during unstable vegetative fusion in *Neurospora*. *Nature* **345:** 177–179.

Colot HV, Park G, Turner GE, Ringelberg C, Crew CM, Litvinkova L, Weiss RL, Borkovich KA, Dunlap JC. 2006. A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc Natl Acad Sci* **103:** 10352–10357.

Copley SD. 2015. An evolutionary biochemist's perspective on promiscuity. *Trends Biochem Sci* **40:** 72–78.

Dumesic PA, Natarajan P, Chen C, Drinnenberg IA, Schiller BJ, Thompson J, Moresco JJ, Yates JR III, Bartel DP, Madhani HD. 2013. Stalled spliceosomes are a signal for RNAi-mediated genome defense. *Cell* **152:** 957–968.

Fox-Walsh KL, Hertel KJ. 2009. Splice-site pairing is an intrinsically high fidelity process. *Proc Natl Acad Sci* **106:** 1766–1771.

Galagan JT, Kennell JC. 2007. Retroplasmids: linear and circular plasmids that replicate via reverse transcription. *Microbiol Monogr* **7:** 163–185.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29:** 644–U130.

Griffiths AJ. 1995. Natural plasmids of filamentous fungi. *Microbiol Rev* **59:** 673–685.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol* **9:** R7.

Haseloff J, Siemering KR, Prasher DC, Hodge S. 1997. Removal of a cryptic intron and subcellular localization of green fluorescent protein are required to mark transgenic *Arabidopsis* plants brightly. *Proc Natl Acad Sci* **94:** 2122–2127.

Hausner G. 2012. Intron, mobile elements and plasmids. In *Organelle genetics: evolution of organelle genomes and gene expresssion* (ed. Bullerwell CE), pp. 329–358. Springer, Berlin.

Kawashima T, Douglass S, Gabunilas J, Pellegrini M, Chanfreau GF. 2014. Widespread use of non-productive alternative splice sites in *Saccharomyces cerevisiae*. *PLoS Genet* **10:** e1004249.

Klassen R, Meinhardt F. 2007. Linear protein-primed replicating plasmids in eukaryotic microbes. *Microbiol Monogr* **7:** 187–226.

Lynch M, Kewalramani A. 2003. Messenger RNA surveillance and the evolutionary proliferation of introns. *Mol Biol Evol* **20:** 563–571.

Pickrell JK, Pai AA, Gilad Y, Pritchard JK. 2010. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet* **6:** e1001236.

Saville BJ, Collins RA. 1990. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell* **61:** 685–696.

Saville BJ, Collins RA. 1991. RNA-mediated ligation of self-cleavage products of a *Neurospora* mitochondrial plasmid transcript. *Proc Natl Acad Sci* **88:** 8826–8830.

Schwartz SH, Silva J, Burstein D, Pupko T, Eyras E, Ast G. 2008. Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res* **18:** 88–103.

Semlow DR, Staley JP. 2012. Staying on message: ensuring fidelity in pre-mRNA splicing. *Trends Biochem Sci* **37:** 263–273.

Sood VD, Collins RA. 2002. Identification of the catalytic subdomain of the VS ribozyme and evidence for remarkable sequence tolerance in the active site loop. *J Mol Biol* **320:** 443–454.

Stajich JE, Harris T, Brunk BP, Brestelli J, Fischer S, Harb OS, Kissinger JC, Li W, Nayak V, Pinney DF, et al. 2012. FungiDB: an integrated functional genomics database for fungi. *Nucleic Acids Res* **40:** D675–D681.

Wahl MC, Will CL, Lührmann R. 2009. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136:** 701–718.

Zhou HL, Luo GB, Wise JA, Lou H. 2014. Regulation of alternative splicing by local histone modifications: potential roles for RNA-guided mechanisms. *Nucleic Acids Res* **42:** 701–713.