

Egyetemi doktori (PhD) értekezés tézisei

**Spliceoszómális iker-intronok (stwintronok)
detektálása és vizsgálata fonalas gomba genomokban**

**Investigation and detection of spliceosomal twin
introns (stwintrons) in filamentous fungi genomes**

Ág Norbert

Témavezető/Supervisor: Dr. Karaffa Levente



DEBRECENI EGYETEM
Juhász-Nagy Pál Doktori Iskola
Debrecen
2016

1. Bevezetés

Az időben 1977-ig kellene visszaigrani, amikor először találhatunk említést olyan rejtélyes szekvenciákról a genomban, melyek fehérjekódoló génekben lokalizálódnak, viszont az mRNS érése során ezek a szekvenciák furcsa módon eltűnnek, kivágásra kerülnek a processzálódnó RNS szakaszból, tehát nem tartoznak a fehérjekódoló régióhoz. Az értekezésemet alkotó egyik sarokkő lerakása Chow és Berget munkájának köszönhető, azzal hogy adenovírusban felismerték és publikálták ezeknek a genomi szakaszoknak a létezését (Berget és munkatársai, 1977; Chow és munkatársai, 1977). A titokzatos szekvenciák gyorsan az állhatatos tudományos érdeklődés középpontjába kerültek, és a történetükben a következő lépcsőfokot az elnevezésük jelentette. A mozzanat az amerikai biokémikus, Walter Gilbert nevéhez köthető, aki a szekvenciák legtömörebb jellemzéséből, az angol intragenic region szavakból javasolta az intron fogalom bevezetést a tudományos szférába (Gilbert, 1978).

A majd másfél évtizeddel később a következő mérföldkő az intronok egyre gyarapodó ismeretanyagában a twintron fogalom definiálása. A twintron szó a twin intron szavak összevont alakja, magyarul iker-intront jelent. Olyan komplex intron szakaszokról beszélünk, ahol több intron együttesen egy nagyobb intront, azaz twintront alkot. Copertino és Hallick nevéhez köthető a definíció megalkotása, Euglena gracilis kloroplasztjában fedeztek fel GroupII/GroupIII intronokból felépülő összetett intront, melyet twintronnak neveztek el (Copertino és Hallick, 1991). Az említett kutatók nevével későbbi, hasonló témájú publikációkban gyakran találkozhatunk.

Az intronok felfedezése óta szűk negyven év telt el, ami a tudomány és a tudományos felfedezések időszámításának mércéjén mérve óriási idő. Elég csak az intronokkal kapcsolatos új ismeretek megjelenésére gondolni, mint az intronok csoportosítása: Group I, Group II, Group III intronok, vagy a spliceoszómális intronok. Megemlíthetjük a több száz fehérjéből és öt snRNS alegységből felépülő, a spliceoszómális intronok eltávolítását végző precíz molekuláris gépezetet, a

spliceoszómát. Az új ismeretek közé sorolhatjuk a splicingot, az alternatív splicingot; vagy a több intronból felépülő összetett intronokat, a twintronokat, melyek közötti elemi strukturális különbségek további csoportosításra adnak lehetőséget. Thomas Cech nevét még meg sem említettük, aki felismerte, hogy a saját szekvenciájuk kivágását katalizáló RNS molekulák is rendelkezhetnek enzimatis tulajdonságokkal, vagyis lehetnek ribozimek. Az általa vizsgált RNS molekulák intronok voltak (Kruger és munkatársai, 1982).

Ezekhez az ismeretekhez tudtam egy „építőelemmel” hozzájárulni, azzal hogy a Tanszékünk laboratóriumában elvégzett kísérletekkel részt vettem a kutatócsoportunk által megalkotott „stwintron” fogalom létrehozásában. Az „stwintron” koncepció alatt olyan spliceoszómális intronokat értünk, ahol egy külső intron kivágásához esszenciális szekvenciát megszakít egy beékelődő, úgynevezett belső intron. A speciális szerkezetből adódóan egymást követően két splicing reakció lejátszódása szükséges, így először a belső, majd a külső intron eltávolítása történik meg az RNS érése közben (Flipphi és munkatársai, 2013).

PhD hallgatóként volt szerencsém több tanulmányi úton is részt venni, az egyik alkalommal Hollandiában, az utrechti CBS-KNAW Fungal Biodiversity Centre működésével ismerkedtem meg. Az intézet rendelkezik a Föld egyik legnagyobb törzsgyűjteményével, több mint 100.000 fajjal, köztük gombákkal, élesztőkkel, baktériumokkal. A tevékenységi körükbe beletartozik a mikroorganizmusok megszekvenálása is. A DNS szekvenálás megjelenésével és elterjedésével arányosan nő az intronok létezése, eredete, vagy funkciója után potenciális kutatási teret adó, ismert DNS szekvenciák száma. Az első teljes megszekvenált genom a phi X 174 bakteriofágé volt, a projekt befejezése 1977-ben, intronok felfedezésének évében történt (Sanger és munkatársai, 1977). Az ezredforduló után a Human Genome Project megvalósításával az emberi genom megszekvenálása is befejeződött. Ez azért lényeges, mert a gerincesekben nagy arányban fordulnak elő intronok, az emberben szinte minden génre jut ~8 intron, ami nem mellékesen az alternatív splicing különböző típusaival a genetikai variabilitást növeli, másrészt ismeretük segíthet

megérteni az intronokkal kapcsolatos kérdéseket (Sakharkar és munkatársai, 2004).

A kutatók mindig is bizonytalanok voltak a spliceoszómális intronok eredetét illetően. Különböző elméletek alakultak ki a létezésük magyarázatára, ilyen az intron „korai” és intron „késői” elmélet, melyek az intronok evolúcióban való megjelenésének időpontját feszegetik (Koonin, 2006). A jövő egyik lehetősége lenne olyan bioinformatikai algoritmusok használata keresőszoftverek segítségével, melyek lehetővé teszik a genomok szisztematikus vizsgálatát, így akár tudományos áttörések elérését ezen a területen.

Az értekezésben új típusú beszűrődő szekvenciákat fogok bemutatni, ahol a twintron elnevezés sensu strictu megfelelő.

2. Anyagok és módszerek

2.1 Adatgyűjtés

Az *Aspergillus nidulans* PIH1, AMFS és BioDA fehérjéket kódoló szekvenciák adták a keresési referenciát a TBLASTN használatával azonosított homológ génekhez és EST információkhoz (Altschul és munkatársai, 1997).

Az átvizsgált adatbázisok:

- National Center for Biotechnology Information nem-redundáns nukleotid (nr/nt),
- MycoCosm az US Dept. of Energy Joint Genome Institute-től (Grigoriev és munkatársai, 2012)
- *Sclerotinia sclerotium* Sequencing Project
- Origin of Multicellularity database of the Broad Institute of Harvard and MIT
- Institut National de la Recherche Agronomique *Botrytis cinerea* T4 genom adatbázisa (Amselem és munkatársai, 2011).

A gén modellek és proteinek manuálisan lettek kikövetkeztetve, lehetőség szerint EST adatokkal megerősítve.

A kutatásban használt *H. solani* B-AC-16A (Mattupalli és munkatársai, 2014) genom szekvenciája elérhető a National Center for Biotechnology Information (NCBI) Whole Genome Shotgun kontigokban Master Accession AWWW000000000.

2.2 DNS illesztés

A genomi *bioDA* szekvenciák illesztése Clustal-W (Thompson és munkatársai, 1994) segítségével történtek. Az illesztések eredménye manuális korrekcióra került majd árnyékolt háttérrel lett elmentve a BOXSHADE (http://www.ch.embnet.org/software/BOX_form) program használatával.

2.3 Filogenetika

A fehérje szekvenciák a MAFT program 6. verziójával (Kato és Toh, 2008) lettek összehasonlítva a G-INS-I algoritmus alkalmazásával. A szekvencia illesztések informatív régiói a Block Mapping and Gathering using Entropy (BMGE) program segítségével kerültek kiválasztásra (Criscuolo és Gribaldo, 2010) (<http://mobyle.pasteur.fr/cgi-bin/portal.py#forms::BMGE>) a BLOSUM 85 hasonlósági mátrixszal. A Maximum

Likelihood fa PhylML (Guindon és munkatársai, 2010) (http://www.phylogeny.fr/version2_cgi/alacarte.cgi) használatával készült, majd adaptálva lett Adobe Illustrator programba egy FigTree (<http://tree.bio.ed.ac.uk/software/figtree>) ábrából. Az Approximate likelihood ratio test (aLRTs) részletesen bemutatva Anisimova és Gascuel publikációjában (2006).

2.4 Gomba törzsek, táptalaj, és növekedési paraméterek

A felhasznált gomba törzsek:

- *Fusarium verticillioides* FRC M-3125 és 57-7-7 (Brown és munkatársai, 2005)
- *Trichoderma reesei* QM9414 (Vitikainen és munkatársai, 2010)
- *Botrytis cinerea* T4 (Amselem és munkatársai, 2011)
- *Helminthosporium solani* B-AC-16A (Mattupalli és munkatársai, 2014)

A *Fusarium verticillioides* törzset Aspergillus Minimal Medium (AMM) 10g/L glükózt és 0,5 g/L bakto-peptont tartalmazó táptalajon tartottuk fenn. A micélium 500 ml térfogatú Erlenmeyer-lombikban növekedett 100 mL Komplex táptalajon (AMM + 2% (w/v) malt extract, 1% bacto-peptone és 2% glükóz) vegetatív spórákról indulva, 25 °C hőmérsékleten, 200 fordulat per perc (rpm) paramétereken, Infors HT Multitron rázóasztalon.

A *Trichoderma reesei* törzsfenntartása malt extract agaron (Difco) történt. A micélium 500 mL térfogatú Erlenmeyer-lombikban növekedett 30 °C és 250 rpm paramétereken. A folyékony táptalaj (Mandels, 1978) 10g/L glükózt tartalmazott szénforrásként.

A törzsfenntartás a *Botrytis cinerea* esetében potato dextrose agaron (Difco) valósult meg. A micélium 25 °C és 200 rpm paraméterek mellett, 500 mL térfogatú Erlenmeyer-lombikban növekedett, ami 100 mL PDB (Potato Dextrose Broth, Difco) táptalajt tartalmazott.

A *Helminthosporium solani* gombát malt extrakt agaron (LAB M Limited, UK) növesztettük. A micélium 100 mL, 1%(w/v) élesztőkivonatot, 0,5% malt extraktot,

0,3% peptont és 0,3% dextrózt tartalmazó folyékony táptalajon (Mimee és munkatársai, 2011) növekedett, 500 mL térfogatú Erlenmeyer-lombikban. A tenyésztés paraméterei rázóasztalon (Infors HT Multitron): 22 °C és 200 rpm.

2.5 Nukleinsav izolálás

Minden egyes gombamicéliumot folyékony táptalajon történő növesztés után szűréssel, desztillált vizes mosással, majd fagyasztással készítettük elő a közvetlen izoláláshoz. A micéliumot folyékony nitrogénben gondosan roncsoltuk és zúztuk, majd a megfelelő mennyiségű és állapotú micéliumtömeget a Kit-ek használatával dolgoztuk fel. Genomi DNS, plazmid DNS, totál RNS izolálásokhoz Promega Kit-eket (Wizard Genomic DNA Purification Kit, PureYield Plasmid Miniprep System, és PureYield RNA Midiprep System) használtunk a gyártó utasításainak szigorú betartásával.

2.6 Reverz transzkriptáz PCR (RT-PCR)

A reverz transzkripció 1 µg totál RNS és Oligo(dT) primer használatával ment végbe a High Fidelity cDNA Synthesis Kit (Roche) instrukciói alapján. A PCR

végterfogatai 25 µl-ek voltak, melyek tartalmaztak 4 µl egyszálú cDNS-t, a génspecifikus oligonukleotidokat, és a DreamTaq DNS polimerázt (Thermo Scientific). A PCR a 2 perc elődenaturálás után 40 ciklussal és a végén egy 5 perces post-elongációval, a következő programmal futott

1. 95 °C 2 min
2. 95 °C 30 sec
3. 60 °C 1 min
4. 72 °C 30 sec
5. 72 °C 5 min.

A reakció befejeztével az amplifikálódott fragmenteket 3% TBE gélben detektáltuk etídium-bromid használatával.

2.7 cDNS szekvenálás

A duplaszálú cDNS a gélből való tisztítás után klónozásra került (pGEM-T Easy Vector System I, Promega). Három egymástól független klónt küldtünk szekvenálásra (MWG-Biotech AG, Ebersberg, Németország). A szekvenciák raktározásra kerültek a GenBank által megadott nyilvántartási számok alatt.

3. Az értekezésben szereplő új tudományos eredmények

Fonalas gomba genomokban spliceoszómális iker-intronok kutatása és karakterizálása közben megalkottuk az „stwintron” fogalmat, és a következő fontosabb, új tudományos eredmények születtek:

1. Kísérletes úton bebizonyítottunk a *Fusarium verticillioides* két, az 57-7-7 és FRC M-3125 törzsek *pih2* génjében található stwintron szakaszok [D1,2] mechanizmuson keresztül való eltávolítását a genomból.
2. A *Trichoderma reesei bioDA* génjében felismert stwintron szakasz belső intronja a külső intron donor helyét a második és a harmadik nukleotid között szakítja meg. Ezt a tényt, és a [D2,3] útvonalon történő stwintron kivágódást beigazoltuk.
3. *Botrytis cinerea* T4 *pih1* génjében egy olyan alternatív splicingot ismertettünk, ahol a két

alternatív intron azonos donor helyet használ, de az akceptor helyeik különböznek.

4. A *Helminthosporium solani aox* génjében lokalizálódó stwintron eltávolítása két, egy [D1,2] és egy a [A2,3] útvonalon is megtörténhet. Mindkét reakció ugyanazt az érett RNS-t eredményezi. Az első olyan stwintron, ahol az akceptor helyet szakítja meg a belső intron.
5. A *Helminthosporium solani* putatív aldóz mutarotáz génjében bizonyítottuk az stwintron [D1,2] mechanizmuson keresztüli kivágódását, illetve egy feltételezett alternatív intron kivágódását az stwintron szakaszból, ami egy korai stop kodonnal rendelkező, nonsense mRNS szálát eredményez.

4. Összefoglalás

Munkánk során bevezettük az „stwintron” fogalmat, amit spliceoszómális iker-intronok (twintronok) rövidítéseként használunk. Az eredményeink első ízben mutatják be az eddig ismeretlen stwintronok létezését, azaz olyan eseteket, ahol egy belső intron kivágása szükséges egy működőképes splicing szekvencia létrehozásához, majd az azt követő külső intron eltávolításához, ami összességében a megfelelően érett mRNS szakaszt eredményezi. A korábban bemutatott, donor vagy akceptor helyen megszakított intronok egymást követő splicingja az intron definíció egy speciális mechanizmusát alkalmazza, többek között az első kivágás után történő újbóli RNS scanning-et és a hasító helyek újraegyesülését. Az általunk leírt esetek az 5' donor és a 3' akceptor szekvenciákat érintik, de nincs elvi kizáró oka olyan stwintron szerkezetek létének, ahol a lariat, más néven elágazási hely szakaszát szakítja meg egy beékelődő intron.

Fusarium verticillioides törzsekben sikerült az első stwintron struktúrát a *pih2* génben felismernünk, és

bizonyítanunk. A szerkezet eltávolítása [D1,2] – a beékelődő intron a külső intron donor helyét az első és második nukleotid között szakítja meg - útvonalon valósul meg.

A következő stwintron megvizsgálásának tárgya a *Trichoderma reesei bioDA* génjében található szekvencia volt. A *Fusarium* eredményekhez hasonló konklúziót vontunk le, azzal a különbséggel, hogy az stwintron kivágása [D2,3] útvonal megy végbe.

A *Botrytis cinerea* gomba *pihl* génjében egy [D4,5] felépítésű stwintron elrendezést feltételeztünk a génszekvencia alapján. A szerkezet pontos felderítésére használt PCR stratégia meglepő kísérleti eredményeket hozott. Megállapításra került, hogy a gén intron szerkezetének eltávolítása nem [D4,5] útvonalon történik, hanem egymást kölcsönösen kizáró, azonos donor helyeket használó alternatív intronok kivágása történik. A rövidebb alternatív intron eltávolítása egy csonka fehérjét eredményezne a szakaszban található korai stop kodon miatt.

Két spliceoszómális stwintronnak igazoltuk a létezését és megvizsgáltuk a kivágását a

Helminthosporium solani gombában. Az alternatív oxidáz génben detektált stwintron kontextusát közelebbről megvizsgálva arra a következtetésre jutottunk, hogy a struktúra eltávolítása két útvonalon is lehetséges. Ezek az útvonalak a [D1,2] és [A2,3] mechanizmusokat jelentik, melyek egyazon érett mRNS szekvenciát eredményezik, amit kísérletesen bizonyítottunk. A mutarotáz gén szakaszában is egy hasonló felépítést véltünk felfedezni azzal a különbséggel, hogy az egyik útvonal egy korai stop kodont tartalmazó fals köztiterméket eredményez. Egy ilyen RNS szárlól nem szintetizálódik működőképes fehérje, transzláció szempontjából teljesen használhatatlan ez a produktum. Elképzelhető, hogy a termék eliminálása az NMD mechanizmus működésének tudható be. Ebből adódóan arra következtethetünk, hogy együttesen az NMD mechanizmus működése és az stwintron kivágási útvonalak preferenciája egy poszt-transzkripciós génszabályozás lehet. Az állítást a *B. cinerea pih1* génjében tapasztaltak is alátámasztják, ahol megfigyeltük, hogy az alternatív splicing eredményezte

korai stop kodont tartalmazó RNS szakasz egy idő után majdnem detektálhatatlanná vált.

A kutatás következő lépcsőfoka lehetne a hozzáférhető genomok szisztematikus átvizsgálása számítógépes algoritmusok létrehozásával. Ily módon egy sokkal komplexebb képet kaphatnánk a spliceoszómális stwintronok elterjedésével kapcsolatban, ami hozzájárulna evolúciós folyamataik felderítéséhez, és megértéséhez.

5. 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. 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, 2003). 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; 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).

The origin of spliceosome introns is a vexing problem. Similarities in splicing mechanism with that of group II introns, supported by recent structural studies, 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., Valles 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. 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).

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.

6. Materials and methods

6.1 Data mining

Homologous genes and ESTs were mined upon TBLASTN (Altschul et al., 1997) screening using the *Aspergillus nidulans pih1*, AMFS and BioDA proteins 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.

6.2 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).

6.3 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) (<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).

6.4 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).
- *B. cinerea* strain T4 (Amselem et al., 2011).
- *H. solani* B-AC-16A (Mattupalli et al., 2014)

F. verticillioides was maintained on Aspergillus Minimal Medium (AMM) plates 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.

6.5 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.

6.6 Reverse transcription PCR (RT-PCR)

Reverse transcription was primed off 1 μ g 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 μ L volume containing 4 μ L of single strand cDNA, using gene-specific oligonucleotides as primers and DreamTaq DNA Polymerase (Thermo Scientific).

Cycling conditions:

1. 95 °C	2 min
2. 95 °C	30 sec
3. 60 °C	1 min
4. 72 °C	30 sec
5. 72 °C	5 min.

Amplified fragments were resolved in a 3 % native agarose gel and stained with ethidium bromide.

6.7 cDNA sequencing

PCR was primed off the primary RT product using gene-specific oligos 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.

7. New Scientific results

We have created the „stwintron” concept while we have researched and characterized spliceosomal twin introns in filamentous fungus and the following new scientific results were born:

1. We shown experimentally the removing of a stwintron sequence through a [D1,2] pathway in the *pih2* gene of the 57-7-7 and FRC M-3125 strains of *Fusarium verticillioides* fungus.
2. The internal intron of the recognized stwintron in the *bioDA* gene of *Trichoderma reesei* interrupts the external intron between the second and third nucleotide. This fact and the splicing of this stwintron by a [D2,3] mechanism are verified experimentally.
3. In the *pih1* gene of *Botrytis cinerea* T4 fungus an alternative splicing was identified where both alternative introns use the same donor site but their acceptor sites are different.

4. The stwintron which is localized in the *aox* gene of *Helminthosporium solani* has two possible pathways, a [D1,2] and a [A2,3] for its splicing mechanism. Both splicing reactions lead to identical matured RNA. The demonstrated [A2,3] instance is the first precedent where the acceptor site is interrupted by the internal intron.
5. The stwintron of the putative aldose mutarotase gene of *Helminthosporium solani* has a similar experimentally confirmed [D1,2] way to splice out the whole structure. Also the excision of a supposed alternative intron from the stwintron configuration is possible but this reaction produces a nonsense mRNA with an early stop codon in its sequence.

8. Summary

In this study we have created the “stwintron” concept which is used as abbreviation of the words spliceosomal twin introns. 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 or acceptor-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- and 3' acceptor sequences, but there seem to be no a priori reasons internal introns interrupting the lariat branchpoint sequences of the host intron could not be extant.

The first stwintron structure has been recognized and confirmed experimentally in the *pih2* gene of *Fusarium verticillioides* strains. The removing of the

structure occurs through a [D1,2] pathway, where the internal intron interrupts the donor sequence of the external intron between the first and second nucleotide.

The subject of the next investigation was a sequence in the bioDA gene of *Trichoderma reesei* fungus. We arrived to similar conclusion as in case of the *pih2* gene of the *Fusarium verticillioides* but the slight difference is the [D2,3] splicing pathway instead of the [D1,2].

According to the gene sequence a [D4,5] stwintron configuration was supposed to have discovered in the *pih1* gene of *Botrytis cinerea* fungus. The PCR strategy for the exploration of the exact sequence led to surprising experimental results. It was established that the complex intron structure is not removed by the proposed [D4,5] pathway but a mutually exclusive alternative splicing event is took place, where the same donor site is used by both alternative introns. The splicing of the shorter alternative intron resulted in a C-terminally truncated protein on account of the early stop codon in the coding sequence.

We have demonstrated the existence and investigated the excision of two new spliceosomal twin introns in *H. solani*. We concluded after a closer look at the context of the detected stwintron sequence in the alternative oxidase gene that there are two possible ways to splice out the whole stwintron structure. These pathways are the [D1,2] and [A2,3] mechanisms, which lead to identical matured mRNA sequence, the hypothesis is confirmed experimentally. A similar stwintron configuration was recognized in the aldose mutarotase coding sequence with the difference that one of the two pathways yields a false intermediate product with an early stop codon in its sequence. No functional protein is synthesized from such an RNA sequence. This kind of RNA is totally useless in the translation. It is possible that the elimination of this product can be attributed to the NMD mechanism. Consequently we concluded that the NMD mechanism combined with the preference of one of the two stwintron pathways can be a post-transcriptional gene regulation. The statement is also confirmed by the outcome in the *pih1* gene of the *B. cinerea*. In this case after the alternative splicing event

the resulted RNA sequence with its stop codon was almost undetectable at some point.

The next step of the research can be the systematic screening of the accessible genomes by computer algorithms. In this way a more complex picture could be gained in connection with the proliferation of the spliceosomal twin introns which would contribute the exploration and understanding of evolutionary processes of this gene sequences.

9. Irodalomjegyzék/References

- Altschul, S. F., Madden, T. L., Schaffer, 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 Research* **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. *Systematic Biology* **55**, 539-552.
- Berget, S. M., Moore, C., Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proceedings of the National Academy of Sciences* **74**, 3171–3175.
- 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 Genetics and Biology* **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.
- Chow, L. T., Gelinas, R. E., Broker, T. R., Roberts, R. J. (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. *The EMBO Journal* **10**, 433–442.
- Copertino, D., Hallick, R. (1993). Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends in Biochemical Sciences* **18**, 467 - 471.

- de Longevialle, A. F., Small, I. D., Lurin, C. (2010). Nuclearly Encoded Splicing Factors Implicated in RNA Splicing in Higher Plant Organelles. *Molecular Plant* **3**, 691-705.
- 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. *Molecular & general genetics* **264**, 682-690.
- 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? *Molecular Biology and Evolution* **15**, 76-86.
- Flippin, M., Fekete, E., Ag, N., Scazzocchio, C., Karaffa, L. (2013). Spliceosome twin introns in fungal nuclear transcripts. *Fungal Genetics and Biology* **57**, 48-57.
- Gilbert, W. (1978). Why genes in pieces? *Nature* **271**, 501-501.
- Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otiillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., Shabalov, I. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Research* **42**, D699-704.
- 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. *Systematic Biology* **59**, 307-321.

- Hallick, R. B., Hong, L., Drager, R. G., Favreau, M. R., Monfort, A., Orsat, B., Spielmann, A., Stutz, E. (1993). Complete sequence of *Euglena gracilis* chloroplast DNA. *Nucleic Acids Research* **21**, 3537-3544.
- Haugen, P., Simon, D. M., Bhattacharya, D. (2005). The natural history of group I introns. *Trends in Genetics* **21**, 111-119.
- Katoh, K., Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* **9**, 286-298.
- Koonin, E. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biology Direct* **1**, 22.
- Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., Cech, T. R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**, 147-157.
- 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 Harbor Perspectives in Biology* **3**, a003616.
- 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. *Molecular & general genetics* **246**, 128-131.
- Mandels, M. A., R.E. (1978). Problems and challenges in the cellulose to cellulase fermentation. *Process Biochemistry* **13**, 6-13.
- Mattupalli, C., Glasner, J. D., Charkowski, A. O. (2014). A Draft Genome Sequence Reveals the *Helminthosporium solani* Arsenal for Cell Wall Degradation. *American Journal of Potato Research* **91**, 517-524.
- Mimee, B., Avis, T. J., Boivin, S., Jabaji, S., Tweddell, R. J. (2011). Effect of iron and nitrogen on the development of *Helminthosporium solani* and potato silver scurf. *Canadian Journal of Plant Pathology* **33**, 506-511.
- Nielsen, H., Johansen, S. D. (2009). Group I introns: Moving in new directions. *RNA Biology* **6**, 375-383.
- 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. *Molecular and Cellular Biology* **32**, 2044-2053.
- 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. *The EMBO Journal* **27**, 122-131.
- Patel, A. A., Steitz, J. A. (2003). Splicing double: insights from the second spliceosome. *Nature Reviews Molecular Cell Biology* **4**, 960-970.

- Ragg, H. (2011). Intron creation and DNA repair. *Cellular and Molecular Life Sciences* **68**, 235-242.
- Rino, J., Carmo-Fonseca, M. (2009). The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends in Cell Biology* **19**, 375-384.
- Rogozin, I. B., Carmel, L., Csuros, M., Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. *Biology Direct* **7**, 11.
- Roy, Scott W., Irimia, M. (2012). Genome Evolution: Where Do New Introns Come From? *Current Biology* **22**, R529-R531.
- Sakharkar, M. K., Chow, V. T., Kanguane, P. (2004). Distributions of exons and introns in the human genome. *In Silico Biol* **4**, 387-393.
- Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, J. C., Hutchison, C. A., Slocombe, P. M., Smith, M. (1977). Nucleotide sequence of bacteriophage ϕ X174 DNA. *Nature* **265**, 687-695.
- 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 Letters* **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 Research* **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 evolutionary analysis of the genus *Euglena*. *Nucleic Acids Research* **23**, 4745-4752.
- 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., Collemare, J. (2012). Birth of new spliceosomal introns in fungi by multiplication of introner-like elements. *Current Biology* **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 Genomics* **11**, 1-16.
- 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. *Journal of Molecular Biology* **167**, 595-605.

10. Publikációs lista

Az értekezés alapjául szolgáló közlemények:

1. Flipphi M., Fekete E., Ág N., Scazzocchio C., Karaffa L. (2013): Spliceosome twin introns in fungal nuclear transcripts.
Fungal Genetics and Biology, 57: 48-57.
DOI:10.1016/j.fgb.2013.06.003
Impakt faktor: 3.262
2. Ág N., Flipphi M., Karaffa L., Scazzocchio C., Fekete E (2015): Alternatively spliced, spliceosomal twin introns in *Helminthosporium solani*.
Fungal Genetics and Biology, 85: 7-13.
DOI:10.1016/j.fgb.2015.10.004
Impakt faktor: 2.587

Az értekezés témájában elhangzott előadások:

1. Michel Flipphi, Erzsébet Fekete, **Norbert Ág**, Claudio Scazzocchio, Levente Karaffa: Spliceosome twin introns revealed by fungal nuclear genomes (Magyar Mikrobiológiai Társaság éves nagygyűlése, 2013. Keszthely).

2. **Ág Norbert:** Fungális stwintronok karakterizálása és azonosítása (Fiatal Biotechnológusok Országos Konferenciája, Szeged, 2014).
3. **Norbert Ág,** Michel Flipphi, Levente Karaffa, Gustavo Cerqueira, Claudio Scazzocchio, Erzsébet Fekete: Spliceosomal twin introns: their evolution and possible relevance to post-transcriptional regulation (13th European Conference on Fungal Genetics, Párizs, 2016. április).

Az értekezés témájában bemutatott posztterek:

1. **Ág N,** Karaffa L, Scazzocchio C, Flipphi M, Fekete E: A mechanism of alternative splicing in filamentous fungi (6th Congress of European Microbiologists, Maastricht, Hollandia, 2015).
2. **Ág N,** Flipphi M, Karaffa L, Scazzocchio C, Fekete E: Spliceosomal twin introns (stwintrons) in fungi (11th Molecular Biology of Fungi Conference, Berlin, Németország, 2015).