AUTHOR QUERY FORM

ELSEVIER	**Journal: YFGBI** **Article Number: 2569**	**Please e-mail or fax your responses and any corrections to:** **E-mail: corrections.essd@elsevier.sps.co.in** **Fax: +31 2048 52799**

Dear Author,

Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list. Note: if you opt to annotate the file with software other than Adobe Reader then please also highlight the appropriate place in the PDF file. To ensure fast publication of your paper please return your corrections within 48 hours.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Click on the 'Q' link to go to the location in the proof.

Location in article	Query / Remark: click on the Q link to go Please insert your reply or correction at the corresponding line in the proof
Q1	Please confirm that given names and surnames have been identified correctly.
Q2	Reference 'Patel and Steitz (2004)' is cited in the text but not listed. Please check.
Q3	This section comprises references that occur in the reference list but not in the body of the text. Please position each reference in the text or, alternatively, delete it. Any reference not dealt with will be retained in this section.
Q4	Please provide captions for supplementary material.

Please check this box if you have no corrections to make to the PDF file ☐

Thank you for your assistance.

Highlights

• First characterisation of twin introns in fungal nuclear transcripts. • Internal intron disrupts donor of pre-extant external intron. • Excision of external intron can only occur after that of internal intron.

Fungal Genetics and Biology xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Fungal Genetics and Biology

journal homepage: www.elsevier.com/locate/yfgbi

Spliceosome twin introns in fungal nuclear transcripts

Michel Flipphi[a,*], Erzsébet Fekete[a], Norbert Ág[a], Claudio Scazzocchio[b,c], Levente Karaffa[a]

[a] Department of Biochemical Engineering, Faculty of Science and Technology, University of Debrecen, 4010 Debrecen, Hungary
[b] Department of Microbiology, Imperial College, London SW7 2AZ, United Kingdom
[c] Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris-Sud, 91405 Orsay, France

ARTICLE INFO

Article history:
Received 23 March 2013
Accepted 9 June 2013
Available online xxxx

Keywords:
Nuclear transcripts
Spliceosomal twin introns
Sequential splicing
Intron late theory
Stwintronisation
Filamentous fungi

ABSTRACT

The spliceosome is an RNA/protein complex, responsible for intron excision from eukaryotic nuclear transcripts. In bacteria, mitochondria and plastids, intron excision does not involve the spliceosome, but occurs through mechanisms dependent on intron RNA secondary and tertiary structure. For group II/III chloroplast introns, "twintrons" (introns within introns) have been described. The excision of the external intron, and thus proper RNA maturation, necessitates prior removal of the internal intron, which interrupts crucial sequences of the former. We have here predicted analogous instances of spliceosomal twintrons ("stwintrons") in filamentous fungi. In two specific cases, where the internal intron interrupts the donor of the external intron after the first or after the second nucleotide, respectively, we show that intermediates with the sequence predicted by the "stwintron" hypothesis, are produced in the splicing process. This implies that two successive rounds of RNA scanning by the spliceosome are necessary to produce the mature mRNA. The phylogenetic distributions of the stwintrons we have identified suggest that they derive from "late" events, subsequent to the appearance of the host intron. They may well not be limited to fungal nuclear transcripts, and their generation and eventual disappearance in the evolutionary process are relevant to hypotheses of intron origin and alternative splicing.

1. Introduction

The discovery of introns (Chow et al., 1977; Berget et al., 1977) resulted in what is arguably the major shift in our conceptualisation of the gene since the demonstration of the colinearity of the genetic message with the amino acid sequence of a given protein (Yanofsky et al., 1964; Sarabhai et al., 1964). Spliceosome introns are exclusive of eukaryotic nuclear genes and they necessitate a complex excision apparatus composed of small nuclear RNAs (snRNAs; U1, U2, U4, U5 and U6) and proteins (for a recent review, see Rino and Carmo-Fonseca, 2009). The snRNAs of the major (U2-type) spliceosome interact with three short sequences within a U2 intron, the 5′ donor site (starting with 5′-GU or occasionally, 5′-GC), the lariat branchpoint sequence (containing the branchpoint adenosine near the 3′ end of the intron) and the 3′ acceptor site (ending with 5′-AG). U2 spliceosome intron excision proceeds by two sequential trans-esterification reactions. Firstly, the 2′ hydroxyl group of the branchpoint adenosine performs a nucleophilic attack on the first nucleotide of the intron donor, an invariant guanosine, forming the lariat intermediate. Then, the free 3′ hydroxyl group of the released exon performs a nucleophilic attack at the last nucleotide of the intron acceptor, another invariant guanosine, thereby joining the exons and releasing the intron in the form of a lariat. The minor (U12-type) spliceosome is very similar to the U2-type spliceosome, however, it splices out introns with rare splice sites using different but functionally analogous snRNAs (U11, U12, U4atac and U6atac, together with U5; see, e.g., Patel and Steitz, 2004). Kupfer et al. (2004) have discussed the consensus sequences of donor-, lariat branchpoint- and acceptor sequences of U2 introns in five model fungi and their subtle differences with those of vertebrate introns. At variance with the latter, poly-pyrimidine tracts between the lariat branchpoint- and acceptor consensus sequences do not appear to be a conserved feature in fungal introns. Degenerated donor and lariat branchpoint sites (5′-GUNNNN and 5′-NYTNAN, respectively) occur quite frequently in filamentous fungal U2 introns.

Differently from spliceosomal introns, group I and group II introns are self-excising ribozymes, which may or not need accessory proteins to complete the splicing process (for reviews, Haugen et al., 2005; Nielsen and Johansen, 2009; Falcon de Longevialle et al., 2010; Lambowitz and Zimmerly, 2011). In eukaryotes, they are primarily restricted to mitochondrial and chloroplast genes, albeit group I introns also occur in the nuclear ribosomal DNA loci of specific eukaryotic microbes (Haugen et al., 2005; Waring et al., 1983). Group III introns are restricted to euglenoid chloroplast genes and are probably abbreviated versions of group II introns (Doetsch et al., 1998; and references therein).

* Corresponding author. Present address: Institut de Génétique et Microbiologie, CNRS UMR 8621, Université Paris-Sud 11, Centre Universitaire d'Orsay, Bâtiment 400, 91405 Orsay Cedex, France. Fax: +33 169 15 72 96.
E-mail address: drir.michelflipphi@gmail.com (M. Flipphi).

1087-1845/$ - see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.fgb.2013.06.003

The origin of spliceosome introns is a vexing problem. Similarities in splicing mechanism with that of group II introns, supported by recent structural studies (Toor et al., 2008), led to the hypothesis that they resulted from an invasion of mobile group II introns originating from the protobacterial ancestor of mitochondria (see Koonin, 2009, for a review of this hypothesis). This "intron early" hypothesis accounts for the striking conservation of intron positions among organisms belonging to widely divergent phyla, but cannot account for episodes of intron gain, which surely have occurred in modern phyla (e.g., Ragg, 2011; Rogozin et al., 2012; van der Burgt et al., 2012; Roy and Irimia, 2012). The "intron early" and "intron late" hypotheses do not necessarily exclude each other. The appearance of new introns in metazoans, where mitochondria are free from introns in general and from group II introns in particular (with very few exceptions in early diverging lineages attributed to horizontal transfer, e.g., Vallès et al., 2008), necessitates mechanisms of *de novo* appearance of introns independent from mobile group II introns.

The term "twintron" was originally applied to complex introns of the group II and III types in the chloroplast genomes of euglenoid species, including 15 introns present in the plastid genome of *Euglena gracilis* (see, e.g., Copertino and Hallick, 1991; Hallick et al., 1993; and most recently, Doetsch et al., 2001) and later identified in a cryptomonad (Maier et al., 1995). Excision of the "internal" or "invading" intron is required for the excision of the "external" or "host" intron as the former interrupts a sequence essential for splicing, such as domains V or VI (but see below). This site of integration of the internal intron leads necessarily to evolutionary stability of the twintron arrangement – the internal intron must remain functional to permit the excision of the external intron (Copertino and Hallick, 1993; Thompson et al., 1995; Doetsch et al., 2001). For the chloroplast *mat2* gene, the "twintron" consist of two introns within the intron-encoded ORF rather than interrupting crucial splicing sequences (Zhang et al., 1995). The presence of group II twintrons has been inferred but not demonstrated in archea and cyanobacteria (Nakamura et al., 2002; Rest and Mindell, 2003; Mohr et al., 2010). The mitochondrial DNA of the mushroom coral *Discosoma* shows an astonishingly complex organisation where the *nad5* gene is interrupted by a group I intron, which carries in its P8 loop 15 mitochondrial genes, including the *cox1* gene, which is itself interrupted by a group I intron (Medina et al., 2006).

Very recently, twintrons have been predicted in fungal mitochondrial genomes (Hafez et al., 2013). Twintrons formed by the insertion of a group I intron within the P1 loop of the external group I intron in the small subunit RNA (*rns*) genes of *Chryphonectria parasitica* and *Ophiocordyceps tricentri* were predicted. The same article describes a group II intron inserted in an open reading frame comprised within a loop in the P9.1 region of a group I intron in the *rns* gene of *Chaetomium thermophilum*. In the latter case, the presence of the group II internal intron does not affect the ribozyme structure of the external intron. It would preclude the excision of the external intron only if the protein encoded by the ORF would be a maturase essential for *in vivo* splicing of the external intron. An intron within an intronic ORF had previously been described in the mitochondrial genome of *Grosmannia piceiperda* by the same laboratory (Rudski and Hausner, 2012).

Some group I introns present in extrachromosomal DNA carrying the gene specifying the small ribosomal RNA subunit in amoeboflagelates (*Naegleria gruberi*) and myxomycetes (*Didymium iridis*) (Johansen et al., 1997; Einvik et al., 1998; Nielsen et al., 2005; Nielsen and Johansen, 2009) were also called "twintrons". These self-splicing introns comprise a conventional group I ribozyme, involved in the splicing reaction, and a second group I ribozyme that catalyses hydrolytic cleavages in the intron RNA, which are presumably necessary for the maturation of the mRNA of the homing endonuclease, whose coding region is contained within the intron.

The third pattern of RNA organisation, which has been called a "twintron" occurs in the *prospero* gene of *Drosophila melanogaster* (encoding a transcription factor expressed in immature neuronal cells), but it is actually a unique case of alternative splicing (Scamborova et al., 2004; Borah et al., 2009). The second intron contains active U2 splicing sites within a U12 spliceosome intron. While the gene and the (U12) intron position are highly conserved, the U2 sites are specific of Diptera and absent in other insects (Mount et al., 2007).

Despite their common denomination, there are fundamental differences between these three classes of "twintrons". In the euglenoid chloroplast twintrons, the excision of the internal intron is essential for the consecutive splicing reaction, and thus for correct exon joining, while in the group I twin-ribozyme introns, each ribozyme has a distinct and specialised function. The *prospero* twintron can in fact be considered as functionally diametrically opposed to the euglenoid group II/III twintrons as in *prospero*, the excision of one intron *precludes* that of the other (Borah et al., 2009), while the removal of chloroplast twintrons *necessitates* "inside out" sequential excision of the constituent introns.

In this article we describe a new type of intervening sequences where the name of twintron *sensu strictu* is appropriate. These are complex spliceosomal introns in which the excision of an "internal intron" is obligatory for the splicing of the "external intron". Thus they are formally splicesomal analogues of the group II/III twintrons of the euglenoid chloroplast, even if their splicing mechanism is necessarily different. We shall call these structures "stwintrons" for "spliceosomal twin introns". We propose to restrict the name stwintron (*sensu strictu*) for those complex spliceosomal introns where the "internal intron" interrupts a sequence essential for the ultimate splicing reaction, so as to differentiate this structure, besides from the alternatively spliced *prospero* intron described above, also from apparently more similar ones such as those processed by recursive splicing in *D. melanogaster* (Burnette et al., 2005), or intrasplicing (Parra et al., 2008, 2012) or nested splicing (Suzuki et al., 2013) in vertebrates (see Section 4).

2. Materials and methods

2.1. Data mining

Homologous genes and ESTs were mined upon TBLASTN (Altschul et al., 1997) screening using the *Aspergillus nidulans* PIH1, AMFS and BioDA proteins (see Results, below) as queries. The screened databases included the National Center for Biotechnology Information non-redundant nucleotide (nr/nt), EST and fungal databases, MycoCosm at the US Dept. of Energy Joint Genome Institute (Grigoriev et al., 2012), the *Sclerotinia sclerotiorum* Sequencing Project and the Origin of Multicellularity database of the Broad Institute of Harvard and MIT, and the *Botrytis cinerea* T4 genome database of the Institut National de la Recherche Agronomique (Amselem et al., 2011). Gene models and proteins were deduced manually and where possible, verified with extant ESTs.

2.2. DNA alignment

Genomic *bioDA* sequences were aligned with Clustal-W (Thompson et al., 1994). The output alignment was manually corrected and printed in shaded backgrounds with BOXSHADE (http://www.ch.embnet.org/software/BOX_form).

Please cite this article in press as: Flipphi, M., et al. Spliceosome twin introns in fungal nuclear transcripts. Fungal Genet. Biol. (2013), http://dx.doi.org/10.1016/j.fgb.2013.06.003

2.3. Phylogeny

Peptidic sequences were aligned with MAFFT version 6 (Katoh and Toh, 2008) (http://mafft.cbrc.jp/alignment/server) using the G-INS-I algorithm. Curation was carried out with Block Mapping and Gathering using Entropy (BMGE) (Criscuolo and Gribaldo, 2010) (http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::BMGE) with a BLOSUM 85 similarity matrix. The Maximum Likelihood tree was obtained with PhyML (Guindon et al., 2010) (http://www.phylogeny.fr/version2_cgi/alacarte.cgi) and adapted in Adobe Illustrator from a FigTree (http://tree.bio.ed.ac.uk/software/figtree) image. Approximate likelihood ratio tests (aLRTs) are described (Anisimova and Gascuel, 2006).

2.4. Fungal strains, media and growth conditions

The fungi used were: *Fusarium verticillioides* strains FRC M-3125 and 57-7-7 (Brown et al., 2005), *Trichoderma reesei* strain QM9414 (Vitikainen et al., 2010), and *B. cinerea* strain T4 (Amselem et al., 2011).

F. verticillioides was maintained on *Aspergillus* Minimal Medium (AMM) plates (Pontecorvo et al., 1953) containing 10 g/L glucose and 0.5 g/L bacto-peptone. Mycelia were grown in 500-mL Erlenmeyer flasks with 100 mL of Complete Medium (i.e., AMM with 2 % (w/v) malt extract, 1 % bacto-peptone and 2 % glucose) seeded with vegetative spore inoculum, at 25 °C in a rotary shaker (Infors HT Multitron) at 200 rotations per min (rpm). *T. reesei* was maintained on malt extract agar (Difco). Mycelia were grown at 30 °C in 500-mL Erlenmeyer flasks with 100 mL of medium at 250 rpm. Liquid medium (Mandels and Andreotti, 1978) contained glucose (10 g/L) as the carbon source. *B. cinerea* was maintained on potato dextrose agar (Difco). Batch cultures were performed in 500-mL Erlenmeyer flasks with 100 mL of PDB (Potato Dextrose Broth, Difco) medium at 25 °C and 200 rpm.

2.5. Nucleic acid isolation

Fungal mycelia were harvested by medium filtration, washed with distilled water, frozen and ground to powder under liquid nitrogen. For the extraction of genomic DNA, plasmid DNA and total RNA, Promega purification kits (Wizard Genomic DNA Purification Kit, PureYield Plasmid Miniprep System and PureYield RNA Midiprep System, respectively) were used according to the manufacturer's instructions.

2.6. 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). PCR reactions were performed in a 25 μL volume containing 4 μL of single strand cDNA, using gene-specific oligonucleotides (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 30 s, followed by one post-cyclic elongation at 72 °C for 5 min. Amplified fragments were resolved in a 3 % native agarose gel and stained with ethidium bromide.

2.7. cDNA sequencing

PCR was primed off the primary RT product using gene-specific oligos (Tables S1 and S2) with cycling conditions as described above. Double strand cDNA was gel-purified and subsequently cloned (pGEM-T Easy Vector System I, Promega). Three independent clones were sequenced (MWG-Biotech AG, Ebersberg, Germany). Sequences were deposited at GenBank under the accession numbers KC019312–KC019315.

3. Results

Inspecting fungal genomes while pursuing unrelated phylogenetic and metabolic work (for example, Magliano et al., 2011), we have found serendipitously four cases where the existence of a stwintron could be predicted. In each of these cases, splicing at the apparent canonical 5′ donor sequence would result in an early frameshift. The actual 5′ donor sequence of these stwintrons is interrupted by a putative second intervening sequence such that only the excision of this second sequence ("internal intron") would

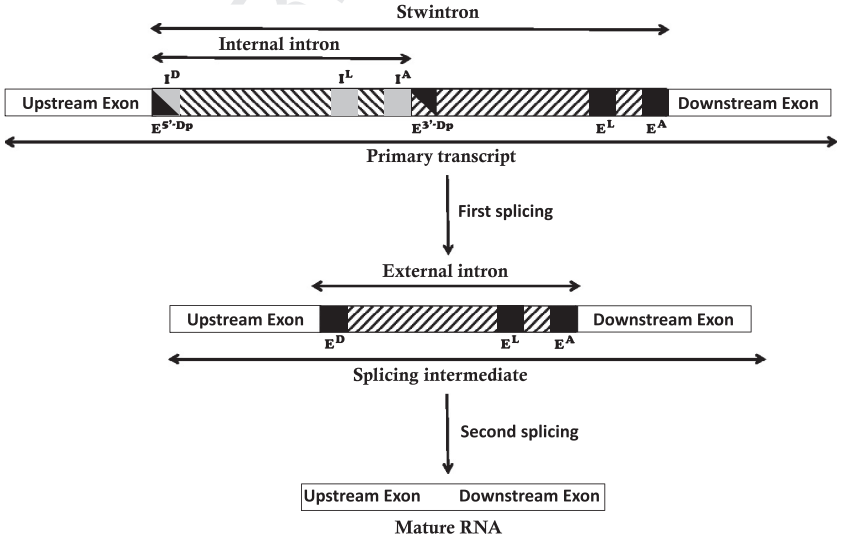

Fig. 1. Schematic representation of stwintron splicing. This occurs in two steps. In the first step, the internal intron is removed and a new, canonical donor sequence is generated. Subsequently, the external intron is excised and the exons are joined. I^D, I^L and I^A indicate the donor-, lariat branchpoint- and acceptor sequences of the internal intron (grey squares, top scheme). E^L and E^A indicate the lariat branchpoint- and acceptor sequences of the external intron (black squares, top and middle schemes). $E^{5'\text{-}Dp}$ and $E^{3'\text{-}Dp}$ indicate the two precursor sequences (black triangles in top scheme) that, when spliced together, generate the donor of the external intron, E^D (black square in middle scheme). The 5′ and 3′ superscripts indicate the relative positions of the two precursor sequences. The internal intron is hatched from left to right; the external intron, from right to left.

Please cite this article in press as: Flipphi, M., et al. Spliceosome twin introns in fungal nuclear transcripts. Fungal Genet. Biol. (2013), http://dx.doi.org/10.1016/j.fgb.2013.06.003

result in a new canonical donor sequence allowing a subsequent splicing event, which then yields the proper continuous open reading frame. The internal intron comprises secondary lariat branchpoint and acceptor sites, allowing its excision and the reconstitution of the external intron's 5′ donor site. Fig. 1 shows the general structure of stwintrons.

3.1. Internal intron splitting the first from the second base of the external intron donor site

A gene encoding a protein similar to bacterial cyclic imidine hydrolases is widely conserved in the Basidiomycota and Pezizomycotina lineages. In *Fusarium verticillioides* and *F. oxysporum*, two closely related paralogues are present (~78% identity at the amino acid level). Phylogenetic analysis indicated that the second gene has arisen through a duplication of an ancestral gene (M. Flipphi and C. Scazzocchio, unpublished results). The second gene (called *PIH2* in Fig. 2, for putative imidine hydrolase paralogue 2) is also present in *F. graminearum* and *F. pseudograminearum*, which lack the ancestral paralogue (NB. Identity of paralogues based on sequence, gene model and synteny, not shown). An intron, absent in the ancestral paralogue *PIH1*, is present in *PIH2*. This intron is a putative stwintron. Its structure is shown in Fig. 2A, while the actual sequences are given in the Supplementary data (Fig. S1).

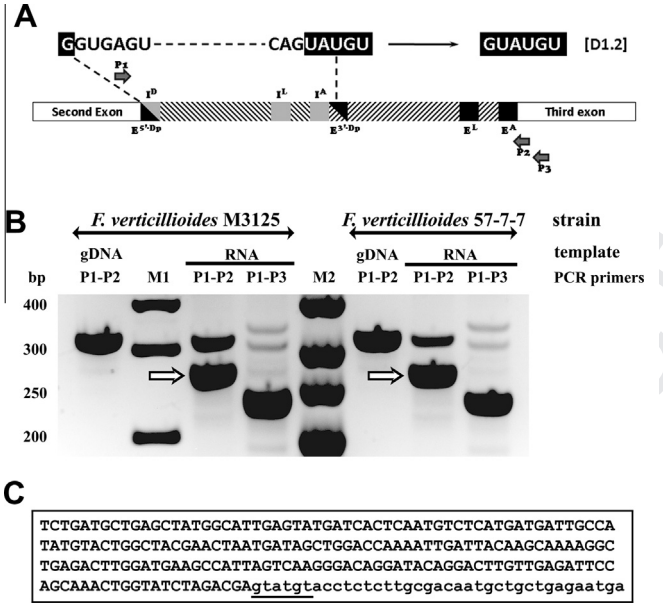

Fig. 2. The [D1.2] stwintron of the *PHI2* gene of *Fusarium verticillioides*. (Panel A) Scheme of the structure of this stwintron, and of the splicing event that generates *de novo* the donor sequence of the external intron. The two precursor sequences are given in white lettering on a black background above the two donor precursor triangles (see Fig. 1). The horizontal arrow points to the outcome of the [D1.2] splicing event which generates the E^{D} sequence (also in white lettering on a black background), where the first nucleotide of the final donor originates from the upstream precursor $E^{5'\text{-}Dp}$ (see Fig. 1). The [D1.2] nomenclature indicates that the splicing together of the ultimate donor sequence occurs between its first and second nucleotide. The positions of the primers used to detect the splicing products are indicated above and below the scheme. P1 is situated on the coding strand within the second (upstream) exon; P2, partly overlapping P3 (see below), on the non-coding strand spanning the stwintron/third exon junction and extending into the external intron; P3, on the non-coding strand within the third (downstream) exon. (Panel B) Experimental results. PCR products were obtained using the primers indicated above. Hollow arrows indicate the fragments corresponding in size to the predicted splicing intermediate. PCR on genomic DNA (gDNA) yields the fragment corresponding in size to the primary transcript, while the most conspicuous band amplified with the P1–P3 primers corresponds to the final splicing product, the mature mRNA. M1 and M2; molecular size markers. (Panel C) Predicted and actual sequence of the splicing intermediate, amplified with the P1–P2 primers. Exonic sequences are printed in capitals, intronic sequences (external intron) in lower case letters. The re-constituted donor sequence of the external intron is underlined. The two *F. verticillioides* strains tested gave identical results.

An extant cDNA clone from *F. verticillioides* (accession number DR668820) shows that the first of the two successive Gs is used in the final donor site (Fig. 2). To account for this departure from the almost universal GU rule at the donor site, we postulated that the second G is in fact the first nucleotide of a consensus donor of an internal intron interrupting the 5′ donor site of the external intron between the first and the second nucleotide of this latter (canonical) site (i.e., 5′-G|UAUGU, for *F. verticilloides* and *F. oxysporum*, G|UAUGC for *F. graminearum* and *F. pseudograminearum*, see Fig. 2A). If we accept that the acceptor and donor splicing sites used are those nearest to each other (i.e., intron definition: Berget, 1995; Romfo et al., 2000), the excision of the predicted internal intron would utilise the most proximal consensus lariat branchpoint-acceptor sequences (5′-GCUAAC – 8 nt – CAG in all species, see Supplementary Fig. S1) located roughly halfway the intervening sequence (as depicted by the grey boxes in the schematic representation in Fig. 2A). Once the internal intron is excised, a second splicing event would occur using the neo-constructed, canonical donor site and the 3′ set of lariat branchpoint- and acceptor sites. This stwintron will be called a [D1,2] stwintron (Fig. 2A).

We endeavoured to detect the postulated splicing intermediate in *F. verticillioides*. Fig. 2 B shows that an intermediate of the predicted size was indeed present in two different strains, 57-7-7 and FRC M-3125. Two clones from the putative intermediate of strain 57-7-7 and one from FRC M-3125 were sequenced, and they all showed the sequence predicted after the splicing of the internal intron (Fig. 2C). A full-length cDNA clone generated from the same 57-7-7 template RNA (GenBank KC019314) showed the predicted sequence obtained after excision of the whole [D1,2] stwintron.

A second, putative [D1,2] stwintron is predicted below. An L-arabinose-inducible gene encoding a putative sugar transporter was identified in the genome of *Aspergillus nidulans* (M. Flipphi, unpublished results). This gene (called provisionally *AMFS* for L-Arabinose-inducible Major Facilitator Superfamily protein) is conserved among many members of the Pezizomycotina, and so are two intron positions in the 5′ moiety. The orthologue genes of two closely related members of the Magnaportaceae family, *Magnaporthe poae* and *Gaeumannomyces graminis*, contain two additional introns downstream, absent from all other orthologue genes in the databases, including that of *Magnaporthe oryzae*. The 5′ of those additional introns is a putative [D1,2] stwintron (Supplementary Fig. S2), analogous in its structure to one demonstrated above (Fig. 2). The GUAUGU donor site extant in the primary transcript (Supplementary Fig. S2) would result in a frame shift, while the donor generated by fusing the preceding G with the internal sequence UGAGA (upon splicing of the proposed internal intron) would yield the correct phase.

3.2. Internal intron splitting the second from the third base of the external intron donor site

The *bioDA* gene (nomenclature from *A. nidulans*) encodes a bifunctional enzyme catalysing respectively the third (dethiobiotin synthase) and second (7,8-diamino pelargonic acid aminotransferase) steps of biotin biosynthesis (Magliano et al., 2011). This fused gene is present in a number of diverse phylogenetic groups including the Pezizomycotina. Most of the Sordariomycete genomes include *bioDA*, with some exceptions where the fused gene is missing (well documented for *Neurospora*, *Sordaria* and *Magnaporthe* species). The *bioDA* gene of the Sordariomycetes typically comprises three introns in conserved positions, two within the 5′ *bioD* homologous sequences separated by a 97-nt long exon and a third within the 3′ *bioA* homologous sequences. The first two intron

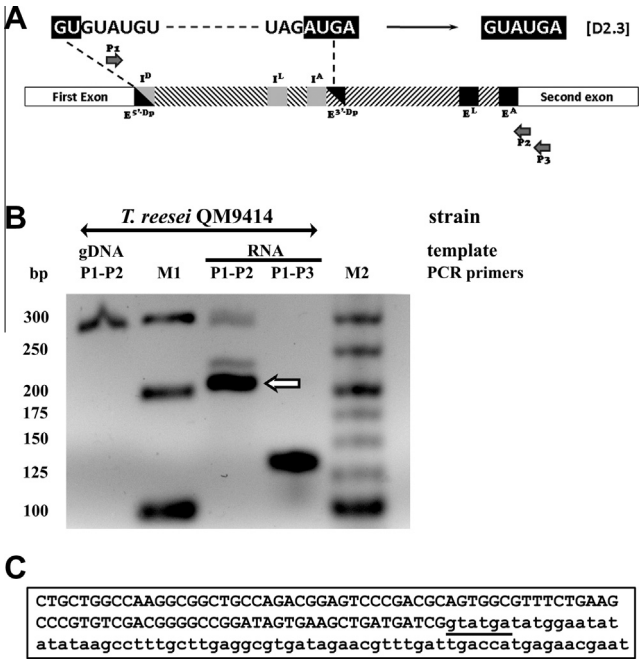

Fig. 3. The [D2.3] stwintron of the *bioDA* gene of *Trichoderma reesei*. All conventions and symbols are as in Fig. 2. The E^{D} sequence (see Fig. 1) is in this case generated by a [D2.3] event, where the first two nucleotides originate from the upstream precursor $E^{5'\text{-}Dp}$. The band amplified with the P1–P3 primers corresponds to the final splicing product.

positions are conserved in the Eurotiomycete and Dothideomycete lineages (not shown).

Perusing the most 5′ intron in the *bioDA* gene of the Sordariomycetes (except *Nectria haematococca*, see below) revealed a paradox similar to the one described above (see Fig. 3 for *Trichoderma reesei*; the sequences of the first *bioDA* intron of all sequenced Sordariomycetes are given in Supplementary Fig. S3). Its 5′ terminus (donor site) contains two successive GU sequences. The second GU and the four nucleotides following it constitute a canonical 5′ donor site. However, if the second GU sequence were functional as the donor in a single splicing reaction, an out-of-frame coding sequence would be generated, while if the first GU was used, a correct open reading frame would result.

In all *bioDA* genes of the sequenced Sordariomycetes (with the exception of *N. haematococca*, see below) we detected upstream from the putative lariat branchpoint site, a second set of consensus lariat branchpoint and acceptor sequences, situated roughly in the middle of the 5′ intron (see Supplementary Fig. S3, for sequences). If we again assume that intron definition applies, we can postulate that in Sordariomycetes, a second intron splits the 5′ donor sequence of a host intron, the second GU doublet functioning as the donor site for this "internal intron" and that the excision of the internal intron generates a new canonical donor site leading to a second splicing event which yields the *bioDA* coding sequence (Fig. 3A). We shall refer to this structure as a [D2,3] stwintron. cDNA sequences are available for the *bioDA* genes of *Trichoderma atroviridae* (accession number GE276214), *Fusarium verticillioides* (a.k.a. *Gibberella moniliformis*) (DR662398) and *Myceliophthora thermophila* (a.k.a. *Sporotrichum thermophilum*) (GT935069) which show that correct splicing occurs when the 5′ GU sequence, rather than the adjacent 3′ GU, serves as a constituent of the final donor and thus support the stwintron concept.

We demonstrated the existence of the predicted splicing intermediate from the [D2,3] stwintron of the *bioDA* gene of *T. reesei*. Fig. 3B shows the presence of a splicing intermediate of the predicted size; the sequences of two independent cDNA clones generated from this intermediate (Fig. 3C) show the excision of the internal intron, exactly as predicted. An extended cDNA clone covering all introns was generated from the same template RNA (GenBank KC019315) and confirmed that excision of the internal intron is followed by that of the reconstituted external intron.

In the cognate intron of the *bioDA* gene of *Nectria haematococca* (*Fusarium solani* f. sp. *pisi*), a perfect donor consensus comprising only one GU doublet can be seen at the 5′ splicing site. This intron is considerably shorter (42 nucleotides) than that of the other Sordariomycetes (Supplementary Fig. S3). A maximum likelihood phylogeny (Supplementary Fig. S4) of all deduced Sordariomycetes BioDA proteins is consistent with the stwintron organisation being basal to the class and the cognate intron in *N. haematococca*, being the result of a secondary (recent) loss of the internal intron.

Suprisingly, a putative stwintron in an identical position as that described above is found in the *bioDA* gene of the two sequenced members of the Botryosphaeriales, *Botryosphaeria dothidea* and *Macrophomina phaseolina*, but not in any other species of the Dothideomycetes class. The BioDA proteins of these two species cluster at the base of the Sordariomycetes clade and not with the other Dothideomycetes (Supplementary Fig. S4), strongly suggesting an episode of lateral transmission of the *bioDA* gene. This could have occurred either from an early member of the Sordariomycetes to an ancestor of modern day Botryosphaeriales or alternatively, from an ancient member of the Botryosphaeriales to a common ancestor of all present day Sordariomycetes.

3.3. A deceptive putative [D4,5] stwintron

The ancestral paralogue of the *Fusarium* gene discussed above (*PHI1* in Fig. 4, for putative imidine hydrolase paralogue 1) comprises an intron, whose position is conserved in four classes of Pezizomycotina. Uniquely in *Botrytis cinerea* and *Sclerotinia sclerotiorum*, this intron features characteristics of a stwintron. In four other sequenced Leotiomycetes, including *Sclerotinia homoeocarpa*, the intron at this position is considerably smaller (see Supplementary Fig. S5 for complete sequences of the cognate intron in the sequenced Helotiales). Here, as above, the 5′ donor of a complex intron could be intron-interrupted. An extant *S. sclerotiorum* EST (Broad G2118P59RE12/FE12) shows that sequential splicing could occur if the 5′ GU of the sequence GUUCGU is used as the ultimate donor, which we may interpret as resulting from the formation of a GUUC|CU donor after excision of an internal intron (5′-GUCAUU–UCUAAU–AAG); a putative [D4,5] stwintron.cDNAs generated in the reaction described below (GenBank KC019312–KC019313) confirmed the corresponding *B. cinerea PIH1* intron as predicted, and its sequence suggested an essentially identical stwintron organisation in that species. The two cloned cDNAs are not identical. In one, two downstream introns are excised while the 5′ intron is not. In the second all three introns are excised. This suggests that the 5′ intron is only removed *after* the excision of the downstream introns and thus that the 5′ intron in *B. cinerea PIH1* – the putative [D4,5] stwintron – is not efficiently spliced out. This inefficient splicing may explain the relative abundance of the amplification product corresponding to the unspliced transcript with both pairs of PCR primers (Fig. 4B). The predicted stwintron is schematically shown in Fig. 4A.

Attempts to detect the predicted splicing intermediate in *B. cinerea* gave surprising results. Fig. 4B shows the presence of a product of intermediate size. However, the sequence of three independent cDNA clones generated from this product (Fig. 4C) is not that predicted by the stwintron hypothesis. The splicing product is generated by using the donor sequence (GUCUGU) present in the primary transcript, and an acceptor sequence (UAG) 37 nt upstream of the acceptor of the proposed internal intron. Thus, rather

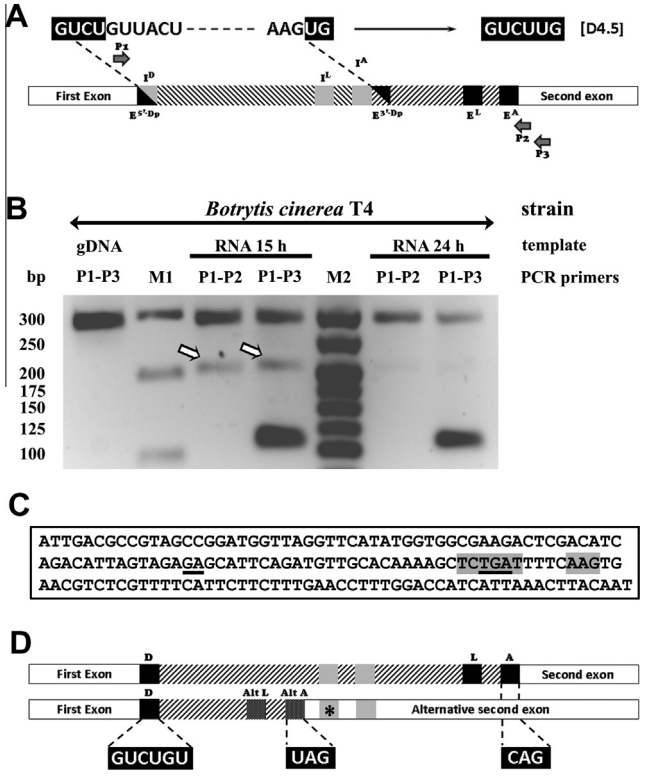

Fig. 4. Alternative splicing in the *PHI1* gene of *Botrytis cinerea*. (Panel A) Hypothetical [D4.5] stwintron as predicted for this gene. Nomenclature as in Figs. 2 and 3. The horizontal arrow points in this case to the predicted generation of the donor sequence of an external intron, where the first 4 nucleotides derive from the upstream precursor ($E^{5'\text{-}Dp}$). (Panel B) Experimental results. RNA samples were extracted after 15 h and 24 h of mycelial growth. The hollow arrows indicate an amplification product of intermediate size that could correspond to the putative splicing intermediate. The major band revealed by the P1–P3 primers is the putative final splicing product. (Panel C) cDNA sequence of the band indicated by the arrows is shown. This does not correspond to the predicted splicing intermediate (see Supplementary Fig. 5B) but reveals an instance of alternative splicing, and includes, overlaid in grey, the lariat branchpoint- and acceptor sequences of the predicted internal intron of the putative [D4.5] stwintron. The exon/exon junction resulting from the excision of the small variant of the alternatively spliced intron is underlined. Note that the first in phase stop codon in the alternatively spliced RNA (underlined TGA) is within the predicted lariat branchpoint sequence of the predicted internal intron. (Panel D) Schematic representations of the two exclusive, alternative splicing events, both using the 5′-GUCUGU donor sequence. Top: A schematic representation of the event that generates an mRNA encoding the full-length protein, derived from the actual sequence of the final splicing product (GenBank KC019312). Bottom: A schematic representation of the alternative splicing event, leading to a protein truncated at the UGA codon (marked with an asterisk) located within the putative lariat branchpoint sequence of the hypothetical internal intron as predicted in Panel A. For both top and bottom schemes: Black squares, donor-, lariat branchpoint- and acceptor sites of the main splicing event; Grey squares, the predicted lariat branchpoint- and acceptor sequences of the hypothetical internal intron; Squares in punctuated black and white pattern, putative lariat branchpoint- and acceptor sequences of the alternative splicing event (tagged "Alt L" and "Alt A", respectively). The sequences of the donor (D) and the two alternative acceptors ("Alt A" and A, respectively) are given below the scheme in white lettering on a black background.

than specifying a stwintron, the *B. cinerea PHI1* transcript undergoes alternative splicing dependent on two, widely separated acceptors, the excision of the "shorter intron" leading to a C-terminally truncated protein. Fig. 4D shows schematically the two splicing alternatives, the one leading to the predicted message encoding the full-length protein and the second, the alternative event leading to a message encoding a truncated protein. In clear defiance to intron definition theory, the excision of the "longer intron", using the same donor as the alternative "shorter intron", appears to be the only way to produce a full-length gene product.

4. Discussion

Our results constitute the first demonstration of the hitherto undetected existence of spliceosome twin introns *sensu strictu* (stwintrons), i.e., instances where an internal intron needs to be removed to generate a functional splicing sequence necessary for the subsequent excision of an external intron to yield properly matured mRNA. Sequential splicing of the donor-disrupted introns described above implies a specific mechanism of intron definition, including a re-scanning of the transcript after the first excision event and adjunction of splice sites. The cases we have described affect the 5′ donor sequence, but there seem to be no *a priori* reasons internal introns interrupting the lariat branchpoint or acceptor sequences of the host intron could not be extant. In [D1,2] stwintrons, splicing of the external intron from the primary transcript could, in principle, occur using the acceptor G of the internal intron (see Fig. 2A or Supplementary Fig. S1 and S2) as its 5′ donor G, with the result that the stwintron could then no be removed properly (unless the first nucleotide behind the [D1,2] stwintron, would also be a G – which is not the case for the *PIH2* and *AMFS* stwintrons described above). The prior excision of the internal intron precludes such futile event, which would lead to an improperly translated mRNA.

A number of instances have been described in large (>100 knt) metazoan spliceosomal introns where the removal of an intron within an intron occurs. We shall discuss the similarities and differences these processes have with the one described in this article. Recursive splicing has been described for a few genes in *D. melanogaster* (Burnette et al., 2005) but so far not in other organisms. The similarities with stwintrons are only apparent and indeed superficial, as in recursive splicing, no new functional splicing sequences (i.e., donor, lariat branchpoint consensus or acceptor sites) are generated. Rather, introns that have abutting acceptor and donor sequences are excised sequentially. In nested splicing, "inside-out", sequential removal of internal introns, brings nearer the final donor and acceptor sites of the external intron resulting in the accurate excision of large introns as shown for the human dystrophin transcript (Suzuki et al., 2013). However, without exception, the internal introns do not disrupt the principal sequence elements necessary for the splicing of the intron in which it is nested. Small nested introns may also be extant in fungal genomes, e.g., the most 3′ intron of the *Neosartorya fischeri mccA* gene encoding the biotin-binding subunit of 3-methylcrotonyl-CoA carboxylase or the second intron of the *Verticillium dahliae acl2* gene encoding the small subunit of ATP citrate lyase (not shown).

The process that is formally most similar to stwintron splicing is "intrasplicing" where the excision of a first intron is necessary for proper removal of a second one. In the vertebrate 4.1R- and 4.1B paralogues, alternative splicing of an intron in the 5′ untranslated region (5′-UTR) is coordinated with the utilisation of alternative promoters, leading to cytoskeletal 4.1 proteins that differ at their N-termini (Parra et al., 2008, 2012). One of the splicing possibilities involves a "weak" acceptor site shortly downstream of a "strong" acceptor. The "weak" site can only be used when the excision of a (constitutive) intron results in both the elimination of the "strong" acceptor and the generation of a structural element joining a poly-pyrimidine tract and a lariat branchpoint site present in the upstream sequence of that intron and the "weak" AG acceptor site downstream of it. This allows the subsequent excision of a second (alternative) intron and the production of the shorter form of the 4.1 proteins uniquely from transcripts controlled from the distal promoter. In the 4.1R gene, the ~200 nt region carrying the poly-pyrimidine tract and the lariat branchpoint site preceding the donor of the "internal" intron has been called an "intraexon" to underline that it behaves as part of an exon in the first splicing

reaction and as an integral part of the second intron in the final splicing reaction (Parra et al., 2008).

It seems clear that the stwintrons uncovered by us, are late appearances in evolution and, moreover, that internal intron generation (leading to a stwintron) and -loss can occur readily. The *bioDA* bifunctional gene is conserved in fungi, plants and öomycetes. The intron position where the stwintron occurs is conserved in a number of Dikarya subphyla, but also in all sequenced Mucoromycotina and in the chytrid *Spizellomyces punctatus* (ESTs: accessions EB744794 and EB747034). However, the *bioDA* stwintron is restricted to the Sordariomycetes and Botryosphaeriales, suggesting that its appearance is more recent than that of the original host intron. The secondary loss of the internal intron in the *bioDA* gene of *N. haematoccoca* has been discussed above. The presence of the putative [D1,2] stwintron in *M. poae* and *G. graminis AMFS* but not in the *Magnaporthe oryzae* orthologue gene, illustrates the same concept. However, the phylogenetic position of the BioDA proteins of the two Botryosphaeriales that also feature a putative [D2,3] stwintron in their *bioDA* gene, argues in this case against an independent stwintron generation in this order of Dothideomycetes (see Results Section 3.3 and Supplementary Fig. S4).

While stwintrons are formally analogous to the chloroplast twintrons of the euglenoids, there are necessarily fundamental mechanistic differences. Group II and presumably group III introns, depend for their excision on a complex secondary structure and have been show in several instances, to be self splicing in eukaryote plastids (Sheveleva and Hallick, 2004; Odom et al., 2004), while two successive reactions involving the whole spliceosome – possibly including spliceosome disassembly and subsequent re-assembly after the initial splice site adjunction and ligation – are essential to excise a stwintron.

While the generation of the euglenoid chloroplast twintrons may involve intron mobility and proliferation (Hallick et al., 1993; Copertino et al., 1994; Zhang et al., 1995), including recent horizontal transmission of cyanobacterial mobile elements (Sheveleva and Hallick, 2004), no similar mechanism could be postulated for stwintrons, where an endogenous origin seem to be the only reasonable alternative (see Catania et al., 2009; Roy and Irimia, 2009; Ragg, 2011; for recent discussions on spliceosomal intron generation). While our present data does not allow to propose a specific mechanism, the internal component of a stwintron could be originated by a mechanism analogous to what has been called "intronisation", the formation of new introns from exonic sequences as it has been proposed for nematode genes (Irimia et al., 2008) where point mutations in exons could result in cryptic splice sites and eventually, the appearance of new introns. Stretches of coding sequences in the *bioDA* gene of the chytrid *Batrachochytrium dendrobatidis* align with two intronic sequences (introns 4 and 5) in *S. punctatus* (Supplementary Fig. S6), which suggests that intronisation can indeed take place in fungi. A similar process, that could be called "stwintronisation", could result in the appearance of an intron within an intron. The alternatively spliced *prospero* "twintron" (see Section 1), where the internal U2 splice sites are unique to Diptera (Mount et al., 2007), may have a similar origin. It may be relevant that all cases of stwintrons *sensu stricto* we have described above, involved an apparent duplication of one or two subsequent base pairs at the 5′ splice site. This may be the consequence of errors in DNA repair (Ragg, 2011).

We have uncovered two experimentally confirmed instances of stwintrons while pursuing unrelated work, suggesting that our findings are just the tip of iceberg. A systematic search of available fungal genomes will be a matter for future work. It remains to be established if stwintrons are a peculiarity of fungi or whether they also occur in species from other kingdoms, especially in those where small intron/large exon gene architecture is extant and splicing is determined by intron- rather than by exon definition (e.g., Berget, 1995; Lorković et al., 2000; Weir and Rice, 2004). For instance, in *Bigelowiella natans* (Rhizaria, Cercozoa) – an unicellular organism featuring short introns which shows surprisingly frequent instances of alternative splicing (Curtis et al., 2012) – a gene encoding malate dehydrogenase contains a putative stwintron where an internal intron could interrupt the donor sequence of the external intron between nucleotides 3 and 4 ([D3,4]) (Supplementary Fig. S7). Two extant cDNA clones (accessions DR039025 and DR040532) demonstrate that ultimate splicing is carried out at the 5′ GU of the sequence 5′-GUAGUGAGU, where GUGAGU may serve as the canonical donor of an internal intron (with internal lariat branchpoint-acceptor couple: 5′-AUUCAU – 10 nt – CAG) disrupting the canonical donor GUA|AGA of a putative external intron, thus supporting a [D3,4] stwintron concept.

Stwintrons may have appeared fortuitously as a result of random mutations within pre-existing introns. Future research may uncover whether they have also been recruited to serve specific functions in alternative splicing or in regulation of gene expression.

5. Uncited reference

Patel and Steitz (2003). Q3

Acknowledgments

We thank Sabine Fillinger for her help accessing the *Botrytis cinerea* T4 database, Eric Espagne and Phillippe Silar for testing the growth requirements of *Podospora anserina*, and Sotiris Amillis and Christian Vélot for help with the design of some figures. Research at the University of Debrecen was supported by the Hungarian Scientific Research Fund (OTKA Grant K1006600), and the TÁMOP-4.2.2/B-10/-1-2010-0024 and TÁMOP-4.2.2.A-11/1/KONV-2012-0043 Projects. E.F. is the recipient of a Bolyai János Research Scholarship (BO/00519/09/8).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fgb.2013.06.003. Q4

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Amselem, J., Cuomo, C.A., van Kan, J.A.L., Viaud, M., Benito, E.P., Couloux, A., Coutinho, P.M., de Vries, R.P., Dyer, P.S., Fillinger, S., Fournier, E., Gout, L., Hahn, M., Kohn, L., Lapalu, N., Plummer, K.M., Pradier, J.-M., Quévillon, E., Sharon, A., Simon, A., ten Have, A., Tudzynski, B., Tudzynski, P., Wincker, P., Andrew, M., Anthouard, V., Beever, R.E., Beffa, R., Benoit, I., Bouzid, O., Brault, B., Chen, Z., Choquer, M., Collémare, J., Cotton, P., Danchin, E.G., Da Silva, C., Gautier, A., Giraud, C., Giraud, T., Gonzalez, C., Grossetete, S., Güldener, U., Henrissat, B., Howlett, B.J., Kodira, C., Kretschmer, M., Lappartient, A., Leroch, M., Levis, C., Mauceli, E., Neuvéglise, C., Oeser, B., Pearson, M., Poulain, J., Poussereau, N., Quesneville, H., Rascle, C., Schumacher, J., Ségurens, B., Sexton, A., Silva, E., Sirven, C., Soanes, D.M., Talbot, N.J., Templeton, M., Yandava, C., Yarden, O., Zeng, Q., Rollins, J.A., Lebrun, M.-H., Dickman, M., 2011. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. PLoS Genet. 7, e1002230.

Anisimova, M., Gascuel, O., 2006. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst. Biol. 55, 539–552.

Berget, S.M., 1995. Exon recognition in vertebrate splicing. J. Biol. Chem. 270, 2411–2414.

Berget, S.M., Moore, C., Sharp, P.A., 1977. Spliced segments at 5′ terminus of adenovirus-2 late mRNA. Proc. Natl. Acad. Sci. U.S.A. 74, 3171–3175.

Borah, S., Wong, A.C., 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.

Brown, D.W., Cheung, F., Proctor, R.H., Butchko, R.A.E., Zheng, L., Lee, Y., Utterback, T., Smith, S., Feldblyum, T., Glenn, A.E., Plattner, R.D., Kendra, D.F., Town, C.D., Whitelaw, C.A., 2005. Comparative analysis of 87,000 expressed sequence tags from the fumonisin-producing fungus *Fusarium verticillioides*. Fungal Genet. Biol. 42, 848–861.

Burnette, J.M., Miyamoto-Sato, E., Schaub, M.A., Conklin, J., Lopez, A.J., 2005. Subdivision of large introns in Drosophila by recursive splicing at nonexonic elements. Genetics 170, 661–674.

Catania, F., Gao, X., Scofield, D.G., 2009. Endogenous mechanisms for the origins of spliceosomal introns. J. Hered. 100, 591–596.

Chow, L.T., Gelinas, R.E., Broker, T.R., Roberts, R.T., 1977. An amazing sequence arrangement at 5′ ends of adenovirus-2 messenger RNA. Cell 12, 1–8.

Copertino, D.W., Hallick, R.B., 1991. Group II twintron: an intron within an intron in a chloroplast cytochrome *b-559* gene. EMBO J. 10, 433–442.

Copertino, D.W., Hallick, R.B., 1993. Group II and group III intron of twintrons: potential relationships to nuclear pre-mRNA introns. Trends Biochem. Sci. 18, 467–471.

Copertino, D.W., Van Hook, F.W., Hall, E.T., Jenkins, K.P., Hallick, R.B., 1994. A group III twintron encoding a maturase-like gene excises through lariat intermediates. Nucleic Acids Res. 22, 1029–1036.

Criscuolo, A., 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.

Curtis, C.A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M.C., Ball, S.G., Gile, G.H., Hirakawa, Y., Hopkins, J.F., Kuo, A., Rensing, S.A., Schmutz, J., Symeonidi, A., Elias, M., Eveleigh, R.J.M., Herman, E.K., Klute, M.J., Nakayama, T., Oborník, M., Reyes-Prieto, A., Armbrust, E.V., Aves, S.J., Beiko, R.G., Coutinho, P., Dacks, J.B., Durnford, D.G., Fast, N.M., Green, B.R., Grisdale, C.J., Hempel, F., Henrissat, B., Höppner, M.P., Ishida, K.-I., Kim, E., Kořený, L., Kroth, P.G., Liu, Y., Malik, S.-B., Maier, U.G., McRose, D., Mock, T., Neilson, J.A.D., Onodera, N.T., Poole, A.M., Pritham, E.J., Richards, T.A., Rocap, G., Roy, S.W., Sarai, C., Schaack, S., Shirato, S., Slamovits, C.H., Spencer, D.F., Suzuki, S., Worden, A.Z., Zauner, S., Barry, K., Bell, C., Bharti, A.K., Crow, J.A., Grimwood, J., Kramer, R., Lindquist, E., Lucas, S., Salamov, A., McFadden, G.I., Lane, C.E., Keeling, P.J., Gray, M.W., Grigoriev, I.V., Archibald, J.M., 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492, 59–65.

Doetsch, N.A., Thompson, M.D., Hallick, R.B., 1998. A maturase-encoding group III twintron is conserved in deeply rooted euglenoid species: are group III introns the chicken or the egg? Mol. Biol. Evol. 15, 76–86.

Doetsch, N.A., Thompson, M.D., Favreau, M.R., Hallick, R.B., 2001. Comparison of *psbK* operon organization and group III intron content in chloroplast genomes of 12 Euglenoid species. Mol. Gen. Genet. 264, 682–690.

Einvik, C., Elde, M., Johansen, S., 1998. Group I twintrons: genetic elements in myxomycete and schizopyrenid amoeboflagellate ribosomal DNAs. J. Biotechnol. 64, 63–74.

Falcon de Longevialle, A., Small, I.D., Lurin, C., 2010. Nuclearly encoded splicing factors implicated in RNA splicing in higher plant organelles. Mol. Plant 3, 691–705.

Grigoriev, I.V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., Kuo, A., Minovitsky, S., Nikitin, R., Ohm, R.A., Otillar, R., Poliakov, A., Ratnere, I., Riley, R., Smirnova, T., Rokhsar, D., Dubchak, I., 2012. The genome portal of the department of energy joint genome institute. Nucleic Acids Res. 40, D26–D32.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Hafez, M., Majer, A., Sethuraman, J., Rudski, S.M., Michel, F., Hausner, G., 2013. The mtDNA *rns* gene landscape in the Ophiostomatales and other fungal taxa: twintrons, introns, and intron-encoded proteins. Fungal Genet. Biol. 53, 71–83.

Hallick, R.B., Hong, L., Drager, R.G., Favreau, M.R., Monfort, A., Orsat, B., Spielmann, A., Stutz, E., 1993. Complete DNA sequence of *Euglena gracilis* chloroplast DNA. Nucleic Acids Res. 21, 3537–3544.

Haugen, P., Simon, D.M., Bhattacharya, D., 2005. The natural history of group I introns. Trends Genet. 21, 111–119.

Irimia, M., Rukov, J.L., Penny, D., Vinther, J., Garcia-Fernandez, J., Roy, S.W., 2008. Origin of introns by 'intronization' of exonic sequences. Trends Genet. 24, 378–381.

Johansen, S., Elde, M., Vader, A., Haugen, P., Haugli, K., Haugli, F., 1997. *In vivo* mobility of a group I twintron in nuclear ribosomal RNA of the myxomycete *Didymium iridis*. Mol. Microbiol. 24, 737–745.

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 286–298.

Koonin, E.V., 2009. Intron-dominated genomes of early ancestors of eukaryotes. J. Hered. 100, 618–623.

Kupfer, D.M., Drabenstot, S.D., Buchanan, K.L., Lai, H., Zhu, H., Dyer, D.W., Roe, B.A., Murphy, J.W., 2004. Introns and splicing elements of five diverse fungi. Eukaryot. Cell 3, 1088–1100.

Lambowitz, A.M., Zimmerly, S., 2011. Group II introns: mobile ribozymes that invade DNA. Cold Spring Harb. Perspect. Biol. 3, a003616.

Lorković, Z.J., Wieczorek Kirk, D.A., Lambermon, M.H.L., Filipowicz, W., 2000. Pre-mRNA splicing in higher plants. Trends Plant Sci. 5, 160–167.

Magliano, P., Flipphi, M., Sanglard, D., Poirier, Y., 2011. Characterization of the *Aspergillus nidulans* biotin biosynthetic gene cluster and use of the *bioDA* gene as a new transformation marker. Fungal Genet. Biol. 48, 208–215.

Maier, U.-G., Rensing, S.A., Igloi, G.L., Maerz, M., 1995. Twintrons are not unique to the *Euglena* chloroplast genome: structure and evolution of a plastome *cpn60* gene from a cryptomonad. Mol. Gen. Genet. 246, 128–131.

Mandels, M., Andreotti, R.E., 1978. Problems and challenges in the cellulose to cellulase fermentation. Process Biochem. 13, 6–13.

Medina, M., Collins, A.G., Takaoka, T.L., Kuehl, J.V., Boore, J.L., 2006. Naked corals: skeleton loss in scleractinia. Proc. Natl. Acad. Sci. U.S.A. 103, 9096–9100.

Mohr, G., Ghanem, E., Lambowitz, A.M., 2010. Mechanisms used for genomic proliferation by thermophilic group II introns. PLoS Biol. 8, e1000391.

Mount, S.M., Gotea, V., Lin, C.-F., Hernandez, K., Makałowski, W., 2007. Spliceosomal small nuclear RNA genes in 11 insect genomes. RNA 13, 5–14.

Nakamura, Y., Kaneko, T., Sato, S., Ikeuchi, M., Katoh, H., Sasamoto, S., Watanabe, A., Iriguchi, M., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., Tabata, S., 2002. Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. DNA Res. 9, 123–130.

Nielsen, H., Johansen, S.D., 2009. Group I introns. Moving in new directions. RNA Biol. 6, 375–383.

Nielsen, H., Westhof, E., Johansen, S., 2005. An mRNA is capped by a 2′,5′ lariat catalyzed by a group I-like ribozyme. Science 309, 1584–1587.

Odom, O.W., Shenkenberg, D.L., Garcia, J.A., Herrin, D.L., 2004. A horizontally acquired group II intron in the chloroplast *psbA* gene of a psychrophilic Chlamydomonas: *in vitro* self-splicing and genetic evidence for maturase activity. RNA 10, 1097–1107.

Parra, M.K., Tan, J.S., Mohandas, N., Conboy, J.G., 2008. Intrasplicing coordinates alternative first exons with alternative splicing in the protein 4.1R gene. EMBO J. 27, 122–131.

Parra, M.K., Gallagher, T.L., Amacher, S.L., Mohandas, N., Conboy, J.G., 2012. Deep intron elements mediate nested splicing events at consecutive AG dinucleotides to regulate alternative 3′ splice site choice in vertebrate 4.1 genes. Mol. Cell. Biol. 32, 2044–2053.

Patel, A.A., Steitz, J.A., 2003. Splicing double: insights from the second spliceosome. Nat. Rev. Mol. Cell Biol. 4, 960–970.

Pontecorvo, G., Roper, J.A., Hemmons, L.M., MacDonald, K.D., Bufton, A.W.J., 1953. The genetics of *Aspergillus nidulans*. Adv. Genet. 5, 141–238.

Ragg, H., 2011. Intron creation and DNA repair. Cell. Mol. Life Sci. 68, 235–242.

Rest, J.S., Mindell, D.P., 2003. Retroids in archaea: phylogeny and lateral origins. Mol. Biol. Evol. 20, 1134–1142.

Rino, J., Carmo-Fonseca, M., 2009. The spliceosome: a self-organized macromolecular machine in the nucleus? Trends Cell. Biol. 19, 375–384.

Rogozin, I.B., Carmel, L., Csuros, M., Koonin, E.V., 2012. Origin and evolution of spliceosomal introns. Biol. Direct 7, 11.

Romfo, C.M., Alvarez, C.J., van Heeckeren, W.J., Webb, C.J., Wise, J.A., 2000. Evidence for splice site pairing via intron definition in *Schizosaccharomyces pombe*. Mol. Cell. Biol. 20, 7955–7970.

Roy, S.W., Irimia, M., 2009. Mystery of intron gain: new data and new models. Trends Genet. 25, 67–73.

Roy, S.W., Irimia, M., 2012. Genome evolution: where do new introns come from? Curr. Biol. 22, R529–R531.

Rudski, S.M., Hausner, G., 2012. The mtDNA *rps3* locus has been invaded by a group I intron in some species of *Grosmannia*. Mycoscience 53, 471–475.

Sarabhai, A.S., Stretton, A.O.W., Brenner, S., Bolle, A., 1964. Co-linearity of the gene with the polypeptide chain. Nature 201, 13–17.

Scamborova, P., Wong, A., Steitz, J.A., 2004. An intronic enhancer regulates splicing of the twintron of *Drosophila melanogaster prospero* pre-mRNA by two different spliceosomes. Mol. Cell. Biol. 24, 1855–1869.

Sheveleva, E.V., Hallick, R.B., 2004. Recent horizontal intron transfer to a chloroplast genome. Nucleic Acids Res. 32, 803–810.

Suzuki, H., Kameyama, T., Ohe, K., Tsukahara, T., Mayeda, A., 2013. Nested introns in an intron: evidence of multi-step splicing in a large intron of the human dystrophin pre-mRNA. FEBS Lett. 587, 555–561.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL-W – improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Thompson, M.D., Copertino, D.W., Thompson, E., Favreau, M.R., Hallick, R.B., 1995. Evidence for the late origin of introns in chloroplast genes from an evolutonary analysis of the genus *Euglena*. Nucleic Acids Res. 23, 4745–4752.

Toor, N., Keating, K.S., Taylor, S.D., Pyle, A.M., 2008. Crystal structure of a self-spliced group II intron. Science 320, 77–82.

Vallès, Y., Halanych, K.M., Boore, J.L., 2008. Group II introns break new boundaries: presence in a bilaterian's genome. PLoS ONE 3, e1488.

van der Burgt, A., Severing, E., de Wit, P.J.G.M., Collemare, J., 2012. Birth of new spliceosomal introns in fungi by multiplication of introner-like elements. Curr. Biol. 22, 1260–1265.

Vitikainen, M., Arvas, M., Pakula, T., Oja, M., Penttilä, M., Saloheimo, M., 2010. Array comparative genomic hybridization analysis of *Trichoderma reesei* strains with enhanced cellulase production properties. BMC Genom. 11, 441.

Waring, R.B., Scazzocchio, C., Brown, T.A., Davies, R.W., 1983. Close relationship between certain nuclear and mitochondrial introns. Implications for the mechanism of RNA splicing. J. Mol. Biol. 167, 595–605.

Weir, M., Rice, M., 2004. Ordered partitioning reveals extended splice-site consensus information. Genome Res. 14, 67–78.

Yanofsky, C., Carlton, B.C., Guest, J.R., Helinski, D.R., Henning, U., 1964. On the colinearity of gene structure and protein structure. Proc. Natl. Acad. Sci. U.S.A. 51, 266–272.

Zhang, L., Jenkins, K.P., Stutz, E., Hallick, R.B., 1995. The *Euglena gracilis* intron-encoded *mat2* locus is interrupted by three additional group II introns. RNA 1, 1079–1088.

Please cite this article in press as: Flipphi, M., et al. Spliceosome twin introns in fungal nuclear transcripts. Fungal Genet. Biol. (2013), http://dx.doi.org/10.1016/j.fgb.2013.06.003