Published online 8 June 2017

Nucleic Acids Research, 2017, Vol. 45, No. 15 ***9085–9092***
doi: 10.1093/nar/gkx520

A mechanism for a single nucleotide intron shift

Erzsébet Fekete[1,*], Michel Flipphi[1], Norbert Ág[1], Napsugár Kavalecz[1], Gustavo Cerqueira[2], Claudio Scazzocchio[3,4] and Levente Karaffa[1]

[1]Department of Biochemical Engineering, University of Debrecen, 4032, Hungary, [2]Broad Institute of MIT & Harvard, Cambridge, 02141 MA, USA, [3]Department of Microbiology, Imperial College London, SW7 2AZ, UK and [4]Institut de Biologie Intégrative de la Cellule, CEA/CNRS/Université Paris-Saclay UMR 9198, 91405 Gif-sur-Yvette, France

Received February 23, 2017; Revised May 29, 2017; Editorial Decision June 01, 2017; Accepted June 01, 2017

ABSTRACT

Spliceosomal introns can occupy nearby rather than identical positions in orthologous genes (intron sliding or shifting). Stwintrons are complex intervening sequences, where an 'internal' intron interrupts one of the sequences essential for splicing, generating after its excision, a newly formed canonical intron defined as 'external'. In one experimentally demonstrated configuration, two alternatively excised internal introns, overlapping by one G, disrupt respectively the donor and the acceptor sequence of an external intron, leading to mRNAs encoding identical proteins. In a gene encoding a DHA1 antiporter in Pezizomycotina, we find a variety of predicted intron configurations interrupting the DNA stretch encoding a conserved peptidic sequence. Some sport a stwintron where the internal intron interrupts the donor of the external intron (experimentally confirmed for *Aspergillus nidulans*). In others, we found and demonstrate (for *Trichoderma reesei)* alternative, overlapping internal introns. Discordant canonical introns, one nt apart, are present in yet other species, exactly as predicted by the alternative loss of either of the internal introns at the DNA level from an alternatively spliced stwintron. An evolutionary pathway of 1 nt intron shift, involving an alternatively spliced stwintron intermediate is proposed on the basis of the experimental and genomic data presented.

INTRODUCTION

Introns are sequences that interrupt open reading frames (ORFs) in RNA. Spliceosomal introns are exclusive of eukaryotic nuclear gene transcripts and they necessitate a complex excision apparatus composed of small nuclear RNAs (snRNAs) and proteins (for a review, see 1). The snRNAs of the major (U2-type) spliceosome interact with three short, conserved sequences within a U2 intron, the 5′ donor site (starting with 5′-GU or occasionally, 5′-GC), the lariat branch point sequence (containing the branch point adenosine, usually near the 3′ end of the intron) and the 3′ acceptor site (ending with 5′-AG) to proceed with the excision reactions (for a recent review, 2). Intron–exon structure is dynamic (3). There are quite a few instances where intron positions within a primary transcript are very close in different organisms, but not identical. The phenomenon giving rise to such discordant introns has been called 'intron sliding', 'intron drift' or 'intron shift'. We have chosen to use the term 'intron shift' throughout this article, which best indicates a process leading to different, adjacent or near adjacent positioning of introns in orthologous genes. This naming implies an ancestral identity for these introns, followed by a subsequent shift of position in the transcript. [There is a degree of arbitrariness here, as 'intron sliding' has been used to indicate differences in position of one nucleotide to as many as 15 nt in the precursor RNA; (4)]. Intron identity by sequence conservation can only be expected in orthologous genes in very closely related organisms, which rarely feature discordant intron positions (5). Thus it is difficult to determine whether a genuine intron shift has occurred rather than a process of intron loss followed by the insertion of a new intron at a nearby position (6). The existence of discordant introns has been reported for individual genes and in families of paralog genes (for recent work see, e.g. 7,8). The fact that significant sequence identity can be found among certain discordant introns (e.g. 9), means that intron shift can occur. Statistical studies indicate that intron shift by 1 nt is a genuine process (4).

The term 'twin intron' (abbreviated 'twintron') has originally been applied to 15 complex group II/III introns in the chloroplast genome of *Euglena gracilis* (10,11). Excision of an 'internal' intron is required for the subsequent excision of an 'external' intron as the former interrupts a sequence essential for the removal of the latter. This site of integration of the internal intron leads to evolutionary stability of the twintron arrangement—the internal intron must remain functional to permit the excision of the external intron. Other instances of complex intervening sequences, including group I introns carrying two ribozymes (12) and a rare form of alternative splicing of an intron in the *prospero*

*To whom correspondence should be addressed. Tel: +36 52 512 900 (Ext. 62488); Fax: +36 52 512 728; Email: kicsizsoka@yahoo.com

transcript of *Drosophila melanogaster* (13), have also been called 'twintrons'. We have discussed in a previous publication the differences between these processes and genuine twintrons (14), and they are also conveniently reviewed and categorized by Hafez and Hausner (15).

Previously, we identified and characterized in fungal nuclear transcripts, four instances of twintrons conceptually analogous to the euglenoid group II/III twintrons that we therefore named spliceosomal twin introns (stwintrons) (14,16). We differentiated stwintron types accordingly to where the insertion of the internal intron occurred—'D', indicating insertions in the donor sequence, and 'A', in the acceptor—followed by the numbers of the nt in the consensus sequence between which the internal intron is present. We have hitherto identified [D1,2] and [D2,3] stwintrons (i.e. internal intron situated between the first and the second nt and the second and the third nt of the donor of the external intron, respectively) and an [A2,3] stwintron (i.e. internal intron separating the acceptor of the external intron between its second and third nt) (see Figure 1A for the stwintron types relevant to this work). Each of the above stwintrons can only be accurately excised with two consecutive U2 reactions requiring 5′- and 3′-splice sites to pair via the intron definition mechanism twice. In fact, the existence of stwintrons provided the evidence that intron definition (cf. 17) applies in filamentous ascomycota (Pezizomycotina). The stwintron arrangement fits the strict definition of a twintron which differentiates it from other known complex intervening sequences of the U2 type, like recursive splicing, the nested intron arrangement or intrasplicing, where no 5′- or 3′ splice site element of one constituent intron is interrupted by another (15).

For one particular [D1,2] stwintron in *Helminthosporium solani*, we demonstrated a complex intronic structure that could undergo two alternative pathways of sequential splicing as it overlaps almost completely with an [A2,3] stwintron (16) (Figure 1A). These two alternative paths will automatically result in two different positions of the resulting canonical 'external' introns, which are shifted by 1 nt with respect to one another but which do not lead to a change in the mature mRNA's ORF. The existence of this particular type of stwintron, the alternative [D1,2]/[A2,3] stwintron, immediately suggests a mechanism for a 1-nt intron shift. Alternative deletion of one of the two internal introns at the DNA level, mimicking each of the two possible splicing paths, would result in two different canonical U2 introns at positions apparently shifted by one nt (Figure 1B). In this paper we show that this is exactly what has happened to an intervening sequence in a gene encoding a transmembrane protein of the DHA1 family in several fungi.

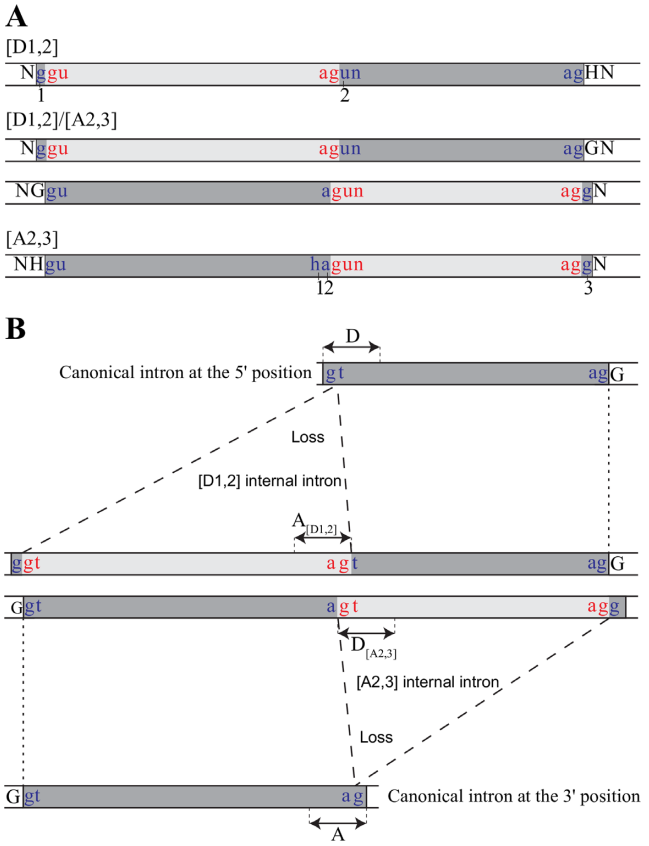

Figure 1. A mechanism for a 1-nt intron shift involving the ancestral existence of an alternatively spliced [D1,2]/[A2,3] stwintron intermediate. (**A**) Structure and nomenclature of stwintrons relevant to this study. Intronic nt are in lower case letter: nt of the internal intron (light gray) in red, and nt of the external intron (dark gray) in blue lettering. The conserved terminal sequences essential for spliceosomal intron excision are indicated: D, donor sequence (gt-) at the 5′ splice site, and A, acceptor sequence (-ag) at the 3′ splice site. In a [D1,2] stwintron, an internal U2 intron interrupts the donor element of an external U2 intron between the first and the second nt. In a [A2,3] stwintron, the internal intron is nested in the (3-nt) acceptor element of the external intron between the second and the third nt. The G upstream the most 5′ donor is a characteristic feature of the [D1,2] stwintron but is exonic for the [A2,3] stwintron; The G downstream the most 3′ acceptor is a characteristic feature of the [A2,3] stwintron but is exonic for the [D1,2] stwintron. (**B**) A mechanism for the evolution of neighboring discordant intron positions. An alternatively spliced [D1,2]/[A2,3] stwintron occurs. When subsequently its [A2,3] internal intron is lost, the remaining canonical intron is localized at the 3′ of neighboring, alternative exon fusion sites utilized by the ancestor [D1,2]/[A2,3] stwintron. However, if its [D1,2] internal intron is lost, the *de novo* canonical intron is localized at the 5′ of those exon fusion sites. The two U2 introns resulting from mutually exclusive internal intron loss events at the DNA level are shown respectively at the top and the bottom, the two intron positions being called 5′ and 3′.

MATERIALS AND METHODS

Fungal strains

Aspergillus nidulans R21 (ATCC 48756) and *Trichoderma reesei* QM9414 (ATCC 26921) were used in this study. Their mantainence and standardized growth medium compositions are described elsewhere (18,19).

Nucleic acid isolation

Aspergillus nidulans and *T. reesei* were grown in 500-ml Erlenmeyer flasks with 100 ml of medium containing 3.0% malt extract (LAB M Limited, UK) seeded with vegetative spore inoculums, in a rotary shaker (Infors HT Multitron) at 200 rotations per min for 24 h. Mycelia were harvested by filtration over sterile cheese cloth. The biomass was washed with distilled water, frozen and ground to powder under liquid nitrogen. For the extraction of genomic DNA and total RNA, Macherey-Nagel NucleoSpin kits (NucleoSpin Plant

II and NucleoSpin RNA Plant, respectively) were used according to the manufacturer's instructions.

Reverse transcription PCR (RT-PCR)

Reverse transcription was primed off 1 μg of total RNA with Oligo(dT) as a primer using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche). Polymerase chain reaction (PCR) reactions were performed in a 25 μl volume containing 4 μl of single strand cDNA, using gene-specific oligonucleotides (Supplementary Table S1) as primers and DreamTaq DNA Polymerase (Thermo Scientific). Cycling conditions after initial denaturation at 95°C for 2 min were: 40 cycles of 95°C for 30 s, 60°C for 1 min and 72°C for 1 min, followed by one post-cyclic elongation at 72°C for 5 min. Amplified fragments were resolved in native agarose gels.

To confirm the existence of the predicted stwintron splicing intermediates, we used the same RT-PCR approach as previously (16) with primer pairs that do not amplify DNA off mature mRNA template. This strategy usually yields two fragments of defined sizes of which the smaller one corresponds to the splicing intermediate and the bigger one, to primary transcript. We processed and sequenced both fragments amplified by RT-PCR. All experiments were done in duplicate, starting with biomass from two independent liquid cultures. Furthermore, we verified the alternative splicing pathways of the *T. reesei* [D1,2]/[A2,3] stwintron in a single RT-PCR reaction using a reverse primer hybridizing to sequences within the cDNA confirmed, canonical U2 intron, 329 nt downstream of the stwintron (see Supplementary Tables S1 and S2).

cDNA sequencing

Double strand cDNA was gel-purified (NucleoSpin Gel & PCR Clean-up) and subsequently cloned (pGEM-T Easy Vector System I, Promega). Plasmid DNA was isolated using the NucleoSpin Plasmid EasyPure kit (Macherey-Nagel). Three independent clones were sequenced over both strands using universal primers hybridizing to the vector (MWG-Biotech AG, Ebersberg, Germany). cDNA sequences were deposited at GenBank under accession numbers KY315812–KY315816.

Tool to search putative stwintrons in whole genome sequences

We defined five degenerated sequence motifs for the donor-, acceptor- and lariat branch point sequence elements within a stwintron, including the two hybrid motifs characteristic for the stwintron type, consistent of nt of the external- as well as of the internal intron. These motifs are based on a statistical consensus for the three conserved elements in *A. nidulans* (20) although we used a more relaxed consensus at the first position of the 6-nt element around the lariat branch point A (D instead of R) and the first position of the 3-nt acceptor (H instead of Y). Furthermore, we defined distance ranges separating these five motifs conforming three principles rooted in our experience in calling intron–exon structure: (i) the minimum length of an *A. nidulans* intron is 42 nt; (ii) The minimum distance between the 6-nt lariat branch point element and the 3-nt acceptor element is 4 nt; and (iii), the distance between the 6-nt donor element at 5′ and the lariat branch point element is always bigger than the distance between the latter and the acceptor at 3′, although usually, the latter distance is considerably smaller than the former. We set the distance range between the donor- and lariat branch point elements from 25 to 120 nt. The mean intron length is reported to be 73 nt in *A. nidulans* (20). For a screen for putative [D1,2] stwintrons, these criteria led to a sequence pattern

GGTRWGYN(25,120)DYTRAYN(4,24)HAGTRWG YN(25,120)DYTRAYN(4,24)HAG which was searched for in whole genome sequences using the Fuzznuc application (21). A screen of the *A. nidulans* FGSC A4 genome (accession: AACD01000000) yielded dozens of candidate [D1,2] stwintrons. These candidates were each manually curated before being subjected to expirimental verification, taking into account several criteria, like: (i) whether the sequence separates (known or putative) coding sequences in *A. nidulans*; (ii) Whether there are expression data available confirming (or dismissing) its existence—principally, by using the *A. nidulans* RNASeq data accessible at the Aspergillus Genome Database (AspGD) (22); [3] Whether the stwintron is present or not in any ortholog gene in other species of *Aspergillus*, primarily, those whose genome sequences are accessible at AspGD. We note that autoannotation is, in the case of [D1,2] stwintrons, always incorrect as the software is not trained to recognize 5′-GGT at the 5′ splice site of a complex intervening sequence. Note that this tool was not designed to discover stwintrons implicated in the new mechanism of intron shift.

Genome mining and phylogenetic analysis

Ascomycete genome sequences accessible at the National Centre of Biotechnology Information (NCBI) and the US Department of Energy Joint Genome Institute (JGI) websites were screened with TBLASTN (23) for genes encoding proteins that are >45% identical to the DHA1 protein specified by the *A. nidulans* stwintron-containing gene at locus AN3270. The correct query sequence had been deduced from full-length cDNA (GenBank Accession KY315812). For each of the more than 500 genes found, the intron–exon structure was manually determined and full-length proteins were subsequently deduced.

For phylogenetic analysis, 256 Pezizomycotina DHA1 proteins were aligned with MAFFT (version 7), using the E-INS-i algorithm and a BLOSUM 45 similarity matrix (24). The alignment was first curated to 459 informative residues with BMGE version 1.12 using BLOSUM 35 and a block size of 3 (25). A Maximum Likelihood tree was then calculated by PhyML 3.0 using the WAG substitution model (26). Approximate Likelihood Ratio Tests provide statistical branch support (27) and were integrally calculated by PhyML with the Chi2-based parametric provided. The tree was visualized with the FigTree drawing program (http://tree.bio.ed.ac.uk/software/figtree) and annotated with Adobe Illustrator. Figure 2 shows the unambiguous clade of the 89 structurally most related proteins taken from the original tree with 256 proteins.

Figure 2. Phylogenetic relations amongst the 89 Pezizomycotina DHA1 proteins most closely related to *Aspergillus nidulans* DhaoS. The figure shows a well-defined clade extracted from much larger Maximum Likelihood tree of Pezizomycotina DHA1 proteins that are at least 45% identical to DhaoS

RESULTS

A stwintron-containing gene in *Aspergillus nidulans*

We devised an informatics tool (described in 'Materials and Methods' section) that enabled us to identify *bona fide* stwintrons in the model organism *A. nidulans* (e.g. 28). We detected a putative [D1,2] stwintron at genome locus AN3270 (AspGD annotation). The gene encodes a membrane protein that belongs to the drug:H+ antiporter family 1 (DHA1) (29). We named it *dhaoS* (drug:H+ antiporter family one gene with stwintron).

We predicted that the primary transcript comprises one complex intervening sequence of 113 nt and confirmed its excision upon sequencing of full-length cDNA specifying a reading frame of 1611 nt (GenBank Accession number KY315812). The predicted protein product is 536 residues long and the stwintron splits the Leu108 UUA codon after the second U. The sequence motif WGPLSELY (amino acids 105–112) is conserved in many fungal DHA1 proteins, often encoded by intronless genes. The exon fusion site confirmed by KY315812 preserves this sequence, at variance with the spurious intron proposed by automated annotation. The predicted *dhaoS* [D1,2] stwintron consists of a 56-nt-long external U2 intron with its 5′ donor sequence split between the first and the second nt by a 57-nt-long internal U2 intron (Supplementary Figure S1), similar to those characterized previously (14,16). An RT-PCR using a forward oligonucleotide primer whose sequence is fully exonic and a reverse primer that terminally overlaps the 3′ distal junction with the predicted external intron (see 'Materials and Methods' section for details), allowed to amplify the splicing intermediate. The shorter amplified fragment was shown to lack the predicted internal intron with the 5′-donor of the external intron reconstructed (GenBank KY315814) whereas the larger amplified fragment was identical to the primary transcript (not shown).

Identification of the orthologs of the stwintron-containing gene

We identified by TBLASTN fungal genes that encode DHA1 proteins at least 45% identical to *A. nidulans* DhaoS (results not shown). We found hundreds of such genes as most Pezizomycotina genomes encode multiple DHA1s. A maximum likelihood tree of DHA1 homologous proteins (of >45% identity) featured one clearly defined clade comprising *A. nidulans* DhaoS and 72 proteins from other species of Pezizomycotina. The 73 clustered proteins are >64% identical at the amino acid level. 16 proteins, which show 56–62% identity with *A. nidulans* DhaoS appear as an outgroup of the putative orthologous clade. Figure 2 shows the evolutionary relationships between these 89 proteins (see Supplementary Table S2 for the cognate genes).

The genes encoding the 73 orthologous proteins all carry intervening sequences at or immediately 3′ to the position of the stwintron in the *A. nidulans dhaoS* gene (coordinates listed in Supplementary Table S2). 21 of these genes comprise a second intron, 329 bp downstream of the first intron position. With one exception amongst the hunderds of other DHA1 genes collected, the 5′ intron position appears specific to the 73 clustered *dhaoS* orthologs. The genes encoding the 16 proteins in the (56–62% identity) paralog branch are all intronless.

In a [D1,2] stwintron, the acceptor sequence of the internal intron, necessarily overlaps with the donor of the external intron. If both flanking exonic positions on each side of the two terminally overlapping U2 introns in the primary transcript are occupied by Gs, a [D1,2] stwintron is also a [A2,3] stwintron (Figure 1A). Amongst the 73 orthologous DHA1 genes, we identified 36 where the stwintron is strictly [D1,2], while 28 others carry a stwintron that can be alternatively spliced (Figure 2 and Supplementary Table S2). In all cases, this coincides with a single A-G transition in the first nt of the exon downstream the [D1,2] stwintron, consistent with a close evolutionary relation between these two (stw)intron organization modes.

The [D1,2]/[A2,3] stwintron in the *Trichoderma reesei* ortholog DHA1 gene

The *Trichoderma reesei* ortholog *dhaoS* gene (misannotated locus TRIREDRAFT_43701) is one of those that feature an alternatively spliced [D1,2]/[A2,3] stwintron, similar to that we characterized in the *H. solani* gene for alternative oxidase (16).

We cloned the cognate cDNA (GenBank KY315813) to confirm a 1548-bp long ORF encoding a protein of 515 residues and confirming the presence of two intervening sequences in the gene. The 5′ of those is the putative [D1,2]/[A2,3] stwintron that splits the Leu125 CUG codon after the second nt (when removed as a [D1,2]). This 154-nt long sequence comprises two alternative internal introns which overlap by one G. The 5′ U2 intron is 66 nt long and is preceded by an exonic G—as expected for a [D1,2] stwintron—while that at 3′ is 88 nt long and followed by an exonic G—consistent with an [A2,3] stwintron (Supplementary Figure S2). We specifically amplified and sequenced cDNAs corresponding to the two splicing intermediates, confirming that the *T. reesei* ortholog primary transcript harbors an alternatively spliced [D1,2]/[A2,3] stwintron (see GenBank accessions KY315815 for the [D1,2] splicing intermediate and KY315816 for the [A2,3] splicing intermediate).

There are more than a hundred expressed sequence tags (ESTs) corresponding to this *T. reesei* gene deposited at NCBI's EST database. Most confirm the

←

(see 'Results' section). Approximate Likelihood Ratio Test values are given at each node. Proteins encoded by 36 orthologous genes that contain a strict [D1,2] stwintron are tagged _D and shown in black lettering. Proteins encoded by 28 orthologous genes that harbor an alternatively spliced, [D1,2]/[A2,3] stwintron at the same position are tagged _A and shown in blue lettering. The proteins encoded by nine genes that harbor a canonical phase-2 intron are tagged _C5 and printed in red lettering while those encoded by the two orthologous genes that carry a phase-0 intron are tagged _C3 and printed in green lettering. The 16 DHA1 proteins in the outgroup (labeled _P and shown in gray lettering) are encoded by intron-less 'nearest paralog' genes; in *Oidiodendron maius*, this gene is duplicated (_Pa and _Pb, respectively). The 21 proteins in bold lettering are encoded by orthologous genes that comprise an additional standard intron, 329 nt downstream of the [D1,2] stwintron position.

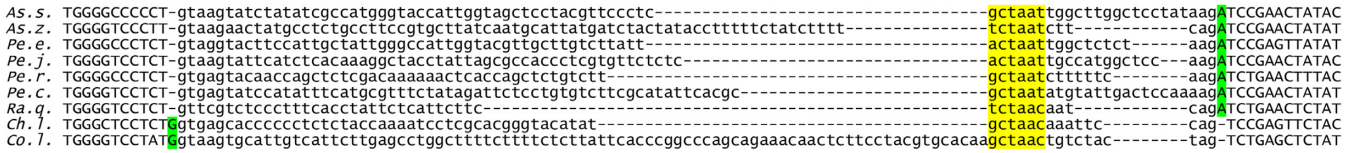

Figure 3. Relevant sequences of the DHA1 genes orthologous to *Aspergillus nidulans dhaoS* that harbor discordant canonical introns rather than a stwintron. The discordant introns interrupt the DNA encoding the conserved amino acid sequence WGPLSELY. Coding nt are in capitals. Species abbreviations: *As.s, Aspergillus sclerotiorum*; *As.z*, *Aspergillus zonatus*; *Pe.e*, *Penicillium expansum*; *Pe.j*, *Penicillium janthinellum*; *Pe.r*, *Penicillium raistrickii*; *Pe.c*, *Penicillium citrinum*; *Ra.q*, *Raffaelea quercivora*; *Ch.l*, *Chalara longipes*; *Co.l*, *Coniochaeta ligniaria*. The corresponding sequence in *Penicillium aurantiogriseum* NRRL 62431 (accession ALJY00000000) is identical to that in *P. expansum* T01 (AYHP00000000). The corresponding sequence in *Hymenoscyphus varicosporoides* CBS 651.66 (LLCF00000000) is identical to that in *P. citrinum* DSM 1997 (LKUP00000000). The 1-nt intron shift is indicated by highlighting in green the third nt of the codon for the first Leu in WGPLSELY. The conserved sequence element around the putative lariat branch point A (6 nt) is highlighted in yellow.

sequence of the mature mRNA as predicted by us. However, two ESTs—accession numbers CB901225 and CF871013—feature the predicted sequence of the [A2,3] splicing intermediate. This could suggest that the *T. reesei* spliceosome would excise the alternative stwintron with a preference for the [A2,3] path. We confirmed the existence of both splicing intermediates by one RT-PCR, using a reverse primer that hybridizes to sequences within the other intron, 329 nt downstream of the stwintron. Using the same template RNA as in our 'pathway-specific' experiments, the majority of the cloned and sequenced RT-PCR products corresponded to the splicing intermediate of the [D1,2] stwintron, but the three other expected amplifications—corresponding to the primary transcript, the [A2,3] splicing intermediate or the RNA from which the complete stwintron was absent—were also found. This implies that under our experimental growth conditions at the timepoint of biomass collection, the [D1,2] path is preferred. It would thus appear that alternative splicing of the [D1,2]/[A2,3] stwintron from the *T.reesei* DHA1 ortholog transcript is a dynamic process.

Alternative loss of either internal intron of the [D1,2]/[A2,3] stwintron leads to discordant canonical introns in orthologous genes

From the identified orthologous DHA1 genes, 11 harbor a canonical intron rather than a stwintron at the corresponding position(s). Our phylogeny of the 73 DhaoS ortholog proteins (Figure 2) strongly suggested that the encoding genes featuring canonical introns, have actually lost an internal intron from a pre-extant stwintron. The sequences of these U2 introns and their immediate exonic context are shown in Figure 3. In two species, *Chalara longipes* and *Coniochaeta ligniaria*, a standard intron separates the codons for Leu or Met, respectively, from Ser in the downstream exon. Their phase-0 intron is consistent with the loss of the [A2,3] internal intron from an alternative [D1,2]/[A2,3] stwintron. In the other genomes, the retained intron splits the Leu codon between the second and third nt. This phase-2 intron is consistent with the loss of the [D1,2] internal intron. Accurate intron loss must have occurred at the genome level, most likely via a reverse transcriptase-mediated mechanism (cf. 30,31), as these mutually exclusive events strictly mimick the excision from one of the constituent internal introns of an alternatively spliced [D1,2]/[A2,3] stwintron in the primary transcript (Figure 1).

DISCUSSION

The apparent shift of an intron position along an ORF—*without* losing or gaining exonic information—cannot be explained with one modification or mutation in the genome but likely involves temporary spaced, compensatory events at either side of the intron (6,32). A mechanism involving the movement of one intron–exon junction before the other seems unlikely to mediate shifts over distances other than codon equivalents. There are instances recorded in which one intron–exon junction has slid—usually due to the utilization of another splice site—changing the length of both intron and exon (see, e.g. 33,34). A mechanism of compensatory indel mutations near either end of the intron (8) would yield a similar intermediate but it has been observed in comparative studies, that alignment gaps near intron–exon junctions are often spurious, largely down to mis-annotation (35). Nevertheless, by including the flexibility provided by alternative splicing into the model, the contribution of intron shift to the proliferation of intron positions may be more substantial than previously thought (36–38). Concrete intermediates have never been identified for shift facilitated by alternative splicing of the primordial intron, due to the low abundance of spliced transcripts detrimental to gene function. An alternative mechanism featuring intron excision, reverse splicing, reverse transcription and homologous recombination (first proposed by Martinez *et al.* 39) is even more complex as it involves four events that necessarily occur in a very short time frame, as the very intron excised has to be 'reverse spliced' into the same RNA before reverse transcription, albeit at a nearby position.

In the current study, we elucidated a novel mechanism of 1-nt intron shift in eukaryotic nuclear transcripts with the alternatively spliced, [D1,2]/[A2,3] stwintron as the key intermediate (Figure 1B). This work demonstrates the relevance of stwintrons for intron evolution. To the best of our knowledge, this is the first instance at which an intermediary stage of an intron shift process has been characterized. This is facilitated by the evolutionary stability of the [D1,2] and [A2,3] stwintron structures (cf. 14) and their smooth interconversion. In contrast to sequential compensatory movement of intron–exon junctions, the stwintron-mediated mechanism allows for a 1-nt intron shift to occur in a temporary phased manner without changing the messenger encoding the gene's product at any intermediary stage. It allows for a stable intermediate phase before in-

tron loss at the DNA level while the key steps of the reverse splicing and transcription mechanism (39) have to proceed immediately as the intron shift effectively takes place at the transcript level before reverse transcription. However, it should be emphasized that the new mechanism via the alternatively spliced stwintron only mediates shifts to neighboring positions and cannot result in movement over larger distances, like e.g. the codon-step sized shifts apparent in some *Drosophila* genes where the introns involved (still) exhibit considerable similarity (8). Nevertheless, discordant introns can always arise after an intron loss event and a subsequent, independent intron gain event, where the two introns are unrelated and the transient 'intermediate' situation is intronless (6).

The present work, which originated from the finding of a [D1,2] stwintron in *A. nidulans*, led us to demonstrate the existence of a phylogenetically related [D1,2]/[A2,3] stwintron and of ortholog genes carrying canonical introns at neighboring positions. The extant genomic data do not allow to decide which was the position of the primordial intron in the basal ortholog of the *dhaoS* gene. Neither can we know whether this hypothetical ancestral intron was a canonical phase-0 or phase-2 intron, a [D1,2], a [A2,3] or a [D1,2]/[A2,3] complex stwintron. Nevertheless, we could verify the prediction that the specific loss at the DNA level of alternatively spliced internal introns from a [D1,2]/[A2,3] stwintron (Figure 1B) is at the origin of discordant introns apparently resulting from a 1-nt intron shift.

ACCESSION NUMBERS

GenBank KY315812–KY315816.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

EU and co-financed by the European Regional Development Fund [GINOP-2.3.2–15-2016–00008]; Hungarian Scientific Research Fund [OTKA NN116519 to L.K.]; Bólyai János Research Scholarship [BO/00548/14 to L.K.]. Funding for open access charge: Hungarian Scientific Research Fund [OTKA NN116519 to L.K.].
Conflict of interest statement. None declared.

REFERENCES

1. Rino,J. and Carmo-Fonseca,M. (2009) The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends Cell Biol.*, **19**, 375–384.
2. Irimia,M. and Roy,S.W. (2014) Origin of spliceosomal introns and alternative splicing. *Cold Spring Harb. Perspect. Biol.*, **6**, a016071.
3. Carmel,L., Wolf,Y.I., Rogozin,I.B. and Koonin,E.V. (2007) Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome Res.*, **17**, 1034–1044.
4. Rogozin,I.B., Lyons-Weiler,J. and Koonin,E.V. (2000) Intron sliding in conserved gene families. *Trends Genet.*, **16**, 430–432.
5. Henricson,A., Forslund,K. and Sonnhammer,E.L. (2010) Orthology confers intron position conservation. *BMC Genomics*, **11**, 412.
6. Stoltzfus,A., Logsdon,J.M., Palmer,J.D. and Doolittle,W.F. (1997) Intron "sliding" and the diversity of intron positions. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 10739–10744.
7. Kyndt,T., Haegeman,A. and Gheysen,G. (2008) Evolution of GHF5 endoglucanase gene structure in plant-parasitic nematodes: no evidence for an early domain shuffling event. *BMC Evol. Biol.*, **8**, 305.
8. Lehmann,J., Eisenhardt,C., Stadler,P.F. and Krauss,V. (2010) Some novel intron positions in conserved Drosophila genes are caused by intron sliding or tandem duplication. *BMC Evol. Biol.*, **10**, 156.
9. Sakharkar,M.K., Tan,T.W. and de Souza,S.J. (2001) Generation of a database containing discordant intron positions in eukaryotic genes (MIDB). *Bioinformatics*, **17**, 671–675.
10. Copertino,D.W. and Hallick,R.B. (1991) Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene. *EMBO J.*, **10**, 433–442.
11. Copertino,D.W. and Hallick,R.B. (1993) Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends Biochem. Sci.*, **18**, 467–471.
12. Nielsen,H. and Johansen,S.D. (2009) Group I introns: moving in new directions. *RNA Biol.*, **6**, 375–383.
13. Borah,S., Wong,A.C. and Steitz,J.A. (2009) *Drosophila* hnRNP A1 homologs Hrp36/Hrp38 enhance U2-type versus U12-type splicing to regulate alternative splicing of the *prospero* twintron. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 2577–2582.
14. Flipphi,M., Fekete,E., Ág,N., Scazzocchio,C. and Karaffa,L. (2013) Spliceosome twin introns in fungal nuclear transcripts. *Fungal Genet. Biol.*, **57**, 48–57.
15. Hafez,M. and Hausner,G. (2015) Convergent evolution of twintron-like configurations: one is never enough. *RNA Biol.*, **12**, 1275–1288.
16. Ág,N., Flipphi,M., Karaffa,L., Scazzocchio,C. and Fekete,E. (2015) Alternatively spliced, spliceosomal twin introns in *Helminthosporiumsolani*. *Fungal Genet. Biol.*, **85**, 7–13.
17. Berget,S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–2414.
18. Pontecorvo,G., Roper,J.A., Hemmons,L.M., MacDonald,K.D. and Bufton,A.W.J. (1953) The genetics of *Aspergillusnidulans*. *Adv. Genet.*, **5**, 141–238.
19. Mandels,M. and Andreotti,R.E. (1978) Problems and challenges in the cellulose to cellulase fermentation. *Process Biochem.*, **13**, 6–13.
20. Kupfer,D.M., Drabenstot,S.D., Buchanan,K.L., Lai,H., Zhu,H., Dyer,D.W., Roe,B.A. and Murphy,J.W. (2004) Introns and splicing elements of five diverse fungi. *Eukaryot. Cell*, **3**, 1088–1100.
21. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, **16**, 276–277.
22. Cerqueira,G.C., Arnaud,M.B., Inglis,D.O., Skrzypek,M.S., Binkley,G., Simison,M., Miyasato,S.R., Binkley,J., Orvis,J., Shah,P. *et al.* (2014) The Aspergillus Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. *Nucleic Acids Res.*, **42**, D705–D710.
23. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
24. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
25. Criscuolo,A. and Gribaldo,S. (2010) BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.*, **10**, 210.
26. Guindon,S., Dufayard,J.F., Lefort,V., Anisimova,M., Hordijk,W. and Gascuel,O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.
27. Anisimova,M. and Gascuel,O. (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.*, **55**, 539–552.
28. Casselton,L. and Zolan,M. (2002) The art and design of genetic screens: filamentous fungi. *Nat. Rev. Genet.*, **3**, 683–697.
29. Pao,S.S., Paulsen,I.T. and Saier,M.H. Jr. (1998) Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.*, **62**, 1–34.
30. Fink,G.R. (1987) Pseudogenes in yeast? *Cell*, **49**, 5–6.
31. Roy,S.W. and Gilbert,W. (2006) The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat. Rev. Genet.*, **7**, 211–221.
32. Lynch,M. (2002) Intron evolution as a population-genetic process. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 6118–6123.

33. Higashimoto,Y. and Liddle,R.A. (1993) Isolation and characterization of the gene encoding rat glucose-dependent insulinotropic peptide. *Biochem. Biophys. Res. Commun.*, **193**, 182–190.
34. Schäfer,U.A., Reed,D.W., Hunter,D.G., Yao,K., Weninger,A.M., Tsang,E.W.T., Reaney,M.J.T., MacKenzie,S.L. and Covello,P.S. (1999) An example of intron junctional sliding in the gene families encoding squalene monooxygenase homologues in *Arabidopsis thaliana* and *Brassica napus*. *Plant Mol. Biol.*, **39**, 721–728.
35. Sêton Bocco,S. and Csűrös,M. (2016) Splice sites seldom slide: intron evolution in öomycetes. *Genome Biol. Evol.*, **8**, 2340–2350.
36. Séraphin,B. and Rosbash,M. (1990) Exon mutations uncouple 5′ splice site selection from U1 snRNA pairing. *Cell*, **63**, 619–629.
37. Brackenridge,S., Wilkie,A.O.M. and Screaton,G.R. (2003) Efficient use of a 'dead end' GA 5′ splice site in the human fibroblast growth factor receptor genes. *EMBO J.*, **22**, 1620–1631.
38. Tarrío,R., Ayala,F.J. and Rodríguez-Trelles,F. (2008) Alternative splicing: a missing piece in the puzzle of intron gain. *Proc. Natl. Acad. Sci. U.S.A.*, **105**, 7223–7228.
39. Martinez,P., Martin,W. and Cerff,R. (1989) Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. *J. Mol. Biol.*, **208**, 551–565.