



1949

# **FUNGÁLIS IKER-INTRONOK (STWINTRONOK) EVOLÚCIÓS ÉS FUNKCIONÁLIS VIZSGÁLATA**

Egyetemi doktori (PhD) értekezés

Kavalecz Napsugár

Témavezető: Dr. Fekete Erzsébet

DEBRECENI EGYETEM

Természettudományi és Informatikai Doktori Tanács

Juhász-Nagy Pál Doktori Iskola

Debrecen, 2020

*Ezen értekezést a Debreceni Egyetem Természettudományi és Informatikai Doktori Tanács Juhász-Nagy Pál Doktori Iskola Biológia doktori programja keretében készítettem a Debreceni Egyetem természettudományi doktori (PhD) fokozatának elnyerése céljából. Nyilatkozom arról, hogy a tézisekben leírt eredmények nem képezik más PhD disszertáció részét.*

*Debrecen, 2020.*

.....  
*Kavalecz Napsugár*

*Tanúsítom, hogy Kavalecz Napsugár doktorjelölt 2016-2020 között a fent megnevezett Doktori Iskola Biológia doktori programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Nyilatkozom továbbá arról, hogy a tézisekben leírt eredmények nem képezik más PhD disszertáció részét. Az értekezés elfogadását javasolom.*

*Debrecen, 2020.*

.....  
*Dr. Fekete Erzsébet*

# FUNGÁLIS IKER-INTRONOK (STWINTRONOK) EVOLÚCIÓS ÉS FUNKCIONÁLIS VIZSGÁLATA

Értekezés a doktori (PhD) fokozat megszerzése érdekében  
a Biológia tudományágban

Írta: Kavalecz Napsugár okleveles biomérnök

Készült a Debreceni Egyetem Juhász-Nagy Pál Doktori Iskolája  
(Biológia doktori program) keretében

Témavezető: Dr. Fekete Erzsébet

Az értekezés bírálói:

Dr. ....  
Dr. ....  
Dr. ....

A bírálóbizottság:

elnök: Dr. ....  
tagok: Dr. ....  
Dr. ....  
Dr. ....  
Dr. ....

Az értekezés védésének időpontja: .....

# Tartalomjegyzék

1.	Bevezetés.....	1
2.	Irodalmi áttekintés.....	4
2.1.	Spliceoszómális intronok.....	4
2.1.1.	Intronok eredete, evolúciója, mobilitása.....	7
2.1.2.	Intronok lehetséges funkciói.....	9
2.1.3.	Alternatív splicing.....	10
2.2.	Twintronok.....	13
2.2.1.	Ribozim típusú twintronok.....	14
2.2.2.	Spliceoszómális twintron-szerű struktúrák.....	18
2.2.3.	Stwintronok.....	22
3.	Célkitűzések.....	25
4.	Anyagok és módszerek.....	26
4.1.	Bioinformatikai módszerek.....	26
4.1.1.	Putatív stwintron struktúrák keresése teljes genom szekvenciákban.....	26
4.1.2.	Ortológ gének vizsgálata (genome mining).....	29
4.1.3.	Másodlagos fehérjestruktúra meghatározása.....	30
4.1.4.	Filogenetikai analízis.....	31
4.2.	Gombatörzsek.....	33
4.3.	Nukleinsav izolálás.....	34
4.4.	Reverz transzkripció PCR (RT-PCR) reakciók.....	35
4.5.	Szekvenálás.....	36
5.	Eredmények.....	38
5.1.	DHA1 ortológok vizsgálata.....	38
5.1.1.	[D1,2] stwintron bizonyítása <i>A. nidulans dhaoS</i> génjében.....	38
5.1.2.	Stwintront tartalmazó <i>dhaoS</i> ortológok azonosítása.....	40
5.1.3.	[D1,2]/[A2,3] stwintron <i>T. reesei</i> DHA1 ortológjában.....	41
5.1.4.	A [D1,2]/[A2,3] stwintron belső intronjának alternatív elvesztése.....	44
5.2.	LipS ortológok vizsgálata.....	46
5.2.1.	Egy új [D1,2] stwintron <i>A. nidulans</i> és <i>A. niger</i> putatív lipáz génjében.....	46
5.2.2.	Ortológ gének vizsgálata.....	49
5.3.	Egy [D2,3] stwintront tartalmazó <i>bioDA</i> gén és BioDA ortológok vizsgálata.....	53
5.4.	Egy [D5,6] stwintron és lehetséges szerepeinek feltárása.....	58
5.4.1.	[D5,6] stwintron <i>A. nidulans rtnA</i> génjében.....	58
5.4.2.	Ortológok vizsgálata.....	60
5.4.3.	A [D5,6] stwintron és a szomszédos exon elvesztése.....	62
5.4.4.	Exon skipping jelenség kimutatása <i>A. niger rtnA</i> génjének transzkriptumában.....	64

5.4.5.	Exon skipping jelenség kimutatása <i>N. crassa rtnA</i> génjének transzkriptumában.....	67
5.4.6.	Egy duális komplex megszakító szekvencia feltárása <i>M. cinnamomea</i> -ban.....	69
6.	Eredmények megvitatása.....	74
6.1.	DHA1 ortológok és az egy nukleotidos intron elcsúszás vizsgálata egy alternatív stwintronon keresztül .....	74
6.2.	A BioDA és LipS ortológok, és stwintronok evolúciójának vizsgálata .....	76
6.3.	Egy stwintronhoz köthető alternatív splicing esemény és egy komplex megszakító szekvencia kimutatása .....	79
7.	Összefoglalás.....	81
8.	Summary.....	84
9.	Irodalomjegyzék.....	87
10.	Tudományos tevékenység.....	102
11.	Köszönetnyilvánítás.....	107
12.	Függelék.....	109
12.1.	A kísérletek során alkalmazott oligonukleotid primerek .....	109
12.2.	A vizsgált fehérjék filogenetikai fái .....	112
12.3.	Egyéb függelék.....	126

# 1. Bevezetés

A XX. század második felétől kezdődően rohamosan nőtt a molekuláris biológia és genetika területén szerzett ismereteink száma, melyek napjainkig intenzíven gyarapodnak. Az elmúlt pár tíz évben a kutatók olyan módszereket és eszközöket fejlesztettek ki, melyek lehetővé teszik, hogy célzottan módosítsuk egy-egy sejt vagy organizmus genetikai állományát, illetve válaszokat találjunk bizonyos genetikai betegségek kialakulásának mechanizmusaira.

Az stwintronok, sőt az intronok felfedezését megelőzően is már számos fontos megállapítás került leírásra a genetikai örökítőanyaggal kapcsolatosan, mely információk elengedhetetlenek voltak ahhoz, hogy a kutatók továbbhaladhassanak a molekuláris biológia, már részben kitaposott ösvényein. Az egyik legjelentősebb felfedezés valószínűleg a DNS szerkezetének meghatározása volt (Watson és Crick, 1953), melyen hosszú éveken át dolgozott James Watson, Francis Crick, Maurice Wilkins és Rosalind Franklin. A négy kutató közül Watson, Crick és Wilkins megosztott ovisi Nobel-díjat vehettek át 1962-ben. Rosalind Franklin munkásságát a tudományos társaság nem honorálta Nobel-díjjal, feltehetőleg részben korai halálának is köszönhetően; annak ellenére, hogy ő készítette el azt a világhírűvé vált röntgendiffrakciós képet, mely alapján a kutatók kikövetkeztethették a DNS kettős hélixének pontos szerkezetét (Maddox, 2013).

1993-ban Richard J. Roberts és Phillip A. Sharp felfedezéseit szintén Nobel-díjjal ismerték el, akik 1977-ben - egymástól függetlenül -

mutatták ki adenovírusokkal végzett kísérleteik során, hogy a gének több, egymástól nem kódoló szakaszokkal elválasztott szegmensből épülnek fel (Berget és mtsai., 1977; Chow és mtsai., 1977). A kutatók azt is megállapították, hogy ezek a DNS darabok nem képezik az érett mRNS részét, valamilyen módon abból eltávolításra kerülnek.

Mivel mindeztáig a kutatók úgy vélték, az mRNS molekulák minden esetben hű másolatai a kódoló DNS szálaknak, az 1977-ben leírt különös jelenség elkezdte foglalkoztatni a tudományos közéletet. Egy évvel később, 1978-ban Walter Gilbert fel is teszi a kérdést: „Why genes in pieces?”; vajon miért állnak darabokból a gének? Gilbert ezeket a kódoló régiókat megszakító szakaszokat „intron”-oknak nevezi, az intragenic region kifejezésből eredeztetve, míg a kódoló szegmenseket „exon”-oknak kereszteli, mint expressed sequences, azaz kifejeződő, átíródo szekvenciák (Gilbert, 1978).

Az intronok különböző típusait és lehetséges funkcióit tagláló publikációk mellett összetett, több intronból felépülő struktúrák leírására is sor került a XX. század végén. 1991-ben Donald W. Copertino és Richard B. Hallick *Euglena gracilis* alga kloroplasztisának genomjában írtak le különleges, két egymásba ékelődött intronból álló szekvenciákat, melyeket „twintron”-oknak neveztek el, a twin (iker) és intron szavak összetételével (Copertino és Hallick, 1991).

A Debreceni Egyetem Biomérnöki Tanszékén jómagam is hasonló struktúrákat vizsgáltam PhD hallgatóként. Mivel az általunk leírt twintronok spliceozómális intronokból épülnek fel „stwintron”-nak kereszteltük őket, mint spliceozómális iker-intronok (Flipphi és mtsai.,

2013). Ezek az összetett szerkezetek (hasonlóan az alga kloroplasztban leírtakhoz) két, egymást követő kivágási (splicing) lépés során kerülhetnek csak teljes eltávolításra az elsődleges mRNS átiratból.

Az elmúlt néhány évben több, különböző típusú stwintront is sikerült kimutatnunk fonalas gombák egyes génjeiben, beleértve a jelen értekezés alapját képező eredményeket is (Flippi és mtsai., 2013; Ág és mtsai., 2015; Fekete és mtsai., 2017; Flippi és mtsai., 2017; Kavalecz és mtsai., 2019). Kezdetben célunk minél több ilyen struktúra leírása volt, majd esetleges funkcióiknak a feltárása. Emellett remek modell rendszereknek tűnnek, melyek segítségével feltérképezhetőek az intronok evolúciós eseményi, valamint egyes alternatív splicing jelenségek lejátszódásához fűződő szerepük.

Úgy vélem, az stwintronok jelenléte nem korlátozódik a gombák országára; nagy valószínűséggel megtalálhatóak magasabb rendű eukarióta szervezetek génjeiben is. Ezen egyszerű organizmusok vizsgálata során elért eredményeink a későbbiekben azonban felhasználhatóak lehetnek akár humán genetikai betegségek, vagy aberráns splicing folyamatok mögött álló jelenségek megértéséhez is.



## 2. Irodalmi áttekintés

### 2.1. Spliceoszómális intronok

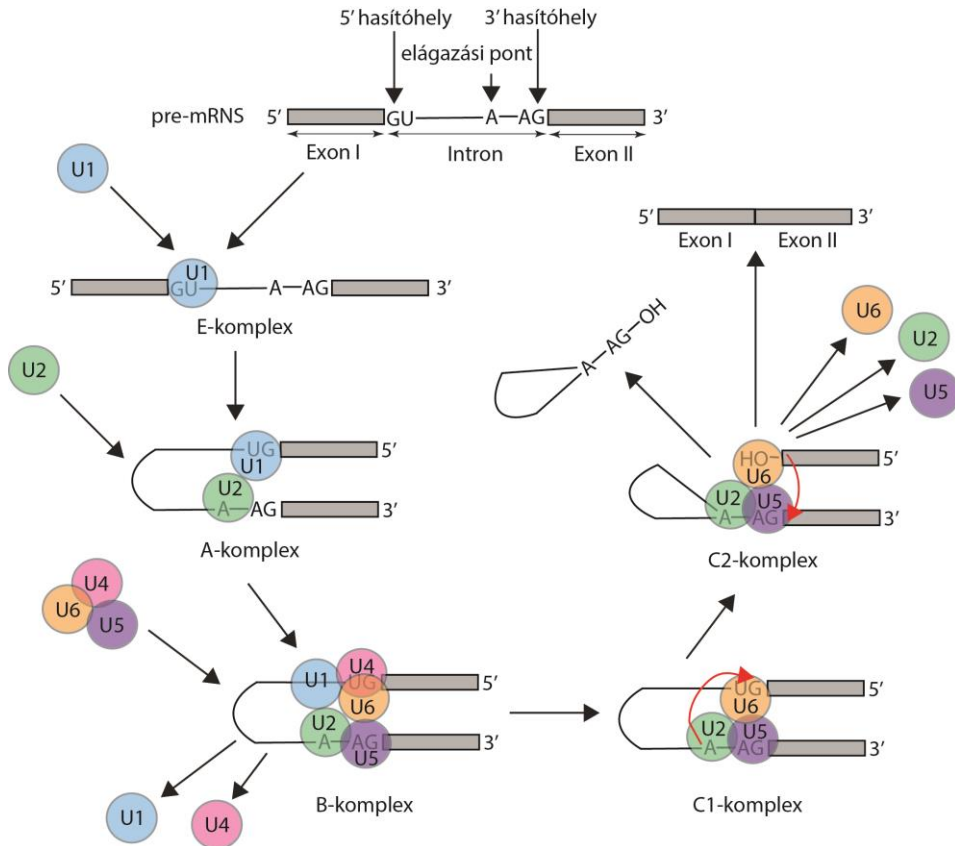
Az eukarióta szervezetek sejtmagi génjei mozaikos felépítésűek; a fehérjekódoló részeket, exonokat, nem kódoló szakaszok, intronok tagolják (Gilbert, 1978). A transzkripció során, az elsődleges mRNS átiratba még mind az exonok, mind az intronok átírásra kerülnek; utóbbiak később, az mRNS érése során, egy „splicing”-nak nevezett folyamat révén kivágásra, eltávolításra kerülnek (Irimia és Roy, 2014). A splicing folyamatot egy rendkívül összetett ribonukloprotein komplex, a spliceoszóma katalizálja (Nilsen, 2003; Rino és Carmo-Fonseca, 2009; Will és Lührmann, 2011; Papasaikas és Valcárcel, 2016; Yan és mtsai., 2019), míg az eukariótákra jellemző, kivágódásukhoz spliceoszómát igénylő megszakító szekvenciákat spliceoszómális intronoknak nevezzük. Az önhasító, ribozim aktivitással rendelkező, Group I és Group II típusú intronok jelenléte eukariótákban a mitokondriális, illetve kloroplasztisz genomra korlátozódik (Bonen és Vogel, 2001).

Kétféle spliceoszómális intront írtak le, az U2 és a jóval ritkábban előforduló U12 típust (Bartschat és Samuelsson, 2010). A két különböző intron fajta splicingját kettő, eltérő felépítésű spliceoszóma katalizálja; az U2-es intronok eltávolítását a fő-, míg az U12-esekét az úgynevezett minor spliceoszóma (Turunen és mtsai., 2013). Minden spliceoszóma öt, különböző snRNP-ből, kis nukleáris ribonukleoproteinből áll; az snRNP-k egy snRNS-ből és különböző fehérjékből épülnek fel. Az U2-típusú spliceoszóma esetében az U1, U2, U4, U5 és U6 snRNP-k (Yan és mtsai.,

2019), míg az U12-típusú, minor spliceoszóma esetén az U11, U12, U4atac, U6atac és U5 vesznek részt a makromolekuláris gépezet kialakításában, illetve a splicing folyamat katalizálásában (Turunen és mtsai., 2013).

A spliceoszóma bizonyos konzervált nukleotidokat, szekvenciaelemeket ismer fel az intronon belül, melyekkel interakcióba lép és megkezdődik a kivágási folyamat. A spliceoszómális intronokra jellemző splicing motívumok az 5' hasítóhely, azaz a donor szekvencia, a 3' hasítóhely, azaz az akceptor szekvencia és az elágazási pont, vagy lariat szekvencia. U2 típusú intronok esetén a leginkább konzervált nukleotidok, melyek a spliceoszómális intronok döntő hányadában megegyeznek, az 5' hasítóhely első két nukleotidja (GT), illetve a 3' hasítóhely utolsó két nukleotidja (AG), továbbá mindenképp jelen kell lennie egy adeninnek (A) az akceptor szekvencia közelében, mely elengedhetetlen a lasszó szerkezet kialakításához (Rogozin és mtsai., 2012).

Az intronikus szakaszok eltávolítása két, egymást követő transzészterifikációs reakció révén valósulhat meg, melyekben a fent bemutatott szekvenciaelemek és ezeket felismerve, a spliceoszóma vesz részt (1. ábra). A spliceoszómális komplexet felépítő, különböző protein komponensek is szerves részt vállalnak a kivágási folyamat lebonyolításában (Yan és mtsai., 2019).



**1. ábra: A spliceozómális intronok spliceozóma általi eltávolításának mechanizmusa** (saját szerkesztésű ábra Will és Lührmann, 2011 alapján)

Mivel elsősorban fonalas gombák genomjában vizsgáltam összetett, spliceozómális intron struktúrákat, fontos megemlíteni, hogy ezen megszakító szekvenciák egyedi sajátosságokat mutatnak a magasabb rendű eukarióta szervezetek örökítőanyagában feltárt intronokhoz képest. Kupfer és munkatársai, öt különböző élesztő-, illetve fonalas gombafaj genomjának, illetve intronjainak vizsgálata során arra a megállapításra jutottak, hogy a fungális intronok általában egészen rövidek, maximum pár száz nukleotid hosszúságúak. Ezenkívül, a polipirimidin traktus jellemzően az 5' hasítóhely és az elágazási pont szekvencia között

található meg bennük (míg például emlősökben a 3' hasítóhely közelében), viszont az 5' GT---AG 3' kanonikus struktúra rájuk is jellemző (Kupfer és mtsai., 2004). A fungális intronokban feltárt egyedi jellemzők miatt a kutatók joggal feltételezik, hogy splicing mechanizmusaikban is mutathatnak eltérést magasabb rendű, gerinces eukariótákhoz képest.

### **2.1.1. Intronok eredete, evolúciója, mobilitása**

A spliceoszómális intronok megjelenését eukarióta genomokban a mai napig rejtély övezi. A kutatók véleménye megoszlik eredetüket illetően; erre vonatkozóan két említésre érdemes elmélet látott napvilágot:

1. Az „introns early”, korai intron hipotézis szerint már az evolúció korai szakaszában is nagy számban jelen voltak intronok az eukarióta génekben. Sőt, jelenlétükkel lehetőséget adhattak rekombinációs események lejátszódására, ezáltal elősegítve új, összetettebb fehérjéket kódoló gének kialakulását.
2. Ezzel szemben az „introns late”, kései intron teória azt állítja, hogy az intronok az evolúció későbbi szakaszaiban jelentek meg az eukarióta genomokban, fehérjekódoló szekvenciákba történő beágyazódás révén (Koonin 2006).

Fontos megjegyezni, hogy a két elmélet alapvetően nem zárja ki egymást.

A spliceoszómális intronok ősei valószínűleg Group II típusú intronok lehettek. A ribozim intronok ugyan nem igénylik spliceoszóma

közreműködését kivágódásukhoz, mégis felfedezhetőek splicing mechanizmusaikban olyan hasonlóságok, melyek erre a feltevésre engednek következtetni (Lambowitz és Zimmerly, 2004). Azt is kimutatták, hogy a spliceoszómális intronok, ortológ gének adott pozícióiból eltűnhetnek, de új intronok is bármikor megjelenhetnek, tehát az intron-exon struktúra állandóan változik. Az intron szerzés és intron vesztés dinamikus egyensúlyban van egymással, az intron evolúció folyamatos (Carmel és mtsai., 2007).

Néhány gén esetén azt is észrevették, hogy az adott intron hasonló, közeli (maximum 15 nukleotiddal távolabbi), de nem azonos pozícióban található meg, különböző organizmusok genomját összehasonlítva. A megfigyelt jelenség alátámasztja, hogy az intronok képesek nemcsak eltűnni, vagy megjelenni, hanem el is „csúszni” a genomban, akár mindössze egy nukleotiddal is. A folyamatot angolul „intron sliding”, „intron drift”, vagy „intron shift”-ként szokták említeni (Rogozin és mtsai., 2000). Közeli rokonfajok ortológ génjeit összehasonlítva nagyfokú konzerváltságot tapasztalhatunk; sokkal ritkábban találkozunk intron elcsúszással, mint evolúciósan egymástól távolabb álló fajok intron pozícióit összevetve. Viszont érdekes, hogy a szekvenciák homológiája nem befolyásolja a jelenség lejátszódásának valószínűségét (Henricson és mtsai., 2010). Persze azt is nehéz megbecsülni, hogy ezekben az esetekben valóban az intron elcsúszásáról van-e szó, vagy intronvesztés, majd újbóli inszertáció történik (Stoltzfus és mtsai., 1997). Mindenesetre, paralóg gének összehasonlításának eredményeképp a kutatók számos esetet találtak intronok „nem-azonos” pozíciójú lokalizációjára (Kyndt és mtsai., 2008; Lehmann és mtsai., 2010). Az intron-sliding mechanizmus

létét továbbá egyértelműen alátámasztja a nem megegyező, de közeli pozícióban talált intronok szignifikáns szekvencia homológiája is (Sakharkar és mtsai., 2001).

### **2.1.2. Intronok lehetséges funkciói**

Annak ellenére, hogy az eukarióta intronokra hosszú éveken át „junk”, azaz szemét, hasztalan DNS szekvenciákként gondoltak, napjainkra egyre inkább bizonyosságot nyer, hogy fontos szerepük lehet számos mechanizmusban, többek között, a génexpresszió szabályozásában és a protein diverzitás növelésében (Parenteau és Elela, 2019).

Fontos megjegyezni, hogy a spliceoszómális intronok átírása, splicingja és az eltávolításuk utáni degradációja jelentős energiaráfordítást igényel a sejt részéről, ami önmagában felveti a kérdést, hogy ugyan miért lennének jelen a genomban, hogyha nincs semmiféle funkcionális hasznuk? Az intronok száma, főképp magasabb rendű eukariótákban számottevő; a humán genom mintegy egynegyedét teszik ki (Sakharkar és mtsai., 2004). Egyes elméletek szerint az intronok mindössze „önző” DNS-darabok, melyek előszeretettel „támadják meg” a fehérjekódoló géneket, eukarióta genomokban, ezzel akár egy populáció méretét is befolyásolva (Lynch, 2002).

A molekuláris biológiai és bioinformatikai módszerek fejlődésével egyre több lehetőségünk nyílik arra, hogy közelebbről is megismerjük a spliceoszómális intronokat, illetve lehetséges szerepeiket.

Az elmúlt időszakban bizonyosságot nyert, hogy intronok direkt módon részt vehetnek:

- alternatív splicing
- génexpresszió
- nonsense-mediated mRNA decay (NMD) mechanizmus
- mRNS transzport és kromatin szerveződés

szabályozásában, de emellett bizonyítottan bírhatnak számos indirekt funkcióval is (Jo és Choi, 2015).

### **2.1.3. Alternatív splicing**

Az alternatív splicing (AS) mechanizmus fontos szerepet tölt be a génexpressziós komplexitás és protein diverzitás növelésében. A teljes humán genom megszekvenálása előtti időkben a kutatók úgy vélték, hogy megközelítőleg 100.000 gén vehet részt az ember örökítőanyagának felépítésében, ám kiderült, hogy a valós szám a korábban becsültnek maximum mintegy negyede (International Human Genome Sequencing Consortium, 2004). A génekről átíródó fehérjék száma, azonban valóban sokszorososa ennek. A meglepő jelenségre a válasz részben a pre-mRNS molekulák alternatív splicingjában keresendő.

A mechanizmus lehetőséget nyújt arra, hogy egyetlen génről többféle fehérje is képződjön; az AS jelenség felfedezésével tehát megdőlt az „egy gén – egy fehérje” hipotézis is. Ma már tudjuk, hogy a több exont tartalmazó humán génekről képződő mRNS-ek mintegy 95%-a áteshet AS folyamatokon, így teoretikusan egynél többféle fehérje transzlálódhat

róluk (Pan és mtsai., 2008; Nilsen és Graveley, 2010). Szélsőséges esetekben egyetlen génről akár több ezer, különböző fehérje is keletkezhet az alternatív splicingnak köszönhetően. *Drosophila Dscam* génje esetén például, több mint 38.000 protein izoformát mutattak ki, melyek mind erről az egyetlen génről transzlálódnak (Schmucker és mtsai., 2000).

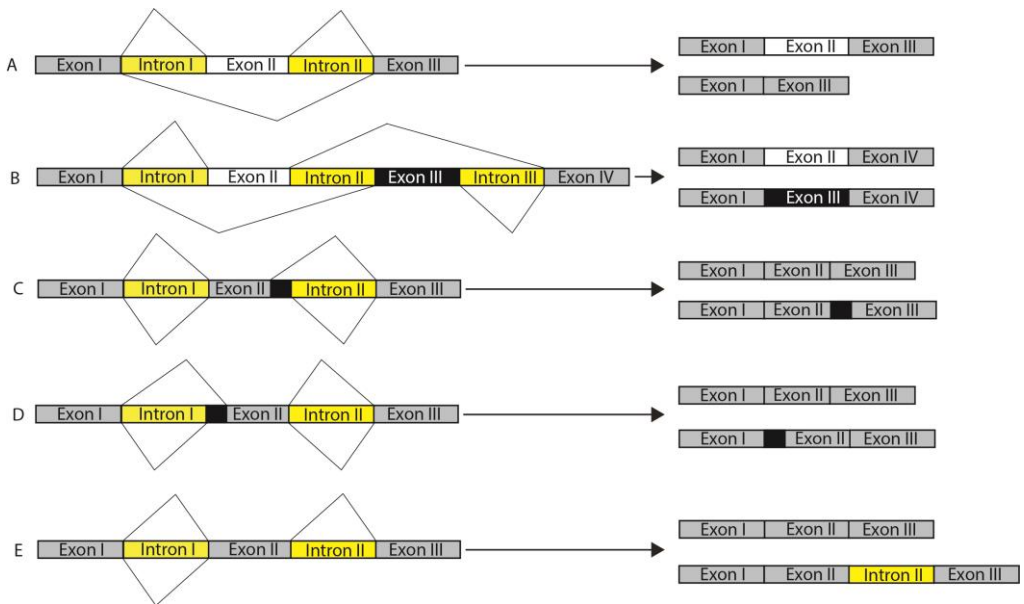
Az AS mechanizmus következtében, úgynevezett korai STOP kodont tartalmazó mRNS-ek is keletkezhetnek, melyekről nem tud fehérje termék transzlálódni. Ezen transzkriptumok többnyire a nonsense-mediated mRNA decay (NMD) mechanizmus által kerülnek felismerésre és lebontásra. AS jelenségek lejátszódásának, az NMD mechanizmussal karöltve tehát fontos szerepük lehet poszttranszkripcionális szintű génexpresszió szabályozásban is (McGlincy és Smith, 2008). Persze felmerül a kérdés, hogy az AS események vajon valóban az eukarióta sejtek egy jól bevált regulációs eszköze, vagy a funkcióképtelen mRNS termékek képződése egyszerűen csak a pontatlan splicing események elkerülhetetlen következménye (Saudemont és mtsai., 2017).

Az AS mechanizmusnak öt főbb típusát különböztethetjük meg (Wang és mtsai., 2015), melyek a következők (2. ábra):

- Exon skipping: az adott exon vagy benne marad az érett mRNS átíratban, vagy kivágásra kerül az őt határoló intronokkal együtt a pre-mRNS-ből.
- Egymást kizáró exonok: egy adott exon jelenléte a végső mRNS átíratban kizárja egy másik exon jelenlétét, és fordítva.



- Alternatív 3' hasítóhely
- Alternatív 5' hasítóhely: alternatív 5' vagy 3' hasítóhelyek használatával különböző méretű mRNS termékek képződhetnek.
- Intron retenció: az exonokkal együtt intron szekvencia is benne marad a végső, érett mRNS átiratban.



**2. ábra: Az alternatív splicing leggyakoribb típusai** (A: exon skipping, B: egymást kizáró exonok, C-D: alternatív 3' vagy 5' hasítóhelyek használata, E: intron retenció) (saját szerkesztésű ábra Wang és mtsai., 2014 alapján)

A felsorolt jelenségek közül magasabb rendű eukariótákra és gerinces élőlényekre az exon skipping a leginkább jellemző, míg az intron retenció a legritkábban előforduló AS esemény; ezzel szemben alacsonyabb rendűek esetén pont az utóbbi típus az, amelynek lejátszódását a legnagyobb számban tudták igazolni (Kim és mtsai., 2007). Alternatív splicing élesztőgombákban nagyon ritkán fordul elő, ami az

élesztőgenomok csekély intron tartalmával is magyarázható (Kempken, 2013). Ezzel szemben kutatócsoportunk vizsgálati alanyai, a fonalas gombák esetén, transzkriptom analízisek elvégzésével, számos AS jelenséget leírtak a közelmúltban (Wang és mtsai., 2010; Kempken, 2013; Grützmann és mtsai., 2014; Xie és mtsai, 2015).

## 2.2. Twintronok

Olyan, egymásba ékelődött, összetett intron struktúrákat nevezünk „twintron”-nak, melyek kivágódása több, egymást követő splicing reakció révén mehet csak végbe. A szóösszetételt Copertino és Hallick alkották meg, a twin (iker) és az intron szavak összevonásával (Copertino és Hallick, 1991). A kutatók *Euglena gracilis* egysejtű alga kloroplasztisz genomjában írtak le twintronokat, melyek group II, illetve group III típusú intronokból épülnek fel (Copertino és Hallick, 1993).

A group II intronok ribozimnak tekinthetők; saját kivágódásukat katalizálják transzészterifikációs reakciók révén. Ilyen típusú intronokat gombák, növények, protiszták sejtorganelleimainak riboszómális-, transzfer-, és messenger RNS-ében, valamint egyes baktériumok mRNS-ében írtak le (Bonen és Vogel, 2001). Group III intronokat gyakorlatilag csak az euglenoid kloroplasztisz genomban találhatunk; szerkezetüket tekintve a group II intronok degradálódott, lerövidült változatainak tekinthetők (Christopher és Hallick, 1989).

A ribozim intronok közé tartoznak a Group I intronok is, melyeket főleg protiszták nukleáris genomjában, baktériumokban, bakteriofágokban, esetleg gombák mitokondriális genomjában találhatunk

meg (Haugen és mtsai., 2005). Group I típusú intronok is részt vehetnek twintronok kialakításában, akár group II intronokkal együtt, „kevert” struktúrákat létrehozva (Zumkeller és mtsai., 2020).

A tRNS intronok főképp Archeák genomjában fordulnak elő, tRNS molekulákban. Kimutattak egymásba ékelődött tRNS intronokból felépülő twintronokat is, például *Thermofilum pendens* Archea fajban (Tocchini-Valentini és mtsai., 2009).

Kizárólag spliceozómális intronokból felépülő twintronokat (stwintronokat) munkacsoportunk írt le először, fonalas gombák genomjában (Flipphi és mtsai., 2013). Az első ilyen felfedezett struktúrát követően számos további, különböző típusú stwintront azonosítottunk.

Sok esetben egyéb, összetett, intronikus szerkezeteket is (tévesen) a twintronok közé sorolnak, mint az U2 / U12 spliceozómális intronokból felépülő, vagy egyéb, alternatív splicingot mutató szekvenciák. Hafez és Hausner egy review cikkükben részletesen összefoglalják a különböző twintronokat és twintron-szerű struktúrákat (Hafez és Hausner, 2015).

### **2.2.1. Ribozim típusú twintronok**

Mint az korábban is említésre került, a twintron, azaz iker-intron szőösszetételt Copertino és Hallick alkották meg 1991-ben, miután *Euglena gracilis* alga kloroplasztisz DNS-én (cpDNS) lokalizált, a citokróm b-559  $\beta$ -alegységét kódoló, *psbF* génjében egy érdekes intron összetételre lettek figyelmesek. Egy 1.024 bp hosszúságú, Group II típusú intron két, egymást követő splicing reakció révén távolítódik el. Először egy 618 bp hosszúságú szakasz vágódik ki, majd a maradék 424 bp

nagyságú; ezek után képződik meg az érett *psbF* mRNS (Copertino és Hallick, 1991).

A *psbF* gén érdekes, egymásba ékelődött intronjairól már Michel és munkatársai is említést tesznek egy korábbi tanulmányukban (Michel és mtsai., 1989). Leírják, hogy a *psbF* gén egyik Group II intronjának V katalitikus doménjébe beágyazódva található egy másik Group II típusú intron; de ők még nem használják a twintron elnevezést. Egy évvel ezelőtt pedig az *E. gracilis* kloroplasztisz genomjának vizsgálata során a kutatóknak szemet szúr a *psbF* gén szokatlanul hosszú, 1.024 bp nagyságú intronja, de „intron az intronban” elrendeződésről még nem tesznek említést (Cushman és mtsai., 1988).

A legelső, felfedezett twintron után a kutatócsoport újabb hasonló struktúra bizonyításának örülhetett. Ezúttal egy Group II/Group III felépítésű twintront mutattak ki, az *Euglena* plasztisz riboszómális proteint kódoló, *rps3* génjében (Copertino és mtsai., 1991 b).

Nem sokkal később sikerült Group III/Group III típusú stwintronok létezését is bizonyítani, szintén *Euglena* cpDNS-ében. Az *rpl6* (L16 riboszómális proteint kódoló) gén első intronjáról, illetve az *rpoC1* (RNS polimeráz alegységet kódoló) gén első, harmadik és tizenegyedik intronjáról derült ki, hogy az említett típusba sorolható, összetett twintronok (Copertino és mtsai., 1992). Mind a négy intronikus szakasz szokatlanul hosszúnak tűnt, kétszer akkora átlagos hosszal (~200 nt), mint más Group III intronok, ezért feltételezték, hogy két intront foglalnak magukba, azaz twintronok lehetnek.

Hamar kiderült, hogy az *E. gracilis* kloroplasztisz genom bővelkedik nem csak intronokban, de twintronokban is. Az első Group II/Group II twintron leírását követően további, Group II és/vagy Group III intronokból felépülő struktúrákat is kimutattak (Copertino és mtsai., 1994; Doetsch és mtsai., 1998). Megfigyelték, hogy mind Group II, mind Group III intronok lehetnek külső és/vagy belső intronok, a twintron szerkezetek kialakításánál (Copertino és Hallick, 1993). Emellett leírtak olyan, többszörösen összetett iker-intronokat is, melyek esetében egy külső intronba beékelődve nem csak egy, hanem több belső intron is található (Drager és Hallick, 1993; Hong és Hallick, 1994).

Twintron, vagy twintron-szerű struktúrák természetesen nem kizárólagosan az *Euglena gracilis* kloroplasztiszának genomjában fordulnak elő. Kimutattak Group II intronokból felépülő twintronokat *Pyrenomonas salina* (Maier és mtsai., 1995), *Monomorphina aenigmatica* (Pombert és mtsai., 2012), illetve *Porphyridium purpureum* (Perrineau és mtsai., 2015) algák cpDNS-ében is. Khan és Archibald pedig „degenerálódott” twintronokról számol be *Rhodomonas salina* alga kloroplaszt genomjában, mely esetben, feltételezésük szerint a belső intron elveszítette splicing képességét és tulajdonképpen egybeolvadt a külső intronnal (Khan és Archibald, 2008).

Prokarióták között is találunk példát Group II intronokból felépülő twintron-szerű struktúrák formációjára. *Trichodesmium erythraeum* és *Thermosynechococcus elongatus* cianobaktériumok esetében bizonyították létezésüket (Pfreund és Hess, 2015; Mohr és mtsai., 2010). Továbbá, egy metántermelő *Archea* faj (*Methanosarcina acetivorans*) és *Desulfitobacterium hafniense* Gram-pozitív baktérium genomjában is

találtak twintron-szerű, Group II intronok egymásba ágyazódása révén kialakult szerkezeteket (Dai és Zimmerly, 2003). A prokarióta twintronokat, helyesebb „nested intron”-oknak nevezni, mivel kísérletes bizonyíték nem áll rendelkezésünkre arra vonatkozóan, hogy ezen szekvenciák esetén is mindenképp szükséges-e a belső intron eltávolítása a külső intron kivágódását megelőzően (Hafez és Hausner, 2015).

Zumkeller és munkatársai korpafüvek és becősmohák mitokondriális genomjában írtak le Group I/Group II, illetve Group II/Group II típusú intron az intronban struktúrákat (Zumkeller és mtsai., 2020). Ezek közül egyeseknél a belső intron eltávolítása nem feltétele a külső intron kivágódásának, ám öt esetben bizonyítottan *sensu stricto* twintronokról van szó.

A szintén ribozim intronok közé tartozó Group I intronok is részt vehetnek twintron-szerű struktúrák kialakításában. Úgynevezett twin-ribozim intronokat mutattak ki *Didymium iridis* nyálkagombában, *Neagleria gruberi* ostoros amőbában, valamint egy *Allovahlkampfia* amőba fajban (Johansen és Vogt, 1994; Einvik és mtsai., 1998; Tang és mtsai., 2014; Nielsen és Johansen, 2009).

Több *Ascomycota* gombafaj mtDNS-ében is találtak egymásba ékelődött, összetett intronokat. *Cryphonectria parasitica* (gesztenyepenész) és *Ophiocordyceps tricenri rns* génjében egy Group I/Group I típusút, míg *Chaetomium thermophilum* ortológ génjében egy „kevert” twintron-szerű struktúrát, ahol egy Group II intront találunk egy Group I intronba beékelődve (Hafez és mtsai., 2013). *Grosmannia*

*piceiperda* mitokondiális genomjában szintén leírtak Group I típusú intront tartalmazó, összetett szerkezeteket (Rudski és Hausner, 2012).

### **2.2.2. Spliceoszómális twintron-szerű struktúrák**

Az eukarióták fehérjekódoló génjei, az exonok mellett jellemzően megszakító szekvenciákat, intronokat is tartalmaznak. Az intronok eltávolítását az elsődleges mRNS átíratból egy ribonukleoprotein komplex, a spliceoszóma végzi (Nilsen, 2003). Kétféle spliceoszómális intron típust különböztethetünk meg, az U2, illetve a sokkal ritkábban előforduló U12 típust (Bartschat és Samuelsson, 2010; Turunen és mtsai., 2013).

Rovarak és gerincesek egyes sejtmagi génjeiben olyan, szintén twintronnak nevezett struktúrákat írtak le, melyek esetén egy U12 típusú intront találunk egy U2 típusba beágyazódva, vagy fordítva (Scamborova és mtsai., 2004; Janice és mtsai., 2013).

A fent említett elrendeződést mutató, spliceoszóma által eltávolításra kerülő twintronok talán legismertebb példája a *Drosophila melanogaster prospero (pros)* génjében található szerkezet, ahol egy U2 típusú intront találunk egy U12 típusba beágyazódva. Az *ecetmuslica pros* génje a Prospero transzkripciós faktort kódolja, mely az éretlen idegsejtekben fejeződik ki főképp az embriogenezis során (Scamborova és mtsai., 2004; Borah és mtsai., 2009).

Az összetett intron szerkezet és annak eltávolítása az alternatív splicing egy ritka és szokatlan esete; az U2 intron splicingja kizárja az

U12 intron kivágódását, így a pre-mRNS-ről két különböző érett mRNS izoforma keletkezhet. Amennyiben az U2 típusú intron távolítódik el, az úgynevezett *pros-L* mRNS képződik, míg az U12 típusú intron splicingjával a *pros-S* izoforma keletkezik (amely 87 nukleotiddal rövidebb, mint a *pros-L*). A kutatók kísérletesen megállapították, hogy a *Drosophila* embriogenezisének korai szakaszában inkább a *pros-L* izoforma a gyakoribb, míg a későbbiekben a *pros-S*; a folyamat felénél arányuk közel azonos (Scamborova és mtsai., 2004). A két splicing útvonal szigorúan szabályozott, melyben számos ribonukleoprotein részt vesz az embriogenezis során (Borah és mtsai., 2009).

*D. melanogaster* mellett, más a *Diptera* rendbe tartozó rovarfajok genomjában is találtak U2/U12 típusú twintronokat, de emellett magasabb rendű eukariótákban (beleértve a humán genomot) is előfordulnak ilyen, egymásba ékelődött intronok (Mount és mtsai., 2007; Janice és mtsai., 2013).

Úgynevezett „nested intron”-okat írtak le a humán disztrofin gén (*DMD*) szokatlanul nagyméretű intronjában. Az emberi genomban megtalálható intronok átlagos hossza nagyjából 5.500 bp (Sakharkar és mtsai., 2004), míg a *DMD* génben talált gigantikus megszakító szekvencia 110.199 bp hosszúságú (Suzuki és mtsai., 2013). A gén kifejeződése során a pre-mRNS-ből először eltávolításra kerül két beágyazódott, „nested” intron, majd ezután történik meg a harmadik, „host” intron splicingja. A beékelődő, belső intronok nem szakítják meg a külső intron esszenciális splicing elemeit, így a struktúra nem tekinthető *sensu stricto* twintronnak, ám valószínűleg fontos szerepe van a két intron



elsődleges eltávolításának, mivel ezáltal jóval közelebb kerül egymáshoz a külső intron 5' és 3' hasítóhelye. Ez nagyban elősegítheti a megszakító szekvencia pontos eltávolítását és az mRNS megfelelő érését (Suzuki és mtsai., 2013).

Twintron struktúrát, pontosabban nested intront találtak Herpes Simplex vírus (HSV-1), latency associated transcript (*LAT*) génjében is. A gén 8.300 bp hosszú; két exonját egy különleges, összetett, mintegy 2.000 bp nagyságú intron szakítja meg. A struktúra érdekessége, hogy az eltávolítása megtörténhet egy lépésben, de két, egymást követő splicing esemény során is. Utóbbi esetben először a beékelődött, nagyjából 500 bp nagyságú „nested” intron kerül eltávolításra, majd ezek után, második lépésként a maradék 1.500 bp-nyi szekvencia (Brinkman és mtsai., 2013).

Szokatlanul nagyméretű intronok eltávolításának egy másik stratégiája lehet a rekurzív splicing. A jelenséget először *D. melanogaster Ultrabithorax (Ubx)* génjében írták le (Hatton és mtsai., 1998). Az ilyen intronok jellegzetességei az RS vagy „ratchet site”-ok, amiket tulajdonképpen két szomszédos intron olyan csatlakozási pontjának írhatunk le, ahol az első intron 3' hasítóhelye közvetlenül kapcsolódik az utána következő intron 5' hasítóhelyéhez (AGGT). Azt is mondhatnánk, hogy a két intron egy nulla nukleotidból álló exont határol, tehát a két szekvencia nincs exon által megszakítva. Ezek a hosszú intronok egymást követő splicing folyamatok révén távolítódnak el, az egybefüggő intron szekvencia 5' végéről kezdődően. Az *ecetmuslica* genomjának vizsgálata során további gének esetén is bizonyították intronok rekurzív splicing általi eltávolítását (Burnette és mtsai., 2005; Duff és mtsai., 2015; Joseph és mtsai., 2018).

A jelenség lejátszódását számos humán intron vizsgálata során is bizonyították (Sibley és mtsai., 2015; Kelly és mtsai., 2015; Georgomanolis és mtsai., 2016; Zhang és mtsai., 2018). A kutatók arra a megállapításra jutottak, hogy humán sejtekben a rekurzív splicing nem kizárja, hanem kiegészíti a kanonikus splicing eseményeket, hosszú intronok eltávolításának egy másik lehetséges stratégiájaként. Bizonyítottan RS-sel rendelkező, nagyméretű intronok eltávolításra kerülhetnek mindkét splicing folyamat révén, a preferált splicing útvonal pedig sok esetben sejtípus függő (Zhang és mtsai., 2018).

A nested és a rekurzív splicing mechanizmusok mellett leírták az úgynevezett intrasplicing jelenségét is, amely szintén twintron-szerű, hosszú intronok eltávolítására szolgálhat. A splicing folyamatot gerincesek *4.1R* és *4.1B* paralóg génjeiben mutatták ki, melyek citoskeletális adaptor fehérjéket kódolnak (Parra és mtsai., 2008; Parra és mtsai., 2012). Az intrasplicing érdekessége, hogy az intron többlépcsős eltávolításakor egy bizonyos szekvenciadarab, „intraexon” az első splicing esemény során exonikus régióként viselkedik, míg a következő lépésben az intron részeként kivágódik.

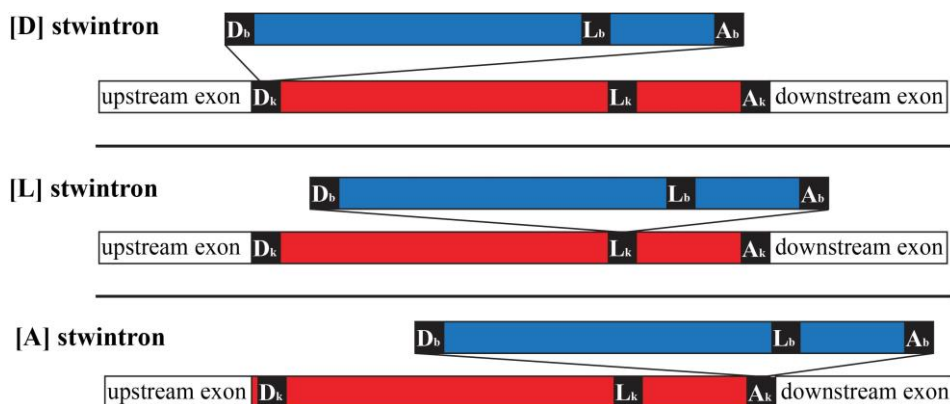
A nested splicing, rekurzív splicing és intrasplicing mechanizmusok tehát lehetőséget adhatnak nagyméretű intronok pontos eltávolítására, illetve alternatív splicing események lejátszódására, és ezzel a protein diverzitás növelésére. Láthatjuk majd, hogy ezek az összetett intron struktúrák felépítésükben és eltávolításuk módjában több-kevesebb hasonlóságot mutatnak a spliceoszómális iker-intronokkal, ám szigorúan véve nem nevezhetőek stwintronnak.

### 2.2.3. Stwintronok

A spliceoszómális iker-intronok, azaz stwintronok olyan különleges megszakító szekvenciák az eukarióta genomban, ahol egy kanonikus U2 típusú spliceoszómális intront találunk egy másik U2 típusú intron valamelyik esszenciális splicing elemébe beágyazódva (Flipphi és mtsai., 2013). A splicing motívum, melybe beékelődhet a belső intron az 5' hasítóhely, vagy másnéven donor szekvencia, a 3' hasítóhely, azaz akceptor szekvencia, továbbá a lariat elágazási pont szekvencia lehet. Megszakítva a külső intron, előbbieken felsorolt egyik szekvenciaelemét, annak splicingja gátolva van mindaddig, amíg a belső intron eltávolításra nem kerül. Kutatócsoportunk által feltárt stwintron struktúrákra tehát abszolút illik a Copertino és Hallick által megfogalmazott twintron definíció, miszerint szigorúan véve akkor nevezhetünk egy összetett intron struktúrát twintronnak, ha a külső intron eltávolításának előfeltétele a belső intron kivágódása, tehát több, egymást követő splicing reakció révén távolítható el a teljes struktúra (Copertino és Hallick, 1991). Ez, az általunk feltárt szerkezetek esetén, igaznak is bizonyul; ám az stwintronok, lévén spliceoszómális intronokból épülnek fel, kivágódásukhoz spliceoszóma közreműködését igénylik.

A belső intron orientációja alapján, mint ahogy az a 3. ábrán is látható, három különböző stwintron típus létezhet; a külső intront, a beékelődő, belső intron a donor, az akceptor vagy az elágazási, másnéven lariat szekvenciájában szakíthatja meg. Ezek alapján megkülönböztethetünk [D], [A], illetve [L] típusú stwintronokat. A megszakított splicing motívumra utaló betű után pedig feltüntetünk két

számot az stwintronok elnevezésekor, mely a belső intron donor-, akceptor-, vagy elágazási hely belső pontos orientációját adja meg. Mivel a donor és lariat szekvenciák hat-hat, az akceptor pedig három nukleotidot foglal magába, elméletileg 12 különböző típusú stwintron létezhet. A következőkben ismertetem, hogy ezek közül mely típusok létezését bizonyítottuk ezidáig.



**3. ábra: A három stwintron típus sematikus ábrája** ([D]: donor szekvenciában megszakított; [L]: lariat szekvenciában megszakított és [A]: akceptor szekvenciában megszakított stwintron; D<sub>b</sub>: a belső intron donor eleme, L<sub>b</sub>: a belső intron elágazási pont szekvenciája, A<sub>b</sub>: a belső intron akceptor eleme, D<sub>k</sub>: a külső intron donor eleme, L<sub>k</sub>: a külső intron elágazási pont szekvenciája, A<sub>k</sub>: a külső intron akceptor eleme; a belső intronok kék, a külső intronok piros színnel jelölve) (Kavalecz és mtsai., 2019)

Fungális genomok vizsgálata során kutatócsoportunknak sikerült több, donor szekvenciában megszakított stwintron létezését bizonyítania, különböző gombafajok és gének esetén. *Fusarium verticillioides* PIH2 (egy putatív imidin hidroláz paralógot kódoló) génjében egy [D1,2] típusú stwintron struktúra került leírásra, ahol a belső intron a külső intron 5' hasítóhelyét, azaz donor szekvenciáját szakítja meg, annak első és második nukleotidja között. A spliceoszómális iker-intron léte az egymást

követő splicing reakciók és az mRNS intermedierek kimutatás révén bizonyítást nyert (Flipphi és mtsai., 2013).

*Trichoderma reesei bioDa* (egy kétfunkciós, a biotin bioszintéziséhez szükséges enzimet kódoló) génjében egy [D2,3] típusú stwintron létét sikerült igazolni, míg *Botrytis cinerea PIH1* (egy putatív imidin hidroláz paralógot kódoló) génjében egy feltételezett [D4,5] típusú spliceoszómális twintront találtak (Flipphi és mtsai., 2013).

*Helminthosporium solani* gomba *aox* (alternatív oxidázt kódoló) génjében egy nagyon érdekes, kétféle útvonalon kivágódni képes, [D1,2]/[A2,3] stwintron került leírásra. A splicing reakciók lejátszódásától függően a struktúra tekinthető [D1,2], illetve [A2,3] típusú stwintronnak is. Mindkét alternatív splicing esemény létezését bizonyítást nyert, továbbá megállapításra került, hogy nincsen egyértelműen preferált útvonal, mind a kettő splicing reakció lejátszódhat *in vivo*. Az stwintron akár [D1,2], akár [A2,3] struktúraként kerül eltávolításra, az mRNS érési folyamatának végén ugyanazon transzkriptumot kapjuk, tehát azonos fehérje fog képződni (Ág és mtsai., 2015). *H. solani* egy másik, putatív aldóz mutarotázt kódoló génjében is hasonló elrendeződésű stwintront találtak, ám itt csak az egyik, a [D1,2] típusú alternatív splicing út vezet olyan érett mRNS termék keletkezéséhez, amely alapján funkcióképes fehérje termék transzlálódhat (Ág és mtsai., 2015).

Még nem publikált kutatási eredményeink egy [L4,5] típusú, tehát lariat szekvenciában megszakított stwintron létezését sejtetik *Aspergillus nidulans* egyik génjében. Így elmondható, hogy minhárom stwintron típusra találunk példát fonalas gomba genomokban.

### 3. Célkitűzések

A legelső spliceoszómális iker-intron struktúrák felfedezését és az stwintron fogalom bevezetését követően célul tűztük ki minél több összetett megszakító szekvencia leírását. Ennek érdekében megpróbáltunk létrehozni egy stwintron modellt, melynek segítségével célzottan vizsgálhatunk át elérhető gomba genomszekvenciákat, egy adott típusú stwintron után kutatva.

A szinte napról napra gyarapodó szekvenciaadatok és elérhető, megszekvenált genomok pedig lehetővé teszik, hogy megvizsgáljunk ortológ géneket több száz, vagy akár több ezer rokonfajban. Ezt kihasználva célunk volt feltárni az stwintronok és ezáltal a spliceoszómális intronok evolúciójának, eltűnésének, megjelenésének rejtélyeit gombagenomokban.

Stwintronok kimutatása és splicing intermedierjeik azonosítása után szerettünk volna választ kapni arra a kérdésre is, hogy vajon milyen funkciót töltenek be ezek a különleges, összetett, intronikus struktúrák a genomban. Bírhatnak-e valamilyen szabályozó funkcióval? Esetleg szerepet játszhatnak-e az alternatív splicinggal kapcsolatos mechanizmusok végbemenetelében?

## 4. Anyagok és módszerek

### 4.1. Bioinformatikai módszerek

#### 4.1.1. Putatív stwintron struktúrák keresése teljes genomszekvenciákban

A fungális intronokra jellemző sajátosságokat figyelembe véve, kutatócsoportunk megalkotott egy stwintron modellt, melynek segítségével teljes gombagenomokban kutathatunk putatív iker-intron struktúrák után. A legelső stwintronokra még teljesen véletlenül bukkantak rá (Flipphi és mtsai., 2013; Ág és mtsai., 2015), gombagének más aspektusú vizsgálata során.

A modell alapját *A. nidulans* esetén megállapított konzervált splicing motívumok szekvenciái képezik (Kupfer és mtsai., 2004). A 4. ábrán bemutatott Kupfer-féle általános intron sémán látható, hogy az 5' donor és a lariat elágazási szekvencia hat-hat nukleotidból áll, míg a 3' akceptorhely csak három nukleotidot foglal magában. A leginkább konzervált helyek a donor első kettő (GT), az akceptor utolsó kettő (AG), illetve az elágazási pont szekvencia ötödik nukleotidja, amely minden esetben adenin (A).

**GTRWGY(58)RYTRAY(15)YAG**

**4. ábra: Kupfer intron modellje *A. nidulans*-ra (Kupfer és mtsai., 2004)** (a konzervált nukleotidokat aláhúzással jelöltem és G: guanin, T: timin, A: adenin R: adenin vagy guanin, W: adenin vagy timin, Y: citozin vagy timin; Comm. IUPAC-IUB, 1970)

A modellt némi „megengedő” változtatás bevezetése (a lariat elágazási szekvencia első nukleotid pozíciójának definiálásakor „D”-t használtunk „R” helyett, az akceptor szekvencia első helyén pedig „H”-t „Y” helyett) után alkalmaztuk (5. ábra).

**GTRWGY(58)**D**YTRAY(15)**H**AG**

**5. ábra: A korigált intron modell** (a konzervált nukleotidokat aláhúzással, a korrekciókat pedig zölddel jelöltem és G: guanin, T: timin, A: adenin R: adenin vagy guanin, W: adenin vagy timin, Y: citozin vagy timin, D: adenin, guanin vagy timin, H: adenin, citozin vagy timin; Comm. IUPAC-IUB, 1970)

A korigált intron séma segítségével manuálisan létrehoztuk a fungális stwintron modellt, ahol két intron egymásba ágyazódását modellezve öt jellemző szekvenciamotívumot kaptunk. Ezek közül kettő „hibrid” elem, ugyanis egy, vagy több nukleotidot mind a külső, mind a belső intron szekvenciájából magában foglal.

A konzervált splicing motívumokon túl meghatároztuk az egyes elemek közti távolságtartományokat is az alábbi információk birtokában:

1. A legrövidebb, *A. nidulans*-ban kimutatott intron 42 nukleotid hosszúságú.
2. A minimális távolság az elágazási pont szekvencia és az akceptorhely között 4 nukleotid lehet; 4 és 24 közötti tartomány értéket állítottunk be.
3. A donorhely és az elágazási pont szekvencia közti távolság mindig nagyobb, mint az elágazási pont és az akceptorhely közötti; ennek az értéknek 25 és 120 nukleotid közötti tartományt adtunk meg.



4. Az átlagos intronhossz *A. nidulans* genomjában 73 nukleotid, míg a 160 nukleotidnál hosszabb kanonikus intronok igencsak ritkák (Kupfer és mtsai., 2004).

Így [D1,2] típusú stwintronok kereséséhez a 6. ábrán bemutatott modellt használtuk, teljes genom szekvenciák átvizsgálásához, a Fuzznuc programot alkalmazva (Rice és mtsai., 2000).

<p><b>GGTRWGYN(25,120)DYTRAYN(4,24)HAGTRWGYN(25,120)DY TRAYN(4,24)HAG</b></p>
---

**6. ábra:** Az általunk használt [D1,2] stwintron modell (kékkel jelöltem a belső intron, pirossal a külső intron splicing motívumait, aláhúzással pedig a konzervált nukleotidokat és G: guanin, T: timin, A: adenin R: adenin vagy guanin, W: adenin vagy timin, Y: citozin vagy timin, N: bármelyik nukleotid, D: adenin, guanin vagy timin, H: adenin, citozin vagy timin; Comm. IUPAC-IUB, 1970)

Egyszerű átalakításokkal (természetesen) más típusú stwintronokra is megalkotható a modell, amit munkánk során ki is használtunk.

*Aspergillus nidulans* FGSC A4 (Accession Number: AACD01000000) genomján lefuttatva a keresést, számos [D1,2] stwintron-jelöltet kaptunk eredményként, melyeket manuálisan is megvizsgáltuk, az alábbi szempontok szerint:

1. A talált szekvencia hol helyezkedik el a genomban, megszakít-e bizonyított vagy putatív kódoló szekvenciadarabokat.
2. Amennyiben rendelkezésre állt, megvizsgáltuk az expressziós, RNASeq adatokat is, melyeket az *Aspergillus* Genome Database (AspGD) adatbázisából értünk el (Cerqueira és mtsai., 2014).

3. Megvizsgáltuk más, az AspGD adatbázisban elérhető, *Aspergillus* törzsek ortológ génjeinek szekvenciáit is. Arra voltunk kíváncsiak, hogy ezekben is kimutatható-e putatív [D1,2] stwintron, azonos pozícióban.

A modellt a későbbiekben, megfelelő módon átalakítva [D5,6] típusú stwintronok keresésére is használtuk. Ebben az esetben a belső intront a külső intron donor szekvenciájának ötödik és hatodik nukleotidja közé helyeztük el (5'-GTRWG|H).

**GTRWG****GTRWGH**(25,120)**DYTR****AY**(4,24)**HAGH**(25,120)**DYTR**  
**AY**(4,24)**HAG**

7. ábra: Az átalakított, [D5,6] típusú stwintronok kereséséhez alkalmazott modell (a szín- és betűkódok megegyeznek a 6. ábrán használtakkal)

A 7. ábrán bemutatott szekvenciamotívum alapján kutattunk potenciális [D5,6] stwintronok után *A. nidulans* genomjában, a Sequence Manipulation Suite-ot (Stothard, 2000) használva.

A kimutatott putatív struktúrákat ebben az esetben is manuális vizsgálat alá vetettük, a korábbiakban ismertetett szempontok alapján.

#### 4.1.2. Ortológ gének vizsgálata (genome mining)

A National Centre of Biotechnology Information (NCBI) és a US Department of Energy Joint Genome Institute (JGI) adatbázisokban nyilvánosan elérhető gomba genomszekvenciákat TBLASTN programmal (Altschul és mtsai., 1997) elemeztük a bizonyítottan stwintront tartalmazó gének ortológjai után kutatva.

A DHA1 ortológok vizsgálatakor olyan gének után kutattunk *Ascomycota* genomokban, melyekről a képződő fehérje termék, aminosav sorrendjét tekintve, legalább 45%-ban azonos *A. nidulans* AN3270-es lókuszon lokalizált, stwintront tartalmazó génjéről átíródó DHA1 proteinnel. A valós kódoló szekvenciákat, amit az ortológok kereséséhez használtunk referenciaként, cDNS adatokból (GenBank Accession Number: KY315812) következtettük. A leírt módszerrel megtalált, több mint 500 gén intron-exon struktúráját manuálisan határoztuk meg, melyek segítségével a későbbiekben következtethettünk a teljes fehérjekódoló szekvenciára.

Az *A. nidulans* AN5404 lókuszon található, 326 aminosavból felépülő, reticulon-like proteint kódoló gén (*rtnA*) ortológjainak felkutatása során is az előzőekben ismertetett módon jártunk el. A feltételezett hét darab, fehérjekódoló exont tartalmazó génmodellt kísérletesen igazoltuk (GenBank Accession Number: MK410458) és referenciaként használtuk az ortológok kereséséhez. A talált gének intron-exon struktúrájának megállapítását ebben az esetben is manuálisan végeztük.

#### **4.1.3. Másodlagos fehérjestruktúra meghatározása**

A reticulon-like proteint (RtnA) kódoló gének vizsgálatakor a fehérjék másodlagos, transzmembrán hélix struktúráját az online elérhető TMHMM (version 2) szerver (Krogh és mtsai., 2001) segítségével prediktáltuk. A strukturális eltérések kimutatásához a DISOPRED3 (Jones és Cozzetto, 2015), továbbá a „coiled-coil” struktúrák

megtalálásához a PCOILS (Gruber és mtsai., 2006) programokat használtuk.

#### **4.1.4. Filogenetikai analízis**

Filogenetikai analízisek elvégzéséhez a fehérje szekvenciákat a MAFFT (Multiple Alignment using Fast Fourier Transform, version7) (Katoh és Standley, 2013) programmal hasonlítottuk össze.

A DHA1 fehérjék esetén, a 256 db *Pezizomycotina* protein összehasonlításához az E-INS-i algoritmust és a BLOSUM 45 hasonlósági mátrixot használtuk. A szekvencia-illesztések 459 informatív régióját BMGE (Block Mapping and Gathering with Entropy, version 1.12) program (Criscuolo és Gribaldo, 2010) segítségével kerestük meg, BLOSUM 35 mátrixot és 3-as blokk méretet használva.

A putatív lipáz ortológok analíziséhez 292 db fehérjét hasonlítottunk össze, E-INS-i algoritmust és BLOSUM 62 hasonlósági mátrixot alkalmazva. A MAFFT program (Katoh és Standley, 2013) segítségével elvégzett illesztés adatait a BMGE szoftver (Criscuolo és Gribaldo, 2010) segítségével rendeztük, BLOSUM 30 hasonlósági mátrixot és 5-ös blokk méretet használva.

A bifunkcionális BioDA fehérje tanulmányozásához 298 db protein szekvenciát hasonlítottunk össze, szintén a MAFFT program (Katoh és Standley, 2013) és E-INS-i algoritmus segítségével, BLOSUM 45, majd a BMGE szoftver (Criscuolo és Gribaldo, 2010) használatkor BLOSUM 55 hasonlósági mátrixot használtunk 5-ös blokk mérettel.

A putatív reticulon-like fehérjék törzsfájának elkészítéséhez 779 db proteint hasonlítottunk össze a MAFFT programot (Katoh és Standley, 2013) és L-INS-i algoritmust használva, amely egy konzervált domén felismerésére alkalmas. A fa gyökerének megtalálásához a vizsgálatba bevontunk kilenc, *Mortierellomycotina* fajokból származó putatív reticulon-like fehérjét. A MAFFT program által elvégzett illesztés rendezését ismét a BMGE programra (Criscuolo és Gribaldo, 2010) bíztuk, különböző hasonlósági mátrixok és blokk-méretetek kipróbálásával.

A maximum likelihood törzsfákat a PhyML 3.0 programot és WAG szubsztitúciós modellt használva készítettük el (Guindon és mtsai., 2010). A reticulon-like fehérjék törzsfájához automatikus szubsztitúciós modell választást (Lefort és mtsai., 2017) alkalmaztunk, a szoftver minden esetben az LG+G+I modellt preferálta.

A fa ághosszainak kiszámításához az Approximate Likelihood Ratio Test-et (Anisimova és Gascuel, 2006) és a PhyML programot használtuk, Chi-2 alapú, parametrikus feltevéssel. A fák megrajzolásában a FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>, version 1.4.3), utólagos szerkesztésében pedig az Adobe Illustrator program volt segítségünkre.

## 4.2. Gombatörzsek

A kísérletes munka során használt gombatörzseket az alábbi táblázatban foglaltam össze:

1. táblázat: A kísérletek során használt gombatörzsek

<b>Törzs neve</b>	<b>Referencia genomszekvenciához</b>	<b>Referencia táptalajhoz és növekedési paraméterekhez</b>
<i>Aspergillus nidulans</i> R21 (ATCC 48756)	Wortman és mtsai., 2009	Pontecorvo és mtsai., 1953
<i>Aspergillus niger</i> ATCC 1015	Andersen és mtsai., 2011	de Vries és mtsai., 2004
<i>Trichoderma reesei</i> QM 9414 (ATCC 26921)	Vitikainen és mtsai., 2010	Mandels és Andreotti, 1978
<i>Helminthosporium</i> <i>solani</i> B-AC-16A	Mattupalli és mtsai., 2014	Mimee és mtsai., 2011
<i>Botrytis cinerea</i> B05.10	Amselem és mtsai., 2011	Fekete és mtsai., 2012
<i>Neurospora crassa</i> OR74A	Galagan és mtsai., 2003	Vogel, 1956
<i>Malbranchea</i> <i>cinnamomea</i> CBS 343.55	Morgenstern és mtsai., 2012	Morgenstern és mtsai., 2012

Az egyes törzsek fenntartásához használt táptalajok pontos összetételét jelen dolgozatban nem részletezem, azok elérhetőek a 1. táblázatban megadott referenciák alapján.

### 4.3. Nukleinsav izolálás

A vizsgált gombatörzseket nukleinsav izoláláshoz 500 ml-es Erlenmeyer-lombikokban növesztettük ki, azokat rázógépben (Infors HT Multitron) 200 rpm-en és az adott törzsre nézve optimális növekedési hőmérsékleten inkubáltuk. A lombikok egyenként 100 ml tápoldatot (1. táblázat) tartalmaztak, melyeket az adott gombatörzs vegetatív spórájával oltottunk le. A törzs növekedésének gyorsaságától függően, 16-24 óras inkubációt követően „gyűjtöttük be” a biomasszát. A lombikok tartalmát steril Miracloth szűrőtextilen (Calbiochem) keresztül leszűrtük. Az összegyűlt gombamicéliumot jéghideg, steril, ioncserélt vízzel mostuk, majd két papírtörülő réteg között víztelenítettük. A mintákat folyékony nitrogénben azonnal lefagyasztottuk, majd felhasználásig  $-80^{\circ}\text{C}$ -os mélyfagyasztóban tároltuk.

A minták porítását dörzsmozsarakban végeztük, folyékony nitrogén folyamatos adagolása mellett. A nukleinsav kivonását Macherey-Nagel NucleoSpin kit-ek (NucleoSpin Plant II-t DNS izoláláshoz és NucleoSpin RNA Plant-et RNS izoláláshoz) segítségével, a gyártó utasításait követve végeztük. A DNS és RNS minták pontos koncentrációját NanoDrop spektrofotométerrel határoztuk meg az izolálást követően.

#### 4.4. Reverz transzkripció PCR (RT-PCR) reakciók

Az RT-PCR reakciók kivitelezéséhez első lépésként cDNS-t szintetizáltunk reverz transzkripció révén, melyhez minden esetben pontosan 1 µg RNS-t, illetve oligo(dT) primert és a First Strand cDNA Synthesis Kit-et (Thermo Scientific) használtunk.

A PCR reakciókhoz 4 µL cDNS-t használtunk templátként (25 µL reakció-végtérfogathoz), melyekről különböző génspecifikus primerek segítségével DreamTaq DNA Polymerase (Thermo Scientific) alkalmazásával szaporítottuk fel a számunkra releváns fragmenseket. Az RT-PCR reakciók során alkalmazott paramétereket a 2. táblázat, míg az oligonukleotid primerek listáját a 12.1. Függelék 3. és 4. táblázata tartalmazza.

**2. táblázat: Az alkalmazott PCR reakciók lépései, paraméterei**

	Lépés	Hőmérséklet	Időtartam	Ciklus
1.	Elődenaturáció	95°C	2 min	1
2.	Denaturáció	95°C	30 sec	40
3.	Anelláció	56/60°C	1 min	
4.	Elongáció	72°C	1 min	
5.	Végső polimerizáció	72°C	5 min	1

Az amplifikált PCR termékeket agaróz gélben választottuk el egymástól, gélelektroforézist alkalmazva.



Annak érdekében, hogy a prediktált splicing intermediereket ki tudjuk mutatni, olyan primerpárokat alkalmaztunk az RT-PCR reakciók során, melyek nem az érett mRNS templátokról szaporítanak fel terméket. Ezen primerek egyikét jellemzően úgy terveztük meg, hogy az intron-exon csatlakozásnál tapadjanak, néhány nukleotidjukkal a feltételezett stwintron külső intronjának 3' végéhez kötődve. Ezekkel az oligonukleotidokkal általában két termék amplifikálódik a PCR során, melyek közül a kisebb méretű a számunkra releváns. Ez ugyanis azon intermedier jelenlétére utal, mely esetén a belső intron kivágásra került, de a második splicing reakció még nem játszódott le, tehát az stwintron külső intronját még tartalmazza a szekvencia. Ezzel szemben a nagyobb méretű PCR termék az elsődleges mRNS átíráról felszaporodó amplikon. A későbbiekben mindkét fragmens megszekvenáltatása megtörtént.

Minden RT-PCR reakciót kétszer végeztünk, egymástól független rázatott lombikos tenyészetekből nyert biomasszából izolált RNS-t használva.

#### **4.5. Szekvenálás**

A PCR termékek agaróz gélben történő elválasztása után már prediktálható, hogy mely esetekben szaporodott fel a keresett fragmens, tehát mely struktúrák esetén beszélhetünk valóban stwintronról. Ahhoz, hogy ezt bizonyítsuk, a méretük alapján megfelelőnek ítélt amplikonokat kitisztítottuk, NucleoSpin Gel & PCR Clean-up (Macherey-Nagel) kit segítségével, majd vektorba klónoztuk pGEM-T Easy Vector System I (Promega) rendszert használva. A plazmid izolálást NucleoSpin Plasmid

EasyPure kit (Macherey-Nagel) segítségével végeztük, követve a gyártó utasításait.

Minden esetben három, független plazmid klónt küldtünk el szekvenálásra (Eurofins Genomics, Ebersberg, Germany). A cég olyan univerzális primerek (pUC/M13 Primer, Forward and Reverse) segítségével végzi a szekvenálást, melyek a vektor konstrukció meghatározott szakaszaival hibridizálnak.

Jelen értekezésben bemutatott kísérletek szekvenálási eredményei mentésre kerültek a GenBank-ban, az alábbi sorszámok alatt: KY315812-KY315816; MF612150-MF612153; MK410458-MK410473; MK421638-MK421641.

## 5. Eredmények

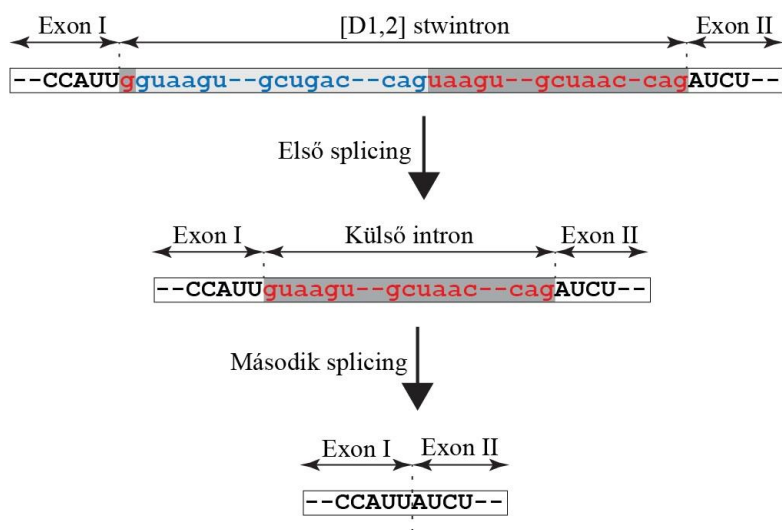
### 5.1. DHA1 ortológok vizsgálata

#### 5.1.1. [D1,2] stwintron bizonyítása *A. nidulans dhaoS* génjében

A korábbiakban megalkotott bioinformatikai eszköz (mely az „Anyagok és módszerek” fejezetben részletes bemutatásra került) segítségével elkezdtünk putatív stwintron struktúrák után kutatni, *Aspergillus nidulans*, *Ascomycota* modellorganizmus genomjában. A gombafaj AN3270-es lókuszon lokalizált génjében detektáltunk egy lehetséges [D1,2] típusú stwintront. A gén egy transzmembrán fehérjét kódol, mely az „1-es típusú” drug : H<sup>+</sup> antiporter családba (DHA1) (Pao és mtsai, 1998) tartozik. Az AN3270-es lókuszu, *A. nidulans* gént *dhaoS*-nek neveztük el, a „drug: H<sup>+</sup> antiporter family one gene with stwintorn” angol elnevezésből rövidítve.

A szekvencia alapján úgy véltük, hogy a *dhaoS* génről képződő elsődleges mRNS átírat tartalmaz egy összetett, 113 nt hosszúságú megszakító szekvenciát. Az intronikus szakasz kivágódását igazoltuk is a gén teljes cDNS-ének megszekvenáltatása révén, mely egy 1.611 nt hosszúságú ORF (GenBank Accession Number: KY315812). A prediktált protein termék tehát 536 aminosavból áll, az stwintron struktúra pedig a 108. kodont (Leu108 UUA) szakítja meg, a második uracil után. A 105. és a 112. aminosav között található szekvenciamotívum (WGPLESELY) nagyfokú konzerváltságot mutat számos fungális DHA1 fehérjében, melyek közül ráadásul gyakori, hogy a kódoló gén egyáltalán nem tartalmaz intronokat.

A feltételezett *dhaoS* [D1,2] stwintron egy 56 nt hosszúságú, U2 külső intronból, illetve egy 57 nt hosszúságú, szintén U2 típusú belső intronból épül fel; utóbbi a külső intront az 5' donor szekvenciájában szakítja meg, annak első és második nukleotidja között. A struktúra felépítése (8. ábra) merőben hasonlít a korábbiakban, kutatócsoportunk által karakterizált stwintronokéhoz (Flipphi és mtsai., 2013; Ág és mtsai., 2015).



**8. ábra:** *A. nidulans* DHA1 fehérjét kódoló *dhaoS* génjében leírt [D1,2] típusú stwintron és kétlépéses kivágódási mechanizmusának bemutatása (az stwintron belső intronját világosszürke háttérrel és kék betűkkel, a külső intronját sötétszürke háttérrel és piros betűkkel jelöltem) (saját szerkesztésű ábra Fekete és mtsai., 2017 alapján)

RT-PCR reakciók révén sikeresen amplifikáltuk a kivágódási köztest; ehhez egy olyan speciális primerpárt használtunk, melyek közül az egyik (forward primer) szekvenciája teljes mértékben exonikus, míg a másik (reverse primer) átfed a feltételezett stwintron külső intronjának 3' végével (a módszer részletes leírásra került az „Anyagok és módszerek” fejezetben). Ezeket az oligonukleotidokat és cDNS templátot használva a PCR reakciókhoz, jellemzően két, különböző méretű terméket tudunk

detektálni. Ezek közül a kisebb méretű ampikon (GenBank Accession Number: KY315814) reprezentálja az stwintron azon állapotát, amikor a belső intron már eltávolításra került, tehát az első splicing reakció már lejátszódott. A nagyobb méretű fragmens ezzel szemben az elsődleges, kivágatlan transzkriptumnak felel meg, ahol az stwintron teljes struktúrája megtalálható még az mRNS szekvenciában.

### **5.1.2. Stwintront tartalmazó *dhaoS* ortológok azonosítása**

Az *A. nidulans* stwintron leírását követően célul tűztük ki a gén ortológjainak feltérképezését és vizsgálatát. Olyan géneket kerestünk, TBLASTN-t használva, melyek DHA1 transzmembrán fehérjét kódolnak és aminosav sorrendjükben legalább 45%-os azonosságot mutatnak *A. nidulans* DhaoS proteinjével. Több száz, a kívánalmaknak megfelelő gént sikerült találnunk, *Pezizomycotina* gombafajok genomjában. A kimutatott DHA1 homológ proteineket felhasználva (összesen 89 darabot) létrehoztunk egy Maximum Likelihood fát, melyen belül észrevehető egy 73 proteint tartalmazó, jól elkülönülő csoport, az *A. nidulans* DhaoS-t is beleértve. Ezek a fehérjék aminosav sorrendjükben, nagyobb, mint 64%-os azonosságot mutattak. Egy 16 fehérjét magába foglaló csoport 56-62%-os megegyezést mutatott az *A. nidulans* DhaoS fehérjével, így ők a putatív ortológ klád kulcsoportját képezik. A vizsgált, 89 protein evolúciós kapcsolatát a 12.2. Függelék 18. ábráján láthatjuk.

A törzsfán elkülönülő, 73 ortológ fehérje kódoló génjeinek mindegyike tartalmaz valamilyen megszakító szekvenciát pontosan abban a pozícióban, vagy közvetlenül annak 3' végénél, ahol az *A. nidulans*

*dhaoS* génjében az stwintront találjuk. A 73 gén közül 21 hordoz egy második intront is, 329 bp távolságra az első intron pozíciótól, downstream irányba. A több száz, összegyűjtött, DHA1 fehérjéket kódoló gént megvizsgálva, egy kivételtől eltekintve, úgy láttuk, hogy az 5' intron pozíció a 73 elkülönülő *dhaoS* ortológ egyedi sajátja. A kulcsoporthoz tartozó 16 fehérje (melyek 56-62%-os azonosságot mutatnak az *A. nidulans* proteinnel) kódoló génei pedig egyáltalán nem tartalmazzak intronokat.

[D1,2] típusú stwintron esetén, a belső intron akceptor szekvenciája szükségszerűen átfedésbe kerül a külső intron donor szekvenciájával. Korábbi megfigyeléseink alapján elmondható, hogy amennyiben az stwintron két oldalán található exonikus régiók intronnal közvetlenül érintkező nukleotidja guanin (G) az elsődleges mRNS átíratban, akkor a szóban forgó [D1,2] stwintron egyúttal egy [A2,3] típusú intron is egyben. A 73 DHA1 ortológ gén közül, 36 esetben szigorúan [D1,2] típusú stwintront azonosítottunk, míg 28 gén esetén olyan stwintronokat találtunk, melyek elméletileg az említett másik, alternatív módon is eltávolításra kerülhetnek.

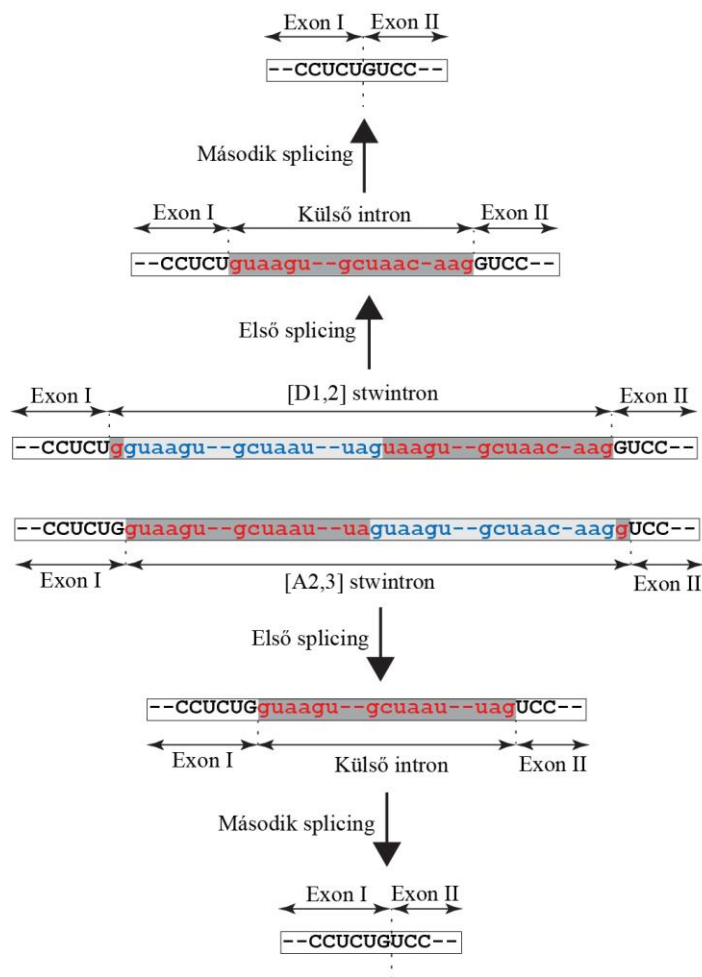
### **5.1.3. [D1,2]/[A2,3] stwintron *T. reesei* DHA1 ortológjában**

*T. reesei dhaoS* ortológ génjében (lókusz: TRIREDRAFT\_43701) talált stwintron egyike azoknak, amely a génszekvenciája alapján egy alternatív kivágódásra képes, [D1,2]/[A2,3] típus, hasonlatosan a korábban *H. solani*-ban detektált stwintronhoz, mely egy alternatív oxidázt kódoló génben (*aox*) került leírásra (Ág és mtsai., 2015).

A kettő megszakító szekvenciát tartalmazó, 1.548 bp hosszúságú ORF egy 515 aminosavból felépülő proteint kódol. Ennek bizonyításához a gén mRNS átíratáról készített, majd PCR-rel felszaporított cDNS fragmenst klónoztuk, majd megszekvenáltattuk (GenBank Accession Number: KY315813). Az első intronikus régió a szekvenciában a putatív [D1,2]/[A2,3] stwintron, ami a 125. kodont (Leu125 CUG) szakítja meg, annak második nukleotidja után, feltéve, hogy [D1,2] típusként kerül eltávolításra. Az stwintront kódoló, 154 nt hosszúságú szekvencia két, alternatív belső intront foglal magában, melyek egyetlen nukleotidban, nevezetesen guaninban fednek át. Az 5' intron 66 nt hosszúságú; a közvetlenül előtte található első, már exonikus régióhoz tartozó nukleotid pedig guanin, amennyiben a struktúrát [D1,2] stwintronnak tekintjük. A 3' intron 88 nt hosszúságú és az utána következő első exonikus nukleotid szintén guanin, hogyha [A2,3] stwintronnak feltételezzük. Sikerült amplifikálnunk mindkét alternatív kivágódási úthoz tartozó splicing intermediert. A fragmenseket klónoztuk és megszekvenáltattuk, bizonyítva ezzel, hogy *T. reesei dhaoS* ortológja valóban egy [D1,2]/[A2,3] alternatív stwintront (9. ábra) hordoz (GenBank Accession Number: KY315815 és KY315816).

Az NCBI EST (expressed sequence tags) adatbázisában többszáz expressziós adatot találhatunk *T. reesei* vizsgált génjéről. Ezek közül a legtöbb az általunk is feltételezett, az ortológ génről képződő, érett mRNS-t reprezentálja. Találtunk azonban két EST-t (Accession Number: CB901225 és CF871013), melyek szekvenciájuk alapján, az [A2,3] típusú splicing intermediereknek felelnek meg. Ennek fényében feltételeztük, hogy

a spliceoszóma preferáltan az [A2,3] útvonalon távolítja el *T. reesei* potenciális alternatív stwintronját.



**9. ábra: *T. reesei*-ben kimutatott alternatív [D1,2]/[A2,3] stwintron felépítésének és kivágódási mechanizmusának sematikus ábrája.** A felfelé mutató nyilak reprezentálják a struktúra [D1,2] stwintronként, míg a lefelé mutató nyilak [A2,3] stwintronként való eltávolítását. (A színek megegyeznek a 8. ábrán használtakkal.) (saját szerkesztésű ábra Fekete és mtsai., 2017 alapján)

Kísérletesen igazoltuk mindkét kivágódási köztes létét, egy olyan, speciális reverz primert alkalmazva az RT-PCR reakciók során, amely a



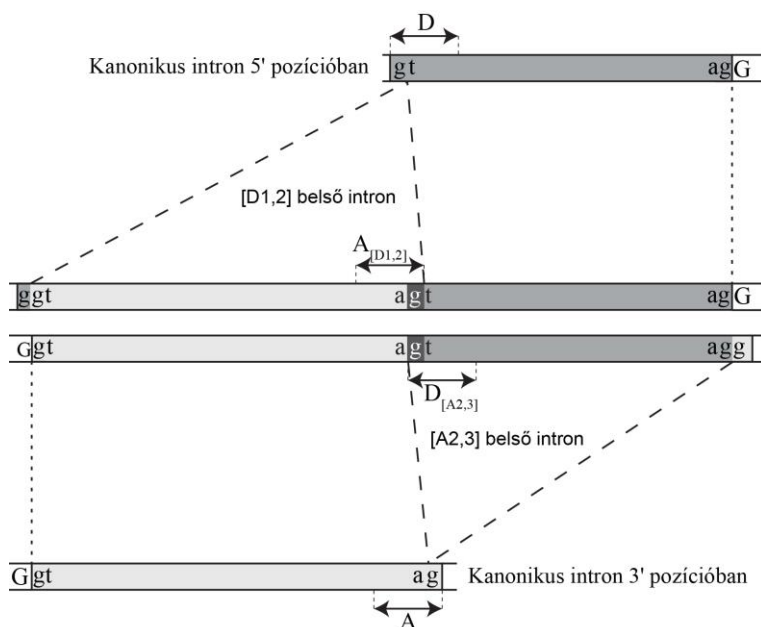
gében található másik intron szekvenciáján belül hibridizál a templáttal, 329 nt távolságra az stwintrontól, downstream irányba. Ugyanazt az mRNS templátot használva, a klónozott és megszekvenált RT-PCR termékek túlnyomó része [D1,2] splicing intermedierek felelt meg, de sikeresen kimutattunk az [A2,3] köztest, az elsődleges transzkriptumot és az érett mRNS-t is, melyből már a teljes stwintron struktúra eltávolításra került. A kapott eredmények azt sugallják, hogy az általunk alkalmazott tenyésztési körülmények mellett, a mintavétel időpontjában a [D1,2]-es útvonal volt a meghatározó, azonban kijelenthetjük, hogy *T. reesei* DHA1 ortológjában leírt [D1,2]/[A2,3] stwintron bizonyítottan eltávolításra kerülhet mindkét, alternatív útvonalon.

#### **5.1.4. A [D1,2]/[A2,3] stwintron belső intronjának alternatív elvesztése**

Az azonosított DHA1 ortológ gének közül 11-ben nem található stwintron, csak egy normál, kanonikus intron a vizsgált pozícióban. A filogenetikai adatok alapján is valószínűsíthető, hogy a 73 DhaoS fehérjét kódoló gén közül azok, amelyek kanonikus intronnal rendelkeznek, korábban szintén stwintront hordozhattak, melynek belső intronját elvesztették. Két faj, nevezetesen *Chalara longipes* és *Coniochaeta ligniaria* esetén azt láttuk, hogy az intron, első esetben leucin (Leu), illetve a második gomba genomjában metionin (Met) kodonját választja el egy szerint (Ser) kódoló triplettől. Mivel két kodon közé találjuk beékelődve, phase-0 intronról van szó, ami az alternatív [D1,2]/[A2,3] stwintron belső intronjának [A2,3]-kénti elvesztésével alakulhatott ki. A

többi, kanonikus intront hordozó genom esetében, a Leu kodonja van megszakítva, annak második és harmadik nukleotidja között, tehát az intron phase-2 intron és az stwintron belső intronjának [D1,2] típusú stwintronkénti elvesztésének következménye.

Az intronvesztés történhet egy reverz transzkriptáz által mediált mechanizmus révén (Fink, 1987; Roy és Gilbert, 2006) is, azonban az általunk vizsgált esetek pontosan reprezentálják az alternatív [D1,2]/[A2,3] stwintron belső intronjának kivágását az elsődleges mRNS átiratból, illetve elvesztését DNS szinten (10. ábra).



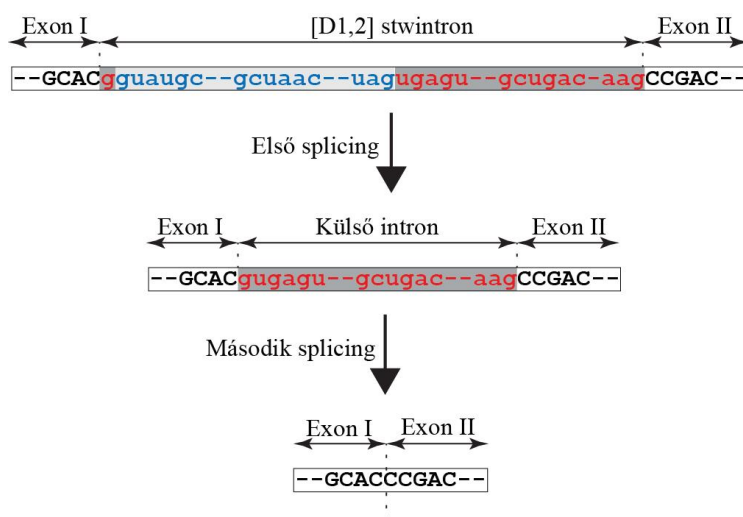
**10. ábra: Az egy nukleotidos intron elcsúzás mechanizmusának bemutatása, egy ősi, alternatív [D1,2]/[A2,3] stwintronból kiindulva.** Az stwintron belső intronjának elvesztése, kivágódásától függően, két eltérő intron pozíciót eredményez DNS szinten. (saját szerkesztésű ábra Fekete és mtsai., 2017 alapján)

## 5.2. LipS ortológok vizsgálata

### 5.2.1. Egy új [D1,2] stwintron *A. nidulans* és *A. niger* putatív lipáz génjében

Az „Anyagok és módszerek” részben ismertetett módon, a létrehozott stwintron modell segítségével, tovább kutattunk [D1,2] típusú stwintronok után *A. nidulans* genomszekvenciájában. Több más jelölt mellett, egy putatív [D1,2] stwintronra (11. ábra) bukkantunk, az említett fonalas gomba AN7524-es lókuszon található génjében. Ez a gén auto-annotáció alapján két megszakító szekvenciát tartalmaz, egyet 69 nt távolságra a start kodontól, mely egy rövid, 52 nt hosszúságú kanonikus intron és egy másik, 289 nt hosszúságú phase-2 intront, az elsőtől 497 nt távolságra downstream irányba, amely egy arginint (Arg) kódoló kodont (CGA) szakít meg. Ezek alapján az érett mRNS pedig egy 824 aminosavból álló fehérjét kódol. Az stwintron séma alapján lefuttatott keresés egy [D1,2] stwintron jelenlétét mutatta a második, nagyméretű intron 5' végén, melynek feltételezett donor szekvenciája az stwintron belső, 53 nt hosszú intronjához tartozhat. Ez a prediktált belső intron egy 46 nt nagyságú külső intronba beágyazódva található, annak donor szekvenciáját megszakítva az első és második nukleotidja között. Ahogyan az a 11. ábrán is látható, a 99 nt hosszúságú, prediktált stwintron két egymást követő U2 splicing reakció révén kerülhet eltávolításra, és egy prolint (Pro) kódoló kodont (CCC) választ ketté, annak első és második nukleotidja közé beékelődve. Az intron definíciót (Berget, 1995) szigorúan véve, a második intron pozícióban található stwintron és a

harmadik pozícióban lévő intron eltávolítása után, a köztük található, 142 nt hosszúságú exon a kódoló szekvenciában marad, melyet az auto-annotáció nem ismert fel. A 289 nt nagyságú, auto-annotált intron másik végén egy rövid (49 nt), kanonikus U2 intron jelenlétét feltételeztük. Az AN7524-es lókuszt általunk prediktált intron-exon struktúrája alapján a gén négy exont tartalmaz.



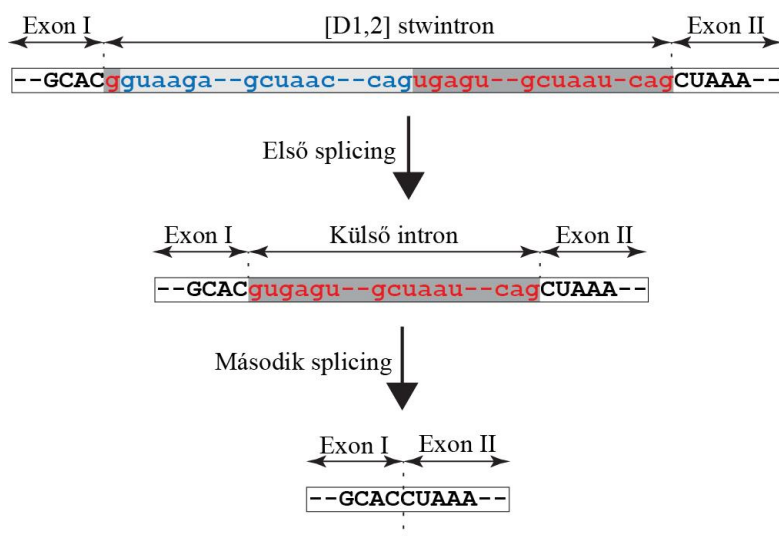
**11. ábra:** *A. nidulans* putatív lipázt kódoló génjében leírt [D1,2] stwintron és kétlépes splicingjának sematikus ábrája. (A színek megegyeznek a korábbi ábrákban használtakkal.) (saját szerkesztésű ábra Flipphi és mtsai., 2017 alapján)

Az elérhető RNA-seq (RNS szekvencia) adatok (Cerqueira és mtsai., 2014) azt mutatták, hogy a vizsgálni kívánt *A. nidulans* gén sokkal nagyobb mértékben expresszálódik, ha a gombát komplex táptalajon növesztjük, minimál táptalajon történő növesztéssel összevetve. Nem találtunk viszont olyan szekvenciaadatot, amelyből a komplett stwintron struktúra, sem olyat, melyből az auto-annotáció alapján prediktált nagyméretű intron hiányozna. Úgy döntöttünk, hogy megpróbáljuk kísérletesen igazolni az *in silico* kimutatott stwintront, illetve a génen

belül fellelhető három megszakító szekvencia splicingját, az „Anyagok és módszerek” fejezetben ismertetett RT-PCR technika segítségével. A kimutatott cDNS (GenBank Accession Number: MF612150) egy 2.616 nt nagyságú ORF-et határoz meg, amely feltehetően egy 871 aminosavból álló fehérjét kódol. Az első intron a His23 és Tyr24 kodonok közé ékelődik be, míg a [D1,2] stwintron a Pro189, a harmadik intron pedig a Ser236 kodont szakítja meg. Kísérleteink eredményeként kimutattuk a [D1,2] stwintron feltételezett kivágódási köztesét is, a korábbiakban ismertetett módon tervezett speciális primerpár segítségével. Az RT-PCR reakciók után detektált termékek közül a kisebb méretű annak az mRNS intermediernek felelt meg, melyből az 53 nt nagyságú belső intron már kivágódott és ezáltal a külső intron donor szekvenciája funkcióképesé vált. A kivágódási köztes általunk meghatározott, pontos szekvenciája megtalálható a GenBank adatbázisában (GenBank Accession Number: MF612152).

A feltételezett, „négyintronos” génmodellünk valódiságát alátámasztja *A. niger* egy elérhető EST klónja (Asn\_02874; Accession Number: DR703192). Bizonyítottuk a [D1,2] stwintron és a két másik intron pozícióját is az ortológ génben (lókusz: ASPNIDRAFT\_53020), cDNS szekvenáltatás révén, beleértve a teljes kódoló régió cDNS-ét (GenBank Accession Number: MF612151), mely egy 909 aminosavból felépülő fehérjét kódol *A. niger* ATCC1015 törzsben. A 109 nt hosszú stwintron a Pro184 kodont (CCU) szakítja meg, az első citozin után (12. ábra). A prediktált stwintron 52 nt hosszúságú belső intronjaaz 57 nt-ból álló külső intront a donor szekvenciájában szakítja meg, az első és a második nukleotidja között. *A. niger* esetében is sikerült kísérletesen igazolnunk a

feltételezett splicing intermediert (GenBank Accession Number: MF612153), melyből a belső intron már eltávolításra került.



**12. ábra:** *A. niger* putatív lipáz kódozó génjében leírt [D1,2] stwintron és kétlépéses splicingjának sematikus ábrája. (A szinkódok megegyeznek a korábbi ábrákon használtakkal.) (saját szerkesztésű ábra Flippi és mtsai., 2017 alapján)

## 5.2.2. Ortológ gének vizsgálata

Az AN7524-es lókuszon lokalizált, stwintront tartalmazó génről képződő fehérjetermék nagyfokú hasonlóságot (46%-os egyezést) mutat egy putatív, 316 aminosavból felépülő, bakteriális lipáz/észteráz (EC 3.1.1-carboxylic ester hydrolase) enzimmal (RefSeq Protein Accession Number: WP\_054929686). Ez a bakteriális fehérje tartalmaz egy jellegzetes  $\alpha/\beta$  hidroláz fold 3 domént (Pfam07859) és a hormone-sensitive\_lipase\_like\_1 enzimesaládba sorolják az  $\alpha/\beta$  hidroláz superfamily-n belül (Lenfant és mtsai., 2013). Az AN7524-es lókuszon található gént ezért *lipS*-ként (**lip**ase gene with **st**wintron) nevezetük el,

utalva rá, hogy egy stwintront tartalmazó, putatív lipáz enzimet kódoló génről van szó. A fungális fehérje jelentősen nagyobb, mint a bakteriális  $\alpha/\beta$  hidrolázok. Az előzetes BLAST keresések azt mutatták, hogy megtalálható egy nagymértékű hasonlóságot mutató ortológ protein *Pezizomycotina*-k és *Taphrinomycotina*-k körében, illetve a főbb *Basidiomycota* osztályokon belül is; az ortológ gén viszont hiányzik *Saccharomycotina*-k genomjából. *Ascomycota*-k esetén nem találtunk nagyfokú hasonlóságot mutató paralógokat, mint például a Drug/H<sup>+</sup> antiporter (DHA1) esetén, ahol 73 *Pezizomycotina* fajban sikerült kimutatnunk a gént és az stwintront.

Az a 46 aminosav hosszúságú szakasz, mely a szekvencia központi részén található és a bakteriális  $\alpha/\beta$  hidroláz „fold 3” doménnel mutat nagyfokú egyezést, a gén harmadik exonja által kódolt, mely a [D1,2] stwintron és a tőle downstream irányban található normál intron között helyezkedik el. A kimutatott stwintron tehát a fehérje egy jól definiált, konzervált elemének kódoló régióját szakítja meg, ami megkönnyíti az ortológ génekben való elterjedtségének vizsgálatát. TBLASTN segítségével, az *A. nidulans* fehérjét véve alapul, több száz potenciális *lipS* ortológot gyűjtöttünk össze mindkét *Dikarya* törzs gombafajainak genomját vizsgálva. 52 *Basidiomycota* ortológ esetén azt vettük észre, hogy az stwintron pozíció nem hordoz megszakító szekvenciát, annak ellenére, hogy a gén, számos faj esetén bővelkedik intronokban; *Basidioascus undulatus* például 22 intront hordoz a *lipS* génjében, *Mrakia frigida* pedig 18-at. Az *Ascomycota*-k, azon belül *Taphrinomycotina*-k közé tartozó *Taphrina* fajok rendelkeznek egy egyedi intron pozícióval, 21 nt távolságra a start kodontól, míg másik öt faj (pl.: *Saitoella*

*complicata*) esetén a gén egyetlen intront sem tartalmaz. Másrészt, számos *Peizizomycotina* faj esetén azt láttuk, hogy az stwintron és az utána következő intron pozíciót is elfoglalja valamilyen megszakító szekvencia; az stwintron pozíció tehát az altörzs egyedi jellemzőjeként írható le. Erre a megállapításra azután jutottunk, hogy mintegy 150 *lipS* homológ gén intron-exon struktúráját manuálisan meghatároztuk nem-*Dikarya* fajokban, melyek közül egyesek akár több tucat *lipS* paralógot is hordoznak (*Conidiobolus incongruus* például 27-et).

A 292 *Ascomycota* faj putatív LipS fehérjéje alapján készült Maximum Likelihood törzsfát a 12.2. Függelék 19. ábrája mutatja be, ahol a *Taphrinomycotina* proteinek szerepelnek kulcsoportként. A fehérjék evolúciós rokonsága nagymértékben tükrözi a gombafajok taxonómiai viszonyait. Azon fajok esetén, melyek a korán divergáló *Peizizomycetes*-ek és *Orbiliomycetes*-ek közé tartoznak, abban az intron pozícióban, melyet *Aspergillus*-okban a [D1,2] stwintron foglal el, egy normál intront találunk. *Tuber melanosporum* esetén az U2 intron jelenlétét ebben a pozícióban EST adatok (Accession Number: FP429675) is megerősítik. *Dactylellina haptotyla* fajnál egy kivételes dolgot vettünk észre: a *lipS* gén duplikálódott a genomban; mindkét kópia tíz intront tartalmaz, illetve két olyan exont, melyek meglepően kicsi méretűek. Ezzel szemben *Pyronema omphalodes* úgy tűnik, hogy egy kivételével az összes intronját elveszítette, beleértve az stwintron pozícióban található is. Az *Aspergillus*-oknál jellemző, harmadik intron pozíció (142 nt távolságra az stwintrontól, downstream irányba) pedig érdekes módon egyáltalán nem hordoz megszakító szekvenciát a két, korán divergáló *Peizizomycotina* osztály fajai között; úgy tűnik, hogy ez a



pozíció a *Leotiomyces*-k (Schoch és mtsai., 2009) egyedi jellemzője. Az első *Aspergillus* intron pozíció (69 nt távolságra a start kodontól) viszont *T. melanosporum* ortológjában is jelen van, tehát valószínűleg ősbibb.

A *lipS* [D1,2] stwintron jelen van *Eurotiomyces*, *Lecanoromyces* és *Leotiomyces* osztályokba tartozó gombafajok ortológ génjeiben és valószínűleg ősbibb struktúra lehet, mint a többi, általunk kimutatott stwintron. Ezekben a *Leotiomyces* fajokban mindhárom intron pozíció, mely az *A. nidulans* génben is jelen van konzerváltnak tűnik, habár számos más osztály tagjai vesztettek el intronikus szekvenciákat. *Sordariomyces* és *Dothideomyces* osztályokban épp az ellenkező esettel találkozhatunk: nem találunk megszakító szekvenciát sem az stwintron, sem a harmadik *A. nidulans* intron pozícióban. Nagy részük hordoz intront, az *Aspergillus*-okra is jellemző első helyen, viszont például *Pleosporales* fajoknál a *lipS* gén intron-exon struktúrája egészen más mintázatot mutat három, az eddig ismertektől teljesen különböző intron pozíciókkal. A legtöbb *Capnodiales* fajban a *lipS* gén pedig egyáltalán nem is tartalmaz intront. *Leotiomyces*-k között nem találunk példát arra, hogy az stwintron pozíciót egy normál, kanonikus intron foglalná el, ez csak a *Pezizomyces* és *Orbiliomyces* fajoknál figyelhető meg. A *LipS* törzsfa és a különböző osztályok gombafajainak intron mintázata alapján azt feltételezhetjük, hogy a [D1,2] stwintron evolúciósan később alakulhatott ki, egy intron (belső intron), egy másik, már meglévő intronba (külső intron) történő ágyazódása révén.

A 12.2. Függelék 20. és 21. ábráin láthatjuk azokat az *Eurotiomyces* és *Leotiomyces* nemzetségeket, melyek bizonyos gombafajai rendelkeznek stwintronnal; a törzsfán jelöltem az stwintron vesztési

folyamatokat is. *Eurotiomycetes*-ek esetén az stwintron teljes mértékben hiányzik (12.2. Függelék, 20. ábra) *Chaetothyriales*, *Verrucariales* és *Phaeomoniellales* rendekbe tartozó, elérhető genomszekvenciával rendelkező gombafajokból, továbbá az *Eurotiales* rendbe tartozó *Penicillium*, *Monascus* és *Xeromyces* nemzetségek egyedeinek genomjából. *Leotiomycetes*-ek közül pedig nyolc fajban nincs jelen (12.2. Függelék, 21. ábra). Érdekes még azt is megjegyezni, hogy az említett fajok esetében, ahol az stwintron eltűnését detektáltuk, azt figyeltük meg, hogy a harmadik *Aspergillus* pozícióból is eltűnt az intron. Egyedül *A. terreus*, *A. apis* és „*G. candidum* 3C” esetében talákoztunk azzal az esettel, hogy a downstream intron benne maradt a génszekvenciában, míg *Marssonina brunnea* és *Calycina herbarum* fajoknál pont az ellenkezőjét tapasztaltuk: egyedül a harmadik intron pozíciójukból veszítették el megszakító szekvenciát.

### **5.3. Egy [D2,3] stwintront tartalmazó *bioDA* gén és BioDA ortológok vizsgálata**

Egy korábbi projektünk alkalmával már azonosítottunk egy másik ősi stwintron struktúrát is, egy, a D-biotin bioszintéziséhez szükséges, bifunkcionális enzimet kódoló struktúrgénben (*bioDA*). A protein mind detiobiotin szintetáz (EC 6.3.3.3 - BioD), mind S-adenozil-L-metionin:8-amino-7-oxononanoát aminosztransferáz (EC 2.6.1.62 – BioA) aktivitást mutat (Magliano és mtsai., 2011). Kimutattuk, hogy a *Sordariomycetes* osztályba tartozó fajok nagy részében (számszerint 31 gombafajban), továbbá kettő, a *Dothideomycetes* osztály, *Botryosphaeriales* rendjébe

tartozó fajban egy [D2,3] stwintron található az első (*A. nidulans bioDA* génje alapján) intron pozícióban. Filogenetikai analízis (Flippi és mtsai., 2013;) eredményeként azt láttuk, hogy a két *Botryosphaerales* protein a *Sordariomycetes* fehérjékkel alkot egy klasztert, nem pedig a többi *Dothideomycetes* (24 faj) eredetű proteinnel. *Nectria haemotococca* (korábbi nevén: *Fusarium solani f. sp. pisi*) fajban egy normál intront találunk abban a pozícióban, amelyet más *Sordariomycetes*-ekben, beleértve más *Fusarium* fajokat is stwintron foglal el. 2013 tavaszától többszáz új *Pezizomycotina* genomszekvencia vált nyilvánosan elérhetővé különböző adatbázisokban. Ezáltal frissíthettük a *bioDA* stwintronnal kapcsolatos korábbi, evolúciós aspektusú analízisünket.

Átvizsgáltuk tehát az adatbázisokat TBLASTN segítségével, az *A. nidulans* fehérjét (Accession Number: ACR44943) véve alapul. A keresés több száz, potenciális *bioDA* gént eredményezett, melyek intron-exon struktúráját manuálisan határoztuk meg. Az új adatok fényében, arra a megállapításra jutottunk, hogy az intron pozíció, melyet egyes esetekben stwintron foglal el, a gombák evolúciójának (Hibbett és mtsai., 2007) már korai szakaszában is jelen volt. A szóban forgó intron pozíció egy standard U2 intront hordoz *Mucoromycotina*, *Mortierellomycotina*, *Entomophthoromycotina* és *Chytridiomycota* fajokban, viszont hiányzik *Rhizophagus irregularis* (*Glomeromycota*) és *Blastocladiomycota* fajokból. Megemlítendő továbbá, hogy mindkét, elérhető genomszekvenciával rendelkező *Batrachochytrium dendrobatidis* törzs ezen intronjának donor eleme egy „nem-kanonikus” szekvencia: 5'-GAAAGA. Az intron pozíció konzervált a *Basidiomycota* törzs mindhárom altörzsén belül. Az stwintron pozícióban normál intront

találtunk továbbá három *Saccharomycetales* faj (*Saprochaete clavata*, *Lipomyces starkeyi* és *Sugiyamaella xylanicola*) esetén is, azok közül, melyek genomjában egyáltalán megtalálható a *bioDA* gén. Másrészt abban a hét, *bioDA* génnel rendelkező *Taphrinomycotina* fajban, melyek genomja elérhető, a vizsgált pozícióban nem találunk semmilyen intronikus szekvenciát.

Filogenetikai analízist végeztünk 298 *Ascomycota* *BioDA* protein bevonásával, melyeket manuálisan határoztunk meg, a *bioDA* ortológ gének szekvenciája alapján. A törzsfán a *Taphrinomycotina* törzs kulcsoportként jelenik meg, míg a *Saccharomycotina* fehérjék a *Pezizomycotina*-k testvér kládjaként (12.2. Függelék, 22. ábra).

A putatív lipáz (LipS) fehérje evolúciós vizsgálata esetén tapasztaltaktól eltérően, a *Pezizomycotina*-k *BioDA* proteinei nincsenek összhangban a fungális taxonómiával; nem mutat monofiletikus jelleget, többek között a *Sordariomycetes* (Maharachchikumbura és mtsai., 2015), *Dothideomycetes* (Hyde és mtsai., 2013) és *Eurotiomycetes* (Geiser és mtsai., 2006) osztályokon belül.

A vizsgált intron pozíció azonban sokkal konzerváltabbnak tűnik a *Pezizomycotina*-k *bioDA* génjében. A [D2,3] típusú stwintron jelen van a legtöbb *Sordariomycetes* és *Botryosphaerales* (*Dothideomycetes*) fajban, melyek ráadásul egy kládba rendeződnek (amint az a 12.2. Függelék, 22. ábráján is látható), habár nem minden faj rendelkezik a *bioDA* stwintronnal a kládon belül. A többi *Dothideomycetes* (beleértve *Pleosporales* és *Capnodiales* rendeket), csakúgy, mint az *Eurotiomycetes* osztály tagjai esetén egy standard U2 intron van jelen az stwintron

pozícióban. Másrészt, a *Leotiomyces* osztály és az *Umbilicariomycetidae* család elérhető genomszekvenciával rendelkező fajaiiban ez az ősi intron pozíció már egyáltalán nincs jelen.

Néhány, a *Sordariomyces* osztály *Ophiostomatales* rendjébe tartozó fajnál úgy tűnik, hogy a *bioDA* gén egy *Leotiomyces* őstől eredeztethető és nem tartalmaz semmilyen intronikus elemet a gén vizsgált pozíciójában. Valószínűleg ez az egyik legmeggyőzőbb bizonyíték a *bioDA* gén laterális transzmissziójára, nem közvetlenül rokon *Pezizomycotina* taxonok között. Ez a különös géntranszfer játszódhatott le *Ophiostomales* fajok, *Rafaelea quercivora* és *Leptographium procerum* esetében is, akik rendelkeznek a *bioDA* stwintronnal és BioDA proteinjük a *Sordariomyces* rendbe tartozó fajok fehérjéivel klasztereződik. Az ilyen események nem egyediek a biotin bioszintéziséhez szükséges struktúrgének tekintetében, ugyanis például a *bioA*, illetve *bioD* gének *Saccharomyces cerevisiae* élesztőgombában bizonyítottan bakteriális eredetűek (Hall és Dietrich, 2007).

A 12.2. Függelék 23. ábráján részletesebben bemutatásra kerültek azon fajok rokonsági viszonyai, melyek hordozzák a *bioDA* stwintront. A putatív lipáz stwintronnal megfigyeltektől eltérően, a *bioDA* stwintron elvesztése legtöbb esetben egy standard intron jelenlétét eredményezi ebben a pozícióban. A belső intron elvesztésére az stwintrontól legalább öt, egymástól független példát találunk a kládban, mely magában foglalja a *Sordariomyces* fajok túlnyomó hányadát. A *Botryosphaeriales* rend 11 vizsgált faja közül öt esetben figyeltük meg a belső intron elvesztését. Ez két független esemény lejátszódását feltételezi. Továbbá, két *Botryosphaeriales* faj, *Phyllosticta capitalensis* és *Neoscytalidium*

*dimidiatum* esetében nem találunk semmilyen, megmaradt intronikus szekvenciát az stwintron pozícióban, annak elvesztését követően. Jelen filogenetikai analízis azt sugallja, hogy a komplett stwintron elvesztése *N. dimidiatum* *bioDA* génjéből egyetlen esemény során következett be, csakúgy, mint a *lipS* gén esetén megfigyelt mechanizmusok során (12.2. Függelék 20. és 21. ábra). Ettől eltérően, *P. capitalensis* stwintron vesztése feltehetőleg két, egymást követő lépésben történt meg, ugyanis közeli rokona, *P. citricarpa* rendelkezik egy standard intronnal ebben a pozícióban.

Számos faj, a *Diaporthales* rendből, azon belül a *Diaporthaceae* és *Valsaceae* családokból, szintén rendelkezik a *bioDA* génnel, melyből az egész stwintron hiányzik. Az öt, egy klaszterbe rendeződő *Diaporthales* proteint kódoló gén úgy tűnik, hogy a *Botryosphaerales* taxonból eredeztethető, mely a fő *Sordariomycetes* ágból divergált. Érdekességként megemlíthető, hogy *Valsa mali*, szintén a *Diaporthales* rendbe tartozó faj, két *bioDA* génnel is rendelkezik (Accession Number: JUIY000000000 és JUIZ000000000), melyek közül az egyikből az stwintron teljes mértékben hiányzik, míg a másik a teljes [D2,3] stwintron struktúrát hordozza. Az elvégzett filogenetikai analízis eredménye arra enged következtetni, hogy a *bioDA* stwintron szekvenciák hiánya az öt *Diaporthales* fajban (pl.: *N. dimidiatum*-ban) egyetlen intron vesztési esemény eredménye, míg a *Botryosphaerales* rend egy másik, divergáló ágába rendeződő fajok (*Lasiodiplodia theobromae*) (még mindig) hordoznak stwintront a vizsgált génben.

## 5.4. Egy [D5,6] stwintron és lehetséges szerepeinek feltárása

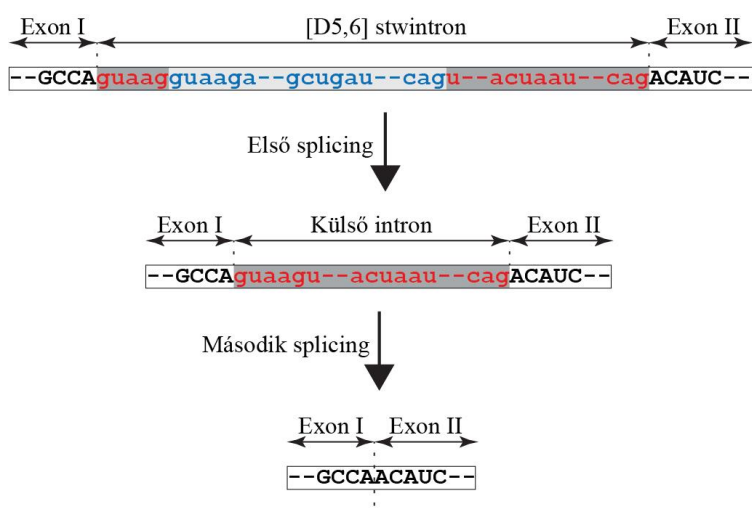
### 5.4.1. [D5,6] stwintron *A. nidulans rtnA* génjében

*A. nidulans* genomjában, az AN5404 lókuszú génben egy lehetséges [D5,6] stwintront azonosítottunk, mely esetén a belső intron a külső intront, annak donor szekvenciájában, az ötödik és a hatodik nukleotid között szakítja meg. A potenciális iker-intron az első megszakító szekvencia a génben, közel az 5' véghez; emellett még tartalmaz további öt U2 intront. A gén teljes, érett mRNS-éről átírt cDNS-t klónoztuk és megszekvenáltattuk (GenBank Accession Number: MK410458). Sikeresen kimutattuk a splicing intermediert is, melyből a 75 nt hosszúságú belső intron már eltávolításra került, viszont a 76 nt nagyságú külső intron még megtalálható a szekvenciában (GenBank Accession Number: MK410459). Az stwintronokra jellemző, kétlépéses eltávolítási mechanizmus a 13. ábrán kerül bemutatásra, a [D5,6] stwintron esetén.

Az AN5404-es lókusz feltételezhetően egy 326 aminosavból álló „reticulon-like” proteint kódol (UniProt Accession Number: C8VGL9; GenBank Accession Number: CBF81965). Az auto-annotáció a PF02453 fehérjecsaldába sorolta, amely olyan proteineket tartalmaz, melyeknek a funkciója ismeretlen, de az endoplazmatikus retikulummal hozhatóak kapcsolatba, valószínűleg ahhoz asszociálva töltik be szerepüket. Ilyen, „reticulon-like”-nak nevezett fehérjék (Rtn) széles körben elterjedtek az *Eukarya* doménen belül (Oertle és mtsai., 2003). *A. nidulans* RtnA proteinje feltételezhetően négy transzmembrán és egy „coiled-coil”

doménből épül fel. A coiled domén (melyet a 250-300 aminosav kódol) egyértelműen kimutatható mind a 779 ortológ szekvenciát összevetve. Viszont azt humán RTN4 fehérjevariánsok esetén is kimutatták, hogy a fehérjeszekvencia végei rendezetlenséget mutatnak (Li és Song, 2007).

Hét fonalas gombafaj (*A. nidulans*, illetve rajta kívül még hat másik faj) esetén vetettük kísérletes vizsgálat alá a reticulon-like ortológokat. Olyan lehetséges Rtn fehérjéket választottunk ki, melyek szekvenciájuk alapján szintén rendezetlenek az N-, illetve C-végükön.



**13. ábra:** A [D5,6] stwintron struktúrájának és kétlépes splicingjának sematikus ábrája. (A színek megegyeznek a korábbi ábrákon használtakkal.) (saját szerkesztésű ábra Kavalecz és mtsai., 2019 alapján)



## 5.4.2. Ortológok vizsgálata

*A. nidulans* *rtnA* génjének ortológjait szeretnénk volna feltárni más *Acomycota* gombák genomjában is. A keresés eredményeként azt láttuk, hogy a [D5,6] stwintron majdnem minden, elérhető genomszekvenciával rendelkező, vizsgált *Pezizomycotina* fajban jelen van, beleértve a korán divergáló *Pezizomyces* és *Orbiliomyces* fajokat is. Összesen 754 gént gyűjtöttünk ki az adatbázisokból és vizsgáltuk, hogy tartalmazzák-e a [D5,6] stwintron struktúrát. Számos gombafaj esetén, melyek a *Dothideomyces*, *Xylonomycetes* vagy *Lecanoromyces* osztályokba tartoznak, feltételeztük egy másik stwintron jelenlétét is, a gén második intron pozíciójában, melyben többek között *A. nidulans* egy kanonikus U2 intront hordoz.

A [D5,6] stwintron struktúra meglétét kísérletesen igazoltuk *A. nidulans* után *Aspergillus niger* (*Eurotiomyces*), *Botrytis cinerea* (*Leotiomycetes*), *Trichoderma reesei* (*Sordariomyces*), *Neurospora crassa* (*Sordariomyces*) és *Helminthosporium solani* (*Dothideomyces*) genomjában is (12.3. Függelék, 5. táblázat). Ezenkívül, az NCBI EST adatbázisában talált szekvenciaadatok (Accession Number: FP417649) egyértelműen bizonyítják, hogy *Tuber melanosporum* (*Pezizomyces*) ortológ génje is rendelkezik [D5,6] stwintronnal.

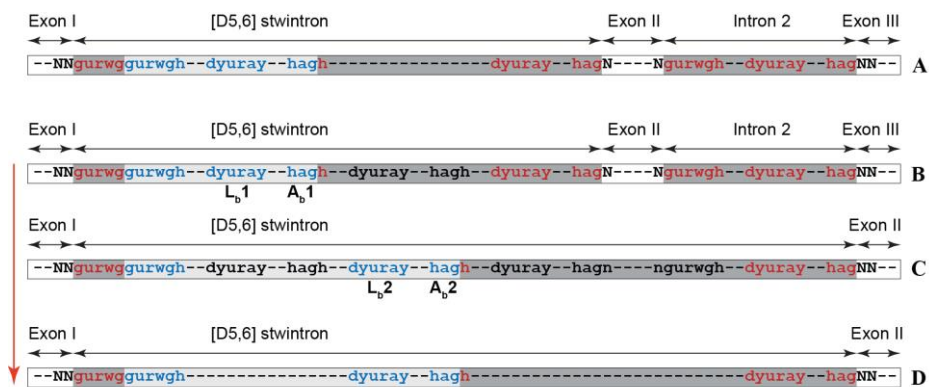
Az ortológ RtnA fehérjék alapján készítettünk egy Maximum Likelihood törzsfát (az ábra, annak komplexitása miatt jelen dolgozatban nem kerül bemutatásra; elérhető Kavalecz és mtsai., 2019 referencia alapján) is, mely jól illusztrálja az evolúciós viszonyokat. Néhány faj

esetén (<3%) azt találtuk, hogy az stwintron belső intronja hiányzik, vagyis csak egy normál U2 intronnal rendelkeznek a vizsgált pozícióban. A belső intron véletlenszerű hiánya többszöri, független (stw)intron vesztési események lejátszódását sugallja. Annak ellenére, hogy a második intron pozícióban (*A. nidulans* alapján) is [D5,6] stwintront találunk *Pleosporales*, *Hysteriales* és *Botryosphaeriales* fajok esetén, *Dothideomycetes* genuszok (pl.: *Capnodiales*, *Myriangiales*, *Aureobasidium*) tagjai egyáltalán nem is hordoznak megszakító szekvenciát a génben. A *Sordariomycetes* osztályba tartozó fajok *rtnA* ortológjai pedig a harmadik, negyedik és ötödik intron pozíciónak (*A. nidulans* alapján) megfelelő helyeken nem tartalmaznak intronokat, viszont rendelkeznek egy, csak erre az osztályra jellemző intronnal, 14 nt távolságra a negyedik intron pozíciótól, downstream irányba. Ezenkívül, *Orbiliomycetes*-ek és *Pezizomycetes*-ek ortológ génjükben hordoznak egy további intront, 70 nt távolságra, downstream irányba a hatodik intron pozíciótól (*A. nidulans* alapján). A gén 3' végéhez közeli intron pozíciók alapján arra következtethetünk, hogy a reticulon-like gének számos, a két másik *Ascomycota* altörzshöz tartozó faj esetén *bona fide* ortológjai *A. nidulans rtnA* génjének. Mindemellett, egyetlen fajt sem találunk a *Taphrinomycotina* vagy *Saccharomycotina* altörzsekben belül, amely hordozna [D5,6] stwintront a reticulon-like génjében. Találtunk (stwintront nem tartalmazó) strukturális homológ géneket nem *Dykaria* törzsekben is, *Glomeromycota*, *Mucoromycotina* és *Mortierellomycotina* fajok esetén; ezzel szemben a *Dykaria Basidiomycota*-kban nem. Úgy tűnik, hogy a kimutatott [D5,6] stwintron evolúciósan stabil a *Pezizomycotina* altörzsön belül.

### 5.4.3. A [D5,6] stwintron és a szomszédos exon elvesztése

A phase-1 típusú megszakító szekvenciákat, melyek az első *A. nidulans* intron pozíciót foglalják el (ilyen például maga a [D5,6] stwintron is), valamint a szekvenciában utána elhelyezkedő, második intron pozíciót, az RtnA protein N-terminálisának közelében találjuk meg. Ez a régió rendezetlen, és mind aminosav sorrendjében, mind hosszában kismértékű konzerváltságot mutat a *Pezizomycotina* altörzsön belül. Ez a variábilis régió nagy valószínűséggel remek lehetőséget nyújt az intron-exon struktúra természetes módon történő átalakulásaira. Sok esetben azt láttuk az *Eurotiomycetes*, *Sordariomycetes*, *Orbiliomycetes*, *Pezizomycetes* osztályokon belül, hogy a második intron pozícióból a megszakító szekvencia eltűnt, a mellette, upstream irányban elhelyezkedő exonnal együtt. Azon gombafajok közül, melyeket kísérletesen is vizsgáltuk (12.3. Függelék, 5. táblázat), *T. reesei* esetén mutattuk ki ezt a jelenséget. Az stwintron mellett, downstream irányban található exon 15 nt hosszúságú a fekete *Aspergilli*-k esetén, míg más *Aspergilli* és *Penicillia* fajokban 24 nt nagyságú (12.2. Függelék, 24. ábra). Az *Aspergillaceae* család más nemzetségeinek körében, mint *Monascus*-ok és *Xeromyces*-ek, a második exon és a második intron is hiányzik az *rtnA* génből; úgy tűnik, hogy a [D5,6] stwintron egy 3' irányba meghosszabbodott változata helyettesíti őket. A *Trichocomaceae* család minden egyedénél, melyek genomszekvenciája elérhető, azt láttuk, hogy szintén hiányzik a második intron és exon. Minden *Eurotiales* faj esetén a harmadik *A. nidulans* intron pozíciótól upstream irányba található exon szintén nagyon konzervált. Megfigyeléseink alátámasztják feltevésünket,

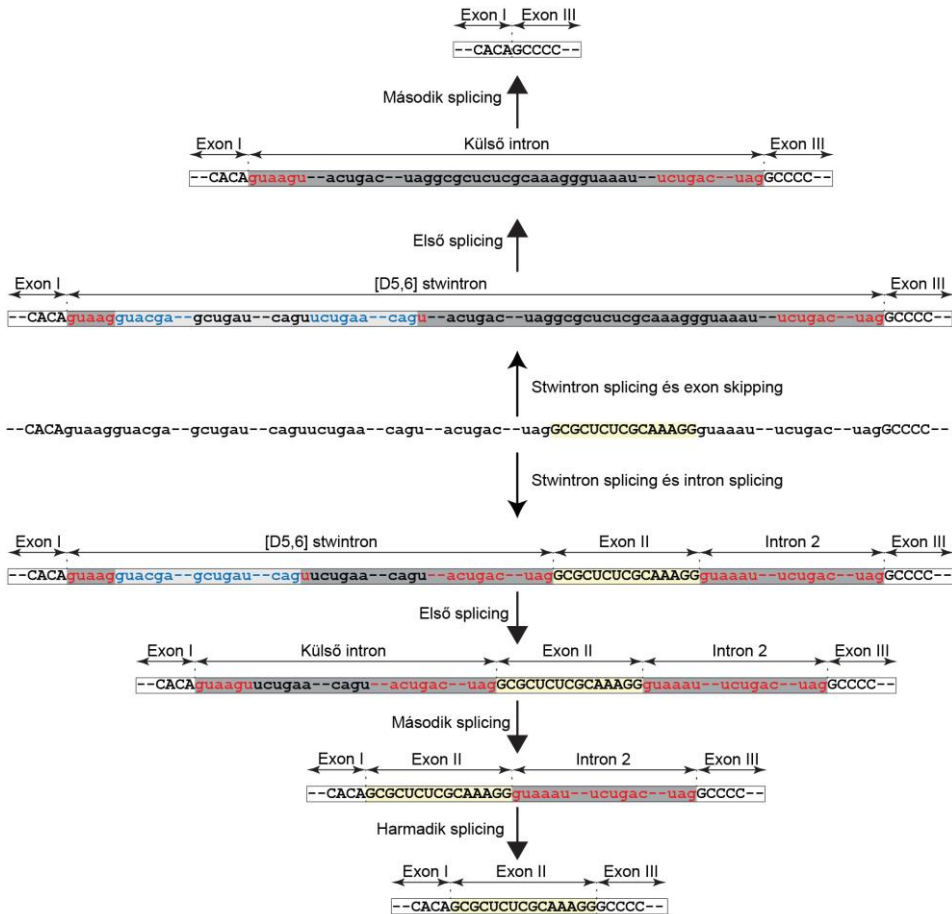
miszerint exonikus szekvencia intronizációja (Irimia és mtsai., 2008) történhetett az exon II esetén, mely eredetileg az stwintron és a tőle downstream irányba található normál intron között helyezkedett el (*A. nidulans* első és második intron pozíciója). Exonikus szekvenciák elvesztése (pl.: *Trichocomaceae*, *Monascus*, *Xeromyces* fajok esetén) megtörténhetett úgy is, hogy egy intron „bekebelezte” azt, mint ahogyan a [D5,6] stwintron külső intronjának részévé vált az exon II. Ezt az intronizációs, evolúciós folyamatot akár egy filogenetikailag állandósult, DNS szinten lejátszódó alternatív splicing eseménynek is felfoghatjuk, amely az exon II kihagyását eredményezte a kódoló szekvenciából (14. ábra).



**14. ábra: A vizsgált [D5,6] stwintron szerepe exon skipping és evolúciós exon vesztés lejátszódásában.** Az *A. nidulans* transzkript, alternatív 3' hasítóhely hiányában, nem képes alternatív kivágódásra, az exon II mindig benne marad az érett mRNS szekvenciában (A). Ezzel szemben *A. niger* elsődleges mRNS átírata alternatív kivágódásra is képes; így exon II vagy benne marad a transzkriptumban, vagy kivágásra kerül (B és C). Végül, *Monascus ruber* transzkript esetén azt láthatjuk, hogy egy hasítóhely elvesztésének következtében, az *A. niger* esetén alternatív kivágódásra képes exon II ezen fajban teljes mértékben intronizálódott és mindig eltávolításra kerül (A színek megegyeznek a korábbi ábrákon használtakkal; a piros nyíl jelzi az exon vesztési folyamat irányát.) (saját szerkesztésű ábra Kavalecz és mtsai., 2019 alapján)

#### **5.4.4. Exon skipping jelenség kimutatása *A. niger rtnA* génjének transzkriptumában**

Vizsgálataink azt mutatták, hogy a [D5,6] stwintron részt vehet az *rtnA* transzkriptum alternatív splicingjában is, exon skipping jelenség lejátszódása révén, melynek következtében az exon II is kivágódik. Az exon skipping úgy mehet végbe, hogy az stwintron belső intronjának kivágódását követően, a második splicing reakcióban a külső intron donor szekvenciája mellett az stwintrontól downstream irányba található, kanonikus intron (intron 2) akceptor szekvenciája vesz részt. Az stwintronhoz kötött alternatív splicing jelenség lejátszódását olyan szempontból is megközelíthetjük, hogy két, alternatív [D5,6] stwintron létezését fogadjuk el. Az egyik (nevezzük elsődleges stwintronnak), melynek splicingja az intron definíciónak szigorúan megfelelően történik és egy második, az előbbieken ismertett, másodlagos stwintron, mely magában foglalja az elsődleges stwintron, exon II és intron 2 szekvenciáját is. Ennek a másodlagos alternatív stwintronnak a kétlépéses eltávolítása exon I és exon III közvetlen kapcsolódását eredményezi, miközben exon II eltávolításra kerül a második splicing reakcióban, az intronikus szakaszokkal együtt.



**15. ábra: Az *A. niger* *rtnA* transzkriptumának alternatív splicingja.** Az elsődleges mRNS átírat számunkra releváns szakaszát az ábra közepén helyeztem el. A lefelé mutató nyilak irányába haladva láthatjuk azt a három, egymás utáni splicing eseményt, melyek következtében, a 15 nt hosszúságú exon II (világossárga háttérrel jelölve) benne marad a végső mRNS termékben. A felfelé mutató nyilak irányába haladva két, egymást követő splicing reakció sematikus ábrázolását és az exon II, intronikus szakaszokkal együtt történő kivágását látjuk. (A szinkódok megegyeznek a korábbi ábrákon használtakkal.) (saját szerkesztésű ábra Kavalecz és mtsai., 2019 alapján)

Az stwintron szekvenciákat átfogó vizsgálat alá vetettük más *Aspergillaceae* fajok esetén is. Kíváncsiak voltunk, hogy az ortológ gének stwintronjai között találunk-e még példát erre a jelenségre, hogy a belső

intron egy alternatív 3' akceptorhelyet használhat kivágódásához. Azt láttuk, hogy az esemény lejátszódhat *A. niger* esetén, ahol egy 5'-CAG található 26 nt távolságra a CAG akceptor szekvenciától, downstream irányba az elsődleges stwintron belső intronjában (15. ábra). Az alternatív 3' hasítóhely asszociálhat egy távolabbi lariat elágazási pont szekvenciaelemmel. Ennek következtében a másodlagos belső intron 29 nukleotiddal hosszabb lesz a 3' vég irányába és az eltávolítása egy kanonikus donor elemet (5'-GUAAG|U) eredményez, ez azonban nem tud az elsődleges stwintron külső intron akceptorjával asszociálni, mivel a putatív elágazási pont szekvenciaelem, a nagyobb méretű belső intron eltávolításának következtében, túl közel kerül a donor elemhez. A külső intron donor szekvenciája így kénytelen a következő lehetséges lariat és akceptorhelyet használni, melyek tulajdonképpen a kanonikus intron 2-höz tartoznak. A [D5,6] stwintron külső intronja így magában foglalja az exon II és intron 2 szekvenciáját is és ezen elemekkel együtt történik meg a kivágódása. Feltételezhető, hogy a másodlagos külső intron kanonikus donor szekvenciája (5'-GUAAG|U) a normál intron 2 donor szekvenciájával (5'-GUAAAU) „verseng” ugyanazon 3' hasítóhelyért (15. ábra). Ebben valószínűleg szerepet játszanak bizonyos asszociációs fehérje komplexek is, mint az Exon Junction Complex (Hauer és mtsai., 2016; Woodward és mtsai., 2017).

Ugyanazokat a primereket használva, mint az elsődleges stwintron splicing intermedierjeinek kimutatásához, azt tapasztaltuk, hogy az RT-PCR reakciók során felszaporodó cDNS klónok kis hányada (<5%) nem tartalmazta a 105 nt hosszú alternatív belső intront (GenBank Accession Number: MK410463), amit a másodlagos stwintron feltételezése esetén

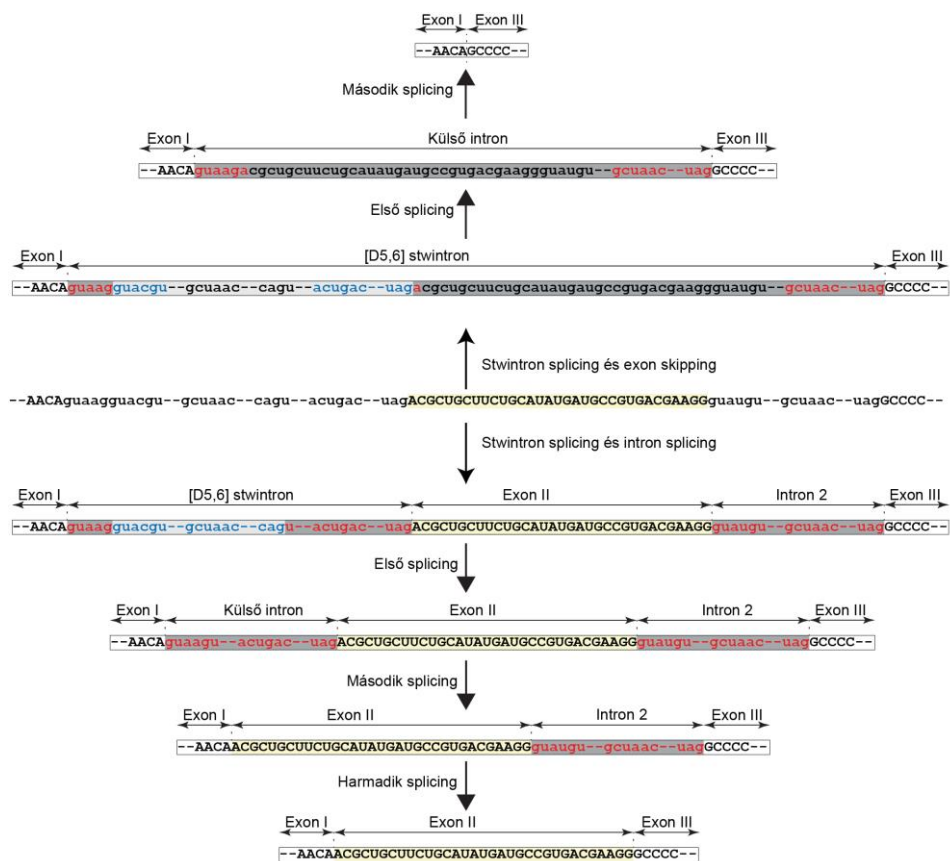
prediktáltunk. Különböző exonikus primereket alkalmazva, kimutattunk az érett mRNS-nek megfelelő, cDNS fragmenseket is, melyekből a 15 nukleotid nagyságú exon II is hiányzik (GenBank Accession Number: MK410462). *A. niger rtnA* transzkriptumának általunk kimutatott alternatív splicingja összhangban van az NCBI Sequence Read Archive (SRA) (Kodama és mtsai., 2012) adatbázisában fellelhető szekvencia adatokkal is.

#### **5.4.5. Exon skipping jelenség kimutatása *N. crassa rtnA* génjének transzkriptumában**

Az NCBI adatbázisát behatóbban vizsgálva, találtunk tíz EST adatot *N. crassa rtnA* génjéről, melyekben a 33 nt hosszúságú exon II bennmaradt az mRNS szekvenciában, viszont további három EST esetén kivágódásra került (Accession Number: GH144932, GH 135593 és GH146396). A talált szekvencia adatok erőteljesen azt sugallták, hogy *A. niger*-hez hasonlóan, *N. crassa rtnA* génje is alternatív splicingon, exon skipping-en eshet át, az mRNS érése során. A másodlagos stwintron kivágódását bizonyító splicing intermediert szintén detektálták; a szekvencia megtalálható az NCBI Sequence Read Archive (SRA) (Kodama és mtsai., 2012) adatbázisában. Ennek az alternatív belső intronnak az 5' vége megegyezik az elsődleges [D5,6] stwintron belső intronjának donor szekvenciájával (16. ábra), a 3' hasítóhelye pedig az elsődleges [D5,6] stwintron külső intronjának akceptor végével. Az alternatív stwintron belső intronjának kivágódása tehát nem teljesen az *A. niger*-nél tapasztaltakkal analóg módon történik, de az eredmény



ugyanahhoz az eseményhez, az exon II eltávolításához vezet. A belső intron kivágódásának következtében egy új donorhely generálódik (5'-GUAAG|A), amely a legközelebbi lehetséges elágazási pont és akceptor szekvenciával fog interakcióba lépni, melyek ebben az esetben az intron 2 szekvenciaelemei. Így a meghosszabbodó, másodlagos külső intronnal együtt exon II is kivágódik, tehát exon skipping játszódik le.



**16. ábra: *N. crassa rtaA* transzkriptumának alternatív splicingja.** A [D5,6] stwintront és az intron 2 kanonikus U2 intront magában foglaló exon skipping jelenség bemutatása *N. crassa*-ban. A kivágódó exon II ebben az esetben 33 nt hosszúságú. (Az ábra felépítése és szinkódjai megegyeznek a 15. ábránál ismertettekkel.) (saját szerkesztésű ábra Kavalecz és mtsai., 2019 alapján)

Eredményeink alátámasztják az *rtnA* gén elsődleges mRNS átíratának alternatív splicingját, *A. niger* és *N. crassa* fonalas gombákban. Továbbá jogosan feltételezhetjük, hogy a jelenség lejátszódásában kulcsfontosságú szerepet játszik a leírt [D5,6] típusú stwintron.

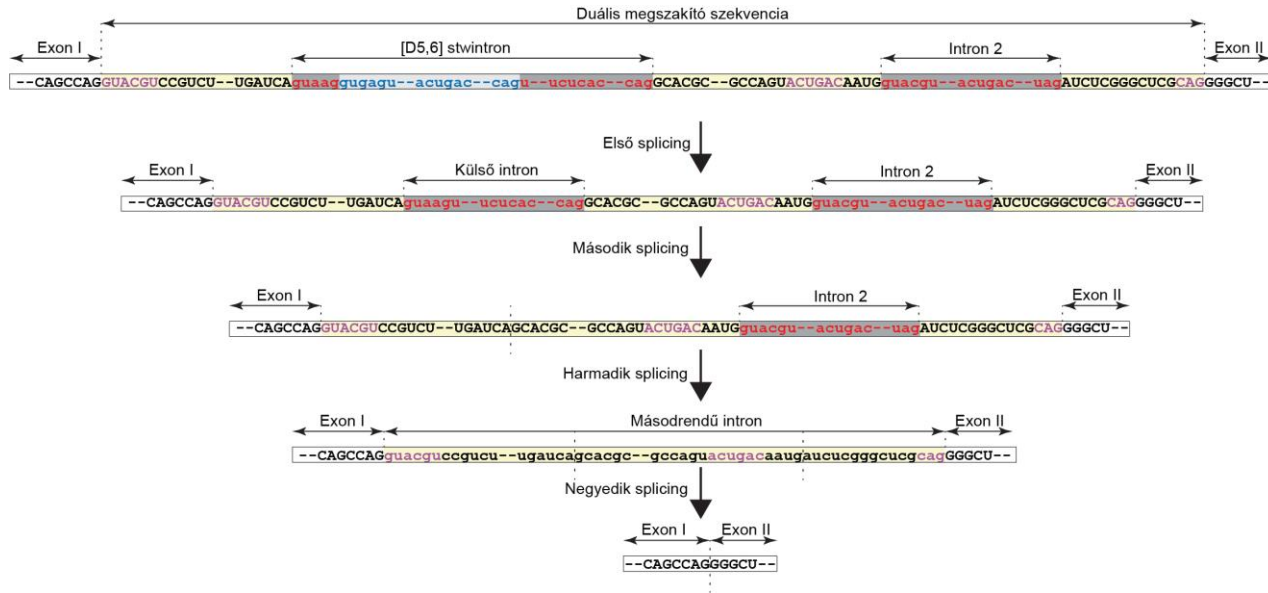
#### **5.4.6. Egy duális komplex megszakító szekvencia feltárása *M. cinnamomea*-ban**

Két kivételtől eltekintve, minden vizsgált, elérhető genomszekvenciával rendelkező *Onygenales* faj rendelkezik komplex, összetett megszakító szekvenciával a reticulon-like gén első intron pozíciójában. Esetükben is egy [D5,6] stwintront tudtunk azonosítani, míg a második intron pozíciót egy kanonikus U2 intron foglalja el. Az *Aspergilli* fajok esetén megállapított génmodell nagyban hasonlít *Onygenales*-ek *rtnA* génjére, az utolsó 4 intron pozíció ezen fajok esetén is erősen konzervált, azonban az stwintron, illetve a második intron nem minden esetben phase-1 helyzetű. Az *Onygenales* fajok csaknem felénél azt találtuk, hogy az exon II nukleotidjainak száma nem osztható hárommal. Ezenkívül, 12 faj esetében, ahol az exon hossza ugyan három többszöröse, a szóban forgó szekvencia tartalmaz egy stop kodont is. Más fajok génszekvenciájának vizsgálata során pedig arra lettünk figyelmesek, hogy amennyiben a teljes [D5,6] stwintron és a második U2 intron is eltávolításra kerül, az mRNS szekvencia nem tartalmaz start kodont.

*Byssoonygena ceratinophila* és *Ophidiomyces ophiodiicola* gombafajok esetén pedig csak öt intront tartalmaz a vizsgált gén, hat helyett; sőt mi több, ezek mindegyike kanonikus, normál intron. Az

említett két faj esetében az első intron pozícióban egy phase-1 típusú U2 intron található, amely feltételezéseink szerint a gén első két megszakító szekvenciájának fúziójával jöhetett létre, míg a további négy intron, 3' irányba, az *Aspergilli* fajoknál leírtakkal azonos pozíciókban találhatóak.

A génszekvenciák alapján feltételeztük, hogy minden más *Onygenales* fajban egy funkcionális [D5,6] stwintront (az első pozícióban), továbbá egy kanonikus U2 intront (a második pozícióban) körülvesz egy komplex megszakító szekvencia, mely a két említett intronikus szakaszt el is választja egymástól, továbbá körülöleli őket mind az stwintrontól 5', mind az U2 introntól 3' irányba (17. ábra). A három, „intronizálódott” szekvencia ősi kódoló régiókból eredeztethető. Ezt a nem mindennapi megszakító szekvenciát „duális” intronnak is nevezhetjük, ugyanis több splicing reakció lejátszódása szükséges, hogy a struktúra teljes mértékben eltávolításra kerüljön az mRNS érése során. *B. ceratinophila* és *O. ophioidiicola* fajokban ez a duális megszakító szekvencia egy kanonikus U2 intronná alakult át, miután elveszítette a belső intron-exon struktúráját. Ezt a hipotézist, különböző *Onygenales* fajok elérhető expressziós adatai is alátámasztják, ahol az érett mRNS-t nem szakítja meg a feltételezett duális intronikus szekvencia. Az NCBI Expressed Sequence Tags és Transcriptome Shotgun Assembly adatbázisaiból származó EST és TSA adatok megtalálhatóak a 12.3. Függelék 6. táblázatában. A lista tartalmaz egy adatot *B. ceratinophila rtnA* mRNS-ére vonatkozóan is, mely bizonyítja egy újonnan kialakult kanonikus intron létét az első pozícióban.



**17. ábra: *Malbranchea cinnamomea rtnA* génjének duális megszakító szekvenciája és a teljes eltávolításához szükséges négy, egymást követő splicing reakció sematikus ábrája.** A duális megszakító szekvencia magában foglalja a [D5,6] stwintront, az Intron 2 kanonikus intront és egy megszakított, másodrendű intront, melybe beágyazódva találjuk az előbbi kettőt. A duális megszakító szekvencia egyes elemeinek feltételezett eltávolítási sorrendje a következő: 1. a [D5,6] stwintron belső intronja (világosszürke háttérrel és kék betűkkel jelölve), 2. a [D5,6] stwintron külső intronja (sötétszürke háttérrel és piros betűkkel jelölve), 3. Intron 2 (szintén sötétszürke háttérrel és piros betűkkel jelölve), 4. Másodrendű intron (sárga háttérrel, fekete, illetve a splicing elemei lila színnel jelölve).

(saját szerkesztésű ábra Kavalecz és mtsai., 2019 alapján)

*Malbranchea cinnamomea* (Morgenstern és mtsai., 2012) reticulon-like fehérjét kódoló génjében kísérletesen is igazoltuk a duális megszakító szekvencia jelenlétét (GenBank Accession Number: MK421638). A 17. ábrán látható, hogyan történik az összetett szekvencia eltávolítása; a [D5,6] stwintron és a második intron pozícióban található kanonikus U2 intron kivágódásának kell először megtörténnie, hogy az utána kialakuló, „másodrendű” intron splicingja is megtörténhessen. Az elsődleges mRNS átíratban a másodrendű intron három, egymástól más szekvenciák által elválasztott darabban van még jelen. Az 5' donor régiója a [D5,6] stwintrontól upstream irányban található, az elágazási pont szekvenciája az stwintron és az intron 2 között, míg a 3' akceptor szakasza az intron 2-t követően helyezkedik el a szekvencián belül. A másodrendű intron így, az stwintron belső és külső intronjának, valamint a kanonikus intron eltávolításának megtörténte után, egy negyedik splicing reakció révén kerülhet kivágásra, kialakítva ezáltal az érett mRNS szekvenciát.

Kísérletesen igazoltuk a három splicing intermediert, valamint kimutattuk az érett mRNS-t is *M. cinnamomea*-ban (GenBank Accession Number: MK421639, MK421640, MK421641, MK410473). Az egyes splicing reakciók lejátszódásának sorrendje esetenként eltérő lehet. Elméletileg az intron 2 eltávolítása megtörténhet az stwintron külső intronjának kivágódása előtt, de akár már a belső intron splicingját megelőzően is. Az sem zárható ki, hogy az egyes splicing események sorrendje időnként különböző, azonban az utolsó eltávolítási lépés, értelemszerűen mindig a „másodrendű” U2 intron kivágása. Habár *Ascomycota* gombákban mindezidáig nem került pontos meghatározásra az exon fogalma (Berget, 1995; De Conti és mtsai., 2013), az *Onygenales*

*rtnA* transzkriptum esetén azt tapasztaltuk, hogy a teljes duális, komplex megszakító szekvencia eltávolítása semmiképp nem történhet meg egyetlen lépésben, ugyanis a „másodrendű” intron három splicing eleme (5'-donor, lariat elágazási és 3'-akceptor szekvencia) egymástól elszeparáltan helyezkedik el, a beágyazódó intronikus szakaszok révén.

*Coccidioides immitis* és *Coccidioides posadasii* fajok EST adatai között találtunk három szekvenciát (GenBank Accession Number: GH365468, GH423005, GH438822), melyek a „másodrendű” intront egy folytatólagos, egybefüggő szekvenciadarabként tartalmazzák. Az első két sequence tag esetén a szekvenciából hiányzik a többi, komplex megszakító szekvenciától downstream irányba található kanonikus intron (intron 3-6). Ebből arra következtethetünk, hogy a „másodrendű” intron az utolsóként eltávolításra kerülő megszakító szekvencia az *rtnA* mRNS-ben, annak ellenére, hogy ez található legközelebb a gén 5' végéhez. A harmadik EST (GH438822) viszont még tartalmazza a negyedik és ötödik kanonikus intront is, a „másodrendű” intron mellett.

Eredményeink révén bizonyítást nyert tehát egy komplex „duális” megszakító szekvencia léte, mely négy, egymást követő splicing reakciót követően kerülhet csak teljes eltávolításra.

## 6. Eredmények megvitatása

### 6.1. DHA1 ortológok és az egy nukleotidos intron elcsúszás vizsgálata egy alternatív stwintronon keresztül

A *dhaoS* gén vizsgálata során leírt, egy nukleotidos intron elcsúszás (exonikus információ elvesztése, vagy szerzése nélkül egy ORF-en belül) nem magyarázható a genom mutáció általi megváltozásával, sokkal inkább egy, az intron egyik végén végbemenő kompenzációs eseménnyel (Stoltzfus és mtsai., 1997; Lynch, 2002). Egy olyan lehetséges forgatókönyv is valószínűtlennek tűnik ebben az esetben, melynek során az intron-exon csatlakozási pont elmozdulása eredményezi az intron elcsúszását. Erre a jelenségre több bizonyított példát is találunk, azonban ezeknél az intron-exon csatlakozási pont elcsúszása általában egy másik hasítóhely használatának eredménye, melynek következtében mind az exonikus, mind az intronikus szekvenciák hossza megváltozik (Higashimoto és Liddle, 1993; Schäfer és mtsai., 1999). Egy „kiegészítő” mutáció lejátszódása az intron egyik végén szintén eredményezhet hasonló intermediert (Lehmann és mtsai., 2010), habár komparatív vizsgálatok alkalmával azt figyelték meg, hogy az egyes szekvenciák esetén, az intron-exon kapcsolódási pontok közelében detektálható eltérések gyakran tévesek (Bocco és Csűrös, 2016). Mindenesetre úgy tűnik, hogy bizonyos intron pozíciók elterjedésében nagyban hozzájárul az alternatív splicing mechanizmus és egyes intron elcsúszási események lejátszódása a genomban (Séraphin és Rosbash, 1990; Brackenridge és

mtsai., 2003; Tarrío és mtsai., 2008). Konkrét intermediereket azonban mindeztáig nem tudtak azonosítani, melyek az alternatív splicing következtében létrejött intron elcsúszást bizonyítanák. Egy másik lehetséges alternatív mechanizmus, amely magában foglalja az intron kivágódását, reverz splicingját, reverz transzkripcióját és homológ rekombinációját (Martinez és mtsai., 1989), pedig túlságosan komplex ahhoz, hogy egy ilyen esemény lejátszódását feltételezzük a jelenség mögött.

A leírt [D1,2]/[A2,3] stwintron és alternatív eltávolítási módjainak vizsgálata segítségével ésszerű magyarázatot adhatunk az egy nukleotidos intron shift jelenségére, eukarióta genomban. Mindemellett bizonyítottuk az stwintronok relevanciáját az intronok evolúciós vizsgálatát illetően.

A szakirodalomban nem találunk rá példát, hogy korábban kimutattak volna intron elcsúszási folyamatot a köztes lépések feltárásával együtt. Erre nekünk a [D1,2] és [A2,3] stwintronok (Flippfi és mtsai., 2013) evolúciós stabilitása és egymásba való átalakulásuknak lehetősége adott módot. Ellentétben az intron-exon csatlakozási pontok közelében lejátszódó szekvencia átalakulásokkal, az stwintron mediált intron shift lehetőséget nyújt a kódoló régiók és így a géntermék szekvenciájának pontos megőrzésére. Egy stabil, köztes állapot alakul ki DNS szinten, mielőtt az intronvesztés bekövetkezne. Fontosnak tartom kiemelni, hogy ez az új, intron elcsúszást előidéző mechanizmus, mely alternatív kivágódásra képes stwintronokból eredeztethető, csak szomszédos pozíciókba történő elmozdulást eredményezhet, nem ad magyarázatot a nagyobb távolságokba történő intron elcsúszási eseményekre.



Az elért kutatási eredmények, melyek egy *A. nidulans* génben leírt [D1,2] stwintron felfedezéséből indultak, segítségünkre voltak, hogy az ortológ gének vizsgálata során rábukkanjunk az alternatív [D1,2]/[A2,3] stwintronra. Több ortológot is bevonva a vizsgálatba, azt láttuk, hogy egyes esetekben, csak normál, kanonikus intronokat találunk szomszédos pozíciókban. A rendelkezésre álló genomadatok alapján azonban nem dönthető el egyértelműen, hogy melyik intron pozíció volt jelen a közös *dhaoS* ortológ ősbén. Az sem állapítható tehát meg teljes bizonyossággal, hogy az ősi megszakító szekvencia egy phase-0 vagy phase-2 kanonikus intron volt-e, vagy egy [D1,2], avagy [A2,3], esetleg egy alternatív [D1,2]/[A2,3] stwintron. Viszont azon feltételezésünk bizonyítást nyert, miszerint a [D1,2]/[A2,3] stwintron belső intronjának elvesztése DNS szinten, intronok egy nukleotidos elcsúszását eredményezheti (10. ábra).

## **6.2. A BioDA és LipS ortológok, és stwintronok evolúciójának vizsgálata**

A BioDA ortológok vizsgálata során kapott eredményeink azt mutatják, hogy a fehérje filogenezise nem idomul a fungális taxonómiához, azonban gombák azon csoportjai, melyek rendelkeznek a [D2,3] stwintronnal, egy jól elkülöníthető klaszterbe csoportosulnak (12.2. Függelék, 22. ábra). A [D2,3] stwintron a legtöbb *Sordariomycetes* esetén és egyes *Botryosphaerales* fajokban azon pozícióban található, melyet *Dothideomycetes* és *Eurotiomycetes* fajokban egy normál, U2 intron foglal el. Úgy tűnik, hogy ez egy rendkívül ősi intron pozíció; emiatt felvetődik a kérdés, hogy vajon a vizsgált [D2,3] stwintron valóban

úgy alakulhatott-e ki, hogy egy már meglévő „host” intron donor elemének második és harmadik nukleotidja közé beékelődött egy másik intron, ezzel kialakítva az stwintron struktúrát a *bioDA* génben? Az is elképzelhető, hogy a belső intron, szekvenciabeli átrendeződések következtében alakult ki a „host” intronon belül, egy általunk korábban „stwintronizáció”-nak elnevezett folyamat révén (Flipphi és mtsai., 2013), mely hasonlatos az exonikus szekvenciák intronizációjához (Irimia és mtsai., 2008). Egy másik lehetséges forgatókönyv lehet a [D2,3] stwintron kialakulását illetően, hogy intron inszerció történt valamilyen váratlan esemény következtében, például a DNS kettős szálának törését követő „non-homologous end joining” javítómechanizmus lejátszódása révén (Farlow és mtsai., 2011). Az stwintron kialakulásának, azaz egy új intron felbukkanásának egy, a genomban már meglévő intron hasítóhelyének belsejében, itt felvázolt két, eltérő módja segíthet általánosságban is megérteni az egyes organizmusok külső forrásokból történő intron szerzését (Yenerall és Zhou, 2012).

A putatív lipáz (LipS) protein filogenezeise (12.2. Függelék, 19. ábra) arra enged következtetni, hogy a [D1,2] stwintron, melyet megtalálhatunk számos, a *Leotiomyces* taxon különböző osztályain belül, intron inszerció révén jöhetett létre, egy kanonikus U2 intront már tartalmazó pozícióban. A korán divergáló *Pezizomyces* és *Orbiliomyces* osztályokban ez utóbbi megmaradt standard intronként, míg *Leotiomyces*-k stwintronjának külső intronja eredeztethető ebből a kanonikus intronból. Összegyűjtött adataink alapján azonban nem egyértelmű, hogy a *lipS* stwintron az előtt vagy után jelent meg, hogy a *Sordariomyces* osztály különvált a többi *Leotiomyces*-től.

Mindenesetre azt láttuk, hogy a *Sordariomycetes*-ek semmilyen intronikus elemmel nem rendelkeznek a *lipS* gén stwintron pozíciójában. Noha ez egy, ebben a pozícióban korábban jelen lévő standard intron, vagy stwintron struktúra elvesztésének következménye is lehet.

A *lipS* ortológok evolúciója során, stwintron veszteségi események nagy valószínűséggel számos alkalommal lejátszódhattak a *Leotiomyceta* osztályokon belül, beleértve más mechanizmusokat, melyek az stwintron hiányához vezettek a *Dothideomycetes* osztály tagjainál, az *Eurotiomycetes* osztály *Chaetothyriomycetidae* szuperrendjénél és a *Penicillium* genusz (*Eurotiomycetes* osztály, *Eurotiales* rend) minden egyes tagjánál. Érdekes, hogy minden esetben, ahol az stwintron eltűnését tapasztaljuk, ott a gén harmadik intron pozíciójából (az *Aspergillus lipS* alapján), 142 bp távolságra az stwintrontól, downstream irányba, szintén eltűnt vagy legalábbis hiányzik az intron. Különböző megszakító szekvenciák egyidejű elvesztése 4 faj kivételével (*A. terreus*, *A. apis*, *C. variabilis* és „*G. candidum* 3C”) minden más esetben megfigyelhető (12.2. Függelék 20. és 21. ábra).

Másrészt, *Sordariomycetes*-ek esetén az stwintron elvesztése a *bioDA* génből, egy kivételtől eltekintve, egy normál U2 intron szekvencia hátramaradását eredményezte (12.2. Függelék, 23. ábra). Úgy tűnik tehát, hogy a legtöbb *Sordariomycetes bioDA* ortológ génben, azok az események, melyek egy belső intron megjelenését idézik elő, fordított esetben a hiányához is vezethetnek.

A vizsgált két, ősi stwintron közti szembetűnő különbség ellenére, minden egyes megfigyelt intron veszteségi esemény egyetlen

mechanizmusra vezethető vissza, mely egyes transzkriptumok reverz transzkripcióját és az így létrejött cDNS és a genom megfelelő lókuszának homológ rekombinációját foglalhatja magában (Derr, 1998; Roy és Gilbert, 2006). Ez a mechanizmus eredményezheti, egyazon exont két oldalról határoló két intron egyidejű eltűnését, ugyanakkor egy stwintron belső intronjának elvesztését is, amennyiben az stwintron köztes kivágódási transzkriptum a reverz transzkripció templátja, nem pedig az érett mRNS. A feltételezett folyamat lejátszódásának esélye, mely során egy stwintron egy normál intronná alakul, nagyban függhet a splicing intermedier sejtmag beli élettartamától is, valamint a belső, illetve a külső intron eltávolításának, illetve a transzport folyamatoknak a gyorsaságától.

### **6.3. Egy stwintronhoz köthető alternatív splicing esemény és egy komplex megszakító szekvencia kimutatása**

Az *rtnA* ortológokban leírt [D5,6] stwintron segítségével közelebb kerültünk a spliceoszómális iker-intronok lehetséges funkcióinak megértéséhez. A 788 vizsgált gombafaj közül, az *Onygenales* rendbe tartozóak esetén új intron alakult ki az stwintron és egy másik, kanonikus U2 intron körül, feltehetőleg intronizációs folyamat révén. A létrejött, összetett struktúra rendkívül komplex. A teljes stwintron és a normál intron splicingjának is végbe kell mennie, ahhoz, hogy a körülöttük kialakult másodrendű intron kivágódhasson. Eredményeinkből arra következtethetünk, hogy az U2 intronok nem feltétlenül ko-transzkripcionálisan kerülnek eltávolításra.

A *Pezizomycotina* altörzs különböző taxonjaiban a [D5,6] stwintrontól downstream irányban található exon hiányzik az *rtnA* génből. Úgy tűnik, hogy az stwintron szekvenciája 3' irányba kibővült, „bekebelezve” a mögötte elhelyezkedő exont és U2 intront is. Ezen exonikus szekvenciák látszólagos intronizációja értelmezhető az exon skipping jelenségének állandósulásaként. Az stwintron szerepet játszhat az alternatív splicing említett típusában, mivel megvan rá a lehetőség, hogy a belső intronja egy alternatív akceptorhelyet használjon kivágódása során. *A. niger* és *N. crassa* gombafajokban bizonyítottuk is a jelenség lejátszódását. Amennyiben a [D5,6] stwintron belső intronja a távolabbi, alternatív 3' hasítóhelyet használja, eltávolítása után a külső intron 5' donorhelyének nincs lehetősége a konvencionális akceptorhelyét használni. Ehelyett a tőle downstream irányba elhelyezkedő kanonikus intron 3' hasítóhelyét veszi igénybe, melynek következtében a köztük lévő exonikus szakasz is eltávolításra kerül.

## 7. Összefoglalás

Az stwintronok olyan speciális megszakító szekvenciák az eukarióta genomban, melyek felépítésüket tekintve két, egymásba ágyazódott spliceoszómális intronból állnak. Az egyik intront (belső), egy másik (külső) intron valamely, a splicing folyamat végbemeneteléhez szükséges szekvenciaelemébe (5' hasítóhely, vagy donor szekvencia / 3' hasítóhely, vagy akceptor szekvencia / lariat, vagy elágazási pont szekvencia) beékelődve találjuk. Mivel a belső intron a külső intron egy, a kivágódás végbemenetele szempontjából esszenciális motívumát szakítja meg, a belső intron eltávolításának mindenképp meg kell előznie a külső intron splicingját. Az stwintronok felfedezését és egyes típusaik karakterizálását követően szeretnénk volna mélyrehatóbban vizsgálni a struktúrák splicing mechanizmusait, funkcióit és evolúcióját.

*A. nidulans dhaoS* génjében bizonyítottuk egy [D1,2] típusú stwintron jelenlétét. Ezen struktúra esetén a belső intron a külső intron donor elemét szakítja meg, annak első és második nukleotidja között. *T. reesei* ortológ génjében, azonos pozícióban pedig egy alternatív kivágódásra képes [D1,2]/[A2,3] stwintront karakterizáltunk, mely esetében két különböző, egymással egyetlen nukleotidban (G) átfedő belső intront fedezhetünk fel. Kísérletesen bizonyítottuk a struktúra [D1,2] és [A2,3] útvonalon történő eltávolítását is, illetve, hogy mindkét út azonos mRNS termékhez vezet. Bizonyítottuk továbbá, hogy az alternatív stwintron egyik, vagy másik belső intronjának alternatív elvesztése közeli, de nem azonos intron pozíciókat eredményezhet a genomban; ezzel egy lehetséges

magyarázattal szolgálva az egy nukleotiddal történő intron elcsúszás jelenségére.

Az egyre nagyobb számban rendelkezésre álló genomszekvencia adatoknak hála, a *lipS* és *bioDA* génekben leírt [D1,2], valamint [D2,3] típusú stwintronokat, azok kialakulását, vagy éppen eltűnését vizsgálni tudtuk gombák egész rendjein és osztályain belül. A *lipS* ortológok vizsgálata során azt tapasztaltuk, hogy az stwintron pozíciót egyes fajok esetén egy normál intron foglalja el. Ebből arra következtettünk, hogy a [D1,2] stwintron egy új intron, már korábban jelenlévő intronba történő beágyazódásával jöhetett létre. *Leotiomycetes*-ek és *Eurotiomycetes*-ek között viszont az stwintron teljes eltűnésére is találunk példát.

Ezzel szemben a *bioDA* ortológok vizsgálatánál azt láttuk, hogy az stwintron pozíciót bizonyos esetekben szintén egy normál intron foglalja el, viszont intron veszítési események lejátszódásával, amit egyes *Sordariomycetes* és *Botryosphaeriales* fajoknál mutattunk ki, a [D2,3] stwintron csak a belső intronját veszítette el, egy normál U2 intront hátrahagyva.

Több fonalas gombafaj ortológ génjében is kísérletesen igazoltuk, az először *A. nidulans rtnA* génjében kimutatott [D5,6] típusú stwintron jelenlétét. *A. niger* és *N. crassa* esetén bizonyítottuk, hogy a struktúra egy alternatív splicing esemény lejátszódásában is részt vehet, mégpedig a szekvenciában mellette elhelyezkedő (downstream irányban) exon kivágása révén. Amennyiben az stwintron belső intronja alternatív 3' hasítóhely használatával kerül eltávolításra, a külső intron, a szekvenciában utána következő normál intron akceptorhelyét veszi

igénybe, ezzel egyetlen nagy intronként kivágódva, magában foglalva az említett exonikus szakaszt is.

Az *rtnA* ortológok további vizsgálata során kimutattuk, hogy 29 *Onygenales* faj esetén egy többlépéses splicing mechanizmus játszódik le a transzkriptum érése során; ugyanis a gén egy komplex megszakító szekvenciát tartalmaz, mely magában foglalja a [D5,6] stwintront, a szomszédos kanonikus U2 intront, és egy ezeket körülölelő másodrendű intront is. A teljes struktúra eltávolításához szükséges négy splicing reakció kivágódási közteseit kimutattuk *M. cinnamomea* fonalas gombában.

Végezetül elmondható, hogy a megalkotott stwintron modell segítségével, fonalas gombák különböző génjeiben *in silico* újabb spliceoszómális iker-intronokat sikerült dekteltatunk és splicing közteseik kimutatása révén létezésüket bizonyítanunk. Emellett bebizonyosodott, hogy ezek a különleges stwintron struktúrák remek modell rendszerek lehetnek a spliceoszómális intronok eredetét, evolúcióját, genomon belüli mobilitását illetően. Részben választ kaptunk arra a kérdésre is, hogy vajon mi lehet a feladatuk, funkciójuk a genomban, hiszen kimutattuk, hogy szerepet játszhatnak alternatív splicing mechanizmusok (jelen esetben exon skipping) lejátszódásában. Biztos vagyok benne, hogy emellett még számos más funkcióval, akár a génexpresszióval kapcsolatos regulációs szereppel bírhatnak, de ezeknek a rejtélyeknek a megoldását már másra hagyom.



## 8. Summary

Stwintrons are unconventional intervening sequences in the eukaryotic genome. They consist of two spliceosomal introns; one of the introns (internal intron) is nested within one of the three canonical splicing motifs (5' splice site or donor sequence / 3' splice site or acceptor sequence / the lariat sequence around the branch point adenosine) of another intron (external intron). While the internal intron interrupts a sequence essential for splicing in the external intron, its removal has to take place first. After the discovery of stwintron structures our aims were to examine the function, evolution and splicing mechanisms of these unique intervening sequences.

We evidenced the existence of a [D1,2] stwintron in the *dhaoS* gene of *A. nidulans*, in which an internal intron interrupts the donor sequence of an external intron, between its first and second nucleotide. An alternative [D1,2]/[A2,3] stwintron has also been characterized in the orthologous gene of *T. reesei*. In this case we can detect two different internal introns, which overlap in one single nucleotide (G). The two alternative splicing pathways of this [D1,2]/[A2,3] stwintron has been experimentally proved and we concluded that both routes result the same mRNA product. Furthermore, the alternative loss of the internal intron can result close, but not identical intron positions in the genome; revealing a possible explanation for the one-nucleotide intron shift phenomenon.

Thanks to the increasing number of accessible genome sequences, we had the opportunity to study the emergence, loss and evolution of the *lipS* [D1,2] and the *bioDA* [D2,3] stwintrons throughout whole taxa and classes of fungi. We showed that the stwintron position is occupied by a canonical intron in some *lipS* orthologues. Thus, we hypothesised that this [D1,2] stwintron emerged by the appearance of a new intron within another ancient intron. However, we can find examples of total disappearance of the stwintron among *Eurotiomycetes*.

We experienced a different scenario regarding to *bioDA* orthologues: the stwintron position can also be occupied by a normal intron, however intron loss can take place among *Sordariomycetes* and *Botryosphaerales* species, which results the presence of a canonical U2 intron in the stwintron position by loss of the internal intron.

We proved the presence of a [D5,6] stwintron (first described in *A. nidulans rtnA* gene) in orthologous gene of many filamentous fungi. We showed the relevance of the stwintron regarding to alternative splicing events, namely exon skipping, in *A. niger* and *N. crassa*. If the internal intron of the stwintron is using an alternative 3' splice site, the external intron has to use the acceptor site of the neighbouring canonical intron in the gene sequence, thus the external intron is cleaved together with this U2 intron and a small exon in between.

In 29 species of *Onygenales* a complex intervening sequence has been revealed including the [D5,6] stwintron, the neighbouring U2 intron and a „second order intron”. The total removal of this complex structure requires

four consecutive splicing reactions as it has been experimentally proved in the transcript of the *M. cinnamomea* orthologous gene.

Summarizing the results, we can say that new stwintrons were found *in silico* using the recently created bioinformatic tool and experimentally confirmed via the detection of splicing intermediates. We proved that stwintrons can serve as excellent model systems regarding to spliceosomal intron loss or emergence, mobility and evolution. Furthermore, we showed that stwintrons can participate in alternative splicing events (in this case exon skipping), so we are closer to solve the mystery about stwintrons and their role in the genome. In my view, these unique structures would have many other functions connected to gene expression and regulation among others. However, I'll let others to solve these mysteries in the future.

## 9. Irodalomjegyzék

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 research*, 25(17), 3389-3402.

Amselem, J., Cuomo, C. A., Van Kan, J. A., Viaud, M., Benito, E. P., Couloux, A., ... & Fournier, E. (2011). Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS genetics*, 7(8), e1002230; 10.1371/journal.pgen.1002230

Andersen, M. R., Salazar, M. P., Schaap, P. J., Van De Vondervoort, P. J., Culley, D., Thykaer, J., ... & Berka, R. M. (2011). Comparative genomics of citric-acid-producing *Aspergillus niger* ATCC 1015 versus enzyme-producing CBS 513.88. *Genome research*, 21(6), 885-897.

Anisimova, M., & Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Systematic biology*, 55(4), 539-552.

Ág, N., Flippi, M., Karaffa, L., Scazzocchio, C., & Fekete, E. (2015). Alternatively spliced, spliceosomal twin introns in *Helminthosporium solani*. *Fungal Genetics and Biology*, 85, 7-13.

Bartschat, S., & Samuelsson, T. (2010). U12 type introns were lost at multiple occasions during evolution. *BMC genomics*, 11(1), 106.

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(8), 3171-3175.

Berget, S. M. (1995). Exon recognition in vertebrate splicing. *Journal of Biological Chemistry*, 270(6), 2411-2414.

Bocco, S. S., & Csűrös, M. (2016). Splice Sites Seldom Slide: Intron Evolution in *Oomycetes*. *Genome Biology and Evolution*, 8(8), 2340.

Bonen, L., & Vogel, J. (2001). The ins and outs of group II introns. *TRENDS in Genetics*, 17(6), 322-331.

- 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. *Proceedings of the National Academy of Sciences*, 106(8), 2577-2582.
- Brackenridge, S., Wilkie, A. O., & Sreaton, G. R. (2003). Efficient use of a 'dead-end' GA 5' splice site in the human fibroblast growth factor receptor genes. *The EMBO journal*, 22(7), 1620-1631.
- Brinkman, K. K., Mishra, P., & Fraser, N. W. (2013). The half-life of the HSV-1 1.5-kb LAT intron is similar to the half-life of the 2.0-kb LAT intron. *Journal of neurovirology*, 19(1), 102-108.
- 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(2), 661-674.
- Carmel, L., Wolf, Y. I., Rogozin, I. B., & Koonin, E. V. (2007). Three distinct modes of intron dynamics in the evolution of eukaryotes. *Genome research*, 17(7), 1034-1044.
- Cerqueira, G. C., Arnaud, M. B., Inglis, D. O., Skrzypek, M. S., Binkley, G., Simison, M., ... & Wymore, F. (2014). The Aspergillus Genome Database: multispecies curation and incorporation of RNA-Seq data to improve structural gene annotations. *Nucleic acids research*, 42(D1), D705-D710.
- Chow, L. T., Gelinis, R. E., Broker, T. R., & Roberts, R. J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12(1), 1-8.
- Christopher, D. A., & Hallick, R. B. (1989). *Euglena gracilis* chloroplast ribosomal protein operon: a new chloroplast gene for ribosomal protein L5 and description of a novel organelle intron category designated group III. *Nucleic acids research*, 17(19), 7591-7608.
- 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(2), 433-442.

- Copertino, D. W., Christopher, D. A., & Hallick, R. B. (1991). A mixed group II/group III twintron in the *Euglena gracilis* chloroplast ribosomal protein S3 gene: evidence for intron insertion during gene evolution. *Nucleic acids research*, 19(23), 6491-6497.
- Copertino, D. W., Shigeoka, S., & Hallick, R. B. (1992). Chloroplast group III twintron excision utilizing multiple 5'-and 3'-splice sites. *The EMBO journal*, 11(13), 5041-5050.
- Copertino, D. W., & Hallick, R. B. (1993). Group II and group III introns of twintrons: potential relationships with nuclear pre-mRNA introns. *Trends in biochemical sciences*, 18(12), 467-471.
- Copertino, D. W., Hall, E. T., Van Hook, F. W., Jenkins, K. P., & Hallick, R. B. (1994). A group III twintron encoding a maturase-like gene excises through lariat intermediates. *Nucleic acids research*, 22(6), 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 evolutionary biology*, 10(1), 210.
- Cushman, J. C., Christopher, D. A., Little, M. C., Hallick, R. B., & Price, C. A. (1988). Organization of the *psbE*, *psbF*, *orf38*, and *orf42* gene loci on the *Euglena gracilis* chloroplast genome. *Current genetics*, 13(2), 173-180.
- Dai, L., & Zimmerly, S. (2003). ORF-less and reverse-transcriptase-encoding group II introns in archaeobacteria, with a pattern of homing into related group II intron ORFs. *RNA*, 9(1), 14-19.
- De Conti, L., Baralle, M., & Buratti, E. (2013). Exon and intron definition in pre-mRNA splicing. *Wiley Interdisciplinary Reviews: RNA*, 4(1), 49-60.
- Derr, L. K. (1998). The involvement of cellular recombination and repair genes in RNA-mediated recombination in *Saccharomyces cerevisiae*. *Genetics*, 148(3), 937-945.

- de Vries, R. P., Burgers, K., van de Vondervoort, P. J., Frisvad, J. C., Samson, R. A., & Visser, J. (2004). A new black *Aspergillus* species, *A. vadensis*, is a promising host for homologous and heterologous protein production. *Applied and Environmental Microbiology*, 70(7), 3954-3959.
- 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(1), 76-86.
- Drager, R. G., & Hallick, R. B. (1993). A complex twintron is excised as four individual introns. *Nucleic acids research*, 21(10), 2389-2394.
- Duff, M. O., Olson, S., Wei, X., Garrett, S. C., Osman, A., Bolisetty, M., Plocik, A., Celniker, S. & Graveley, B. R. (2015). Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature*, 521(7552), 376-379.
- Einvik, C., Elde, M., & Johansen, S. (1998). Group I twintrons: genetic elements in myxomycete and schizopyrenid amoebflagellate ribosomal DNAs. *Journal of biotechnology*, 64(1), 63-74.
- Farlow, A., Meduri, E., & Schlötterer, C. (2011). DNA double-strand break repair and the evolution of intron density. *Trends in genetics*, 27(1), 1-6.
- Fekete, É., Fekete, E., Irinyi, L., Karaffa, L., Árnayasi, M., Asadollahi, M., & Sándor, E. (2012). Genetic diversity of a *Botrytis cinerea* cryptic species complex in Hungary. *Microbiological research*, 167(5), 283-291.
- Fink, G. R. (1987). Pseudogenes in yeast? *Cell*, 49(1), 5-6.
- 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.
- Galagan, J. E., Calvo, S. E., Borkovich, K. A., Selker, E. U., Read, N. D., Jaffe, D., ... & Rehman, B. (2003). The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*, 422(6934), 859-868.

- Geiser, D. M., Gueidan, C., Miadlikowska, J., Lutzoni, F., Kauff, F., Hofstetter, V., ... & Aptroot, A. (2006). *Eurotiomycetes: Eurotiomycetidae and Chaetothyriomycetidae*. *Mycologia*, 98(6), 1053-1064.
- Georgomanolis, T., Sofiadis, K., & Papantonis, A. (2016). Cutting a long intron short: recursive splicing and its implications. *Frontiers in physiology*, 7, 598.
- Gilbert, W. (1978). Why genes in pieces? *Nature*, 271(5645), 501-501.
- Gruber, M., Söding, J., & Lupas, A. N. (2006). Comparative analysis of coiled-coil prediction methods. *Journal of structural biology*, 155(2), 140-145.
- Grützmann, K., Szafranski, K., Pohl, M., Voigt, K., Petzold, A., & Schuster, S. (2014). Fungal alternative splicing is associated with multicellular complexity and virulence: a genome-wide multi-species study. *DNA research*, 21(1), 27-39.
- 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(3), 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 genetics and biology*, 53, 71-83.
- Hafez, M., & Hausner, G. (2015). Convergent evolution of twintron-like configurations: one is never enough. *RNA biology*, 12(12), 1275-1288.
- Hall, C., & Dietrich, F. S. (2007). The reacquisition of biotin prototrophy in *Saccharomyces cerevisiae* involved horizontal gene transfer, gene duplication and gene clustering. *Genetics*, 177(4), 2293-2307.
- Hatton, A. R., Subramaniam, V., & Lopez, A. J. (1998). Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Molecular cell*, 2(6), 787-796.



- Hauer, C., Sieber, J., Schwarzl, T., Hollerer, I., Curk, T., Alleaume, A. M., Hentze, M. W. & Kulozik, A. E. (2016). Exon junction complexes show a distributional bias toward alternatively spliced mRNAs and against mRNAs coding for ribosomal proteins. *Cell Reports*, 16(6), 1588-1603.
- Haugen, P., Simon, D. M., & Bhattacharya, D. (2005). The natural history of group I introns. *TRENDS in Genetics*, 21(2), 111-119.
- Henricson, A., Forslund, K., & Sonnhammer, E. L. (2010). Orthology confers intron position conservation. *BMC Genomics*, 11(1), 412.
- Hibbett, D. S., Binder, M., Bischoff, J. F., Blackwell, M., Cannon, P. F., Eriksson, O. E., ... & Lumbsch, H. T. (2007). A higher-level phylogenetic classification of the Fungi. *Mycological research*, 111(5), 509-547.
- Higashimoto, Y., & Liddle, R. A. (1993). Isolation and characterization of the gene encoding rat glucose-dependent insulinotropic peptide. *Biochemical and biophysical research communications*, 193(1), 182-190.
- Hong, L., & Hallick, R. B. (1994). A group III intron is formed from domains of two individual group II introns. *Genes & development*, 8(13), 1589-1599.
- Hyde, K. D., Jones, E. G., Liu, J. K., Ariyawansa, H., Boehm, E., Boonmee, S., ... & Diederich, P. (2013). Families of *Dothideomycetes*. *Fungal diversity*, 63(1), 1-313.
- International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931.
- Irimia, M., & Roy, S. W. (2014). Origin of spliceosomal introns and alternative splicing. *Cold Spring Harbor perspectives in biology*, 6(6), a016071.
- 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 in Genetics*, 24(8), 378-381.

- Comm. IUPAC-IUB. (1970). Abbreviations and symbols for nucleic acids, polynucleotides and their constituents. *Biochimica et biophysica acta. Nucleic acids and protein synthesis. Biochemistry*, 9(20), 4022-4027
- Janice, J., Jąkowski, M., & Makołowski, W. (2013). Surprisingly high number of Twintrons in vertebrates. *Biology direct*, 8(1), 4.
- Jo, B. S., & Choi, S. S. (2015). Introns: the functional benefits of introns in genomes. *Genomics & informatics*, 13(4), 112.
- Johansen, S., & Vogt, V. M. (1994). An intron in the nuclear ribosomal DNA of *Didymium iridis* codes for a group I ribozyme and a novel ribozyme that cooperate in self-splicing. *Cell*, 76(4), 725-734.
- Jones, D. T., & Cozzetto, D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*, 31(6), 857-863.
- Joseph, B., Kondo, S., & Lai, E. C. (2018). Short cryptic exons mediate recursive splicing in *Drosophila*. *Nature structural & molecular biology*, 25(5), 365-371.
- Kastner, B., Will, C. L., Stark, H., & Lührmann, R. (2019). Structural insights into nuclear pre-mRNA splicing in higher eukaryotes. *Cold Spring Harbor perspectives in biology*, 11(11), a032417, <https://doi.org/10.1101/cshperspect.a032417>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780.
- Kelly, S., Georgomanolis, T., Zirkel, A., Diermeier, S., O'Reilly, D., Murphy, S., Längst, G., Cook, P., R. & Papantonis, A. (2015). Splicing of many human genes involves sites embedded within introns. *Nucleic acids research*, 43(9), 4721-4732.
- Kempken, F. (2013). Alternative splicing in *Ascomycetes*. *Applied microbiology and biotechnology*, 97(10), 4235-4241.
- Khan, H., & Archibald, J. M. (2008). Lateral transfer of introns in the cryptophyte plastid genome. *Nucleic acids research*, 36(9), 3043-3053.

- Kim, E., Magen, A., & Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic acids research*, 35(1), 125-131.
- Kodama, Y., Shumway, M., & Leinonen, R. (2012). The Sequence Read Archive: explosive growth of sequencing data. *Nucleic Acids Research*, 40(D1), D54-D56.
- Koonin, E. V. (2006). The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biology direct*, 1(1), 22.
- Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology*, 305(3), 567-580.
- 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. *Eukaryotic cell*, 3(5), 1088-1100.
- Kyndt, T., Haegeman, A., & Gheysen, G. (2008). Evolution of GHF5 endoglucanase gene structure in plant-parasitic nematodes: no evidence for an early domain shuffling event. *BMC Evolutionary Biology*, 8(1), 305.
- Lambowitz, A. M., & Zimmerly, S. (2004). Mobile group II introns. *Annual Review Genetics*, 38, 1-35.
- Lefort, V., Longueville, J. E., & Gascuel, O. (2017). SMS: smart model selection in PhyML. *Molecular biology and evolution*, 34(9), 2422-2424.
- Lehmann, J., Eisenhardt, C., Stadler, P. F., & Krauss, V. (2010). Some novel intron positions in conserved *Drosophila* genes are caused by intron sliding or tandem duplication. *BMC evolutionary biology*, 10(1), 156.
- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., & Chatonnet, A. (2013). ESTHER, the database of the  $\alpha/\beta$ -hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Research*, 41(D1), D423-D429.

- Li, M., & Song, J. (2007). The N-and C-termini of the human Nogo molecules are intrinsically unstructured: Bioinformatics, CD, NMR characterization, and functional implications. *Proteins: Structure, Function, and Bioinformatics*, 68(1), 100-108.
- Lynch, M. (2002). Intron evolution as a population-genetic process. *Proceedings of the National Academy of Sciences*, 99(9), 6118-6123.
- Maddox, B. (2003). The double helix and the 'wronged heroine'. *Nature*, 421(6921), 407-408.
- Maharachchikumbura, S. S., Hyde, K. D., Jones, E. G., McKenzie, E. H., Huang, S. K., Abdel-Wahab, M. A., ... & Hongsanan, S. (2015). Towards a natural classification and backbone tree for *Sordariomycetes*. *Fungal Diversity*, 72(1), 199-301.
- Magliano, P., Flippi, 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 genetics and biology*, 48(2), 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. *Molecular and General Genetics MGG*, 246(1), 128-131.
- Mandels, M. M. & Andreotti, R. E. (1978) Problems and challenges in the cellulose to cellulase fermentation. *Process Biochemistry*, 13, 6–13.
- Martinez, P., Martin, W., & Cerff, R. (1989). Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. *Journal of molecular biology*, 208(4), 551-565.
- 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(5), 517-524.

- McGlinchy, N. J., & Smith, C. W. (2008). Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? *Trends in Biochemical Sciences*, 33(8), 385-393.
- Michel, F., Kazuhiko, U., & Haruo, O. (1989). Comparative and functional anatomy of group II catalytic introns - a review. *Gene*, 82(1), 5-30.
- 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(4), 506-511.
- Mohr, G., Ghanem, E., & Lambowitz, A. M. (2010). Mechanisms used for genomic proliferation by thermophilic group II introns. *PLoS biology*, 8(6).
- Morgenstern, I., Powlowski, J., Ishmael, N., Darmond, C., Marqueteau, S., Moisan, M. C., Quenneville, G. & Tsang, A. (2012). A molecular phylogeny of thermophilic fungi. *Fungal Biology*, 116(4), 489-502.
- Mount, S. M., Gotea, V., Lin, C. F., Hernandez, K., & Makałowski, W. (2007). Spliceosomal small nuclear RNA genes in 11 insect genomes. *RNA*, 13(1), 5-14.
- Nielsen, H., & Johansen, S. D. (2009). Group I introns: moving in new directions. *RNA biology*, 6(4), 375-383.
- Nilsen, T. W. (2003). The spliceosome: the most complex macromolecular machine in the cell? *Bioessays*, 25(12), 1147-1149.
- Nilsen, T. W., & Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463(7280), 457-463.
- Oertle, T., Klinger, M., Stuermer, C. A., & Schwab, M. E. (2003). A reticular rhapsody: phylogenic evolution and nomenclature of the *RTN/Nogo* gene family. *The FASEB Journal*, 17(10), 1238-1247.
- Pan, Q., Shai, O., Lee, L. J., Frey, B. J., & Blencowe, B. J. (2008). Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature genetics*, 40(12), 1413.

- Pao, S. S., Paulsen, I. T., & Saier, M. H. (1998). Major facilitator superfamily. *Microbiology and Molecular Biology Reviews*, 62(1), 1-34.
- Papasaikas, P., & Valcárcel, J. (2016). The spliceosome: the ultimate RNA chaperone and sculptor. *Trends in biochemical sciences*, 41(1), 33-45.
- Parenteau, J., & Elela, S. A. (2019). Introns: Good Day Junk Is Bad Day Treasure. *Trends in Genetics*, doi: 10.1016/j.tig.2019.09.010.
- Parra, M. K., Tan, J. S., Mohandas, N., & Conboy, J. G. (2008). Intraspllicing coordinates alternative first exons with alternative splicing in the protein 4.1 R gene. *The EMBO journal*, 27(1), 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. *Molecular and cellular biology*, 32(11), 2044-2053.
- Perrineau, M. M., Price, D. C., Mohr, G., & Bhattacharya, D. (2015). Recent mobility of plastid encoded group II introns and twintrons in five strains of the unicellular red alga *Porphyridium*. *PeerJ*, 3, e1017.
- Pfreundt, U., & Hess, W. R. (2015). Sequential splicing of a group II twintron in the marine cyanobacterium *Trichodesmium*. *Scientific reports*, 5, 16829.
- Pombert, J. F., James, E. R., Janouškovec, J., & Keeling, P. J. (2012). Evidence for transitional stages in the evolution of euglenid group II introns and twintrons in the *Monomorpha aenigmatica* plastid genome. *PLoS One*, 7(12).
- Pontecorvo, G., Roper, J. A., Chemmons, L. M., MacDonald, K. D., & Bufton, A. W. J. (1953). The genetics of *Aspergillus nidulans*. *Advances in genetics* 5, 141-238.
- Radtke, M., Srdic, I., & Schroeder, R. (2017). Genome-wide identification of intrasplicing events in the human transcriptome and hints to their regulatory potential. *BioRxiv*, 159350.

- Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics*, 16, 276-277.
- Rino, J., & Carmo-Fonseca, M. (2009). The spliceosome: a self-organized macromolecular machine in the nucleus? *Trends in cell biology*, 19(8), 375-384.
- Rogozin, I. B., Carmel, L., Csuros, M., & Koonin, E. V. (2012). Origin and evolution of spliceosomal introns. *Biology direct*, 7(1), 11.
- Rogozin, I. B., Lyons-Weiler, J., & Koonin, E. V. (2000). Intron sliding in conserved gene families. *Trends in Genetics*, 16(10), 430-432.
- Roy, S. W., & Gilbert, W. (2006). The evolution of spliceosomal introns: patterns, puzzles and progress. *Nature Reviews Genetics*, 7(3), 211-221.
- 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(6), 471-475.
- Sakharkar, M. K., Chow, V. T., & Kanguane, P. (2004). Distributions of exons and introns in the human genome. *In silico biology*, 4(4), 387-393.
- Sakharkar, M. K., Tan, T. W., & de Souza, S. J. (2001). Generation of a database containing discordant intron positions in eukaryotic genes (MIDB). *Bioinformatics*, 17(8), 671-675.
- Saudemont, B., Popa, A., Parmley, J. L., Rocher, V., Blugeon, C., Necsulea, A., Meyer, E. & Duret, L. (2017). The fitness cost of mis-splicing is the main determinant of alternative splicing patterns. *Genome biology*, 18(1), 208.
- 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. *Molecular and cellular biology*, 24(5), 1855-1869.

Schäfer, U. A., Reed, D. W., Hunter, D. G., Yao, K., Weninger, A. M., Tsang, E. W., Reaney, M. J. T., MacKenzie, S. L. & Covello, P. S. (1999). An example of intron junctional sliding in the gene families encoding squalene monooxygenase homologues in *Arabidopsis thaliana* and *Brassica napus*. *Plant molecular biology*, 39(4), 721-728.

Schmucker, D., Clemens, J. C., Shu, H., Worby, C. A., Xiao, J., Muda, M., Dixon, L. E. & Zipursky, S. L. (2000). *Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell*, 101(6), 671-684.

Schoch, C. L., Sung, G. H., López-Giráldez, F., Townsend, J. P., Miadlikowska, J., Hofstetter, V., ... & Gueidan, C. (2009). The *Ascomycota* tree of life: a phylum-wide phylogeny clarifies the origin and evolution of fundamental reproductive and ecological traits. *Systematic biology*, 58(2), 224-239.

Séraphin, B., & Rosbash, M. (1990). Exon mutations uncouple 5' splice site selection from U1 snRNA pairing. *Cell*, 63(3), 619-629.

Sibley, C. R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., Trabzuni, D., Ryten, M., Weale, M., E., Hardy, J., Modic, M., Curk, T., Wilson, S., W., Plagnol, V. & Ule, J. (2015). Recursive splicing in long vertebrate genes. *Nature*, 521(7552), 371-375.

Stoltzfus, A., Logsdon, J. M., Palmer, J. D., & Doolittle, W. F. (1997). Intron “sliding” and the diversity of intron positions. *Proceedings of the National Academy of Sciences*, 94(20), 10739-10744.

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*, 28(6), 1102-1104.

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(6), 555-561.



- Tang, Y., Nielsen, H., Masquida, B., Gardner, P. P., & Johansen, S. D. (2014). Molecular characterization of a new member of the lariat capping twin-ribozyme introns. *Mobile DNA*, 5(1), 25.
- Tarrío, R., Ayala, F. J., & Rodríguez-Trelles, F. (2008). Alternative splicing: a missing piece in the puzzle of intron gain. *Proceedings of the National Academy of Sciences*, 105(20), 7223-7228.
- Tocchini-Valentini, G. D., Fruscoloni, P., & Tocchini-Valentini, G. P. (2009). Processing of multiple-intron-containing pre-tRNA. *Proceedings of the National Academy of Sciences*, 106(48), 20246-20251.
- Turunen, J. J., Niemelä, E. H., Verma, B., & Frilander, M. J. (2013). The significant other: splicing by the minor spliceosome. *Wiley Interdisciplinary Reviews: RNA*, 4(1), 61-76.
- 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), 441, 10.1186/1471-2164-11-441
- Vogel, H. J. (1956). A convenient growth medium for *Neurospora crassa*. *Microb. Genet. Bull.*, 13, 42-43.
- Wang, B., Guo, G., Wang, C., Lin, Y., Wang, X., Zhao, M., Guo, Y., He, M., Zhang, Y. & Pan, L. (2010). Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic acids research*, 38(15), 5075-5087.
- Wang, Y., Liu, J., Huang, B. O., Xu, Y. M., Li, J., Huang, L. F., Lin, J., Zhnag, J., Min, Q. H., Yang, W. M. & Wang, X. Z. (2015). Mechanism of alternative splicing and its regulation. *Biomedical reports*, 3(2), 152-158.
- Watson, J. D., & Crick, F. H. (1953). Molecular structure of nucleic acids. *Nature*, 171(4356), 737-738.
- Will, C. L., & Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harbor perspectives in biology*, 3(7), a003707, <https://dx.doi.org/10.1101/cshperspect.a003707>

Woodward, L. A., Mabin, J. W., Gangras, P., & Singh, G. (2017). The exon junction complex: a lifelong guardian of mRNA fate. *Wiley Interdisciplinary Reviews: RNA*, 8(3), e1411.

Wortman, J. R., Gilsenan, J. M., Joardar, V., Deegan, J., Clutterbuck, J., Andersen, M. R., ... & Von Döhren, H. (2009). The 2008 update of the *Aspergillus nidulans* genome annotation: a community effort. *Fungal Genetics and Biology*, 46(1), S2-S13.

Xie, B. B., Li, D., Shi, W. L., Qin, Q. L., Wang, X. W., Rong, J. C., ... & Chen, X. L. (2015). Deep RNA sequencing reveals a high frequency of alternative splicing events in the fungus *Trichoderma longibrachiatum*. *BMC genomics*, 16(1), 54.

Yan, C., Wan, R., & Shi, Y. (2019). Molecular mechanisms of pre-mRNA splicing through structural biology of the spliceosome. *Cold Spring Harbor Perspectives in Biology*, 11(1), a032409, <https://doi.org/10.1101/cshperspect.a032409>

Yenerall, P., & Zhou, L. (2012). Identifying the mechanisms of intron gain: progress and trends. *Biology direct*, 7(1), 29.

Zhang, X. O., Fu, Y., Mou, H., Xue, W., & Weng, Z. (2018). The temporal landscape of recursive splicing during Pol II transcription elongation in human cells. *PLoS genetics*, 14(8).

Zumkeller, S., Gerke, P., & Knoop, V. (2020). A functional twintron, 'zombie' twintrons and a hypermobile group II intron invading itself in plant mitochondria. *Nucleic Acids Research*. doi: 10.1093/nar/gkz1194

## 10. Tudományos tevékenység

### Az értekezés alapját képező publikációk:

1. **Kavalecz N**, Ág N, Karaffa L, Scazzocchio C, Flipphi M, Fekete E (2019): A spliceosomal twin intron (stwintron) participates in both exon skipping and evolutionary exon loss. *Scientific Reports*, 9(1), 1-11.

**Impakt faktor: 3,998**

2. Flipphi M, Ág N, Karaffa L, **Kavalecz N**, Cerqueira G, Scazzocchio C, Fekete E (2017): Emergence and loss of spliceosomal twin introns. *Fungal Biology and Biotechnology*, 4:7.

3. Fekete E, Flipphi M, Ág N, **Kavalecz N**, Cerqueira G, Scazzocchio C, Karaffa L (2017): A mechanism for a single nucleotide intron shift. *Nucleic Acids Research*, 45: 9085-9092.

**Impakt faktor: 11,561**

## Egyéb publikációk:

1. Ág N, **Kavalecz N**, Péntes F, Karaffa L, Scazzocchio C, Flipphi M, Fekete E (2020): Complex intron generation in the yeast genus *Lipomyces*. *Scientific Reports*, 10: 6022.

**Impakt faktor: 3,998**

2. Fekete E, Orosz A, Kulcsár L, **Kavalecz N**, Flipphi M, Karaffa L (2016): Characterization of a second physiologically relevant lactose permease gene (*lacpB*) in *Aspergillus nidulans*. *Microbiology-SGM*, 162: 837-847.

**Impakt faktor: 2,151**

## Az értekezés témájában bemutatott poszterek:

1. **Kavalecz N**, Ág N, Karaffa L, Scazzocchio C, Flipphi M, Fekete E: Involvement of spliceosomal twin introns in instances of alternative splicing in *Aspergillus* (*15<sup>th</sup> European Conference of Fungal Genetics, Rome, Italy, 2020*)
2. Ág N, **Kavalecz N**, Péntes F, Karaffa L, Claudio S, Flipphi M, Fekete E: A spliceoszómális iker-intronok (stwintron) szerepe az alternatív splicing két típusában (*Biotechnológia a Debreceni Egyetemen Tudományos Szimpózium, Debrecen, 2019*)
3. **Kavalecz N**, Ág N, Karaffa L, Flipphi M, Fekete E: Role for spliceosomal twin introns in two modes of alternative splicing (*8<sup>th</sup> FEMS Congress of European Microbiologists, Glasgow, UK, 2019*)

4. **Kavalecz N**, Ág N, Karaffa L, Scazzocchio C, Flipphi M, Fekete E: A spliceosomal twin intron (stwintron) participates in both exon skipping and intron evolution (*A Magyar Mikrobiológiai Társaság XVIII. nemzetközi nagygyűlése, Budapest, 2019*)
5. **Kavalecz N**, Ág N, Karaffa L, Scazzocchio C, Flipphi M, Fekete E: A spliceosomal twin intron conserved across all classes of the *Peizizomycotina* subphylum (*30th Fungal Genetics Conference, Pacific Grove, California, USA, 2019*)
6. **Kavalecz N**, Flipphi M, Ág N, Karaffa L, Cerqueira G, Scazzocchio C, Fekete E: Genesis and loss of spliceosomal twin introns (“stwintrons”) (*5<sup>th</sup> Central European Forum for Microbiology [CEFORM], Keszthely, Hungary, 2017*)
7. Rimaszombati F, **Kavalecz N**, Ág N, Fekete E: Spliceoszómális iker-intronok (stwintronok) detektálása és vizsgálata *Aspergillus nidulans* fonalas gombában (*Szent-Györgyi Albert Konferencia, Budapest, 2017*)
8. Fekete E, Flipphi M, Ág N, **Kavalecz N**, Cerqueira G, Scazzocchio C, Karaffa L: Pivotal role for the alternatively spliced, spliceosomal twin intron in one-bp intron drift (*29<sup>th</sup> Fungal Genetics Conference, Pacific Grove, California, USA, 2017*)

### **Egyéb posztterek:**

1. Ág N, **Kavalecz N**, Péntes F, Karaffa L, Scazzocchio C, Flipphi M, Fekete E: Formation of a new intron within an extant intron: how can stwintronisation happen? (*15<sup>th</sup> European Conference of Fungal Genetics, Rome, Italy, 2020*)

2. **Kavalecz N**, Ág N, Karaffa L, Flipphi M, Fekete E: A new class of spliceosomal twin intron revealed in *Aspergillus nidulans*: A role in gene expression through alternative splicing? (*8<sup>th</sup> FEMS Congress of European Microbiologists, Glasgow, UK, 2019*)
3. **Kavalecz N**, Ág N, Karaffa L, Flipphi M, Fekete E: A lariat branch point motif-interrupted spliceosomal twin intron in *Aspergillus nidulans* (*2nd Splicing Conference, Costa da Caparica, Portugal, 2018*)
4. **Kavalecz N**, Ág N, Karaffa L, Flipphi M, Fekete E: A lariat branch point motif-interrupted spliceosomal twin intron in *Aspergillus nidulans* (*14<sup>th</sup> European Conference of Fungal Genetics, Haifa, Israel, 2018*)
5. Kulcsár L, Orosz A, **Kavalecz N**, Flipphi M, Fekete E, Karaffa L: Lactose uptake is mediated by differentially regulated permeases in *Aspergillus nidulans*, one of which is also involved in cellobiose catabolism (*13<sup>th</sup> European Conference of Fungal Genetics, Paris, France, 2016*)

#### **Az értekezés témájában elhangzott előadások:**

1. **Kavalecz N**, Ág N, Karaffa L, Flipphi M, Péntes F, Fekete E: Fungális stwintronok (spliceoszómális iker-intronok) szerepe alternatív splicinghoz köthető mechanizmusokban (*Biotechnológia a Debreceni Egyetemen Tudományos Szimpózium, Debrecen, 2019*)

## Egyéb előadások:

1. **Kavalecz N, Ág N, Karaffa L, Flipphi M, Fekete E:** Egy új típusú spliceoszómális iker-intron leírása *Aspergillus nidulans* gombában (*Magyar Mikrobiológiai Társaság Nagygyűlése, Eger, 2018*)
2. **Kavalecz N, Ág N, Karaffa L, Flipphi M, Fekete E:** A lariat branch point motif-interrupted spliceosomal twin intron in *Aspergillus nidulans* (*2<sup>nd</sup> International Caparica Conference in Splicing, Costa da Caparica, Portugal, 2018*)
3. **Kavalecz N:** The first instance of a lariat branch point motif-interrupted spliceosomal twin intron in *Aspergillus nidulans* (*7th Interdisciplinary Doctoral Conference, Pécs, 2018*)

## 11. Köszönetnyilvánítás

Mindenek előtt szeretnék köszönetet mondani témavezetőmnek, **Dr. Fekete Erzsébet**, egyetemi docensnek, aki mindvégig motivált és figyelemmel kísérte tudományos munkámat egyetemi éveim alatt. Bármilyen problémával bizalommal fordulhattam hozzá, szakmai hozzáértésével és ötleteivel pedig nagyban elősegítette a kísérletek sikerességét. Köszönöm a rengeteg támogatást és türelmet!

Köszönettel tartozom **Dr. Karaffa Levente** egyetemi docens, az MTA doktora, tanszékvezetőnek, akitől lehetőséget kaptam, hogy a Biomérnöki Tanszék laboratóriumaiban végezzem a disszertációm elkészültéhez szükséges kísérleteimet. Ehhez nem csak az anyagi és technikai feltételeket biztosította, de figyelemre méltó szakmai tudásával is támogatta a kutatási projekteket.

Hatalmas köszönet illeti az stwintron kutatócsoportunk további tagjait is. **Dr. Michel Flippi** holland kutatónak szeretném megköszönni a publikációk elkészültében nyújtott nélkülözhetetlen segítségét; **Dr. Ág Norbert** egyetemi adjunktusnak pedig elsősorban a releváns molekuláris biológiai módszerek elsajátításában, valamint az ábrák elkészítésében nyújtott segítségét.

Természetesen a Biomérnöki Tanszék minden korábbi, illetve jelenlegi dolgozójának hálás vagyok, aki valamilyen úton-módon segítette a munkámat. Külön kiemelném **Dr. Németh Zoltánt**, akihez bármilyen kritikus helyzetben fordulhattam tanácsért, illetve **Fekete Zoltán** tanszéki



mérnököt, aki segítségemre volt a felmerülő technikai problémák áthidalásában.

Szeretném megköszönni a rengeteg segítséget és támogatást PhD hallgató társaimnak, különösképp **Kolláth István Sándornak**, **Pénzes Fruzsínának**, **Márton Alexandrának** és **Bíró Viviennek**, akikre nem csak munkatársként, de barátként is számíhattam, számíthatok.

Köszönöm azon volt, illetve jelenlegi biomérnök BSc és MSc hallgatók munkáját, akik **szakdolgozóként** részt vettek az stwintron kutatócsoport projektjeiben és hozzájárultak a kísérletek sikerre viteléhez.

Végezetül szeretném megköszönni **családomnak** és **barátaimnak**, hogy hittek bennem és mindvégig támogattak, illetve **Lipták Noémi Csillának** és **Rubisco-nak**, hogy toleráltak és mellettem voltak még a leggyengébb pillanataimban is!

Köszönöm!

A kutatás a GINOP-2.3.2-15-2016-00008 számú projekt keretében, az Európai Unió támogatásával, az Európai Regionális Fejlesztési Alap társfinanszírozásával valósult meg, valamint az EFOP-3.6.1-16-2016-00022 számú projekt keretében az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával.

## 12. Fűggelék

### 12.1. A kísérletek során alkalmazott oligonukleotid primerek

3. táblázat: Splicing intermedierek azonosításához használt primerek

Gombatorzs	Lókuszt	Primer neve	Szekvencia (5'-)
<i>A. nidulans</i> R21 (ATCC 48756)	AN3270	P1	CTGTGGCTATACTGGCTGGT
<i>A. nidulans</i> R21 (ATCC 48756)	AN3270	P3	CAGTTCAGATCTGATCATGT
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	P1	GACGACAAGAGCAACCCCAT
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	P3	GTTCGGACCTTGTAGATTGTA
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	P2	GTCCTCTGGTAAGTTGTCT
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	P4	CCAACCGATATACTCAGTCG
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	P1	GACGACAAGAGCAACCCCAT
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	P5	CATCCAACAGGTTTTGAGCT
<i>A. nidulans</i> R21 (ATCC 48756)	AN7524	P1_F	ATTGCGAGGCGAGTGGATTG
<i>A. nidulans</i> R21 (ATCC 48756)	AN7524	P2_R	GCGGTAGTCGGCTTCTATTC
<i>A. nidulans</i> R21 (ATCC 48756)	AN7524	P3_R	TAGCGGCAGTCCTTGGTCAC
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_53020	P1_F	GTGGATTGAGATGAAGGATG
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_53020	P2_R	GCGCCAGGCGATATTTAGCTG
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_53020	P3_R	AGGGAGAAATTTAAATAGCAC
<i>A. nidulans</i> R21 (ATCC 48756)	AN5404	Anid_D5.6_P1	CTTGCACTTTCATCGCTTCT
<i>A. nidulans</i> R21 (ATCC 48756)	AN5404	Anid_D5.6_P2	CGACTCCTTGATGTCTGCAA
<i>A. nidulans</i> R21 (ATCC 48756)	AN5404	Anid_D5.6_P3	GTGGTGGTGAAGGAATCAC
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_198312	Anig_D5.6_P1	CTTTGCGAGAGCGCCTAAGC
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_198312	Anig_D5.6_P2	GCAGTGGTAGATGAAGGGCTA
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_198312	Anig_D5.6_P3	TTGAAGAACCAACGGAGGAG
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_44705	Trire_D5.6_P1	CACAGCTCTCGACCTCTCTT

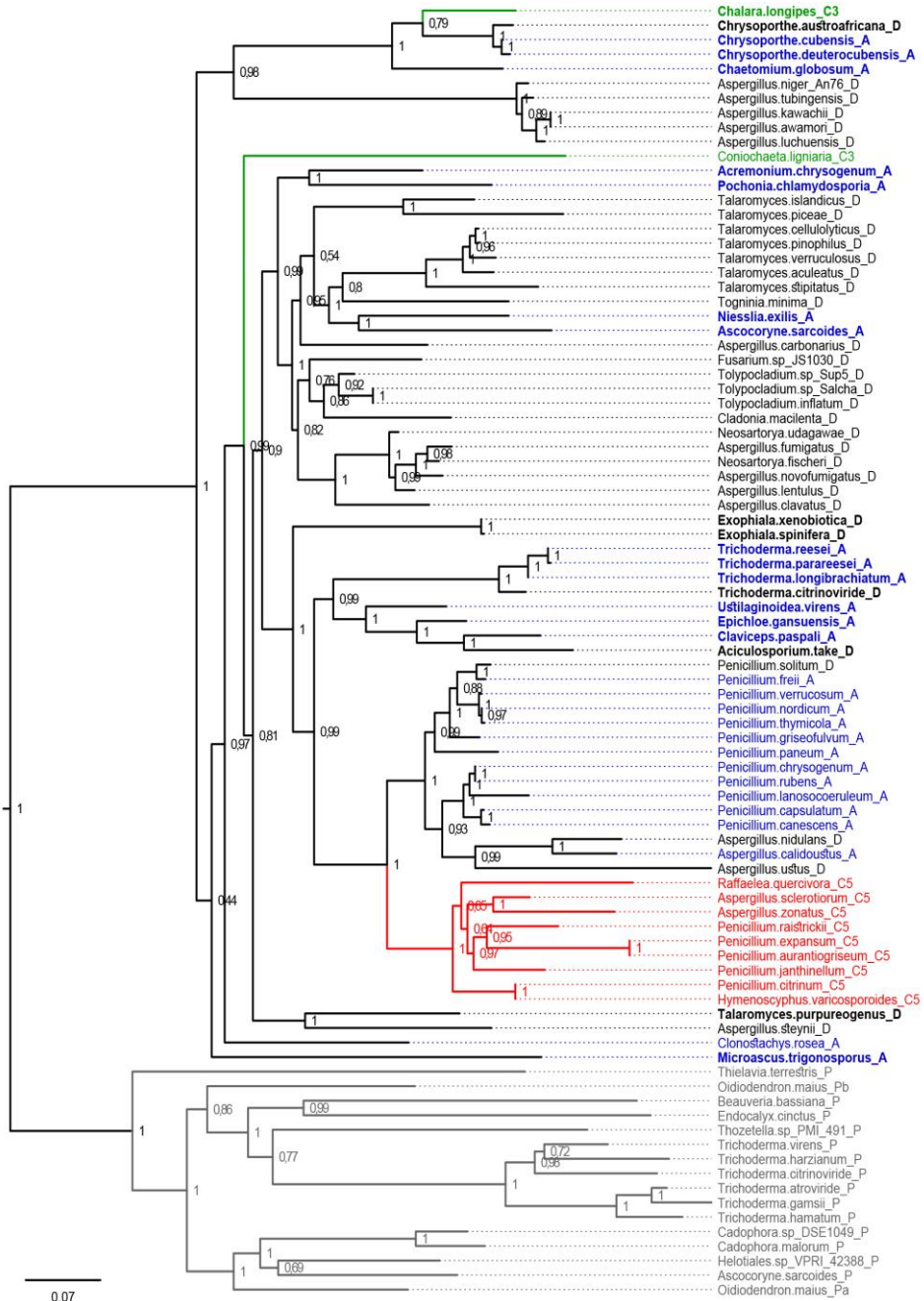
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_44705	Trire_D5.6_P2	CGGCTTCTAGAGCTCATCGT
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_44705	Trire_D5.6_P3	TCTCCCAGGAGAAGAGATGG
<i>B. cinerea</i> B05.10	BCIN_07g03170	Bcin_D5.6_P1	ACACAACCGCCTTTCCATCA
<i>B. cinerea</i> B05.10	BCIN_07g03170	Bcin_D5.6_P2	GGCAGCTGCAGATGCATCTAAA
<i>B. cinerea</i> B05.10	BCIN_07g03170	Bcin_D5.6_P3	CGTTCACGACACTTTGAGTGA
<i>H. solani</i> B-AC-16A	HELSOL_RTNA	Hsol_D5.6_P1	TTGCGGTTGTTAGAGTGTGC
<i>H. solani</i> B-AC-16A	HELSOL_RTNA	Hsol_D5.6_P2	CCGTTTCAAACCACTCGACCA
<i>H. solani</i> B-AC-16A	HELSOL_RTNA	Hsol_D5.6_P3	CAGGACCTGTGGGGAATTTA
<i>H. solani</i> B-AC-16A	HELSOL_RTNA	Hsol_D5.6_P4	GGGCATTTGCAAGATCAGAA
<i>N. crassa</i> OR74A	NCU06361	Ncra_D5.6_P1	GACTTGACAGCTTCATCCTC
<i>N. crassa</i> OR74A	NCU06361	Ncra_D5.6_P2	GTCTATTGGGATGCAAGTCAG
<i>N. crassa</i> OR74A	NCU06361	Ncra_D5.6_P3	TCACGGCATCATATGCAGAAG
<i>N. crassa</i> OR74A	NCU06361	Ncra_D5.6_P4	CTAACAGAGAAGATGGCAGCG
<i>M.cinnamomea</i> CBS 343.55	Malci1p7_008099	Malci_D5.6_P1	GTACGTCCGTCTCCTTTGTTT
<i>M.cinnamomea</i> CBS 343.55	Malci1p7_008099	Malci_D5.6_P2	GAATGTGTGAGACGACTATC
<i>M.cinnamomea</i> CBS 343.55	Malci1p7_008099	Malci_D5.6_P3	AGAGACAACAGACGCATCGA

#### 4. táblázat: cDNS szekvenálathoz használt primerek

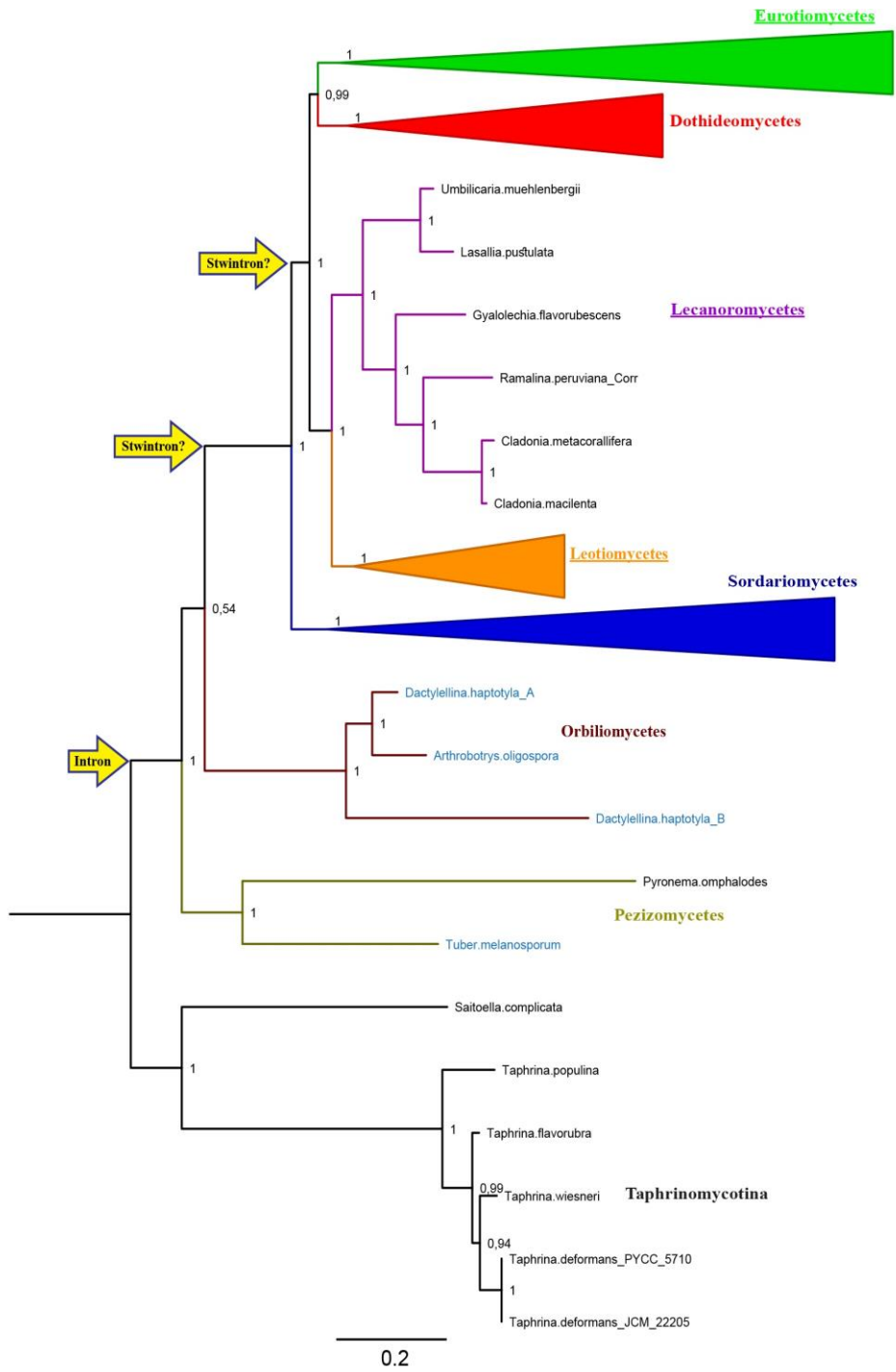
Gombatórzs	Lókusz	Primer neve	Szekvencia (5'-)
<i>A. nidulans</i> R21 (ATCC 48756)	AN3270	AN3270_seq_Forw1	CTGTGGCTATACTGGCTGGT
<i>A. nidulans</i> R21 (ATCC 48756)	AN3270	AN3270_seq_Rev1	ACTGTCCCGTAGATGATGGC
<i>A. nidulans</i> R21 (ATCC 48756)	AN3270	AN3270_seq_Forw2	TCGGTGGATTTCGTCTCGCAA
<i>A. nidulans</i> R21 (ATCC 48756)	AN3270	AN3270_seq_Rev2	TCGCCGTATGTGCTCTAAAC
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	Trire_seq_Forw1	CATCTTTGCTTCGTTGTCCAG
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	Trire_seq_Rev1	GCTTCTTGCCCTGGTCTTTC
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	Trire_seq_Forw2	CATCATTGGCGTCATCTTCG
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_43701	Trire_seq_Rev2	GCAATCTAAGGTAAGTCGTC
<i>A. nidulans</i> R21 (ATCC 48756)	AN7524	AN7524_F1	GCAATGAGCATCCTTGACG
<i>A. nidulans</i> R21 (ATCC 48756)	AN7524	P3_R	TAGCGGCAGTCCTTGGTCAC
<i>A. nidulans</i> R21 (ATCC 48756)	AN7524	AN7524_F2	GAATAGAAGCCGACTACCGC

<i>A. nidulans</i> R21 (ATCC 48756)	AN7524	AN7524_R	AGGTATCAATCTAGCTCAA
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_53020	53020_F1	GGTCATCTTTACCTCCTCTAG
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_53020	P3_R	AGGGAGAAATTAATAAGCAC
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_53020	53020_F2	CAGCTAAATATCGCCTGGCGC
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_53020	53020_R	CGACAACCACTGACTGCCAAC
<i>A. nidulans</i> R21 (ATCC 48756)	AN5404	AN5404_seq_F1	GCTTCCGCACAGACTCAAAC
<i>A. nidulans</i> R21 (ATCC 48756)	AN5404	AN5404_seq_R1	TCATGTCAAAGAAGCAGATT
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_19831 2	Anig_seq_F1	GGTCGTTCCCTTCGTCTTCC
<i>A. niger</i> ATCC 1015	ASPNIIDRAFT_19831 2	Anig_seq_R1	CCGCAATTCTCACAGTCCATC
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_44705	Trire_seq_F1	CTCCCACAGCCAGAAGAAAG
<i>T. reesei</i> QM9414 (ATCC 26921)	TRIREDRAFT_44705	Trire_seq_R1	GAAGGAAAGAACGTGTCGATA
<i>B. cinerea</i> B05.10	BCIN_07g03170	Bcin_seq_F1	TCGACTGTTCCCTTCTTCC
<i>B. cinerea</i> B05.10	BCIN_07g03170	Bcin_seq_R1	CGTGATTCTCGCTGACTCCTC
<i>H. solani</i> B-AC-16A	HELSOL_RTNA	Hsol_seq_F1	GAAACCCCTCTGTCCGCATTG
<i>H. solani</i> B-AC-16A	HELSOL_RTNA	Hsol_seq_R1	AGCCCAGCCTGCGTTGATACT
<i>N. crassa</i> OR74A	NCU06361	Ncra_seq_F1	GACTTGACAGCTTCATCCTC
<i>N. crassa</i> OR74A	NCU06361	Ncra_seq_R1	GGTATCGTGATCATGGAATCC
<i>M. cinnamomea</i> CBS 343.55	Malci1p7_008099	Malci_seq_F1	TCGAGTCTCCCCTCAGCAGC
<i>M. cinnamomea</i> CBS 343.55	Malci1p7_008099	Malci_seq_R1	TCATCCGATAGTCACAGGGCA

## 12.2. A vizsgált fehérjék filogenetikai fája

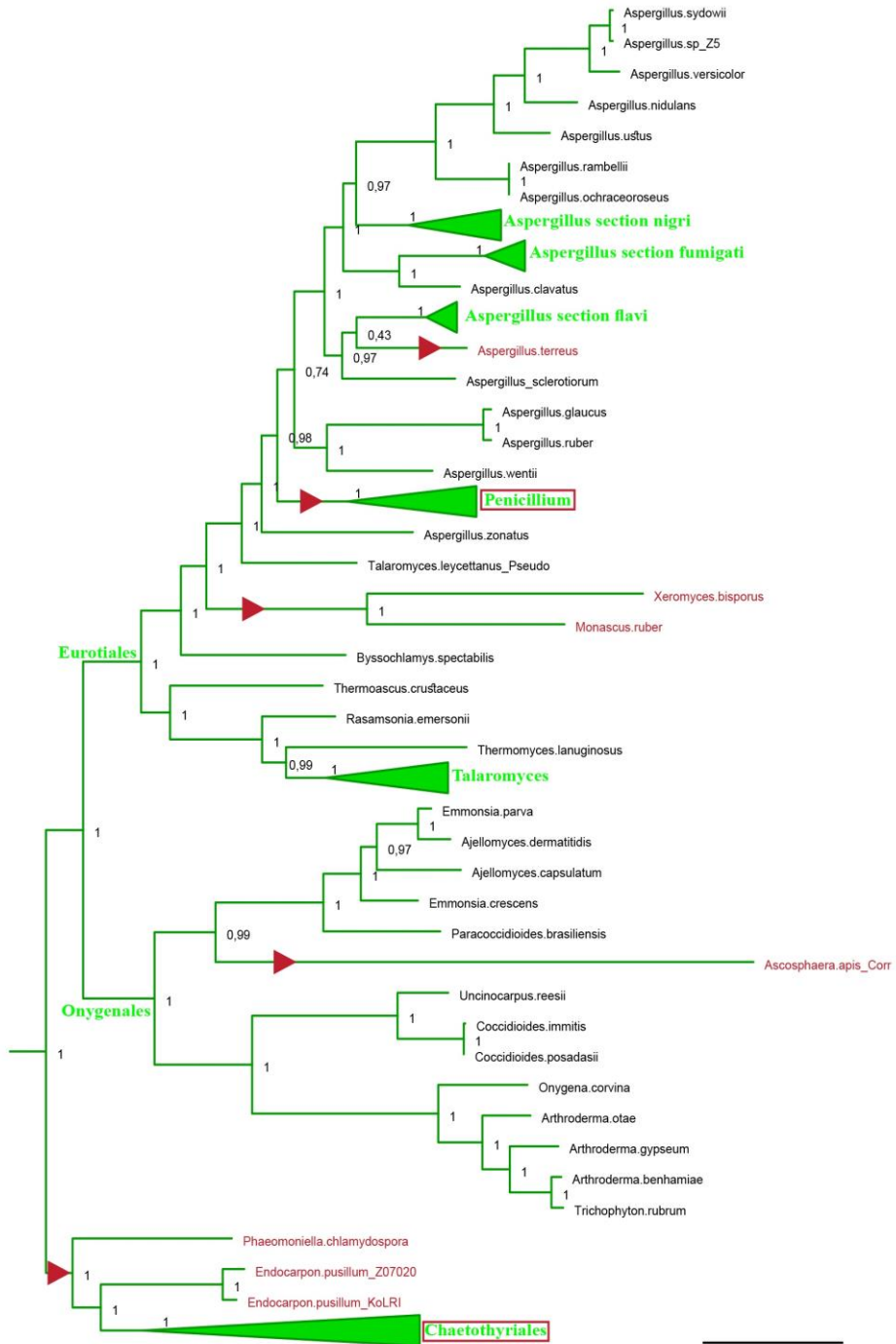


**18. ábra: *Pezizomycotina* DHA1 fehérjék filogenetikai kapcsolatát bemutató Maximum Likelihood fa.** Az *A. nidulans* DhaoS proteinjével nagyfokú hasonlóságot (legalább 45%-os) mutató *Pezizomycotina* DHA1 fehérjék (89 db) alapján készült törzsfát láthatjuk az ábrán. A Maximum Likelihood Ratio teszt értékei feltüntetésre kerültek az elágazásokban. Az a 36 ortológ gén, melyek tartalmazznak [D1,2] stwintront, a fajnév mögött egy „**D**” jelölést kaptak és fekete színnel szerepelnek az ábrán. Azon fehérjék (28 db), melyek kódoló génje egy alternatív [D1,2]/[A2,3] stwintront tartalmaz azonos pozícióban, „**A**” jelölést kaptak és kék színnel szerepelnek az ábrán. A proteinek (9 db), melyek kódoló génje egy kanonikus phase-2 intront hordoz az *A. nidulans* stwintron pozícióban, „**C5**” utótaggal lettek ellátva és piros színnel vannak feltüntetve, míg az a két ortológ, mely phase-0 intronnal rendelkezik, „**C3**” jelölést kapott és zöld színnel kerültek feltüntetésre. A külcsoportba tartozó 16 darab DHA1 proteint („**P**” utótaggal és szürke színnel jelölve) kódoló gének egyáltalán nem tartalmazznak intront; *Oidiodendren maius* genomjában a gén ráadásul két kópiában is megtalálható („**Pa**”-val és „**Pb**”-vel jelölve az ábrán). Az a 21 fehérje, melyek félkövér betűkkel kerültek feltüntetésre, olyan ortológ gének által kódoltak, melyek tartalmazznak még egy normál intront, 329 nt távolságra a [D1,2] stwintron pozíciótól downstream irányba. (saját szerkesztésű ábra Fekete és mtsai., 2017 alapján)

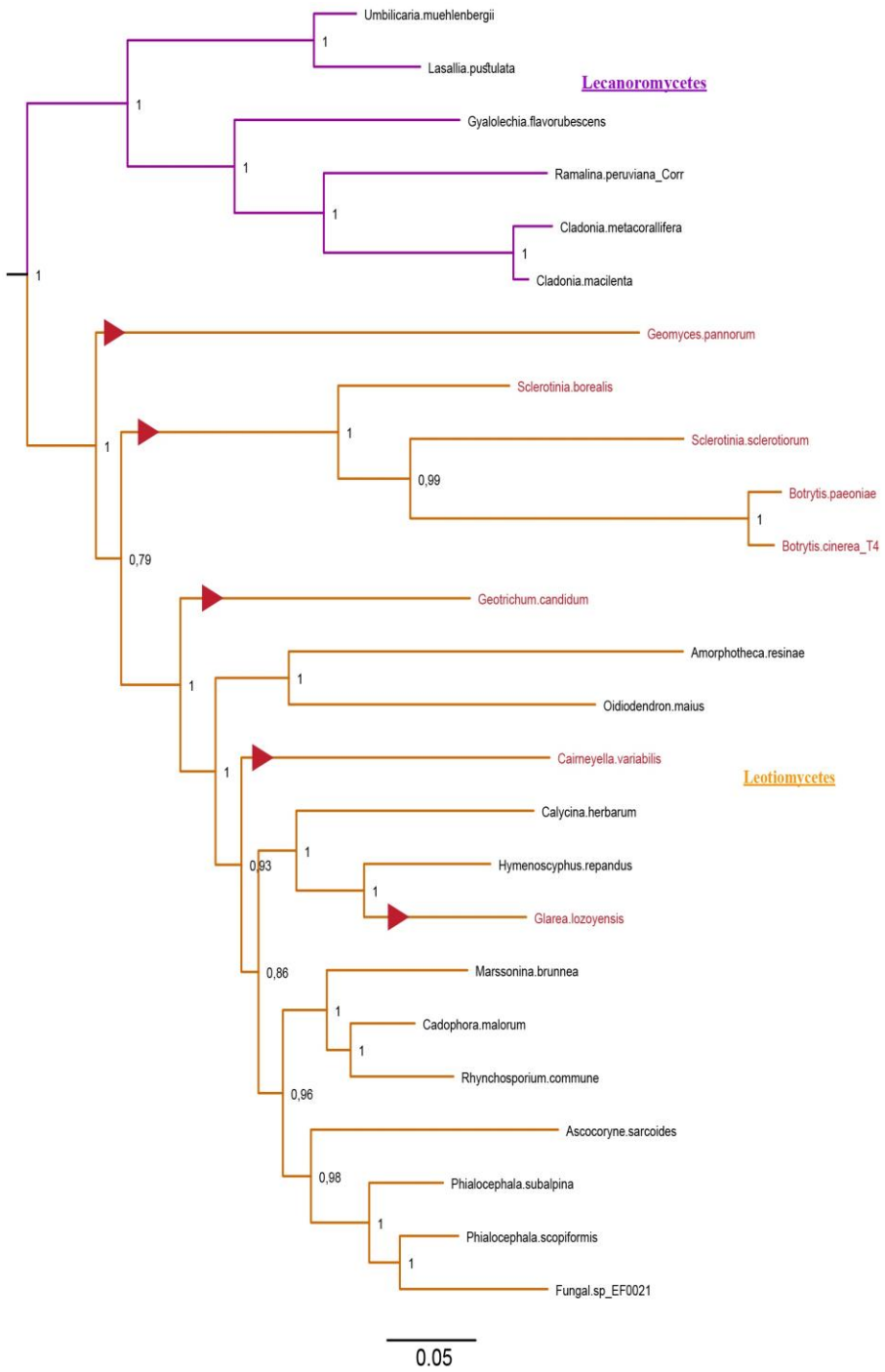


**19. ábra: Az Ascomycota LipS ortológok filogenetikai kapcsolatát bemutató Maximum Likelihood fa.** A *Taphrinomycotina* fehérjék képezik a törzsfá kulcsoportját. A Maximum Likelihood Ratio teszt alapján kapott értékék (0-1) feltüntetésre kerültek az elágazásokban. A fa ágainak színezésével különböztettük meg az egyes *Pezizomycotina* osztályokat, a következők szerint: olivazöld: *Pezizomyces*, barna: *Orbiliomyces*, kék: *Sordariomyces*, zöld: *Eurotiomyces* (beleértve *A. nidulans*-t is), piros: *Dothideomyces*, narancssárga: *Leotiomyces*, lila: *Lecanoromyces*. A felsoroltak közül, az utolsó öt osztály a *Leotiomyces* szuperosztályba tartozik. A könnyebb áttekinthetőség érdekében, négy osztály (*Sordariomyces*, *Eurotiomyces*, *Dothideomyces* és *Leotiomyces*) fajait osztályonként egybevontuk. Azokat az osztályokat, melyek gombafajainak genomjában megtalálható a vizsgált stwintron, aláhúzással jelöltük. Azon *Pezizomyces* és *Orbiliomyces* fajok nevét, melyek az stwintron pozícióban (az *A. nidulans* második intron pozíciója) egy normál intront hordoznak az ortológ génben, világoskék színnel tüntettük fel. Az stwintron létrejöttének feltételezett idejét sárga nyilak jelzik. A megvizsgált adatok alapján nem dönthető el egyértelműen, hogy az stwintron az előtt, vagy után jelent meg, hogy a *Sordariomyces* osztály divergált. (saját szerkesztésű ábra Flipphi és mtsai., 2017 alapján)

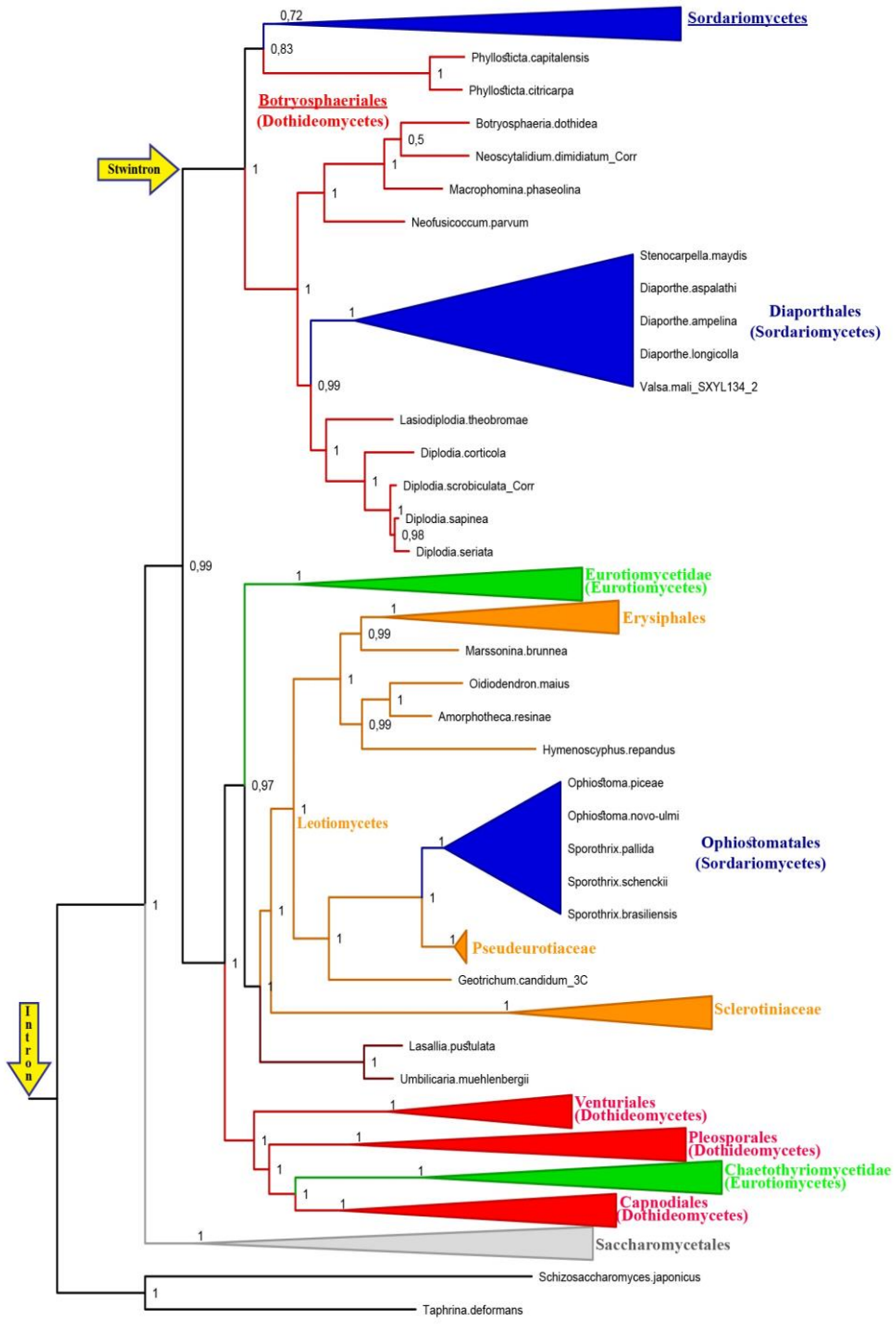




**20. ábra: A LipS protein Maximum Likelihood törzsfája: Stwintron veszteségi események *Eurotiomyces*-ek között.** Az elvégzett filogenetikai analízis alapján ábrázolt fa (19. ábra) „al”-fáját láthatjuk az ábrán, melyen az egyes *Eurotiomyces* rendek stwintron veszteségi eseményt részletesebben ábrázoltuk. A szinkódok megegyeznek a 19. ábrán használtakkal. Azon fajok nevei, melyek *lipS* génjükből elvesztették stwintronjukat, piros színnel lettek az ábrán feltüntetve. Az ábra könnyebb átláthatósága érdekében gombák egyes csoportjait összevontuk, melyek esetén azonos intron veszteségi/szerzési eseményeket tudunk kimutatni. Az összevont *Chaetothyriales* klád, illetve a *Penicillium* nemzetség egyik tagja sem hordoz intronikus szekvenciát az stwintron pozícióban (második *Aspergillus* intron pozíció a génen belül); az ábrán piros négyzettel jelölve, utalva az stwintron vesztesésre. Az stwintron elvesztésének további, független eseteit piros háromszögek jelzik. (saját szerkesztésű ábra Flippi és mtsai., 2017 alapján)

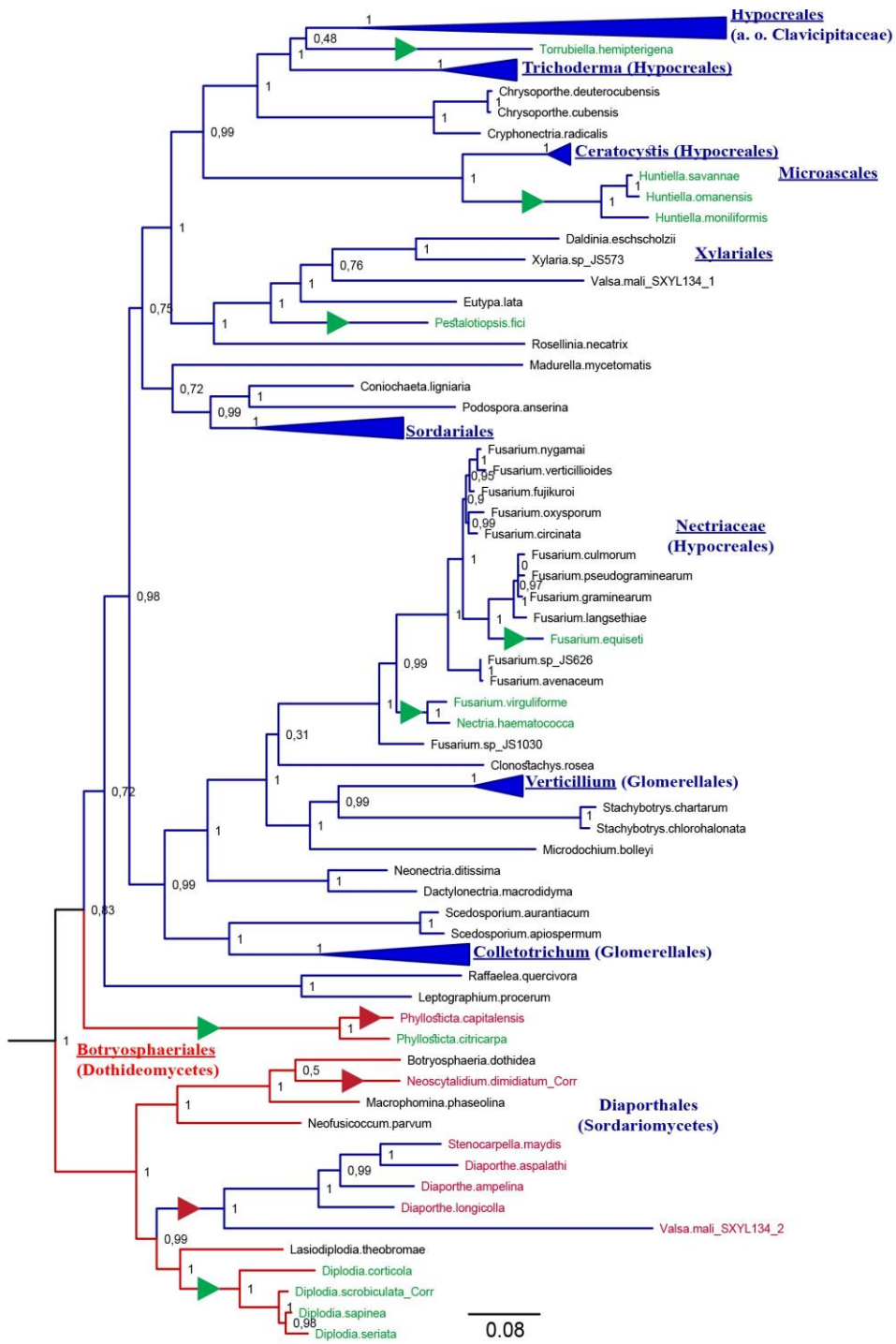


**21. ábra: A LipS protein Maximum Likelihood törzsfája: Stwintron veszteségi események *Leotiomycetes*-ek között.** Az elvégzett filogenetikai analízis alapján ábrázolt fa (19. ábra) „al”-fáját láthatjuk az ábrán, melyen az egyes *Leotiomycetes* rendek stwintron veszteségi eseményeit részletesebben ábrázoltuk. A színek megegyeznek a 19. ábrán használtakkal. Azon fajok nevei, melyek *lipS* génjükből elvesztették stwintronjukat, piros színnel lettek az ábrán feltüntetve. Az stwintron elvesztésének további, független eseteit piros háromszögek jelzik. (saját szerkesztésű ábra Flippi és mtsai., 2017 alapján)



0.2

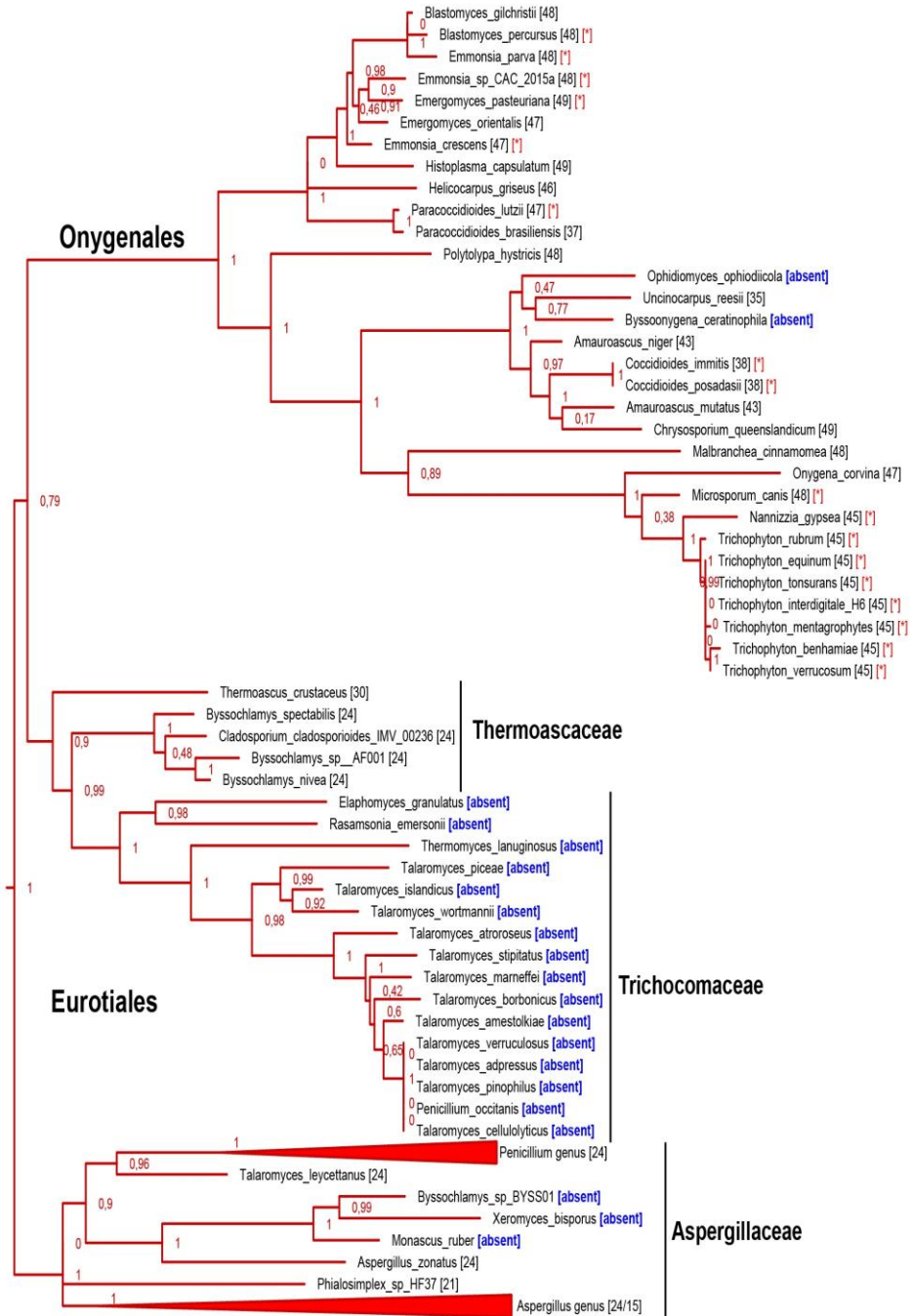
**22. ábra: Az *Ascomycota* BioDA protein Maximum Likelihood törzsfája: egy [D2,3] stwintron megjelenése.** A fán a *Taphrinomycotina* fehérjék képezik a kulcsoportot. Az Approximate Likelihood Ratio Teszt értékei (0-1) az elágazási pontokban kerültek feltüntetésre. A különböző színű ágak az egyes *Pezizomycotina* osztályokat reprezentálják, a színkódok megegyeznek a 19. ábrán használtakkal. A jobb áttekinthetőség érdekében egyes gombataxonok összevonásra kerültek az ábrán. Azok az osztályok, rendek, melyekben az stwintron megjelenését detektáltuk (*Sordariomycetes* és *Botryosphaeriales*), aláhúzással lettek jelölve. A belső intron felbukkanását, egy már meglévő másik intronon belül, sárga nyíl jelöli. Ez a „host” intron jelen van több, nem-*Dikarya* génben is. (saját szerkesztésű ábra Flippi és mtsai., 2017 alapján)



**23. ábra: A BioDA protein Maximum Likelihood törzsfája: Példák intron vesztesésre.**

Ezen az ábrán bemutatott fán részletesebben láthatóak az stwintron vesztesési események, a *Sordariomycetes* és *Botryosphaeriales* taxonokon belül. A színek megegyeznek a 19. ábrán használtakkal. Azon fajok neveit, melyek elveszítették *bioDA* stwintronjuk belső intronját, a pozícióban egy normál intront hátrahagyva, az ábrán zöld színnel tüntettem fel. Az egyszerűség kedvéért egyes gombafajokat/csoportokat összevontunk, melyek azonos módon viselkedtek (stw)intron vesztesés szempontjából. Az egymástól független, belső intron vesztesési eseményeket zöld háromszög jelöli. Az intronikus szakasz teljes hiánya a *bioDA* gén stwintron pozíciójából (az elkülönülő *Diaportales* kládba tartozó, illetve két *Botryosphaeriales* faj esetén – a fajnevek az ábrán piros színnel jelölve), egyetlen, vagy két, egymást követő intron vesztesési esemény következménye is lehet. A teljes stwintron elvesztésének eseményeit piros háromszögek jelzik az ábrán. (saját szerkesztésű ábra Flippi és mtsai., 2017 alapján)





0.09

**24. ábra: A reticulon-like fehérjét (RtnA) kódoló gén második exonjának evolúciója az Eurotiomycetidae alosztályon belül.** Az ábra a *Peizomycotina* fehérjék alapján készített Maximum Likelihood törzsfa egy részletét mutatja be, magában foglalva *Eurotiales* és *Onygenales* rendeket. Az Approximate Likelihood Ratio Teszt értékei a fa ágainak elágazási pontjaiban kerültek feltüntetésre. Az *Eurotiales* renden belül három, elérhető genomszekvenciával rendelkező család látható az ábrán; ahol az *Aspergillus* és *Penicillium* rendbe tartozó fajok összevonásra kerültek. Minden egyes gombafaj (esetleg nemzetség) neve után egy szám került zárójelben feltüntetésre. Ezek a számok az egyes fajokhoz tartozó exon-hosszakat adják meg, nukleotidban. *A. nidulans* esetén ez a kisméretű exon a [D5,6] stwintron és a második intron között helyezkedik el. A legtöbb *Peizomycotina* taxon esetén, exon II nukleotidjainak száma három többszöröse, avagy „bekebelezésre” került a [D5,6] stwintron 3' végének kiterjedése révén. Azon esetekben, ahol exon II hiányzik a génszekvenciából, egy [absent] jelölést használtunk a fajnév után. Minden *Penicillium* faj esetén, exon II mérete 24 nt, csakúgy, mint *Aspergillus*-okban, kivéve a *Nigri* szekció tagjait és *A. sclerotium*, *A. persii* és *A. steynii* fajokat (esetükben 15 nt). Az *Onygenales* rendbe tartozó fajok csaknem felénél (14 gombafaj) exon II hossza nem többszöröse háromnak. A fennmaradó 15 fajnál az exon mérete 48 vagy 45 nt, tehát hárommal osztható, azonban három gombafajt leszámítva egy korai stop kodont tartalmaz a szekvencia. Azon fajokat, melyek exon II-ben korai stop kodont hordoznak, az ábrán [\*] jelöléssel láttuk el. Két *Onygenales* faj, *B. ceratinophila* és *O. ophioidicola* esetén, az exon II hiányzik a génből és az stwintron (intron 1), valamint intron 2 (más fajok esetén intron 3) egy új, kanonikus intronnal van helyettesítve. Az exon hiányára ezekben az esetekben is [absent] jelöléssel utaltunk. (saját szerkesztésű ábra Kavalecz és mtsai., 2019 alapján)

## 12.3. Egyéb függelék

**5. táblázat: Kísérletesen igazolt [D5,6] stwintronok az *rtnA* transzkriptumban, öt *Peizomycotina* osztályon belül.** Az stwintron belső intronjának konzervált splicing elemei kékkel, a külső intron splicing elemei pirossal jelölve; a közöttük feltüntetetésre kerülő számok a köztük lévő távolságot adja meg, nukleotidban.

Fajnév	Accession number (stwintron splicing intermedier)	Accession number (érett mRNS)
<i>Aspergillus nidulans</i>	MK410459	MK410458
<b>guaagguaaga-45-gcugau-15-cagu-49-acuaau-12-cag</b>		
<i>Aspergillus niger</i>	MK410461	MK410460
<b>guaagguacga-47-gcugau-14-cagu-42-acugac-15-uag</b>		
<i>Botrytis cinerea</i>	MK410470	MK410469
<b>guaagguaagc-34-gcuaau-7-uagu-37-acuaau-9-uag</b>		
<i>Helminthosporium solani</i>	[1]:MK410465 [2]: MK410466	MK410464
[1]	<b>gugagguaggu-37-gcuaac-11-cagu-38-ucuaac-7-aag</b>	
[2]	<b>gugagguaggu-66-gcugac-15-cagu-172-gcuaaa-7-cag</b>	
<i>Malbranchea cinnamomea</i>	MK421639	MK421640
<b>guaagguaggu-61-acugac-32-cagu-58-ucucac-8-cag</b>		
<i>Neurospora crassa</i>	MK410472	MK410471
<b>guaagguacgu-61-gcuaac-8-cagu-69-acugac-11-uag</b>		
<i>Trichoderma reesei</i>	MK410468	MK410467
<b>guaagguaugc-111-gcugac-12-cagu-55-gcugac-9-uag</b>		
<i>Tuber melanosporum</i>	FP417649	FP444195; FP402467; FP393188; FP389271; FP402157; FP414860; FP429142; FP445968; FP396067; FP398543; FP425519; FP449366
<b>gugagguuagu-49-gcuaac-14-uagc-64-acugau-6-uag</b>		

**6. táblázat: *Onygenales* fajok EST és TSA adatainak elérése, melyek bizonyítják a „duális megszakító szekvencia” jelenlétét a reticulon-like fehérjét kódoló ortológ gének transzkriptumaiban.**

Fajnév	Accession Number (EST/TSA)
<i>Amauroascus mutatus</i>	GDQZ01000714
<i>Amauroascus niger</i>	GDRA01006022
<i>Bysoonygena ceratinophila</i> (canonical U2)	GDRB01009133
<i>Chrysosporium queenslandicum</i>	GDRC01001023
<i>Coccidioides immitis</i>	GH391836; GH382243; GH384133; GH376123; GH384898; GH337422
<i>Coccidioides posadasii</i>	GH414948; GH428981; GH446754; CO007951; GH435960; GH428556; BF253074; GH437168; GH428674; GH448138; BF251975; BF252780; CO001624; CO021040; BF251988; CO003313; GH407932; GH423405; GH448484; CO023342; GH449542
<i>Paracoccidioides brasiliensis</i>	CA583413
<i>Trichophyton rubrum</i>	DW696795; AJ884154; AJ883503; AJ881778; AJ882612; AJ881400; DW687691