



1949

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

Egyetemi doktori (PhD) értekezés

Szabó-Pénzes Fruzsina

Témavezető: prof. Dr. Fekete Erzsébet

DEBRECENI EGYETEM
Természettudományi és Informatikai Doktori Tanács
Juhász-Nagy Pál Doktori Iskola
Debrecen, 2023

Ezen értekezést a Debreceni Egyetem Természettudományi és Informatikai Doktori Tanács Juhász-Nagy Pál Doktori Iskola Fermentációs Biotechnológiai és Biomérnöki program (korábban Biológia Doktori program) 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, 2023

*.....
Szabó-Pénzes Fruzsina*

Tanúsítom, hogy Szabó-Pénzes Fruzsina doktorjelölt 2019- 2023 között a fent megnevezett Doktori Iskola Fermentációs Biotechnológiai és Biomérnöki Doktori programjának (korábban Biológia Doktori programjának) keretében irányítással 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 javaslom.

Debrecen, 2023

*.....
prof. Dr. Fekete Erzsébet*

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

Értekezés a doktori (Ph.D.) fokozat megszerzése érdekében
a Biológia tudományágban

Írta: Szabó-Pénzes Fruzsina okleveles biomérnök

Készült a Debreceni Egyetem Juhász-Nagy Pál doktori iskolája Fermentációs
Biotechnológiai és Biomérnöki Doktori programja
(korábban Biológia Doktori programja) keretében

Témavezető: prof. Dr. Fekete Erzsébet

Az értekezés bírálói:

Dr.

Dr.

A bírálóbizottság:

elnök: Dr.

tagok: Dr.

Dr.

Dr.

Dr.

Az értekezés védésének időpontja: 20... ..

Tartalom

1. BEVEZETÉS.....	1
2. IRODALMI ÁTTEKINTÉS.....	4
2.1 INTRONOK	4
2.2 SPLICEOSZÓMÁLIS INTRONOK	7
2.3 ALTERNATÍV SPLICING	10
2.4 TWINTRONOK.....	13
2.4.1 Ribozim típusú twintronok.....	14
2.4.2 Spliceoszómális twintron-szerű struktúrák	18
2.5 STWINTRONOK	21
3. CÉLKITŰZÉSEK.....	25
4. ANYAGOK ÉS MÓDSZEREK.....	26
4.1 BIOINFORMATIKAI MÓDSZEREK.....	26
4.1.1 Intron modellek, potenciális stwintronok keresése	26
4.1.2 Ortológ gének vizsgálata, génbányászat	28
4.1.3 Filogenetikai analízis, és egyéb informatikai módszerek.....	30
4.1.3.1 A retikulum fehérje analízise	30
4.1.3.2 Sister stwintronok analízise	30
4.1.3.3 Egyedi (nem-sister) stwintronok analízise.....	32
4.2 FELHASZNÁLT GOMBA TÖRZSEK	32
4.3 NUKLEINSAV IZOLÁLÁS.....	33
4.4 RT-PCR ÉS A PCR TERMÉKEK KLÓNOZÁSA.....	34
5. EREDMÉNYEK ÉS MEGBESZÉLÉSÜK.....	37
5.1 RETUKULUM FEHÉRJE (<i>RTN</i>) ORTOLÓGOK VIZSGÁLATA	37
5.1.1 Stwintronizáció <i>Lipomyces</i> fajokban	37
5.1.2 Három új típusú CIS <i>Lipomyces</i> ekben	39
5.2 SISTER (STW)INTRONOK <i>HYPOXYLON</i> GOMBÁKBAN	48
5.2.1 Sister (stw)intronok <i>Hypoxylon sp. CO27-5</i> fajban	48

5.2.2	Sister (stw)intronok szimmetriai tulajdonságai.....	53
5.2.3	Hipotézis: sister intronok proliferációja.....	55
5.2.4	Sister (stw)intronok rokon fajokban.....	57
5.3	EGYEDI ÉS ISMÉTLŐDŐ STWINTRONOK A <i>HYPOXYLACEAE</i> CSALÁDBAN	61
5.3.1	Több, mint 100 [D1,2] stwintron felkutatása <i>Hypoxylon sp.</i> <i>CO27-5</i> -ben.....	61
5.3.2	[D1,2] stwintronok előfordulása/megoszlása <i>Hypoxylaceae</i> családban.....	65
5.3.3	117 [D1,2] stwintron szekvencia elemzése.....	67
5.3.3.1	Intron-exon határok, szekvencia logo.....	67
5.3.3.2	Szekvencia hosszúságok vizsgálata.....	68
5.3.3.3	A stwintronok AU tartalma.....	70
5.3.4	Belső szimmetria bemutatása 117 [D1,2] stwintronban.....	70
5.3.5	[D1,2] stwintronok másodlagos szerkezeteinek felépítése.....	73
6.	ÖSSZEFOGLALÁS.....	77
7.	SUMMARY.....	81
8.	IRODALOMJEGYZÉK.....	85
9.	TUDOMÁNYOS TEVÉKENYSÉG.....	98
10.	KÖSZÖNETNYILVÁNÍTÁS.....	102
11.	MELLÉKLETEK.....	103
11.1	GENOM ADATBÁZISOK ÉS SRA ADATOK.....	103
11.2	FELHASZNÁLT PRIMEREK.....	104
11.2.1	<i>Lipomyces</i> fajokhoz használt primerek.....	104
11.2.2	<i>Hypoxylon sp. CO27-5</i> sister (stw)intronokhoz használt primerek	105
11.2.3	<i>Hypoxylon sp. CO27-5</i> egyedi/ismétlődő stwintronjaihoz felhasznált primerek.....	109
11.3	SISTER (STW)INTRONOK FÜGGELÉKEI.....	112
11.3.1	Teljes sister stwintronok (23).....	112
11.3.2	Sheared sister stwintronok (13).....	115

11.3.3	Sister kanonikus (cropped) intronok (11).....	117
11.3.4	<i>Hypoxylon sp. CO27-5</i> sister (stw)intronok filogenetikai törzsfái és szimmetriai tulajdonságai	118

1. BEVEZETÉS

A molekuláris biológia és genetika megjelenése óta rengeteg információval bővült globális tudásunk. Watson és Crick az 1950-es évek elején munkásságukkal mérföldkövet értek el, amikor a DNS kettős szálú spirál felépítését publikálták (1). Néhány nappal e felfedezést követően rájöttek, hogy bázispárosodási szabályoknak megfelelően egy egyszálú DNS alkalmas lehet arra, hogy replikálható legyen, tehát lehetséges a DNS-ről történő másolat készítése (2). Áttörésüket később Nobel-díjjal jutalmazták. A gének mozaikosságát és az intronok jelenlétét Chow és Berget egymástól függetlenül végzett hasonló témájú kutatásaikkal adták tudunkra 1977-ben, amikor adenovírus kísérleteikkel kimutatták ennek a szekvencia darabnak a létezését (3) (4).

A doktori értekezésem alapjául szolgáló soron következő mérföldkő már az eukariótákban keresendő. Chow és Berget munkáját követően joggal kezdte el foglalkoztatni a kutatókat a tény, hogy a DNS-ről történő pre-mRNS másolat nem teljes egészében felel meg a DNS templát szekvenciájának, hanem kisebb darabok távolítódnak el belőlük az idő múlásával. Gilbert nem is sokkal később felfedezte, hogy a gének szegmentálva vannak, szekvenciájuk darabokra osztható. A darabok megkülönböztethetőségéért Gilbert elnevezte őket. A megszakító (eltávolított) szekvenciák az **intron** nevet kapták, az „intragenic region” kifejezésből eredően, míg a később kifejezésre kerülő mRNS darabokat **exon**oknak nevezte el (5).

A gének mozaikosságáról tudomást szerezve, az intronok és exonok még nagyobb figyelmet kaptak. Próbálták kideríteni, hogy a gének miképpen keletkeztek: Vajon az intronok hogyan kerültek a génekbe és

mikor? 1987-ben Walter Gilbert megírta a „The Exon Theory in Genes” című cikkét (6), amiben a „korai intron” teóriát állítja fel, miszerint az intronok már az első gének létrejöttekor léteztek. Ezzel ellentétben a másik oldal a „késői intron” teóriát látja igazabbnak. Eszerint az intronok az evolúció során később kerültek be az exonok közé (7) (8).

Az intronokról egyre több információhoz jutva Thomas Cech amerikai kémikus felfedezte, hogy az intronok rendelkeznek önmaguk kivágódását katalizáló ribozim aktivitással, amiért Nobel-díjat kapott 1989-ben (9) (10). Ezt követő munkásságával pedig jobban megismerhettük a Group I intronokat (11) (12). Az intronok különbözőségük miatt csoportokba oszthatók. Ez alapján a korábban felállított mindkét intron teória igaznak bizonyulhat. A prokarióták génjeiben ribozim intronok lehettek jelen, mutatja ezt az evolúciós megfigyelés, és a mitokondriális, valamint kloroplasztok DNS-ében jelenlévő intron tartalom. Azonban később spliceoszómális intronok alakultak ki (13).

Copertino és Hallick voltak, akik először írtak egy speciális intron struktúráról, a twintronokról. 1991-ben *Euglena gracilis* kloroplasztjának citokrom 559-b génjében olyan összetett intron struktúrát találtak, ami nem csak a gén kifejeződésben jelent egy újabb evolúciós lépést, de jelenlegi kutatásom alappilléret is adta. Az általuk leírt szekvencia két intron egymásba ágyazódásaként volt jelen az exonok között. A twintron szót az angol „twin intron” szavakból alkották meg (14).

A doktori disszertációm témaköre a twintronokhoz fűződik, ugyanis a Debreceni Egyetem Biomérnöki Tanszékének twintronokról szóló első publikációja 2013-ban volt (15). E kutatás során a twintronok között is egy speciálisabb szekvenciát véltek felfedezni, amit stwintronoknak neveztek

el. A stwintronok olyan „intron-az-intronban” megszakító szakaszok, amiket csak spliceoszómális intronok alkotnak, ebből adódóan az elnevezésük a *spliceosomal twintron* szavakból ered.

2018-ban sikeresen becsatlakoztam MSc-s hallgatóként a tanszéki stwintron kutatócsoporthoz. Mesterdiplomám megszerzését követően pedig 2019 februárjától immár PhD hallgatóként folytathattam a kutatásokat. Kezdetben a tanszéki kutatások célja minél több stwintron igazolása volt különböző fonalas gomba fajokban, amit több tanszéki cikk is bizonyít (16) (17) (18) (19). Ehhez hozzájárult a jelenlegi értekezés alapjául szolgáló valamennyi cikk is (20) (21) (22). Emellett megfigyeltünk több esetben alternatív kivágódást és írtunk a stwintronizáció jelenségéről is. Egyre több fajban és egyre érdekesebb funkcióval rendelkező stwintronokat találunk. Segítségükkel megfigyelhetjük a gombák evolúciós változását és felfedhetjük filogenetikai tulajdonságaikat is.

2. IRODALMI ÁTTEKINTÉS

2.1 Intronok

Az intronok eukarióta géneket megszakító, fehérjét nem kódoló szekvencia darabok. 45 évvel ezelőtti felfedezésük két kutatócsoport munkásságának köszönhető (3) (4). Régebben úgy vélték, hogy a gének kizárólag olyan szekvenciát tartalmaznak, amiből később fehérje szintetizálódik, ahogy az a prokarióták esetében is történik. Azonban Philip A. Sharp és Richard J. Roberts a DNS és mRNS hibridizációjával igazolták, hogy vannak szakaszok a szekvenciában, amik kilógnak, az érett mRNS már nem tartalmazza őket (23) (4). Felfedezésükkel a genetikai ismereteink rohamosan bővülni kezdtek, és a splicing (kivágódás) eseményére is fény derült.

Az intronok vizsgálata során láthatóvá vált, hogy különböző csoportra lehet őket osztani. A legfőbb dolog, ami közös az intronokban az az, hogy mindegyik eltávolításra kerül az mRNS-ből, viszont kivágódási eseményeik eltérőek lehetnek. Ez alapján a következő csoportokba lehet őket sorolni:

- Az első csoportba a ribozimként működő intronok tartoznak. Ebben az esetben nincs szükség külső enzimatiság segítségére az exonok közül történő kivágódáshoz, ugyanis önmaguk kivágódását képesek katalizálni, ribozimként (ribonukleinsav enzim) viselkednek. Ezen belül még két intron csoportot tudunk elkülöníteni, a **Group I** és **Group II** intronokat. Jelenlétük tRNS, rRNS és mRNS kódoló génekben, vírusokban, bakteriofágokban, valamint eukarióták kloroplasztisz és mitokondrium génjeiben fordul elő (24) (25).

- A **Group III** csoportba tartozó intronok nagyon hasonlítanak az előbbieken említett ribozimekre, azonban végül külön csoportba sorolták őket, mivel jelenlétük kizárólag az *Euglena gracilis* kloroplasztra korlátozódik (26). Szerkezetileg a Group II intronok degradálódott, rövidebb változatai. Azonban kivágódásukhoz fontos konszenzus elemekben különböznek. A Group III intronok 5'-végükön (5'-**NTNNG**) egy T és egy G nukleotid, a 3'-végükön (**ANNTNNNN**-3') egy T és egy A konzerválódott szigorúan. Az adenin pontosan 8 nukleotidnyira helyezkedik el a vágási helytől. Ezzel szemben a Group II intronok több konzervált nukleotiddal rendelkeznek ezeken a helyeken (27).
- Az önkivágódó intronokon kívül viszont léteznek olyan megszakító szekvenciák is, amik nem képesek saját splicing folyamatukat katalizálni, ezért külső enzimatikus segítségre van szükségük. Ez a segítség egy nagy, fehérjékből és ribonukleoproteinekből álló komplex révén valósul meg, amit spliceoszómának nevezünk. A spliceoszóma által kivágott intronok a **spliceoszómális intron** elnevezést kapták. Ezek a legelterjedtebb intronok az eukarióta magi genomok körében. Róluk részletesebben a 2.2 pontban fogok írni.
- Az általam bemutatott utolsó csoport a **tRNS intronok** csoportja, amely a legritkábban előforduló csoport.

A spliceoszómális intronok eredetéről ma már többet tudunk, bár a teljes igazságot még mindig rejtély övezi. Alapvetően két teóriát állítottak fel a kutatók:

- Az „**introns late**” teória szerint az intronok és a spliceoszóma jóval később jelent meg az eukariótákban. Akkumulálódtak, beékelődtek a fehérjekódoló génekbe (13).

- Az „**introns early**” teória szerint az intronok jelen voltak már a mikroorganizmusok legrégebbi közös ősében is, viszont az archeákból és baktériumokból fokozatosan eltűntek, amit cDNS rekombinációk okozhattak.

Valamint 2006 elején egy harmadik lehetőséget is leírtak, ami sokban az „introns early” teóriához hasonlít:

- Az „**introns first**” teória szerint az intronok és a spliceoszóma az RNS világ maradványa. Ezt a megfigyelést snoRNS-ek (small nucleolar RNAs) gyakran intronok által kódoltságának tudták be, ugyanis az RNS világban csak RNS-ek által történt a fehérje összeszerelődése és az snoRNS-ek is részt vettek a pre-riboszómák kialakulásában, ezeket pedig gyakran intronok kódolták (28).

Az elméletek egymást nem zárják ki, ugyanis mindhárom elméletre találtak igazolást. Egy dolog biztos, a spliceoszómális intronok a prokariótákban nem találhatóak meg, az eukarióták esetében pedig még mindig újak és újak jelenhetnek meg, vagy tűnhetnek el. Az intronok méretükből adódóan az exonokhoz viszonyítva egész biztos, hogy jelentőséggel bírnak a genomra nézve. Ez megcáfolja a kutatók korai gondolatait, miszerint az intronok használhatatlan, „szemét” régiók, és önzők módjára ékelődnek be a génekbe, ezzel növelve a sejt energiaigényét, ami az eltávolításukhoz szükséges (29).

Az intronok pozíciója a génekben segíti az evolúciós kutatásokat. A különböző mikroorganizmusok ortológ génjeikben az intron pozíciója elég konzervatív. Ezt a tényt nyilván befolyásolja a mikroorganizmusok evolúciós egymáshoz viszonyított távolsága, rokonsági viszonya. Ezt a tulajdonságot kihasználva a kutatók észrevették, hogy az intronok

eltűnhetnek és új intronok pedig beékelődhetnek, valamint „intron sliding” esemény is bekövetkezhet, azaz az intron elcsúszhat a szekvencián belül. Az elcsúszás viszonylag ritkábban fordul elő, inkább egy eltűnés, majd egy újabb intron megjelenése az eltűnt intron melletti pozícióban a jellemző (30) (31) (32).

2.2 Spliceoszómális intronok

A spliceoszómális intron a legelterjedtebb intron típus az élővilágban, ugyanis az eukarióta sejtek magi DNS-ében kizárólag ez az intron típus található meg.

Az eukarióta szervezetek génjei mozaikos elrendezésűek, exon (kódoló régió) és intron (nem-kódoló régió) szekvenciák felváltva tagolják (5). Ennek következményeképpen kialakult az eukariótákban egy kivágódási – „splicing” – mechanizmus. A génkifejeződés során az átíródott messenger RNS prekursorából (pre-mRNS) a nem kódoló (intron) részek eltávolításra kerülnek egy hatalmas RNS-fehérje komplex által, amit a szakirodalomban spliceoszómaként ismerünk. A spliceoszóma által kivágott intronokat spliceoszómális intronoknak nevezzük. Az intronok kivágódása után, a folyamatot az exonok összekapcsolódása zárja, amivel az érett messenger RNS jön létre (33).

Az intronok szekvenciájában három konzervatív régió található, melyeknek a splicing eseményben fontos szerepe van (donor, lariat és akceptor). A spliceoszóma képes felismerni ezeket és hozzájuk kötődni, ezáltal az intront kivágni az exonok közül. A donor az intron 5' végén egy 6 nukleotidból álló szakasz, az akceptor az intron 3' végénél található 3 nukleotidból álló szakasz, a lariat pedig az akceptor és a donor között elhelyezkedő, az akceptorhoz közelebb eső 6 nukleotidnyi szekvencia szakasz (1. ábra).



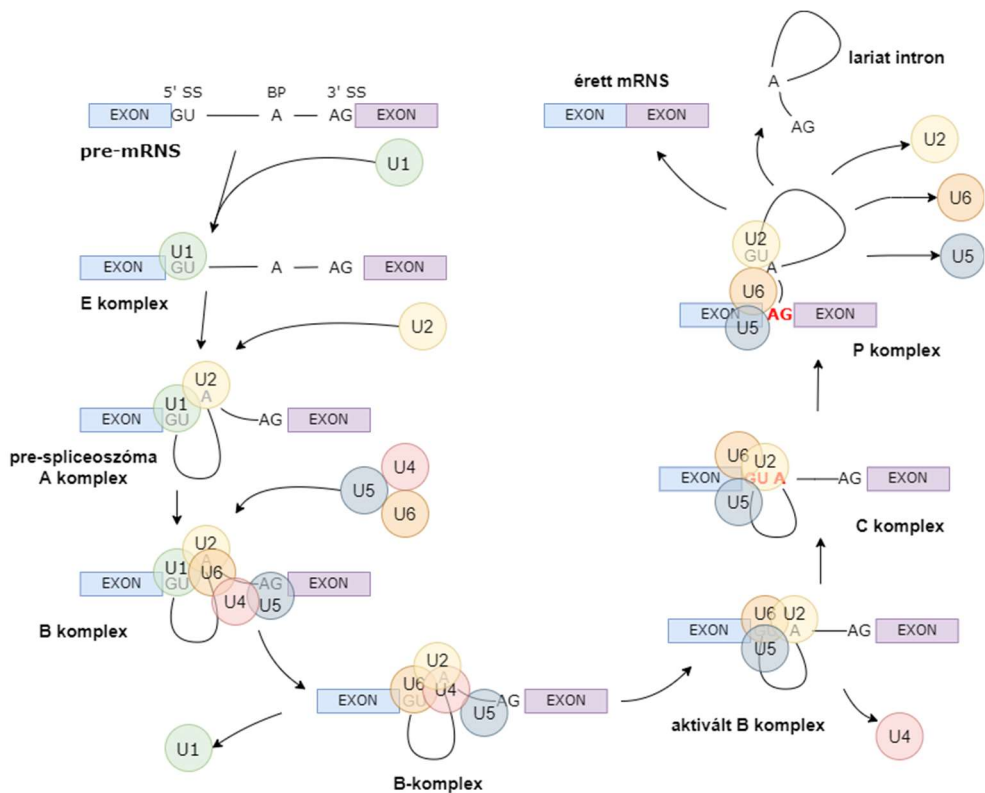
1. ábra U2 típusú spliceoszómális intron konzervatív régióinak általános sematikus ábrázolása élesztőkben. Donor – 5' hasítási hely, lariat – elágazási pont, akceptor – 3' hasítási hely (saját készítésű ábra (35) forrás alapján).

A spliceoszóma által katalizált folyamatoknak 2 típusát különítjük el. A leggyakrabban >99% gyakorisággal előforduló U2 típusú intronokat a „major” (fő) spliceoszóma, míg az U12 típusú kevesebb, mint 1%-ban előforduló intronokat a „minor” (kicsi) spliceoszóma katalizálja. A kétfajta spliceoszóma ugyanúgy több kicsi nukleáris ribonukleoproteinből épül fel (snRNP), azonban az általuk kivágott intronok hosszúságukban és konzervált elemeiben is eltérőek. Így a spliceoszómák az intronokhoz való kötődésükben is nagy különbséget mutatnak (34) (35).

A legtöbb szakirodalomban általánosságban az élesztőkben előforduló intronok konzervatív régióiról a következőket lehet elmondani: az 5' végen GUAUGU, az elágazási helyen (lariat/BP) UACUAAC, a 3' végen YAG szekvencia elemek szerepelnek, ahol az Y valamely pirimidinnek felel meg. A BP elem és 3'-YAG között 18-40 nt távolság van.

Erre az egyik legjobb megfigyelést Kupfer és munkatársai adták (36), mely során 5 gomba faj került összehasonlításra. A fungális intronok rövidek, ez az 5'SS (splice site) és a BP elem közti távolságból adódik. Az elágazási pont adeninje a 3'SS-től számítva a 13-36 nukleotidok között, 5'SS-től számítva 52-220 nukleotidok között helyezkedhet el, ami az intron teljes hosszától és persze a fajtól is függ. Az 5 faj vizsgálata során a donorra kapott 3 osztály: 5'GU—AG3'; 5'GC—AG3' és 5'AU—AC3', a donor további 4 nukleotidja A vagy U volt. Emellett a gombákra jellemző még egy polipirimidin rész a BP elem és a 3'SS régió között.

A spliceozóma a sejt egyik legnagyobb és legbonyolultabb gépezete, ami minden alkalommal összeszerelődik, ahányszor csak egy mRNA érés történik. Öt fő alkotó részből tevődik össze, amik maguk is komplex elemek. Ezek az U1, U2, U4/U6 és U5 snRNP-k. Mindegyikük kis nukleáris RNS-ből és a hozzájuk kötődő fehérjékből állnak. Ezen felül több, mint 100 fehérje alkotja még a spliceozóma szerkezetét. A kivágódást nem-snRNP típusú fehérjék, egyéb fehérje komplexek, splicing-ot segítő faktorok és 8 ATP-függő helikáz támogatja. Az snRNP-k transzkripciós faktorok jelenlétében RNS polimeráz II és RNS polimeráz III enzimek segítségével szintetizálódnak (37).



2. ábra Spliceozómális intron kivágódás sematikus ábrázolása (143) és (35) alapján.

Az intronok a pre-mRNS-ből két konzekutív transzészterifikációs lépésen keresztül fognak kivágódni. Először a lariat adenzin 2'-OH csoportja az 5' SS vágási helyet fogja támadni ezzel egy lasszó (lariat) alakot eredményezve. Másodszor, az 5' végen lévő exon 3'-OH csoportja fogja támadni a 3' SS helyet, amit a két exon összeligálódása követ, majd végül és az intron kivágódik (35) (38). A folyamat részleteit a 2. ábra segítségével mutatom be.

In vitro megfigyelések alapján ez a rengeteg splicing mechanizmusban résztvevő alkotó elem 15-60 perc alatt gyűlik össze és készül fel a folyamatra, míg maga a splicing *in vivo* kevesebb, mint 3 perc alatt végbemegy. Ez is bizonyítja, hogy a splicing tekintetében sokkal bonyolultabb szabályozás-jellemzi az eukariótákat, mint a prokariótákat (39). A génkifejeződést, beleértve az mRNS érését és az intron kivágódását is végig ellenőrzése alatt tartja a sejt és sok-sok szabályozási faktor játszik szerepet abban, hogy minden megfelelően működjön. Ám így is előfordulhat, hogy nem teljesen a vártnak megfelelő események zajlanak le és hiba csúszik a folyamatba.

2.3 Alternatív splicing

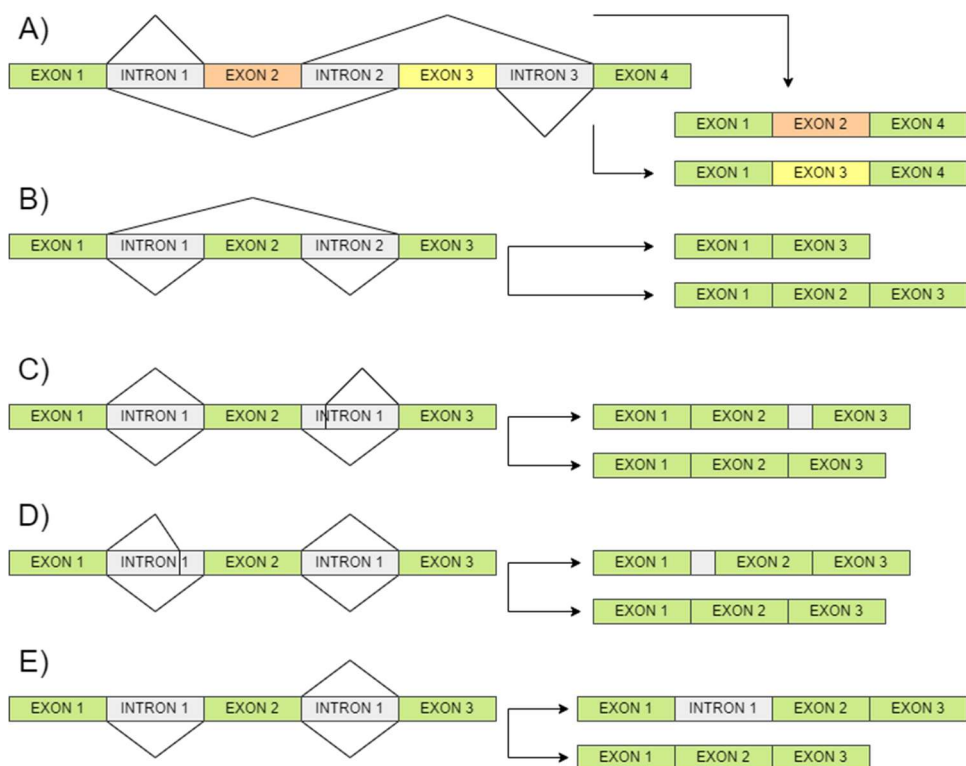
Az alternatív splicing az intron kivágódás során következik be, ha a folyamat nem teljesen a szabályoknak megfelelően történik. Azonban ez a „hiba” nem hibának fogható fel a sejt evolúciójának és genetikai sokszínűségének szempontjából. Alternatív splicing során az intron nem szabályosan vágódik ki az exonok közül, ezáltal a mRNS más nukleotid sorrenddel/összetétellel fog rendelkezni, aminek következtében előfordulhat, hogy egy teljesen más fehérje fog képződni róla.

Az alternatív splicing (AS) felelős az adott mRNS-ről képződő fehérjék izoformáiért, amik rendelkezhetnek hasonló tulajdonságokkal, ám

eltérhetnek funkciókban, felépítésben és egyéb jellemzőkben is. Az amőboid protozoák csak 40%, a növényekben 60%, míg az emberek génjeinek 95%-a érintett az alternatív splicingban (40). Ez a magyarázat arra, hogy az emberben lévő 20.000 génről 50.000-100.000 fehérjevátozat is képződhet (41). Egy kutatás során megfigyelték, hogy *Drosophila* DSCAM génjéről 38.016 fehérjevátozat tud képződni (42). Fonalas gombákban és élesztőkben azonban ez jóval ritkábban fordul elő. Ezt az intronok denzitásának köszönhetik, ugyanis sokkal kisebb intron készlettel rendelkeznek, mint fejlettebb eukarióta társaik (43). A legkevesebb AS esemény az élesztő típusú fajokban történik, míg a fonalas gombákban 4 féle alternatív splicing típus is megfigyelhető. Az emlős sejtekben exon skipping, a gombákban az intron retenció a leggyakoribb típus (41).

Alapvetően az alternatív splicing 5 fő típusát különböztetjük meg (3. ábra):

- A. Exon skipping: ebben az esetben az exon vagy benne marad az érett mRNS-ben, vagy pedig kivágódik a mellette lévő intronokkal együtt. Mindkét esetben működő fehérjét kaphatunk.
- B. Egymást kizáró exonok esetében, ha egy adott exon benne marad az mRNS-ben, akkor egy másik exon biztosan nem kerül az mRNS-be, egymás jelenlétét kizárják.
- C. Intron retenció során az intron az érett mRNS-ben marad. A tömlősgombák köreiből ez a legelterjedtebb forma.
- D. Alternatív 3'- és alternatív 5' hasítóhely használata akkor fordul elő, ha kettő vagy több hasítóhely kerül felismerésre az exon egyik végén, amely következtében az exonok fehérjekódoló kapacitása változik (43) (44).



3. ábra Az alternatív splicing 5 fő típusa: A) egymást kizáró exonok, B) exon skipping, C) alternatív 5' hasítási hely használata, D) alternatív 3' hasítási hely használata, E) intron retenció (saját készítésű ábra (43) alapján)

Az alternatív splicing eseményét számos dolog befolyásolhatja, pl. nemi genotípus, sejt differenciálódás, különböző jelátviteli útvonalak aktiválódása. Új funkcióval rendelkező domének adódhatnak hozzá a transzkriptumhoz, vagy akár a gén végétől számítottan hamarabb kialakuló stop kodon vethet véget a transzlációnak, ugyanis ez esetben egy labilis mRNS molekula képződik. Az alternatív splicing okozta változások meghatározzák a fehérjék növekedési faktor receptorainak és a sejtadhéziós molekulák ligand kötődéseinek lokalizációját, valamint a kinázok általi foszforilációs tulajdonságukat is. Az AS olyan sejt működésekbe is

beleszólhat, mint az apoptózis, neuronok közti kapcsolatok kialakítása, vagy a tumoros sejtek befolyásolása (42).

A spliceoszómális intronoknak az utóbbi időkben több funkciójára derítettek fényt. Szerepük van az alternatív splicing regulációjában azáltal, hogy cis-acting elemeket tartalmazhatnak a határoló részeiken. Kimutatták, hogy emlős gének esetében egy intron visszamaradása a transzkriptumban az mRNS produkció emelkedését mutatja (45). Az UTR (untranslated region) intronok szintén regulációs befolyással bírnak az NMD (nonsense-mediated decay) folyamatokban, mely során segítenek felismerni a potenciális pre-mRNS-eket (46). Ezenkívül az mRNS nukleusz és citoplazma közötti transzportját (47). Állatokban és növényekben talált megfigyelés alapján a géneken belüli első intron mindig a leghosszabb és konzerváltabb (48). Carvinus és társainak hipotézise alapján az intronok új gének kialakulásához is vezetnek. Egyfajta környezetben képesek a proto-gének potenciális működő génekké fejlődni (49). Az intronok továbbá védelmi funkciókat is ellátnak, főleg a fejlettebb eukarióták esetében, ahol az intronok (pl. emberben 25%) nagy számmal szerepelnek a genomban, ezáltal a bekövetkező pontmutációk vagy SNP-k jobb eséllyel az intron régiót érintik. Az intronok többféle, nem kódoló funkcionális RNS-gént tartalmaznak (50).

2.4 Twintronok

A gének megszakító szekvenciái – az intronok – összetett szerkezettel is rendelkezhetnek. Két vagy több egymásba ékelődött intron egy egységként fordul elő az exonok között. Ezeket a struktúrákat twintronnak keresztelték el. A twintront alkotó intronok egymás utáni lépésekben távolítódnak el teljesen az mRNS-ből (14). A twintronokat időrendbeli megjelenésük alapján mutatom be. Ez tükrözi az intronok lehetséges

időrendbeli megjelenését, ugyanis az első leírt twintronok ribozim intronokból tevődtek össze, majd azt követték a később megjelent spliceoszómális intronokból álló twintron struktúrák. Kizárólagosan spliceoszómális intronokból álló és a konszenzus régiókba ékelődő belső intronnal rendelkező twintron struktúrák a stwintron nevet kapták, amely felfedezését a tanszékünk tette meg 2013-ban (15).

A twintronokról szóló szakirodalmak az első felfedezésük óta sorra jelennek meg. Azonban egy alternatív splicinghoz, vagy egy elcsúszáshoz is társítanak twintron kifejezést tévesen. Hafez és Hausner 2015-ben foglalta össze és csoportosította az addig megjelent twintron típusokat kivágódásuk alapján (51). Ebből adódóan jelen értekezésemben hasonlóan csoportokra különítem el a twintronokat: önkivágódó/ribozim típusú twintronokra, melyeket értelemszerűen ribozim intronok alkotnak; és a spliceoszóma által kivágódó twintron szerű struktúrákra, melyeket U2 vagy U12 típusú spliceoszómális intronok alkothatnak, ám a beékelődés helye nem a konszenzus szakaszra tehető, nincs meg a stwintron jelleg.

A ribozim típusú twintronok esetében a kivágódás sorrendje nincs igazolva, nem mondhatjuk, hogy a belső intron splicingjának meg kell előznie a külsőét, így az ebben a struktúrában résztvevő belső intront leginkább „nested intronok”-nak nevezhetjük (51).

2.4.1 Ribozim típusú twintronok

A twintronok története 1991-ben Coppertino és Hallick felfedezésével kezdődött. Az első észlelt twintronok plasztisz DNS-ben szerepeltek és ribozim aktivitással rendelkező intronokból tevődtek össze, ebből adódóan ribozim típusú twintronokról beszélhetünk.

A struktúra az *Euglena gracilis* kloroplasztjának *psbF* génjében található. Megfigyelésük egy hosszabb, 1042 nt nagyságú Group II intronra

irányult, ami jobban szemügyre véve két intront tartalmazhatott, amik egymásba ékelődve voltak az exonok között. Az RNS cDNS-sé szintetizálása, megfelelő primer használat és northern hibridizációs eljárásnak köszönhetően kiderült, hogy a nagy intronból először egy 618 nukleotidnyi szakasz vágódik ki, majd a maradék 424 nt is önálló Group II intron formájában szintén eltávolítódik. A twintron kifejezés ez alapján két iker **Group II/Group II** típusú intronból tevődött össze. Megfigyelések alapján első lépésnek a belső intron kivágódása volt a sejt számára preferáltabb és csak azt követően vágódott ki a külső intron is (14).

Majdnem egy év elteltével megtalálták a twintronok első kevert **Group II/Group III** struktúráját is szintén *Euglena gracilis* kloroplasztjában. A riboszómális protein S3 génjén (RPS3) egy belső Group II intron ékelődött be egy Group III intronba. A twintronok létezésének alapja lehetnek az evolúciós fejlődés következtében beépült Group II intronok, melyek közös ősrre vezethetők vissza az eu-, ciano- és archeabakteriákkal, valamint a kloroplasztiszokkal (52). Egyre több intron-az-intronban megszakító szekvencia került leírásra az *E. gracilis* cpDNS-ében. Az *rpl23* riboszómális protein operonon egy Group III intronon belül egy Group II intron található, ami twintron-szerű struktúrát alkotva szakítja meg a *roaA* gént (53). Az *atpE* és *psbD* gének Group II/Group II twintronokat (54), a *psbK* gén pedig Group III twintront tartalmaznak (55). Hasonló kevert twintronokat találtak rokon fajokban is (56).

Az *E. gracilis* kloroplasztjának további kutatása kapcsán négy újabb típusú twintront is felfedeztek (**Group III/Group III**), amit csak Group III intronok alkotnak. Érdekeségükhöz tartozik, hogy az egyik egy cirkuláris elrendezésű intronban fordul elő. (57).

A következő érdekes felfedezés 1993-ban következett, amikor többszörösen összetett twintronokat találtak. Egy intronon belül több belső intron található, amik egymás utáni lépésekben távolítódnak el (58). Az *E. gracilis* kloroplaszt genomjának teljes szekvenálását követően kiderült, hogy számos intron-az-intronban struktúra alkotja a megszakító szekvenciákat. Ezek között Group II és Group III intronok egyaránt alkotnak azonos típusból vagy kevert típusokból twintronokat. Lehetnek külső vagy akár belső intronok is és többszörösen beépült összetettebb twintronok is előfordulnak (59) (60).

Twintronok nem csak az *E. gracilis* plasztisz DNS-ére korlátozódnak. Erről számol be a Maier és munkatársai által íródott 1995-ben megjelent cikk (61). Kriptomonád egysejtű alga, *Pyrenomonas salina* plasztiszában talált megszakító szekvencia nemcsak hogy twintron, de az első intron elem, amit felfedeztek ebben a genomban. *Rhodomonas salina* alga plasztisz genomjában degenerált twintronokra lettek figyelmesek, ahol a belső intron elvesztette kivágódásra képességét és a külső intronjával keveredve alkot megszakító szekvenciát (62). A hasadó- és sarjadzó élesztőkben 1994-ben írták le az első intron-az-intronban struktúrát (63). *Monomorpha aenigmatica* plasztiszában twintronokat találtak euglenoidok intronjainak evolúciós feltérképezésekor (64). *P. purpureum* plasztisz DNS-eiben is találtak twintronokat miközben mobilis genetikai elemeket vizsgáltak (65).

Deng és munkatársai *Hypomyces aurantius* mikoparazita mitokondriális *cox* génjében észleltek twintron-szerű struktúrákat, melyek közvetlen egymás után helyezkednek el (66). Hasonló struktúrák Zubaer publikálásában tandem intronokként kerültek leírásra *Endoconidiophora resinifera* mitokondriális DNS-ében (67). *Annulohyphoxylon stygium* fonalas gomba mitokondriális DNS-ében 9 twintron struktúrát írtak le, melyek közt

egy-, vagy több parazita intron ékelődött be a külső intronokba (68). Mikopatógén *Mycogone pernicioso* mitokondriális genomja vizsgálatával is evolúciós változásokat figyeltek meg mely bizonyítja az intronok gyors reagálását, miszerint számuk a genomban könnyen növekszik és csökkenhet is. Az evolúciós folyamat során számos rövid fragmentumot érintő inszerciós/deléciós esemény, valamint mutáció történt, ami a vizsgált klád közötti twintronok különbségéhez vezetett (69).

A twintron kutatások segítségével egyre jobban megérthetjük ezen genetikai elemek evolúcióját. Dabbagh és munkatársai 2017-ben végzett kutatásai euglenoid kloroplasztjában intron multiplikációról számolnak be. A cpDNS intronjainak és twintronjainak vizsgálata alapján érdekes észrevételeket tettek: az 51 intron struktúrából több mint fele magas szintű homológiát mutat egymással, valamint Group III intronoknak nyomát sem találták még twintront alkotó intronokként sem. Ez arra engedett feltételezni, hogy az euglenoidok között első körben a Group II intronok terjedtek el, majd ezekből alakulhattak ki a Group III típusú intronok (70).

A ribozimekhez tartozó **Group I** intronok is alkothatnak twintronokat. *Didymium iridis* nyálkapenészben találtak twin-ribozim típusú struktúrát (71). Több *Ascomycota* fajban **Group I/Group I** típusút *Cryphonectria parasitica* és *Ophyocordyceps tricenri* fajokban, valamint egy **Group II/Group I** típusút *Chaetomium thermophilum*-ban találtak (72). 2014-ben fungális homing endonukleáz kifejeződésében szerepet játszó twintron splicingot írtak le *Chaetomium thermophilum* gombában. Az mtDNS-ben található twintron külső Group I intronját megszakítja egy Group II intron. A belső intron kivágódásával kezdődhet el a homing endonukleáz szintézise (73). Az I-es típusú organelláris intronok gyakran hordoznak ORF (Open Reading Frame) elemet, amely endonukleázt kódol.

Zumkeller és munkatársai „zombi” twintronokról számol be. Kutatásuk során Group I/Group II és Group II/Group II twintronokat fedeztek fel korpafüvek és becősmohák mitokondriális genomjában, melyek közt előfordulnak *sensu stricto* (szigorúan vett/korlátozott) és *sensu lato* (szélesebb/tágabb) struktúrák is (74). (A *sensu stricto* kifejezést a szigorú szabályok alapján kivágódó struktúrákat értjük. Mivel a konszenzus elemekben található a belső intron, így a külső intron nem vágódhat ki. Ezzel szemben a *sensu lato* azokra az esetekre vonatkozik, ahol nincs ez a sorrend szabály.)

Metántermelő *Methanosarcina. mazei* Archea baktériumokban retroelemek vizsgálata közben találtak twintronokat (75). Ezek olyan transzpozabilis genetikai elemek, melyek reverz transzkripció bevonásával „ugrálnak” a genomon belül. *Trichodesmium erythraeum* tengeri cyanobaktérium Group II twintron struktúrákat hordoz Rnáz HI-vel kapcsolatos génjében (76). Myxomycetes valódi nyálkagombák rDNS S516 pozíciójában egy kisméretű spliceoszómális intron struktúrákat találtak, amely a His-Cys homing endonukleáz ORF-jét szakítja meg (77).

Az I. és II. csoportba tartozó intronok, komplex intronok (twintronok) az evolúció példáinak tekinthetők, ahol a komplex rendszerek nem adaptív mechanizmusok révén fejlődtek ki. Valószínűleg az RNS-világból származnak, az organelláris ribozim-alapú intronok ősei. Az I. csoportba tartozó intronok a tRNS (BHB) intronok potenciális ősei, a II. csoportba tartozó intronok pedig a nukleáris spliceoszómális intronok potenciális ősei (78).

2.4.2 Spliceoszómális twintron-szerű struktúrák

Az első spliceoszómális intronokhoz köthető twintron struktúrák említése a *Drosophila melanogaster pros* génjéhez köthető. Ez a gén az

embriógenézis során játszik kulcs szerepet és alternatív splicing miatt kétféle fehérje izoforma képződhet róla. Olyan twintron található a génben, mely során egy U2 intron egy U12 típusúba van ékelődve. A két különböző típust a hozzájuk tartozó különböző spliceoszómák képesek vágni. Kiderült, hogy a belső U2 intron kivágódása kizárja a külső U12 intron kivágódását. Ugyanis az U2 intron kivágódásakor a *pros-L* forma, míg az egész twintron, az U12 intron eltávolítását követően a *pros-S* izoforma fog képződni az mRNS-ről. Ezek aránya befolyásolja az embriógenézis folyamatát, a korai szakaszra a *pros-L*, a késői szakaszra a *pros-S* izoforma keletkezése a jellemzőbb (79).

Gerincesek génjeiben is írtak le U12 és U2 típusú intronokból összetevődött twintron struktúrákat, ahol a spliceoszómák versengése befolyásolja a képződő mRNS-t. A különböző spliceoszómák 5' és 3' SS helyei eltolódva helyezkednek el az intronokban, így az intronok átfedést mutatnak (80).

Twintronokra jellemző kivágódást fedeztek fel *Herpes simplex* 1-es típusú vírus LAT génjének egy óriási intronjában. A 2 kb hosszúságú intron kétféleképpen tud kivágódni az exonok közül, ami történhet egy lépésben, vagy a twintronokra jellemzően két lépésben is, melynek során először egy 0,5 kb nagyságú szakasz vágódik ki, majd a megmaradt 1,5 kb nagyságú rész távolítódik el (81). Amikor rövid intronok ékelődnek be egy külső intronba, azzal nem befolyásolva a kivágódás sorrendjét, „**nested**” intronokról beszélhetünk, ami egy twintron-szerű struktúraként fogható fel. Ekkor a külső intron esszenciális spliceoszóma általi felismerő szekvencia szakaszok a belső intron által érintetlenek maradnak, tehát *sensu lato* twintron struktúrákról van szó (51). Ekkor viszont egy hatalmas megszakító szekvencia található két exon között, aminek az eltávolítása bonyolult lehet

a sejt számára, így a kisebb intronok hamarabb kivágódnak. Tökéletes példa erre Suzuki Hitoshi és munkatársainak 2013-ban tett publikálása, mely során két „nested” intront találtak emberi disztrofin (DMD) génben. Az ebben található hatalmas struktúra 110.199 nukleotidból áll és két kisebb intron kivágódását vélték felfedezni a harmadik külső intron eltávolítódását megelőzően (82).

Nagy intronok kivágásának egy másik módját is leírták a *Drosophila ultrabithorax* (Ubx) génjében először 1998-ban (83). Később „**rekurzív splicing**”-ként elnevezve a kivágódást, a szerzők tovább folytatták a kutatást. Tulajdonképpen két intron egymást követően helyezkedik el, a 3' és 5' végeik közvetlenül összeérnek (AGGT) és egy „ratchet point”-ot alkotnak, ami egy 0 nukleotidos exonként fogható fel. Az intronok sorban távolítódnak el egymást követően amíg a teljes intron struktúra ki nem vágódik (84). Egyéb *Drosophila* génekben is igazolták a rekurzív splicing jelenségét (85), valamint *Homo sapiens* agyi génjeivel kapcsolatosan is írtak erről (86).

Hafez és Hausner 2015-ös publikációjában a nested és rekurzív splicingokon kívül még összefoglalták az „**intrasplicing**” mechanizmusát is (51), melyet Parra és munkatársai mutattak be még 2008-ban. Gerincesek citoszkeletális adaptor fehérjéinek 4.1R- és 4.1B paralógjaiban figyelték meg. Intrasplicing során a megszakító szekvencia többlépéses folyamat során úgy vágódik ki, hogy egy szomszédos szekvenciadarab exonként (intraexon) viselkedik, így az visszamarad a mRNS-ben, majd a legközelebbi splicing során intronként viselkedve eltávolítódik (87).

Az intrasplicing, nested splicing és rekurzív splicing jelenségei mellett spliceoszómális twintron struktúrák kivágódására egy újabb

lehetőség került feltáráásra, amely immár stwintronként került bemutatásra a tanszékünk első *sensu stricto* twintronokról szóló cikkjében (15).

A twintron és twintron-szerű struktúrákkal kapcsolatos további vizsgálatok indokoltak, mivel megértésükkel többet tudhatunk meg az organelláris, mind a nukleáris genom evolúciójáról. Az intronikus régiók hatással lehetnek a gének expressziójára, és biotechnológiában is alkalmazhatók.

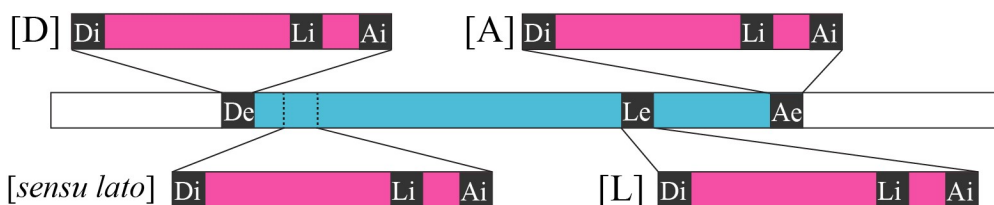
2.5 Stwintronok

A stwintronok az eddig említett twintron struktúrákhoz hasonlóan több intronból álló megszakító szekvenciák. Az őket alkotó intronok mindegyike spliceoszómális U2 intron, tehát a spliceoszóma aktivitás nélkülözhetetlen a kivágódásukhoz. A résztvevő intronok egymásba ékelődve helyezkednek el, tulajdonképpen az euglenoid kloroplasztban található Group II/Group III twintron spliceoszómális analógjai. Kivágódásuk így négy egymást követő transzészterifikációs folyamat, mely során a belső intron kivágódása megelőzi a külsőét (15).

2013-ban kerültek leírásra először spliceoszómális iker intron struktúrák a Debreceni Egyetem Biomérnöki Tanszékén. Genomok filogenetikai vizsgálata során akadt rá a kutatócsoport egy keret eltolódásra („frameshift”), majd jobban megvizsgálva, látható volt, hogy az intron donorjában egy másik intron található. Laboratóriumban végzet kísérletek alapján bizonyítást nyert, hogy a *Fusarium verticillioides* ciklikus imidin hidrolázt kódoló *PIH2* génjében a donor első és második nukleotidja közé ékelődve található egy belső intron egy *sensu stricto* stwintront eredményezve (15).

A stwintronok kivágódása kétlépcsős folyamat, a belső intron kivágódása nélkül a külső intron sem tud kivágódni. A spliceoszómális

intronok esetében három konzervatív szekvencia található meg az intronban. Az 5'-donor régió az intron első 6 nukleotidja; a lariat vagy elágazási régió az intron 3' végéhez közelebb található meg és 6 nukleotidból áll; valamint az akceptor régió, ami az intron 3 utolsó nukleotidja. A spliceoszóma képes felismerni ezeket a konzervatív szekvenciákat, majd kivágni a komplett intront az mRNS-ből. A belső intronok a három szekvencia valamelyikébe beékelődve találhatóak meg, ez alapján három stwintron típusról beszélhetünk: donor [D]; lariat [L]; vagy akceptor [A] helyre ékelődő stwintronok. Mivel a külső intron konszenzus szekvenciái egy másik (belső) intron által megszakítva vannak, a spliceoszóma nem képes felismerni és ezáltal kivágni a külső intront. Helyette a belső intront fogja első lépésben kivágni. Ezt követően viszont a külső intron konszenzusai már helyreállítódnak, a spliceoszóma számára felismerhetővé válnak és második lépésben kivágódhatnak az mRNS-ből.



4. ábra: Stwintron típusok. [D] donor régióba, [L] lariat régióba, [A] akceptor régióba vagy [sensu lato] az 5' véghez közel, de nem konszenzus szekvenciába ékelődött belső intron alapján (22). Az „i” betű jelöli a belső intront (internal), az „e” betű pedig a külső intront (external).

A stwintronok elnevezéséhez tartozó nomenklatúra egy betűt és 2 számot tartalmaz kapcsos zárójelben. A betű a három konszenzus szekvencia egyikére utal, míg a számok azt adják meg, hogy a belső intron melyik 2 nukleotid közé ékelődött be a konszenzus szekvencián belül (15).

A kutatócsoportunk általi első felfedezett [D1,2] típusú stwintron mellett egy [D2,3] típusú is publikálva lett *Trichoderma reesei bioDA* biotin szintézishez szükséges enzimet kódoló génjében (15), azonban később ezt

a kutatást számos fajra kivetítve újra elvégezték. Ortológ *bioDA* géneket vizsgáltak filogenetikai elemzéssel összevetve. Ugyanazon intron pozíciókat vizsgálva láthatóvá vált, hogy az intronok megjelenhetnek, és/vagy eltűnhetnek a genomokból. Számos családon belül a stwintron struktúra már régen jelen volt a génben, viszont más fajokban a gén nem hordozza a stwintront, helyette normális kanonikus intron van (18).

Két évvel később 2 új stwintron került publikálásra. *Helmithosporium solani* alternatív oxidáz génjében egy különleges [D1,2]/[A2,3] típusú stwintront találtak. Ennek érdekessége, hogy a struktúra belső intronja úgy helyezkedik el, hogy [D1,2] vagy [A2,3] típusú stwintronként is értelmezhető, ezt bizonyítja a kétfajta kivágódási módja is. A kivágódási útvonalak viszont egyenrangúként zajlanak és a kapott érett mRNS szekvenciája ugyanaz, tehát ugyanaz a fehérje képződik róla (88). A másik *Helmithosporium solani* aldóz mutarotáz enzimet kódoló génjében talált stwintron hasonló kivágódási mechanizmusokkal tudja eltávolítani intronjait, azonban csak a [D1,2] típusú kivágódás eredményeként kapott transzkriptumról fog működőképes fehérje képződni. Az [A2,3] kivágódásakor a belső intron eltávolítását követően egy STOP kodon képződik az átiratban, így arról nem képződik működő fehérje (88).

Aspergillus nidulans genomjában történő keresések után bukkantak rá egy drug H⁺ antiporter membrán fehérjét kódoló *DHAI* génben lévő potenciális stwintronra, ami hasonló [D1,2] típusú, mint a korábbi publikációkban résztvevő stwintronok. A kutatásba beletartozott ezen *DHAI* gén ortológjainak vizsgálata is több gomba fajban, ugyanis nagy mértékű konzervativitása révén ez alapján több stwintront lehetett találni más fajokban is. *Trichoderma reesei*-ben sikerült is detektálni egy [D1,2]/[A2,3] típusú stwintront, ami hasonlóképpen két úton képes

kivágódni, mint a *H. solaniban* (88) (17). *A. nidulans* és *Aspergillus niger* fonalas gombákban végzett keresések alapján új [D1,2] típusú stwintronokat találtak lipáz enzimet kódoló ortológ génekben (18). *A. nidulans* *Rtn* génjében „reticulon-like protein” - egy új típusú [D5,6]-os stwintront talált kutatócsoportunk, amely 2019-ben került publikálásra (19). További fajok ortológ génjeinek kutatásával kiderült, hogy majdnem minden elérhető *Pezizomycotina* altörzsbe tartozó gomba genomjában jelen vannak az stwintronok. Kísérletes bizonyítást nyertek *Botrytis cinerea*, *A. niger*, *T. reesei*, *Neurospora crassa* és *H. solani* genomjaiban. *A. niger* és *N. crassa* génjeiben viszont egy exon skipping jelenségét is megfigyeltek, melyet a stwintron, és egy nem sokkal mellette elhelyezkedő kanonikus intronnal együttes kivágódás jellemez. A két megszakító szekvencia közti rövid exon így szintén kivágódásra kerül a transzkriptumból. Ez a jelenség evolúciós intron vesztéshez vezetett (19).

3. CÉLKITŰZÉSEK

A kutatócsoportunk célja az első stwintron struktúra feltárását követően kezdetben minél több stwintron leírása volt. Az évek során számos fonalas gomba genomot vizsgáltunk át és sikeresen azonosítottunk stwintronokat. Így ez alkalommal a következő célokat tűztük ki a tervezett kísérletsorozathoz:

1. Fonalas gomba genomtól eltérő, élesztőgomba fajokban található stwintronok igazolása molekuláris biológiai módszerekkel.
2. A stwintron kereső modellünk átdolgozása és finomhangolása [D1,2] típusú struktúrákra.
3. Több ugyanazon [D1,2] típusú stwintronokat tartalmazó genom keresése, majd igazolása laboratóriumi körülmények között és *in silico* analízissel.
4. Az összegyűjtött stwintronok szekvencia elemzése, szimmetriai vizsgálata, valamint a másodlagos szerkezetének lehetséges felépítésének meghatározása.
5. Az adatgyűjtést követően az elemzett stwintronok lehetséges stwintronizációs folyamatainak kikövetkeztetése, feltárása.
6. Tényezők meghatározása, amelyek befolyásolhatják/kialakíthatják a stwintronok proliferációját.

4. ANYAGOK ÉS MÓDSZEREK

4.1 Bioinformatikai módszerek

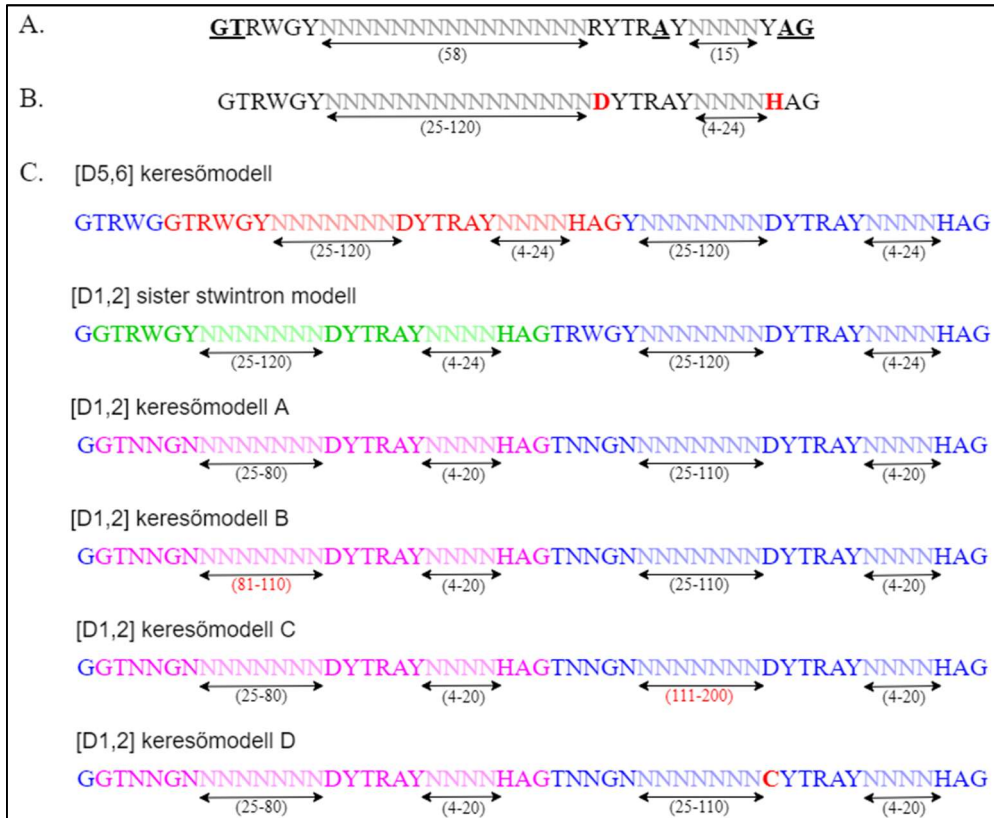
4.1.1 Intron modellek, potenciális stwintronok keresése

A twintronok keresésére a Kupfer által publikált spliceoszómális intronokra jellemző modellt (36) alkalmaztuk, a külső és belső intronokra egyaránt. Kupfer *A. nidulans*-ra alkotott modelljét 2017-es cikkünkben kezdtük el összetett intronokra használni, amikor a lariat első betűjelét, „R”-t „D”-re és az akceptor szekvencia első betűjelét „Y”-ről „H”-ra cseréltük, ezzel bővítve a potenciális intron jelöltek számát a keresés során (17).

Kupfer cikkét alapul véve az *A. nidulans* spliceoszómális intronjainak konzervatív részei: a donor első két nukleotidja: **GT**; a lariat elágazási pontja: **A**; és az akceptor utolsó két nukleotidja: **AG**. A konszenzus szekvenciák közötti távolságok meghatározásához számításba vettük az alábbi szabályokat: a genom átlagos intron hossza 73 nt, a minimum intronhossz 43 nt, a maximum intronhossz 160 nt. A donor és lariat közötti szekvencia szakasz nagyobb, mint a lariat és akceptor közötti, így a donor és lariat között 25-120 bázispár, a lariat és akceptor között 4-24 bázispár távolságot állítottunk be (17) (36).

A későbbiekben bemutatásra kerülő egyedileg előforduló (UO) [D1,2] stwintronok (5.3 bekezdés) keresésére az általunk korábban felállított modelltől indultunk ki, amely az 5. ábrán a „[D1,2] sister intron modell”. A későbbiekben még többet „engedtünk” a találatok számának azáltal, hogy a donorok potenciális jelöltjeit bővítettük. Ezekben a hibrid modellekben mind a külső- és belső intronokhoz tartozó donorok 3. 4. és 6. nukleotidját „N”-re cseréltük, ami mind a négy nukleotid lehetőséget engedi 5'-GTNNGN. Azonban az összes, 64 lehetséges intron variációból csak 24

félét alkalmaztunk, aminek az alapja egy korábbi, nem publikált eredmény. Azt a 24 lehetőséget használtuk, amelyek egy korábbi *Peizomycotin* ábrán végzett statisztikai analízis során >0,6% valódi intron donort képviselnek.



5. ábra A.) **Kupfer intron modellje *A. nidulans-ra*:** G: guanin, T: timin, R: adenin vagy guanin, W: adenin vagy timin, Y: citozin vagy timin (36). B.) **Általunk kibővített intron modell:** G: guanin, T: timin, R: adenin vagy guanin, W: adenin vagy timin, Y: citozin vagy timin, D: adenin, guanin vagy timin, H: adenin, citozin vagy timin (17). C.) **Stwintron modellek a kereséshez:** kék – külső intron; magenta – belső intron. A keresőmodell A, B, C és D közötti változtatások piros színnel vannak jelölve. C: citozin, N: bármely nukleotid (saját készítésű ábra).

Az eredmények könnyebb kezelhetősége érdekében kizártuk a kevésbé kanonikus 5'-GC típusú donorokat és az 5'-GAG akceptor szekvenciákat a találatok közül. Ezek alapján a „keresőmodell A” volt a kiinduló pont, majd a konszenzus szekvenciák közti távolságokat is módosítottuk annak

reményében, hogy több stwintron jelöltet találunk. A távolságok meghatározásához szintén a *Peizomycotinákban* történt analízist használtuk fel. A *Sequence Manipulation Suit DNA Pattern Find* (89) programjában alkalmaztuk a különböző kereső motívumunkat.

A keresőmodell A 24 lehetséges donor hexamerekkel (a piros szín a belső intront, míg a kék szín a külső intron elemeit reprezentálja):

```
G (GTAAGT | GTATGT | GTGAGT | GTACGT | GTAAGC | GTATGC | GTAGGT | GTA  
AGA | GTGCGT | GTACGC | GTGAGC | GTATGA | GTAAGG | GTTAGT | GTCAGT | G  
TGAGA | GTGTGT | GTTTGT | GTACGA | GTATGG | GTTCGT | GTGGGT | GTAGGC  
| GTACGG) . {25, 80} [AGT] [CT] T [AG] A [CT] . {4, 20} [ACT] AGT (AAG  
T | ATGT | GAGT | ACGT | AAGC | ATGC | AGGT | AAGA | GCGT | ACGC | GAGC | AT  
GA | AAGG | TAGT | CAGT | GAGA | GTGT | TTGT | ACGA | ATGG | TCGT | GGGT | A  
GGC | ACGG) . {25, 110} [AGT] [CT] T [AG] A [CT] . {4, 20} [ACT] AG
```

A *Hypoxylon sp. CO27-5* (MDCL01) genom szekvenciáját, ami 580 contigot tartalmaz, a National Center for Biotechnology Information (NCBI) (90) adatbázisából töltöttük le.

4.1.2 Ortológ gének vizsgálata, génbányászat

A retikulum fehérje őseinek felkutatása céljából, amely egy korábbi publikációnkból ered (19), homológ géneket kezdtünk keresni különböző gombafajokban, melyhez az adatbázisokat az NCBI szolgáltatta számunkra.

A nagy hasonlóságot mutató [D1,2] stwintronok gyűjtése a *Hypoxylon sp. CO27-5* genomjából történt (Gene Bank Master Accession number MDCL00000000) BLASTN használatával NCBI hivatalos online adatbázisában. *A. nidulans* (GenBank AACD01000130, CDS koordináták: start-stop: 208790-210069, publikálatlan munka) egy korábbi kutatásból származó jól konzervált mitokondriális carrier proteinjének ortológját kerestük számos gomba genomban, amikor figyelmesek lettünk két találatra *Hypoxylon sp. CO27-5* és *EC38* gombákon belül. A *CO27-5* [D1,2] stwintronja a HCOc017A nevet kapta. Később ezt a szekvenciát

felhasználva TBLASTN segítségével Genome Shotgun contig (WGS) adatbázisokat kutattunk át és rengeteg teljes- vagy parciálisan hasonló szekvenciát kaptunk, amik azt sejtetik, hogy ez a stwintron része egy új struktúráisan kapcsolódó megszakító szekvencia csoportnak. Kilenc gomba genomjában találtunk nagymértékű hasonlóságot mutató stwintronokra, amelyeket „sister” stwintronoknak neveztünk el. A felhasznált WGS adatokat a *11.1 melléklet* tartalmazza.

A *Xylariales* rendbe és *Hypoxylaceae* családba tartozó fajok közül ortológ és paralóg gének egyaránt kigyűjtésre kerültek, hogy az intron pozíciókat jobban meg tudjuk határozni. Az intron-exon struktúrákat manuálisan határoztuk meg és minden génre végeztünk analízist, így az összes szekvenciaeredetű információt összegyűjtöttük. A *Hypoxylon sp. CO27-5*, *Hypoxylon sp. EC38* és *Daldinia sp. EC12* legtöbb génmodelljét SRA adatokkal megerősítettük, ugyanis az NCBI Sequence Read Archive (SRA) a legtöbb stwintron köztes szekvenciáját (splicing intermedier, „splinter”) igazolja. A splinter szekvencia azért különleges, mert az RNS-ből történő stwintron kivágódás során először a belső intron vágódik ki, majd egy következő lépésben a külső intron, így az SRA ellenőrzésnél 60-60 nukleotidnyi szakaszokat vizsgáltunk a belső intron kivágódási helyétől 5'- és 3' irányban. Ekkor az olvasatban a külső intronnak még benne kell lennie, ez az állapot lényegesen ritkábban kimutatható, mint a teljesen kivágódott stwintron struktúra olvasata. Ehhez tartozó SRA adatok a *11.1 melléklet* alatt találhatóak.

Az eredményeinket nem használtuk automatikus annotáló programokban, ugyanis a stwintron struktúráját ezen programok nem ismerik fel, így a transzláció után funkcionálisan annotáltuk a legtöbb sister stwintront hordozó gént a Protein Families (Pfam) adatbázisból (91).

4.1.3 Filogenetikai analízis, és egyéb informatikai módszerek

4.1.3.1 A retikulum fehérje analízise

A Multiple sequence Alignment with Fast Fourier Transform (MAFFT: version 7) (92) programmal többszörös illesztéseket végeztünk L-INS-i algoritmust alkalmazva, BLOSUM80 pontozási mátrixot használva. Ez megfelelő a PF02453 domén számára. 902 valószínű fehérjét hasonlítottunk össze, ebből 136 informatív találat került kiválasztásra a Block Mapping and Gathering with Entropy (BMGE, version 1.12) (93) szoftver segítségével BLOSUM40 hasonlósági mátrixot és 4-es blokkméretet használva. A maximum likelihood törzsfákat PhyML (version 3.0) program automatikus szubsztitúciós modelljét (94) használva, alapértelmezett módban készítettük. A Smart Model Selection (SMS) szoftver választotta az LG+G+I+F szubsztitúciós modellt (95) az Akaike Information Criterion (AIC) becslését használva. A fa ágainak hosszát az Approximate likelihood-ratio test (96) biztosította. A törzsfa rajzolása FigTree (version 1.4.3: <http://tree.bio.ed.ac.uk/software/figtree>) program segítségével történt a PhyML Newick kimeneteléből. A további szerkesztéseket Adobe Illustrator (Adobe Inc., San Jose, CS, USA) programmal készítettük.

4.1.3.2 Sister stwintronok analízise

A törzsfa készítéséhez a *Xylariales* 37 fajának ortológ mitokondriális szállító fehérjéit (329 aminosav, *Hypoxyton sp. CO27-5*; cDNA: GenBank MW498246, HCOc017A) használtuk fel az általunk létrehozott génmodelleket alapul véve. A MAFFT (92) program G-INS-i algoritmusával és BLOSUM62 pontozó mátrixával többszörös illesztéseket végeztünk. 316 informatív adatot választottunk ki a BMGE (93) szoftver segítségével a BLOSUM62 hasonlósági mátrix és 4-es blokkméret

paraméterekkel. A maximum likelihood törzsfa készítése SMS-PhyML LG+G+I+F szubsztitúciós modell (95) alapján FigTree program segítségével történt. A fa ágait Shimodaira-Hasegawa-like approximate likelihood-ratio tests (SH-like aLRT) (96) adta. A törzsfát összekötöttük 6 olyan fajjal is, melyek nem tartoznak a *Hypoxylaceae* vagy a *Xylariaceae* csoportba. (A fa nem irányadó a *Hypoxylaceae* molekuláris taxonómiájához.)

A maximum likelihood törzsfa készítése 19 paralóg sister stwintron és 10 2-es típusú cropped sister intronból történt *Hypoxylon sp. CO27-5* fajban. MAFFT (92) program E-INS-i algoritmusával és 20PAM pontozó mátrixával többszörös illesztéseket végeztünk. Két vagy három intronikus szekvencia összehasonlítására a Clustal Omega (97) szekvenciaillesztését alkalmaztuk. A fa 37 fajból származó ortológ mitokondriális szállító fehérje alapján lett elkészítve FigTree programban. A sister stwintronok és a fehérjék filogenetika fájához a BMGE (93) szoftver segítségével válogattunk adatot, majd a PhyML programot használva az SMS (95) szoftverrel készítettük el azt úgy, hogy az SMS-PhyML 5000 ismétlést futtatott Felsenstein-féle bootstrap (98) alapján az ágak kikövetkeztetéséért. A szubsztitúciós modellt automatikusan választottuk: GTR+G+I. Adobe Illustrator segítségével végeztük el a további módosításokat. A fordított terminális ismétlődésekkel rendelkező egyszálú elágazó RNS intronok másodlagos szerkezetének a kalkulálásához az RNAfold (99), míg a teoretikus hibridizációs energia- és bázispárosodásból következő duplaszálú molekula kalkulációja ugyanazokból a (stw)intron RNS-ekből az RNAcifold (99) (100) programmal történt. A sister (stw)intronok pre-mRNS-einek 5'- és 3'-exon-intron találkozásainak közvetlen

környezetének szekvencia logóját a Skyline (101) program adta alapértelmezett beállításokkal.

4.1.3.3 Egyedi (nem-sister) stwintronok analízise

A stwintronok pre-mRNS-einek 5'- és 3'-exon-intron találkozásai közvetlen környezetének szekvencia logóját a Skyline (101) program adta alapértelmezett beállításokkal. A *Hypoxylon sp. CO27-5* gombából 117 stwintron lett azonosítva és a rájuk jellemző paramétereket – szekvenciahossz, fázis, minimális szabadenergia felhasználás a másodlagos szerkezet kialakításához, AU arány – összegyűjtöttük egy Excel táblázatba, ami nem lett mellékelve az értekezéshez (Microsoft Office Professional Plus 2016; Microsoft Cooperation, Redmond, WA, USA). MAFFT (92) program E-INS-i algoritmusával szekvenciaillesztéseket végeztünk az egyszálú stwintronok saját fordított komplementer szálával. Az RNAfold (100) (99) segítségével képeztük a másodlagos szerkezeteket az izolált bázispárok elkerülése nélküli beállítás mellett, és a legkisebb minimum szabad energia felhasználási adatokat is feljegyeztük (ΔG). Ezeket összehasonlítottuk az Mfold (UNAFold weboldal) (102) (103) által kapott alternatív eredményekkel. A sister stwintronokkal összehasonlítást végezve, a másodlagos szerkezet javítására a MAFFT eredményeken manuálisan korrigálni kellett egy-két sister stwintron esetében. Ezek másodlagos szerkezetét RNAalifold (ViennaRNA webhely) (104) adta.

4.2 Felhasznált gomba törzsek

1. táblázat A felhasznált törzsek, a törzsek gyűjtemény számai és az alkalmazott táptalajok

Kísérletben használt törzsek			
Faj	Gyűjteményszám	Táptalaj	Táptalaj összetevők
<i>Lipomyces suomiensis</i> (105)	CBS 7251	GPYA	4 % glükóz 0,5 % pepton 0,5 % élesztőkivonat

			(1,5 % agar)
<i>Lipomyces kononenkoae</i> (105)	CBS 2514	MEX	3 % maláta kivonat (1,5 % agar)
<i>Lipomyces lipofer</i> (105)	CBS 944 (NCAIM Y.00351)	MEX	3 % maláta kivonat (1,5% agar)
<i>Lipomyces starkeyi</i> (106)	CCY 33-1-1	YEPD	1 % élesztőkivonat 2 % pepton 2 % dextróz (1,5 % agar)
<i>Hypoxylon sp.</i> CO27-5 (107)	besorolatlan	PDA	4 % burgonya kivonat 2 % dextróz (1,5 % agar)

Az élesztőket a származási helyük által ajánlott megfelelő körülmények között tartottuk, neveltük az 1. táblázat alapján 26 °C-on.

A *Hypoxylon sp.* CO27-5 törzset Potato Dextrose Agar (PDA) (Neugen Culture Media, Lansing, MI, USA) táptalajon neveltük 7 napig, szobahőmérsékleten (25 °C).

4.3 Nukleinsav izolálás

Az élesztő törzseket 100 ml táptalajt tartalmazó 500 ml-es Erlenmeyer lombikokban növesztettük ki. 16 órán keresztül 200 rpm-en rázógépből (Infors HT Multitron, Infors AG, Lonay, Switzerland) tartottuk őket. A biomasszát centrifugáltuk 10.000 rpm-en 5 percig, majd sejtfeltáráshoz kerámiagyöngyöket tartalmazó centrifugacsőbe (MagNa Lyser Green Beads; Roche) helyeztük és 4000 rpm-en 30 mp-ig homogenizáltuk MagNa Lyser Instrument (Rocher) segítségével. A homogenizátumból a Macherey-Nagel NucleoSpin kitek segítségével történt a nukleinsav izolálás. Genomi DNS esetében a NucleoSpin Plant II-, az RNS esetében NucleoSpin RNA Plant kitek használtunk (Macherey-Nagel GmbH & Co. KG, Düren, Germany). A mennyiségek

meghatározásához NanoDrop 2000 készüléket használtunk (Thermo Scientific).

Hypoxylon sp. CO27-5 süllyesztett tenyészetének készítéséhez 100 ml potato dextrose broth (HiMedia Laboratories GmbH, Einhausen, Germany) táptalajt tartalmazó Erlenmeyer lombikokat rázógépben (Infors HT Multitron, Infors AG, Lonay, Switzerland) rázatunk 25 °C-on 24 órán át 200 rpm-en. A lombikokat a 7 napos Petri-csésze felületéről lekapart és steril 1/10⁴ Tween 80 (VWR International LLC, Debrecen, Hungary) oldatba vitt micéliummal oltottuk le. A biomassza micéliumát Miracloth-on (Millipore, Merck KGaA, Darmstadt, Germany) keresztül szűrtem és steril jéghideg disztillált vízzel mostam. A micéliumot folyékony nitrogénben hirtelen lefagyasztottam, majd kerámia dörzsmozsárban folyamatos nitrogén adagolás mellett feltörtem. A genomi DNS és RNS izolálása a Macherey-Nagel NucleoSpin Plant II- és NucleoSpin RNA Plant kitéket használtam (Macherey-Nagel GmbH & Co. KG, Düren, Germany).

4.4 RT-PCR és a PCR termékek klónozása

A cDNS szintetizálását az RNS templát és Oligo(dT) primer hozzáadásával a RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific, Thermo Fisher Scientific, Waltham, MA, USA) segítségével készítettem. Az RT-PCR (Reverse Transcription Polimerase Chain Reaction) elvégzéséhez a cDNS-t templátként, célzottan tervezett primereket (Integrated DNA Technologies, Leuven, Belgium, 11.2 *bekezdés*) és Dream Taq DNS polimerázt (Thermo Scientific, Thermo Fisher Scientific, Waltham, MA, USA) használtam. A reakció T100™ Thermal Cycler (Bio-Rad, Bio-Rad Hungary Ltd., Budapest, Hungary) készülékben folyt a 2. táblázat szerinti körülmények között.

Az RT-PCR-hez használt primerek felhasználás előtt gradiens PCR-ben ellenőrizve lettek, amihez genomi DNS templátot alkalmaztam. Ezáltal a primerek betapadási hőmérséklete is megválasztásra került, melyet a 2. táblázat szemléltet. A splinter szekvenciák kimutatásához a primereket úgy kellett megtervezni, hogy a teljes stwintron kivágódás köztes termékét kiszűrjük, így a reverz primert a külső intronon belülre tervezzük. Ez kétféle méretű terméket fog adni, melyek közül a kisebbik lesz az általunk keresett köztes szakasz.

2. táblázat Az alkalmazott RT-PCR reakciók paraméterei

	Lipomyces fajok		Hypoxylyon sp. CO27-5		
	95 °C	2 min	95 °C	2 min	
40X	95 °C	30 sec	95 °C	30 sec	35X
	56 °C	1 min	60 °C	1 min	
	72 °C	1 min	72 °C	0,5-1 min	
	72 °C	5 min	72 °C	5 min	

A PCR reakciók eredményének a detektálására agaróz-gélelektroforézist végeztem (SeaKem LE Agarose; Lonza Group Ltd., Basel, Switzerland). A futtatás végén elválasztott terméket gélextrakciós folyamattal tisztítottam (NucleoSpin Gel & PCR Clean-up, Macherey-Nagel GmbH & Co. KG, Düren, Germany), majd klónoztam pGEM-T Easy (pGEM-T Easy Vector System I, Promega Corporation, Madison, WI, USA) vektor segítségével. Minden RT-PCR reakciót kétszer végeztem el különböző inokulum tenyészetekből. A plazmid DNS a NucleoSpin Plasmid EasyPure kittel lett izolálva (Macherey-Nagel GmbH & Co. KG, Düren, Germany). Három egymástól független klónt küldtünk szekvenálásra minden termék esetében (Eurofins Genomics, Ebersberg,

Germany). A szekvenálás során univerzális primereket használnak, melyek a vektor meghatározott szakaszaival hibridizálnak.

Az általunk GenBank-ba feltöltött SRA adatok a következők:

- *Hypoxylon sp. CO27-5* sister (stw)intronok esetében: MW490712–MW490722, MW498245–MW498262 és MW530466–MW530509;
- *Hypoxylon sp. CO27-5* egyedi/ismétlődő stwintronok esetében: OL539745–OL539746, OL624519–OL624535, OL672706–OL672707, OM256448, OM541588–OM541592, OM719008–OM719015 és OM837808–OM837820;
- *Lipomyces starkeyi*: MN689085- MN689088;
Lipomyces kononenkoae: MN689089- MN689090;
Lipomyces lipofer: MN689083-MN689084;
Lipomyces suomiensis: MN689081-MN689082.

5. EREDMÉNYEK ÉS MEGBESZÉLÉSÜK

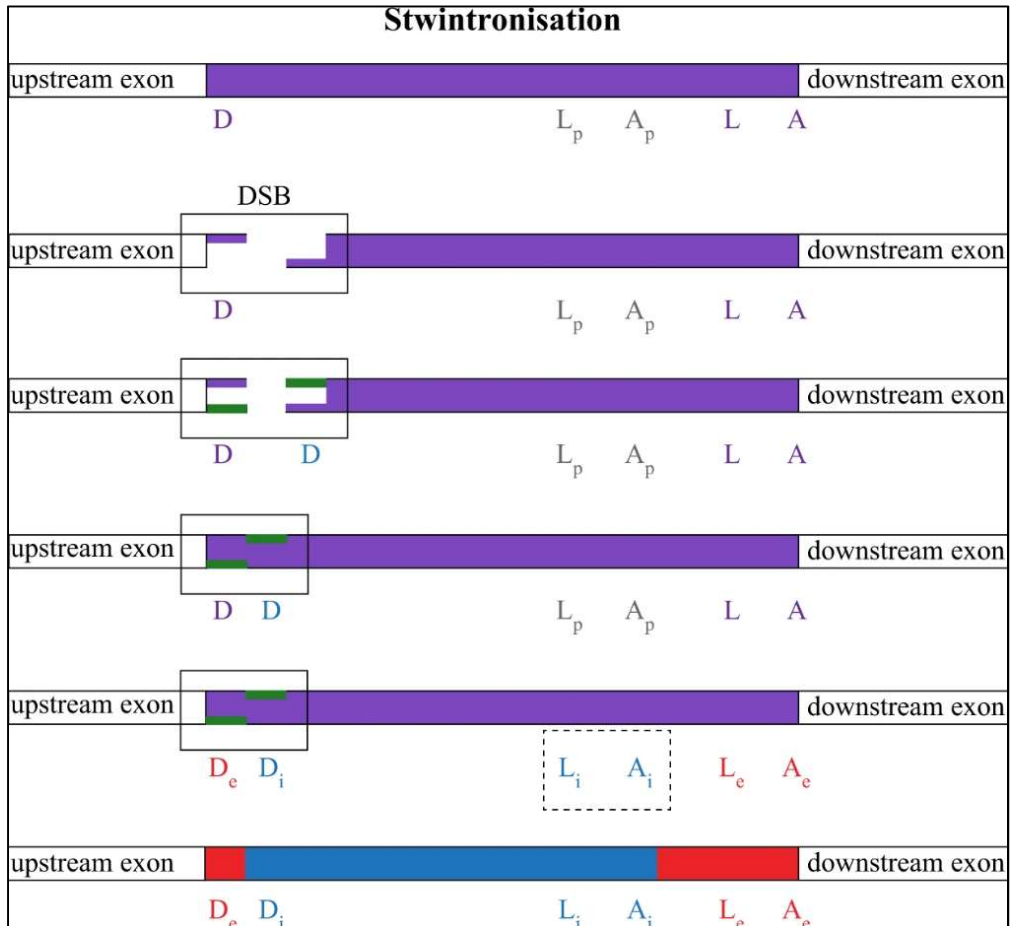
5.1 Retikulum fehérje (RTN) ortológok vizsgálata

5.1.1 Stwintronizáció *Lipomyces* fajokban

Korábbi kutatásaink révén fonalas gombák evolúciós megfigyelése közben két [D] stwintronra lettünk figyelmesek *Pezizomycotina* taxonon belül. Ezek olyan pozíciókban helyezkednek el, ahol korábban elkülönült fajokban még csak kanonikus intronok szerepeltek, így hát feltehetően egy „stwintronizációs” lépés, avagy stwintron kialakulása folyhatott le az idők folyamán (18). A stwintronizáció kifejezést az intronizáció szó után vett analóggként használjuk (108). Egy új [D] stwintron keletkezéséhez elengedhetetlen egy új 5'-vágási hely kialakulása az eredeti intronon belül vagy a donor régiója mellett. Ez a donor- (5'-GURWGY), vagy a donor egy részének a duplikációjával következhet be, amit egy aszimmetrikus duplaszálú törést eredményező javító mechanizmus, a DSB (double-strand breaks) okozhat 3' irányú túlnyúló véggel. Egy ettől teljesen függetlenül bekövetkező mutáció az eredetileg jelenlévő intron belsejében 3'-vágási helyet eredményezhet, ez által egy belső intron konszenzus szekvenciái alakulhatnak ki, ami így egy kétlépésből kivágódó stwintronhoz vezet.

A stwintronizáció folyamatát a 6. ábra szemlélteti. Ha egy U2 kanonikus intron hossza engedi, két funkcionálisan kivágható intronná képes formálódni, melyek egymásba ékelődve – stwintronként – lesznek jelen a genomban. Az intront a lila szakasz jelöli. A DSB mechanizmus következményeképpen létrejövő túlnyúló részt zölddel jelöltük. Mielőtt a ligálás végbemehetne, a törést követően egy kanonikus nem-homológ vég összekapcsoló (canonical nonhomologous DNA end-joining, cNHEJ) mechanizmus feltölti a szálát. Egy ettől független mutáció

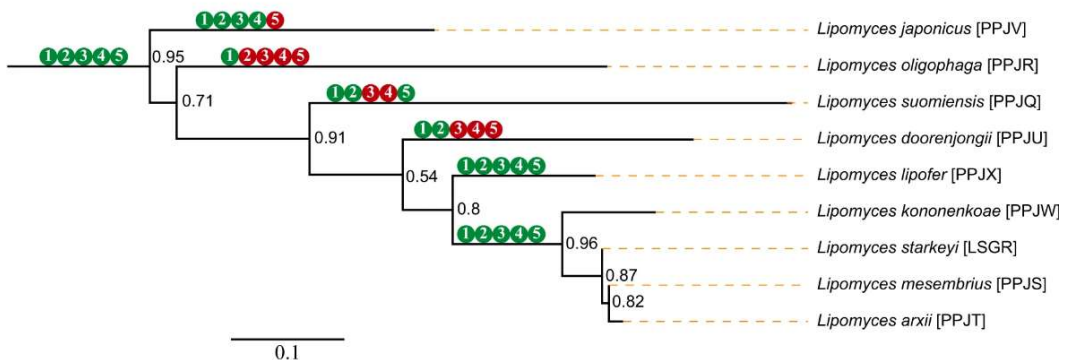
eredményeképpen az eredeti intronon belül található prekursor („p”) szekvenciák hasonlóságuk miatt egy lariat („L”) és akceptor („A”) 3’-végi motívumokat eredményeznek („i”).



6. ábra Stwintronizáció lépéseinek sematikus ábrázolása: Új, funkcionális 5'- és 3'- splicing motívumok megjelenése egy már létező U2 intronon belül. Lila: eredeti kanonikus U2 intron; zöld: DSB következménye, lelógó szekvencia darab; kék: keletkezett belső intron; piros: az eredeti intron külső intronná alakult (20).

5.1.2 Három új típusú CIS *Lipomyces*ekben

A retikulum fehérje őseinek felkutatása céljából, amely egy korábbi publikációnkból ered (19), homológ géneket kezdtünk keresni különböző gomba fajokban. Találtunk homológ géneket az *Ascomycota* (Tömlősgombák) törzs *Saccharomycotina* és *Taphrinomycotina* altörzseiben, a *Blastocladiomycota*, a *Glomeromycota*, a *Mucoromycotina* és a *Mortierellomycotina* törzsekben. Majdnem az összes élesztő intron nélküli homológ gént tartalmaz, azonban néhány faj, a *Cephaloascaceae*, *Alloascoideaceae*, *Pichiaceae* és *Debaryomycetaceae* családokból kanonikus intront hordoz az *Aspergillus nidulans rtnA* génjének 5. intron pozíciójának megfelelő helyén (19). Ennek a pozíciónak a konzerváltsága filogenetikai kapcsolathoz vezet, ám a *Pezizomycotina* altörzsön belül nem találtunk [D5,6] stwintronokat.

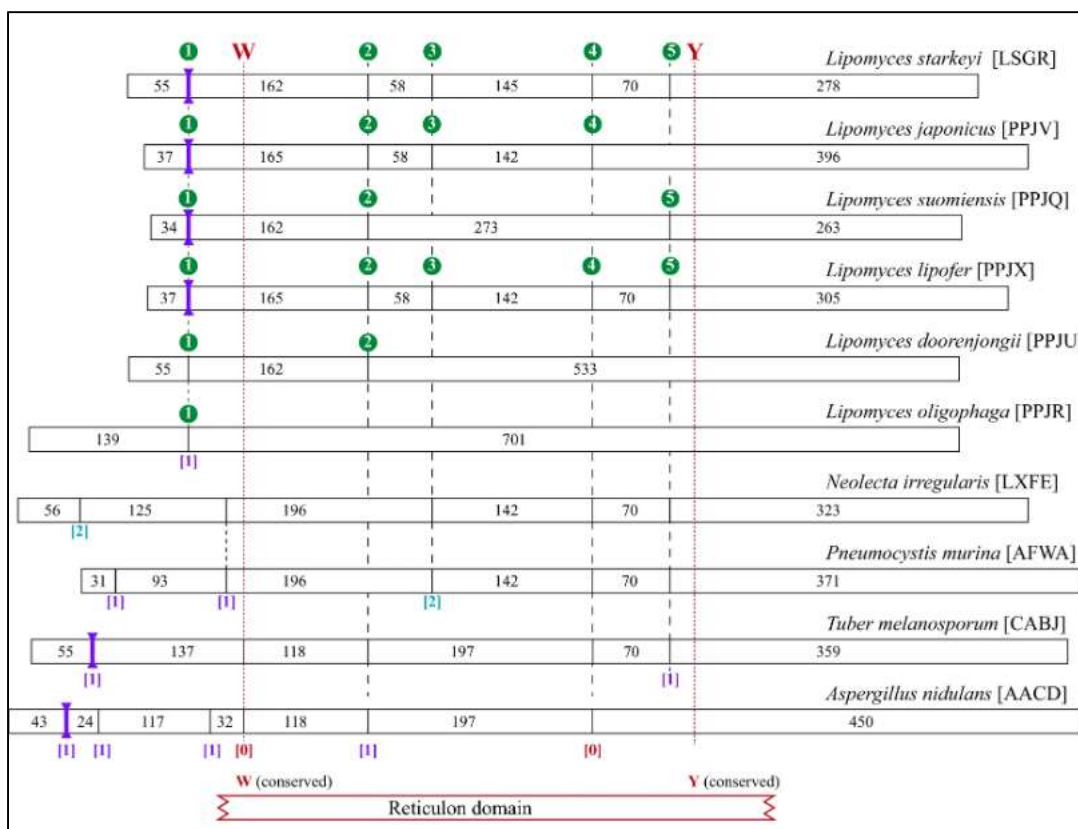


7. ábra *Lipomyces* nemzetség monofiletikus klasztere a 902 *Ascomycota* faj retikulum fehérjéjéről készült maximum likelihood törzsfából. A törzsfá készítése részletesebben a 4.1.3 címsor alatt olvasható. Az ágakat az approximate likelihood-ratio test (96) adta, a skálázás 0.1 aminosav szubsztitúció alapján történt. A színezett körök számokkal az 5 konzervált intron jelenlétét (zöld) vagy hiányát (piros) mutatja a különböző fajokban a gén elsődleges átíratában. A fajok nevei egyedi 4 betűből álló jelzéssel lettek ellátva a Whole Genome Shotgun Master Accession alapján.

Egyedül a *Lipomyces* fajok homológ génjében találhatóak meg többszörös intronok, ahol a 2. intron pozíció az eredeti *A. nidulans* 5. intron pozíciója. Az összes RTN fehérje kódoló génjét felhasználtuk egy törzsfá

készítéséhez a tömlősgombákban belül, hogy ellenőrizzük a *Lipomycetaceae*, mint a *Saccharomycotina* legkorábban elváló családjának státuszát (105) (106). A használt *Lipomyces* genom szekvenciákat a 11.1 melléklet szemlélteti.

A beazonosított [D5,6] típusú stwintron a retikulum fehérje kódoló génen belül található az *Ascomycoták* *Pezizomycotina* altörzsébe tartozó gombáknál (19). A gén ismeretlen funkciójú, de az endoplazmatikus retikulumhoz köthető feladattal rendelkezik (protein family PF02453). A



8. ábra *Lipomyces*- és egyéb *Ascomycota* fajok összehasonlításának sematikus ábrázolása a génmodelljeik intron-exon struktúrájának alapján. A fehér téglalapok jelölik a géneket. Bennük az exonok téglalapjai szerepelnek számmal ellátva (nukleotidok). Az intronokat a függőleges folyamatos vonalak jelzik. Kapcsos zárójelben láthatóak az intronok fázis számai, a CIS elemeket a vastag lila vonalak jelölik. A konzervált retikulum régió a génmodellek alatt látható piros keretben, valamint a konzervált intronok (2-5. pozícióban) szaggatott vonalakkal vannak összekötve (20).

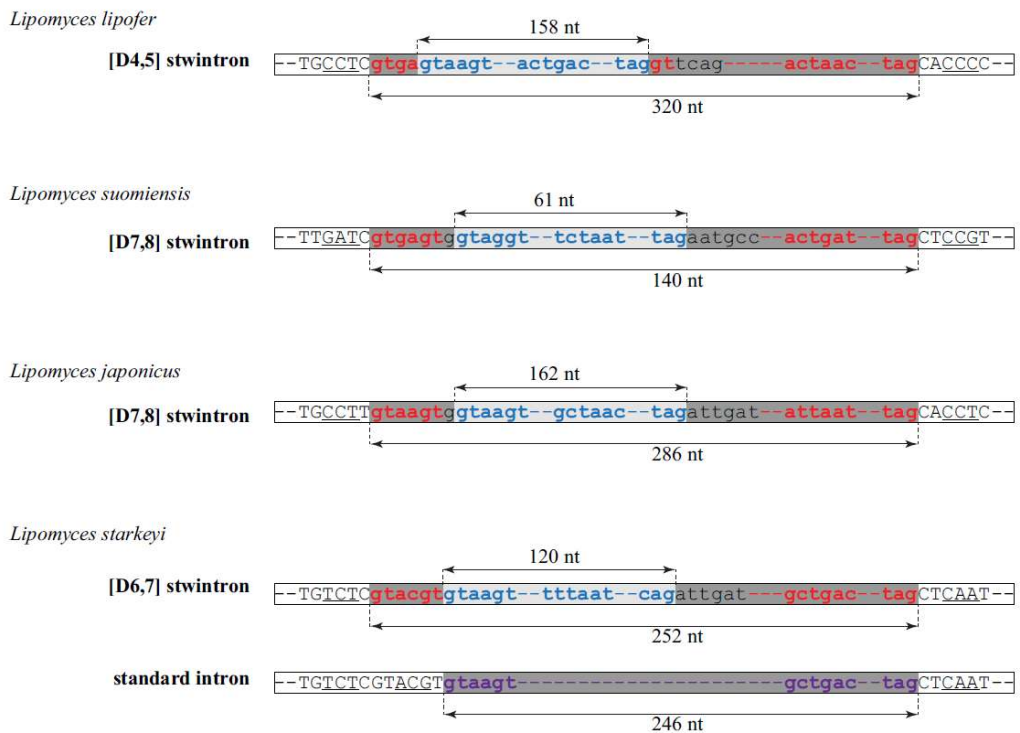
stwintron származásának vizsgálata során U2 típusú intronokból álló különböző komplex megszakító struktúrákat (CIS) fedeztünk fel a vizsgált gén 7 *Lipomyces* fajban található ortológjában, ami egy másik altörzse az *Ascomycotának*. A 3 különböző CIS elem párhuzamos evolúciós előfordulása szigorúan meghatározott intron pozíciókban ugyanazon nemzetség fajai között bizonyíthatja a stwintronizációs jelenséget, így ezek a stwintronok egy újabb evolúciós lépést képviselnek.

Az *Aspergillus nidulans* és más *Pezizomycotina* taxonhoz tartozó fajok retikulum fehérjéje *Taphrinomycotina* alosztály több fájában, valamint a *Lipomycetaceae* család tagjain belül erősen konzervált intron pozíciókkal rendelkeznek. A retikulum fehérje *Lipomyces* nemzetség 9 fajáról készült monofiletikus klasztert a 7. ábra szemlélteti. A *Lipomycetaceae* család egyedülálló filogenetikai pozíciót foglal el a *Saccharomycotina* kládban (105). A *Lipomyces starkeyi* kiemelkedően több ORF (open-reading frame) régiót megszakító intronnal rendelkezik, mint a *Lipomycetaceae* családon kívüli élesztő fajok. Az intron sűrűségét inkább a *Pezizomycotina* tagjainak az intron sűrűsége jellemzi.

A *Saccharomycotinában* kivételesen, a retikulum fehérje génjének elsődleges átírata több intront tartalmaz nyolc genom-szekvenált *Lipomyces* törzs esetében, habár az intron tartalom jelentősen eltér a *L. starkeyi*-, másik 3 közeli rokon faj és *L. lipofer* maximális 5 intronjától (az intron-exon struktúrát a 8. ábra szemlélteti). *L. japonicus* legszélső intronja 3' irányból hiányzik, *L. suomiensis*-ben a 3. és 4. intron pozícióban hiányoznak az intronok. *L. doorenjongii* csak az első két helyen lévő intronnal rendelkezik, *L. oligophaga* pedig egy intront tartalmaz átíratának 5' végén. A 4 intron 3' irányból *L. starkeyi*-ben megszakítja a PF02453 kódoló domént. Ezekből az intronokból a 2. pozícióban lévő a *Pezizomyctinában*-, a 3. pozícióban lévő

Taphrinomycotina fajokban, a 4. és 5. pozícióban lévő intronok pedig mindkét altörzsben konzerváltak jelen. Ezen megfigyelések alapján megállapítható, hogy a korán divergáló *Lipomyces* fajok vesztették el az intronjaikat, és nem intron képződés jellemző más taxonokra. Ettől teljesen független a *Saccharomycotina* egyéb családjaiban belüli masszív intronvesztés.

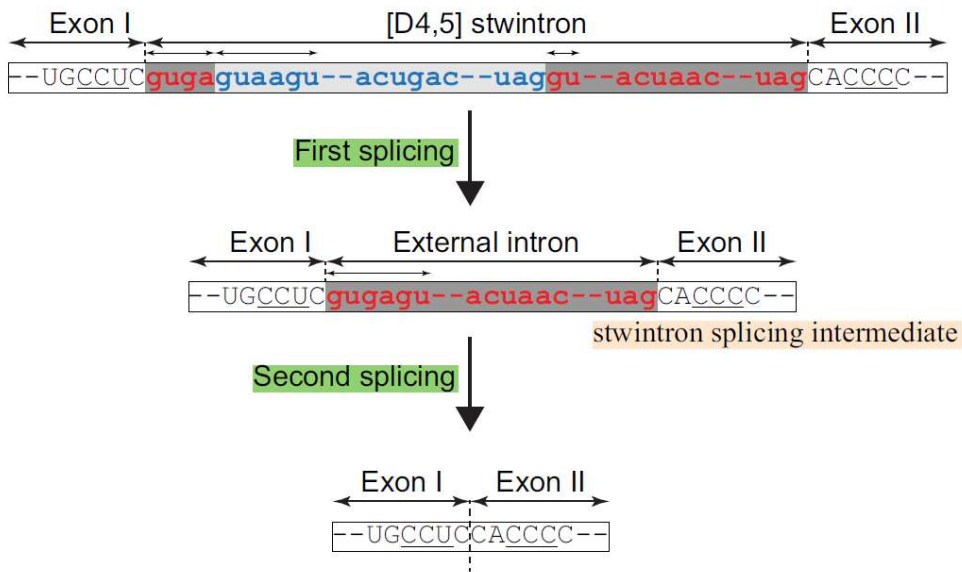
A 9 *Lipomyces* fajra jellemző az első konzervált intron pozícióban lévő CIS (komplex megszakító szekvencia) retikulum fehérjét kódoló gén esetében. A gén 2. exonja így 162 nt vagy 165 nt hosszúságú (*L. japonicus* és *L. lipofer*). Azaz extra kodon az erősen konzervált PF02453 domén Trp (W, 8. ábra) aminosavtól 5' irányban található. *L. oligophaga* és *L.*



9. ábra A 3 komplex megszakító szekvencia [D4,5], [D6,7], [D7,8] *Lipomyces* fajok retikulum fehérje génjében. *L. starkeyi* alternatív splicingja egy önálló nagy kanonikus U2 intronként szintén ábrázolva (20).

doorenjongii első intron pozíciójában nem CIS elem, hanem 78 nt és 46 nt hosszúságú kanonikus U2 intron található, azonban a 7 másik fajban jóval hosszabb – *L. suomiensis* 140 nt, *L. lipofer* 320 nt – két egymásba ékelődött U2 intron struktúra található. Ezek a szerkezetek 3 módon szerveződtek. Mindegyik esetben a belső intron az 5' végén szakítja meg a külső intront minimális eltérésekkel. A 9. ábra illusztrálja a kísérletesen igazolt CIS elemeket 3 fajban. RT-PCR segítségével kimutattuk a stwintronok splinter, és a teljes gén intronmentes mRNS átíratát cDNS formában (4.4 bekezdés).

Lipomyces lipofer

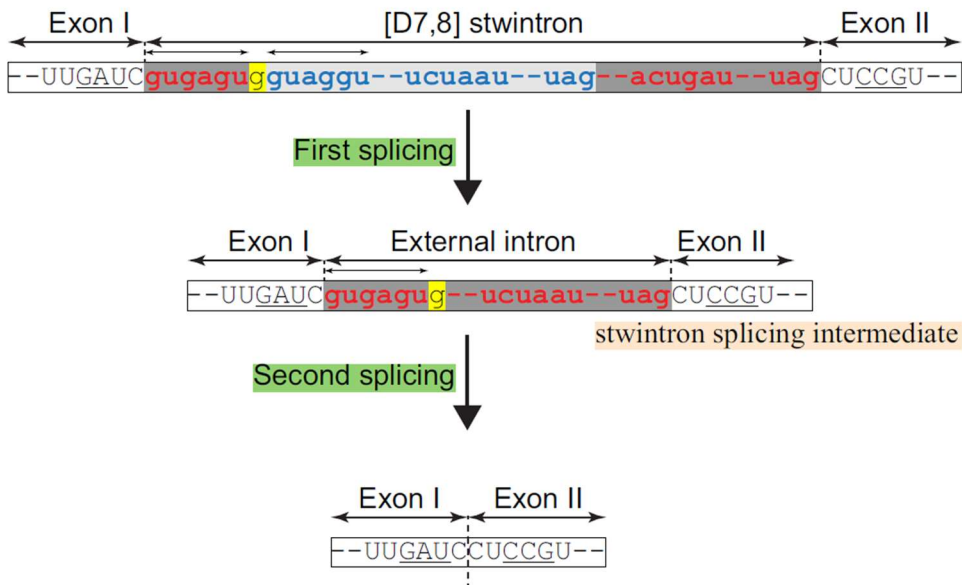


10. ábra *L. lipofer* retikulum fehérjét kódoló génjében található [D4,5] stwintron két lépéses kivágódásának sematikus ábrázolása (20).

L. lipofer-ben a CIS elem 320 nt hosszú. Egy 158 nt hosszúságú belső intron (5'-GUAAGU – 131 nt – ACUGAC – 12 nt – UAG) megszakítja az 5'-donor szekvenciáját egy 162 nt hosszúságú külső intronnak (5'-GUGA | GU – 137 nt – ACUAAC – 10 nt – UAG) a donor 4. (A₄) és 5. (G₅) nukleotidja között. A kivágódási mechanizmust a 10. ábra mutatja. Ezen

[D4,5] típusú stwintron splinter, belső intron nélküli és az ép külső intront tartalmazó köztes PCR termék, valamint a teljesen kivágódott CIS elem gyakorlatban történő igazolása megtörtént és a GenBank MN689083, MN689084 számokon regisztráltak. Ez az első donorjában megszakított stwintron (*sensu stricto*) *L. lipofer*-ben és ez az első stwintron, ami a *Pezizomycotina* altörzsön kívül esik.

Lipomyces suomiensis



11. ábra *L. suomiensis* retikulum fehérjét kódoló génjében található [D7,8] (stwintron *sensu lato*) CIS két lépéses kivágódásának sematikus ábrázolása (20).

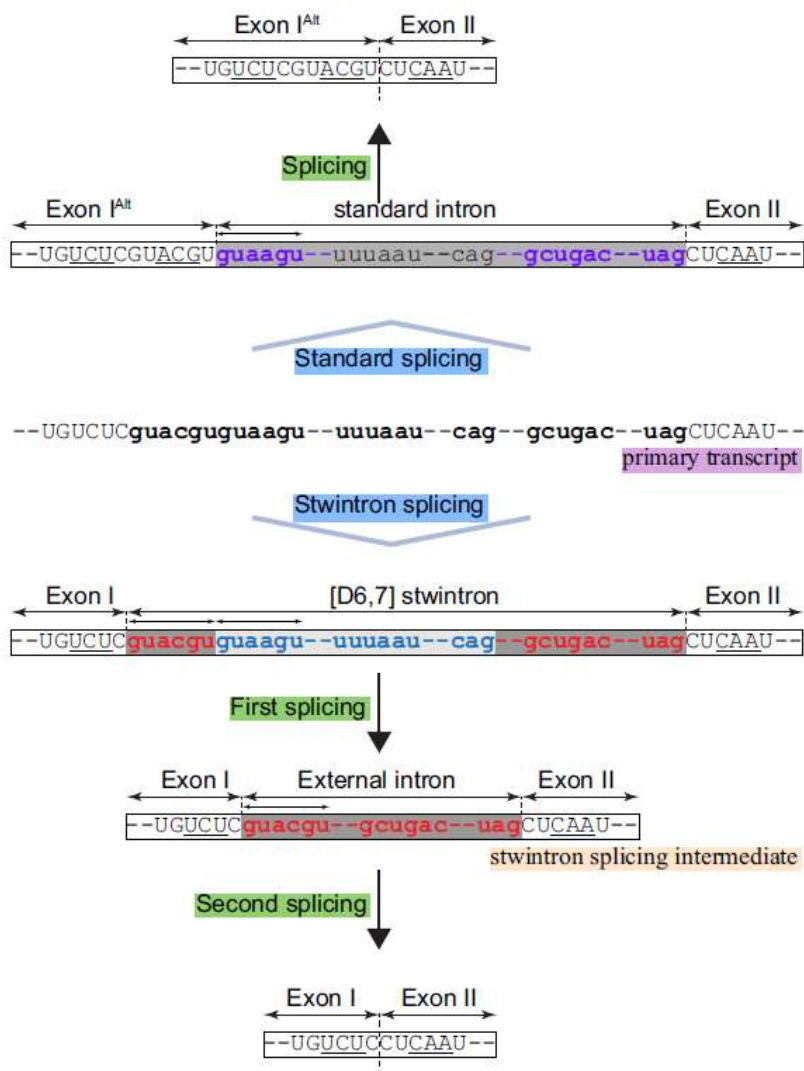
Két faj – *L. suomiensis* és *L. japonicus* – esetében két potenciális 5'-donor elemet észleltünk egy nukleotidnyi távolságra egymástól a szokásos 1. intron pozícióban lévő CIS elemen belül. 3'-akceptor és hozzátartozó lariat elemek is megtalálhatóak. *L. suomiensis* érett mRNS-ének szekvencia analízise (GenBank MN689082, 11. ábra) mutatta meg, hogy a szélső donor elem (5'-GUGAGU) felel a teljes 140 nt hosszúságú CIS struktúra teljes kivágódásáért. További kísérletekkel bebizonyítottuk, hogy a 3' irányú

donor (5'-GUAGGU) felel a belső 61 nt hosszú intron kivágódásáért (5'-GUAGGU – 38 nt – UCUAAC – 8 nt – UAG), ami egy 79 nt-ből álló külső intront (5'-GUGAGUG | A – 45 nt – ACUGAC – 17 nt – UAG) szakít meg a 7. (G₇) és 8. (A₈) nukleotidja között (11. ábra). Ehhez hasonlóan *L. japonicus*-ban is kimutattunk egy 162 nt hosszúságú belső intront (5'-GUAAGU – 135 nt – ACUAAC – 12 nt – UAG), ami megszakít egy 124 nt hosszú külső intront (5'-GUAAGUG | A – 97 nt – AUUAAU – 10 nt – UAG) G₇ és A₈ között. A teljes CIS 286 nt hosszú és két ugyanolyan szekvenciájú (5'-GUAAGU) funkcionális donor régió található benne (9. ábra). A két faj közötti különbség a CIS elemeket alkotó intronjaik közti méret különbség. Az *L. suomiensis*-ben lévő intronok fele akkorák, mint az *L. japonicus*-ban.

Ezeknek a [D7,8]-as CIS elemeknek viszont ugyanúgy kétlépéses egymást követő splicing eseménynek kell lezajlania ahhoz, hogy teljesen kivágódjanak. Ez esetben a beékelődés nem *sensu stricto* (109) stwintront hozott létre ugyanis a belső intron nem szakítja meg a külső intron kivágódáshoz esszenciális donor szekvenciáját, 1 nukleotidnyi távolságra található a két donor egymás mellett. Az *L. suomiensis* retikulum fehérjéjének elsődleges átiratában a spliceoszóma által elsődlegesen választott 5' és 3' vágási helyek egyértelműen a lehető legkisebb belső intron kivágódását preferálják (intron definíció), így az ehhez hasonlóan beépült intronok példái lehetnek egy új osztályú „*sensu lato*” stwintronnak.

L. starkeyi-ben és 3 másik rokon fajban (7. ábra) szintén egy „*sensu lato*” típusú CIS struktúrára bukkantunk. A belső intron beékelődése egy [D6,7] típusú stwintront alakított ki (12. ábra). Az átiratban közvetlenül egymás mellett helyezkedik el két feltételezhetően donor elem. Azonban az 5' irányban lévő donor szekvenciája (5'-GUACGU) kevésbé kanonikus,

Lipomyces starkeyi



12. ábra *Lipomyces starkeyi* retikulum fehérjé kódoló génjében található [D7,8] (sensu lato) CIS kétlépéses és egy lépésben történő alternatív kivágódásának sematikus ábrázolása (20).

mint a 3' irányban mellette lévő (5'-GUAAGU). A kivágódási lehetőségeket először *L. starkeyi*-ben teszteltük (törzs CCY 33-1-1: gén szekvencia, GenBank MN689085) és kétfajta cDNS transzkriptumot kaptunk. Az egyik transzkriptum normál, kanonikus intronként vágódik ki

úgy, hogy az 5' iránybeli donor régiót a spliceozóma nem veszi figyelembe és az érett mRNS szekvenciájában ez a plusz 6 nt (GTACGT) benne marad, a másik mRNS klón pedig egy stwintron kivágódási köztes eredményét mutatta (GenBank MN689086 és MN689087). A kísérleti körülményeink során a stwintronokra jellemzően kivágódott szekvencia volt a több. Utólagosan két egymáshoz közel eső (5'-AUUCAU és 5'-UUUAAU) tökéletlen szekvenciájú lariat elemet fedeztünk fel, melyek közül valamelyik 3' irányban elhelyezkedő akceptor (5'-CAG) régió kíséretében belső intronként eltávolítódik a spliceozóma által. Ez esetben a megszakító [D6,7] szekvencia összesen 252 nt hosszú, melyben egy 121 nt hosszúságú U2 intron (5'-GUAAGU – 75 nt – AUUCAU – 14 nt – UUUAAU – 10 nt – CAG) megszakít egy 132 nt hosszúságú külső intront (5'-GUACGU | A – 109 nt – GCUGAC – 7 nt – UAG) U₆ és A₇ nukleotidok között (12. ábra). Ha a spliceozóma nem veszi figyelembe a belső intron tökéletlen konszenzus szekvenciáit, ez a szekvencia a 3' irányú donorral egy 246 nt nagy kanonikus U2 intronként (5'-GUAAGU – 224 nt – GCUGAC – 7 nt – UAG) egy lépésben is ki tud vágódni, ami egy alternatív splicing eredménye (12. ábra) és a transzkriptum 6 nukleotiddal hosszabb lesz (GenBank MN689086).

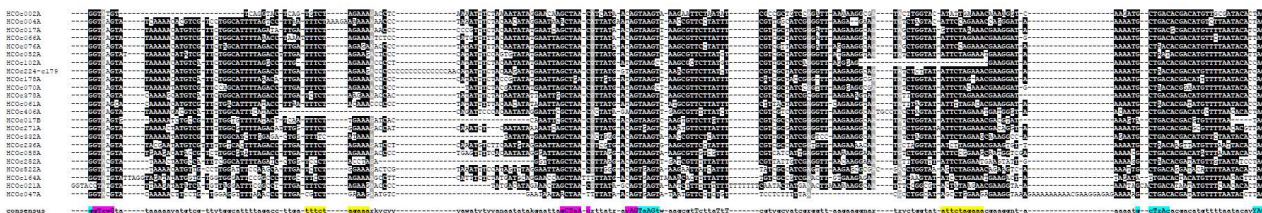
Az *L. starkeyi* retikulum fehérjét kódoló génje egy [D6,7] típusú *sensu lato* stwintront tartalmaz, amelyben két közvetlen egymás mellett álló 5' donor elem található. Ez a CIS elem képes stwintronokra jellemző eltávolítódásra és képes egy hosszú önálló kanonikus U2 intronként is kivágódni. Mindkét donor funkcionálisan részt vesz az alternatív splicingban. A stwintron féle kivágódási mechanizmus mindkét donor elemet felhasználja. A 3' irányú kanonikusabb szekvenciájú donor elem a CIS felénél lévő tökéletlen 3'-splice hellyel szemben a kanonikusabb 3' vágási

hellyel vágódik ki első lépésben. Az U2 spliceozóma nem a közvetlenül szomszédos funkcionális 5'-donorok között válogat, hanem a rendelkezésre álló lariatok közül azt választja, amelyik nagyobb távolságra helyezkedik el.

5.2 Sister (stw)intronok *Hypoxyylon* gombákban

5.2.1 Sister (stw)intronok *Hypoxyylon sp. CO27-5* fajban

A. nidulans jól konzervált mitokondriális szállító fehérje génjének exon-intron struktúrájának vizsgálata során (*A. nidulans* GenBank AACD01000130, CDS koordináták start-stop: 208790-210069, nem publikált eredmény) tipikus [D1,2] típusú stwintronokra bukkantunk *Hypoxyylon sp. CO27-5* és *Hypoxyylon sp. EC38* gombák ortológ génjeiben, ami az összes többi vizsgált gombából hiányzott. A *CO27-5* stwintronjának szekvenciáját visszaneztük a teljes genomon BLASTN segítségével és újabb teljesen- vagy részlegesen hasonló megszakító szekvenciákat azonosítottunk a genomon belül. Ez arra enged következtetni, hogy ez a [D1,2]-es stwintron része egy struktúráisan hasonló megszakító szekvencia csoportnak, így felkeltette az érdeklődésünket.

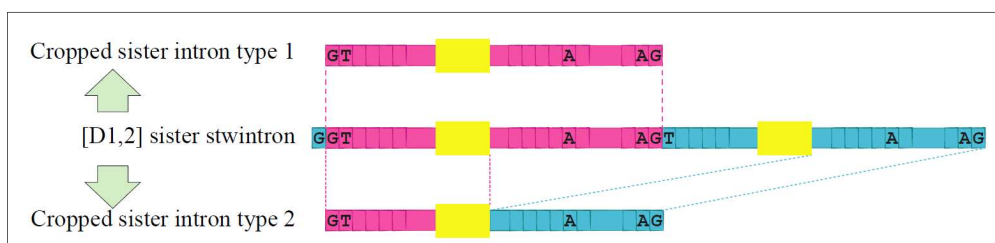


13. ábra 23 [D1,2] típusú magas hasonlóságot mutató stwintron egymás alá igazítva *Hypoxyylon sp. CO27-5* gombából (GenBank MSCL00000000: 580 contigs) BLASTN kereséssel történt a szekvenciák gyűjtése egy konzervált mitokondriális szállító fehérjében lévő stwintron alapján (HCOc017A) (21)

Egy egyszerű BLASTN kereséssel 23 olyan [D1,2] típusú stwintront találtunk *Hypoxyylon sp. CO27-5* genomjában, amelyek teljes hosszúságukban hasonlóságot mutatnak egymással. Ezt a 13. ábra

szemlélteti, ahol a szekvenciák egymás alá vannak illesztve és a fekete háttér jelöli az egyforma nukleotidokat. Az összehasonlítást a legelső talált stwintronhoz végeztük (mitokondriális szállító fehérje), amit HCOc017A névvel láttunk el, ahol a „HCO” a gomba fajra utal, a c017 a kontigra (GenBank MDCL01000017), a nagybetű a végén (A) pedig a kontigon belüli stwintronok megkülönböztetésére szolgál. A 22 stwintron hasonlóságának mértéke 58-90%. Ezen 23 stwintront „sister” (testvér) stwintronoknak neveztük el, magas mértékű homológiájuk és a valószínűsíthető közös őstől való származásuk miatt.

Egy sister stwintron (HCOc224-179) két, nem átfedő kontigon ível át, aminek teljes szekvenciáját szintén megszekvenáltuk (GenBank accession number MW477887). Ismétlődő intron szekvencia elemek – másnéven „intronerek” – korábban publikálásra kerültek különböző eukarióta országokon belül, köztük a gombákéban is, azonban összetett megszakító szekvenciák, mint a stwintronok esetében még nem történt proliferációról szóló jelentés (110) (111) (112) (113) (114) (115) (116) (117) (118) (119) (120) (121).

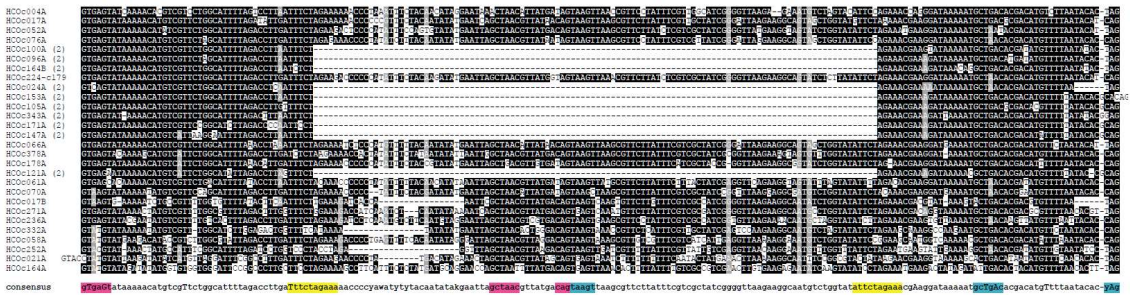


14. ábra *Cropped sister intronok típusainak származtatása a sister stwintronokból. (Magenta: belső intron, türkiz: külső intron, sárga: palindrom szekvencia 5'-WTTCTAGAAA.) (21)*

Ezen kívül találtunk még 13 [D1,2] stwintront, amik a sisterekhez regionálisan hasonlítanak. A HCOc103A külső intronja nagy hasonlóságot mutat a sister intronok külső intronjával, azonban a belső intronja különböző. Mivel ezek a megszakító szekvenciák igaz, csak parciálisan, de

magas szintű hasonlóságot mutatnak „sheared sister” stwintronoknak nevezük őket. Közülük 7 konzervált régiót tartalmaz a szekvencia középső részén, ahol a két intron átfed egy guaninban (AGT). Mind a 36 sister stwintron szekvenciája (23 sister és 13 sheared sister) a 11.3.1 és 11.3.2 bekezdésekben láthatóak. A sheared sister stwintronok szekvenciáján belül sárga háttérrel és aláhúzással lett jelölve HCOc017A-hoz való homológikus rész. 2 kivételével, az összes sister szintén szerepel *EC38*-ban is.

A stwintronok tipikus kétlépéses kivágódását az NCBI-on elérhető RNS SRA adatokból és a párhuzamosan folyó RT-PCR reakciók eredményeivel és szekvenálással igazoltuk. 31 sister kivágódási köztes termékeit sikeresen bizonyítottuk, viszont 4 esetben kísérletes úton nem tudtuk a splicing intermediereket megtalálni, valószínűleg az általunk alkalmazott növesztési körülmények nem tették lehetővé ezen gének kifejeződését.



15. ábra A 2-es típusú cropped sister intronok illesztése a 18 leghasonlóbb sister stwintronhoz. Belső intron kivágódáshoz szükséges elemei magenta színnel, a külső introné türkizzel vannak jelölve. Sárga színnel a palindrom szekvenciák vannak jelölve (21).

A sister és sheared sister stwintronok mellett további 12 kanonikus intronra is felfigyeltünk, amelyek szintén magas szintű homológiát mutatnak a sister stwintronokhoz. 2 típusba soroltuk őket az alapján, hogy a stwintronok mely régióihoz hasonlítanak, így **1-es- és 2-es típusú cropped sister intronoknak** kereszteltük el őket. Az egyes típusba csupán 2 intront soroltunk (HCOc047B, HCOc229A), ezek a belső intronokhoz

hasonlítanak nagyon, valószínűleg a külső intron deléciójával jöhettek létre (17) (18). A 2-es típusú cropped sister intronok viszont a stwintron 5' végi belső intron szekvenciájának és a 3' végi külső intron szekvenciájának a fúziójaként véve mutat magas szintű hasonlóságot (14. ábra). Valószínűleg a sister stwintronok középső régiójának deléciója következtében jöhettek létre, hiszen sokkal magasabb mértékű homológiát mutatnak, mint a sister stwintronok egymáshoz viszonyítva (15. ábra).

A 15. ábrán 10 2-es típusú cropped sister intron a 18 legnagyobb homológiát mutató sister stwintronnal van egymás alá illesztve. Az ábrából feltételezhetően a stwintronok középső része az azt határoló két palindrom szekvencia rekombinációjával vágódhatott ki (az ábrán sárga rész), ezzel létrehozva egy középső részen palindromot tartalmazó kanonikus intront. A 10 nukleotidból álló palindrom szekvencia 5'-WTTCTAGAAA 5' oldali másolata, ami a belső intron középső részén helyezkedik el, a legtöbb esetben a 5'-W=T, viszont a 3' oldali másolata, ami a külső intron középső részén helyezkedik el, a legtöbb esetben tökéletlenül 5'-W=A (egy kivétellel). Így a feltehetően ebből a rekombinációból keletkezett sister intronok egy tökéletes palindromot tartalmaznak a középső részükön: 5'-TTTCTAGAAA. 19 sister stwintron és 10 2-es típusú cropped sister intron maximum likelihood szerinti filogenetikai kapcsolata *Hypoxylon sp. CO27-5*-ben a 11.3.4 bekezdésben látható (24. ábra).

Ezt a mechanizmust két ortológ gén 5. pozícióiban lévő megszakító szekvenciát összehasonlító megfigyelésünk alátámasztja. *Hypoxylon sp. EC38* törzsben egy membrán fehérje 5. intron pozíciójában egy 2-es fázisú 2-es típusú cropped sister intron van jelen (HECc066A), aminek *CO27-5* törzs ortológ génjének ugyanebben a pozícióban egy nukleotiddal elcsúszva, egy 1-es fázisú sister stwintron (HCOc066A) található. A két

szekvencia összeillesztésekor látható, hogy az *EC38*-ban lévő kanonikus intron a kivágódásnak megfelelően alakulhatott ki, hiányzik belőle a stwintron középső része.

A két palindrom közötti intron szekvencia delécióját a DSB (double-stranded DNA break) mechanizmus és az azt követő MMEJ (microhomology-mediated end-joining) javító lépés okozhatja (122) (123). Ehhez a palindromok szolgálnak mikrohomológokként és egy kulcsenzim segítségével történik, ami a *Hypoxylon sp. CO27-5* esetében egy valószínű POLQ-like helikáz (HelQ) enzim (MWMW530496, MW530509). Metazoák esetében Polimeráz θ (POLQ), több *Pezizomycotina* taxonban pedig superfamily-II DNA/RNS helikáz segíti a deléciót, ami homológ a metazoákéban lévővel, csak polimeráz domén nélkül (124) (125) (126).

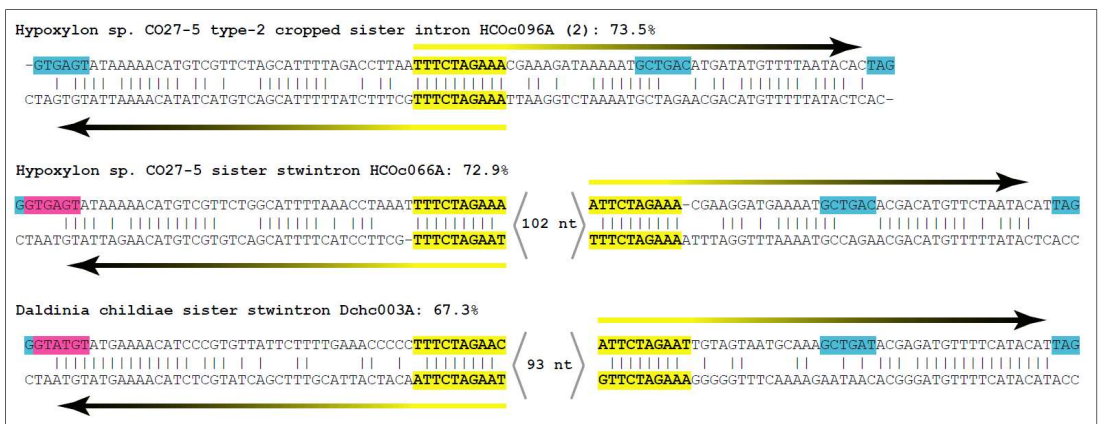
Ez a stwintronból intronná történő átalakulás arra utal, hogy a stwintron külső intronjának az 5'-donor G_1 nukleotidja exonná válik, és egy másodlagos mutáció szükséges így az érett mRNA olvasási keretének helyreállításához. Ebben az esetben (HECc034A), bármelyik intron-exon csomópont melletti másodlagos mutációk transzverziót eredményeztek és az exonná alakult G_1 timinné alakult az intron pozíciójának egyidejű eltolódásával. A Thr kodonban megszakított (A|CC) 1-es fázisú stwintron *CO27-5*-ben Ile kodonná válik (AT|C), amit azután már egy 2-es fázisú kanonikus intron szakít meg *EC38*-ban egy exon C elvesztésével az eredeti stwintronon belül. Más *Hypoxylon* fajokban, ebben a pozícióban és környékén nincs intron, viszont majdnem mindegyik esetben jelen lévő Thr kodon azt sejteti, hogy az *EC38* cropped sister intron az ortológ *CO27-5*-ben lévő sister stwintronból alakult ki, és nem fordítva.

Külön csoportot alkot hat 2-es típusú cropped sister intron (HCOc105A; HCOc121A; HCOc147A; HCOc153A; HCOc171A;

HCOc343A, 11.3.3 bekezdés), melyek egy plusz guaninnal rendelkeznek 5' irányban a 3' akceptoruktól (kivéve HCOc153A). A stwintronokban nem található meg ez a plusz G. Továbbá 7 nukleotiddal 5' irányban a plusz guanintól, ez a 6 cropped intron rendelkezik egy timinnel, viszont a többi 4 intron és a stwintronok itt egy adeninnel rendelkeznek. Végezetül a kimaradt 4 cropped intron TAG akseptorral rendelkezik, míg a hat intront tartalmazó csoport tagjai (egy kivételével HCOc343A kevésbé kanonikus 5'-GAG) CAG akseptorral bírnak. Ez a hat tagú csoport megerősíti az elgondolást, miszerint a tagjai a sister stwintronokból proliferálódtak.

5.2.2 Sister (stw)intronok szimmetriai tulajdonságai

A sister megszakító szekvenciák nem csak egymáshoz viszonyítva mutatnak hasonlóságot, hanem önmaguk reverz komplementer szekvenciájukkal is. A láncvégi régiók magas szintű hasonlóságot mutatnak, ahol terminális fordított ismétlődések (terminal inverted repeat; TIR) vannak jelen. Ebben a homológiában szimmetriát véltünk felfedezni.



16. ábra *Cropped sister intronok és sister stwintronok saját reverz komplementer szekvenciájukkal végzett illesztések a láncvégi szimmetria szemléltetésére. A vízszintes nyilak a TIR régiót jelölik. Magentával a belső, türkizzel a külső intron konszenzus elemei, sárgával a palindrom szekvencia van jelölve (21).*

A 16. ábra néhány példát szemléltet a cropped sister intronok és sister stwintronok saját reverz komplementer szekvenciájukkal végzett illesztésekre, ahol láthatóak a láncvégi szimmetriák. A HCOc096A (93 nt) intront a tökéletlen TIR- (49 nt) és a palindrom szekvenciájával átfedésben van, így a hasonlóságuk egymáshoz viszonyítva 73,5%-os. A HCOc066A stwintronban (207 nt) 48 nukleotid hosszan találtunk azonosságot, ami 72,9%-ot jelent.

A 23-ból 9 stwintronnál találtunk az illesztés segítségével TIR (45-55 nt) régiót. A 11.3.4 bekezdésben látható sister (stw)intronok filogenetikai törzsfáján látható a szekvenciaillesztéssel megbecsült palindromok és TIR szekvenciák jelenlétének jelzése az (stw)intronok neve mellett. A palindromokat legalább 7 nukleotidnyi konzerváltság mellett igazoltuk. A 10 legkevésbé különböző (azaz valószínűleg a legutoljára kifejlődött) CO27-5 sister stwintronból összesen 9 rendelkezik egyértelmű TIR szekvenciával és a 10 nt hosszú palindrom mindkét másolatával, míg a HCOc061A-ban az egyik palindrom valamelyest elkorcsosult. A filogenetika fa (11.3.4 bekezdés) 9 korábban leágaztatott stwintron többségénél egyértelmű TIR elemeket nem lehetett azonosítani a reverz komplementer szekvenciák egyszerű illesztésével, ugyanis gyakran az egyik, vagy mindkét palindrom degenerált, vagy hiányzik. (A szimmetriához kigyűjtött információk a 11.3.4 bekezdésben láthatóak.)

A *Hypoxyton sp.* CO27-5-ben található 33 nagyon hasonló sister (stw)intron RNS-ek az Rfam adatbázisban (14.3-as verzió: 3446 család) (127) egyetlen ismert RNS-családhoz sem rendelhetőek hozzá. Nem mutatnak elsődleges szekvencia-hasonlóságot egyik leírt gomba introner-szerű szekvenciák családjával sem a *Capnodiales* rendbe tartozó

Dothideomycetes fajokban (113) (114) (115) (116) (117) (118) (119) (120) (121).

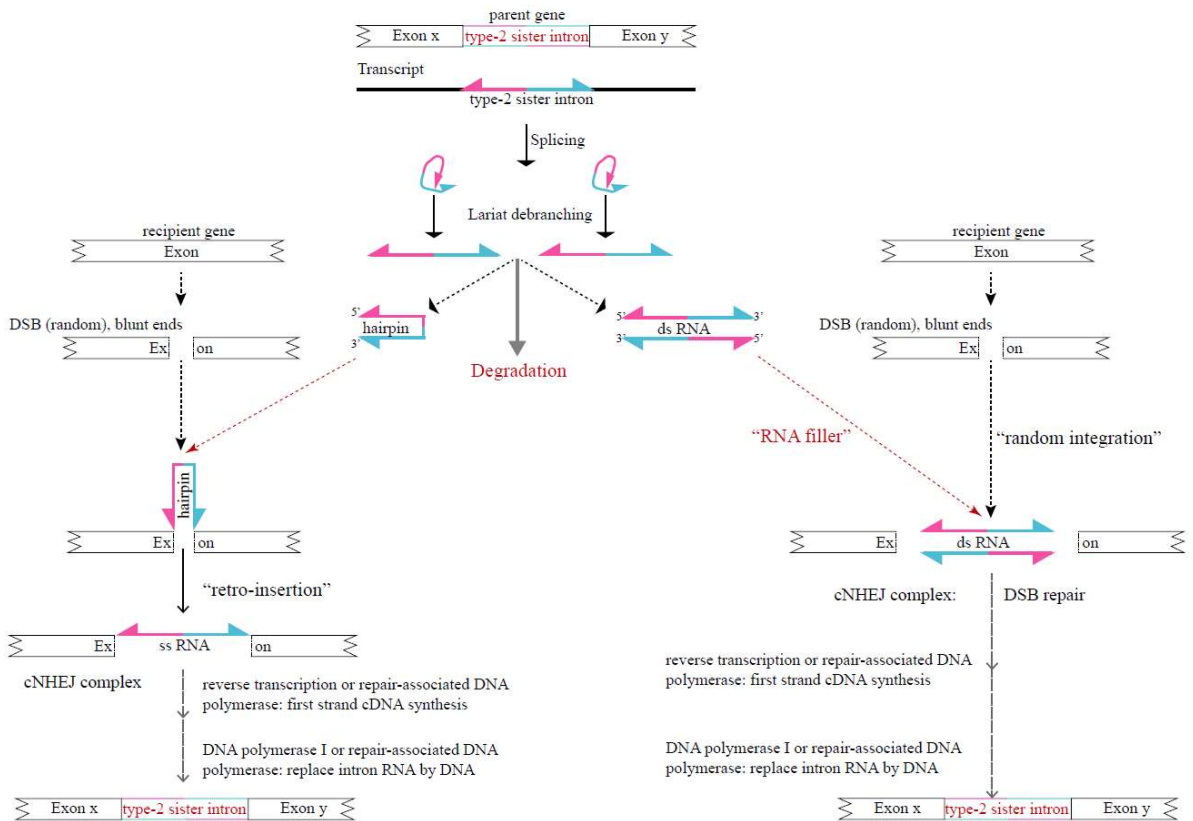
Fontos megjegyezni, hogy az expressziós adatok (SRA/RT-PCR) alapján a terminális szimmetria és a palindrom szekvenciák nem kitételei a sisterek kivágódásának. A TIR régiók figyelemre méltó jellemzője, hogy gyakran a három invariáns nukleotid, ami részt vesz az U2 kivágódás két átészterezési reakciójában (5'-terminális donor G, lariat elágazási A és a 3' terminális akceptor G), gyakran nem bázispárosodik a kalkulált hajtúszerkezet egyszálú (ss) RNS-ben. A valódi (stw)intron szekvencia csak a kódoló DNS-szálon van jelen.

5.2.3 Hipotézis: sister intronok proliferációja

Több megfigyelésünk is alátámasztja, hogy a sister (stw)intronok hasonlóságot mutatnak az introner-szerű elemekkel (115) (116) (117). Ezek az elemek (17. ábra), amik mára már felismerhetetlenségig elfajultak, random helyekre képesek voltak az exonok közé integrálódni. Az introner-szerű elemek stabil másodlagos szerkezettel és TIR régiókkal rendelkeznek. Magas belső szimmetriájuknak köszönhetően a 2-es típusú cropped sister intronok szintén képesek lehetnek duplikációra, ugyanis magas szintű hasonlósággal rendelkeznek a saját reverz komplementer szálukkal, amit a TIR régiók is erősítenek. Így képesek stabil másodlagos szerkezet kialakítására.

A hosszú láncvégi ismétlődések a transzpozonok szekvenciájában is tipikusan jelen vannak (128), ami azt sugallja, hogy a sister intronok proliferációját is az egyik leggyakoribb intronképző mechanizmus, a transzpozon inszerció segíthette elő (129) (130) (131). Azonban tandem hely duplikációk (tandem site duplication; TSD) nincsenek jelen egyik sister szekvenciában sem, ami elengedhetetlen a DNS transzpozícióhoz. Az

intronok terjedésével kapcsolatban egy olyan teória került felállításra, miszerint olyan reverz transzkripció folyamatok játszhatnak szerepet



17. ábra Sister intronok lehetséges duplikációt követő integrációs mechanizmusai sematikus ábrázolással (21)

ebben a folyamatban, mint a retro-homing a group II intronokban (132).

Eddigi tapasztalataink és szakirodalmak alapján 2 lehetőséget gondoltunk ki, a sister intronok duplikációjára, melyet a 2-es típusú cropped sister intronokkal szemléltetünk (17. ábra). Az intron duplikációt egy tompa-végű kettős-szálú DNS törés (DSB) következtében lezajló javítómechanizmus, a cNHEJ, vagy egy alternatív end-joining mechanizmus okozhatja (126) (133). Az új intronok integrációs helyeit a DSB határozza meg, ami korrelál a sister stwintronjaink integrálódott

helyeivel. Megfigyelésünk alapján egy 2-es típusú cropped intron (HCOc121A) esetében az intron inszerciója 3 nukleotid deléciójához köthető közvetlenül az integrációs helytől 5' irányban (5.2.4 bekezdés). Jelentős szerepe van az RNS-nek és az RNS-kötő fehérjéknek az NHEJ rendszerek DSB javításában (134) (135) (136). Nemrégiben beszámoltak RNS szakaszok átmeneti beépüléséről a DNS-szekvenciákba mitokondriális genom esetében és feltételezik, hogy ez előfordul a nukleáris genomokban is (137).

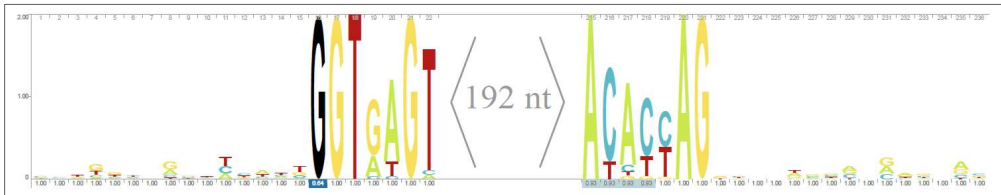
Az egyszálú sister intron RNS-ek hajtúkanyart képeznek, ami megvédi őket a degradációtól. Ebben a formában a cNHEJ komplex ki tud alakulni és az intron ön-ligációt tud végrehajtani az egyszálú DNS-be, majd hasonlóan a reverz splicinghoz az intron képes integrálódni a genomba.

Alternatív mechanizmusként a DSB alatt bekövetkezhet egy RNS-általi random integráció (138) (139). A DSB növeli a DNS integrációk gyakoriságát. Egy duplaszálú (ds) intron RNS képes lehet kialakulni két egyszálú (ss) és magas szimmetriai tulajdonságot mutató intron összetapadásával a lariat forma elágazásánál (140). Ez kitöltő szerepet tölt be a cNHEJ vezérelte DSB mechanizmus kezdeti szakaszában, végül integrálódik a genomba, ahogy azt *Micromonas* esetében láthatjuk (112).

5.2.4 Sister (stw)intronok rokon fajokban

A *Hypoxylon sp. CO27-5*-ben talált megszakító szekvenciákat rokon fajok – *Hypoxylaceae* és *Xylariaceae* – ortológ génjeiben is elemeztük. Az összesen 48 *CO27*-ben azonosított megszakító szekvencia közül 2 stwintron (HCOc076A és HCOc102A) és egy 1-es típusú cropped sister intron (HCOc047B) egyedi *CO27-5* és *EC38* gombákban. Egy „sheared” sister stwintron (HCOc046A) a géntől 29 nukleotiddal 5' irányban nem-

transzlálódó régióban található. A HCOc021A stwintron egy rövid láncú dehidrogenáz/reduktáz pszeudogénben található. A HCOc121A egy 2-es típusú sister intron a *CO27-5/EC38*-ban, viszont a többi faj esetében ugyanebben a pozícióban ez a szekvencia 3 nukleotidot helyettesít.



18. ábra Az *(stw)*intron-exon csomópontok szekvencia logója 28 sister stwintron alapján *Hypoxylon CO27-5*-ben.

Ezeknek a megszakító szekvenciáknak a határoló exon régióit megvizsgáltuk, hogy találunk-e valami mintázatot, ami alapján, bizonyos helyeken potenciálisan kialakulhatnak. Azonban az őket közvetlenül határoló (15 nt) exon régióban nem találtunk mintázatot vagy kettős kanonikus elemeket. 33 stwintronból 7 (~20 %) olyan génekben fordul elő, ahol más megszakító szekvenciák nincsenek. Kiemelkedő példa erre az oxidoreduktáz enzimet kódoló génben található HCOc153A, másrészt egy SIR2-szerű doménnel rendelkező fehérjét kódoló gén a második pozícióban egy sister stwintront (HCOc164A), a harmadik intron pozícióban pedig egy 2-es típusú cropped sister intront (HCOc164B) tartalmaz. Ez a gén így kétszer különböző intron proliferációs eseményen esett át.

A sister stwintronok és a 2-es típusú cropped sister intronok közül összesen 27 van jelen új intron pozíciókban, ha az ortológ gének génmodelljeit a *Xylariales* rendre vetítjük. 20 sister stwintron és a 2-es típusú intronok többsége *CO27-5* és *EC38* törzsekben található, azonban *Hypoxylon sp. E7406B* és *H. pulvicidum* (és a távolabbi rokon *Xylariales* fajokban) nincsenek jelen. A HCOc103A sheared sister stwintron szintén csak a *CO27-5/EC38* fajokban fordul elő, viszont a maradék 12 sheared

sisterből 9 mind a 4 fent említett *Hypoxylon* genomban jelen van (11.3. bekezdés). A három kimaradt sheared stwintron (HCOc046A, HCOc091B és HCOc304A) ortológjai pedig hiányoznak a *H. pulicicidum* és *E7406B* fajok genomjából. Ez az előfordulási mintázat arra utal, hogy a magas homológiát mutató 33 sister stwintron és a 2-es cropped sister intronok a *CO27-5/EC38* és *E7406B/H. pulicicidum* fajok szétválása után épültek be a genomba, a sheared sister stwintronok viszont már korábban megjelentek. Bár nem zárhatjuk ki, hogy a *CO27-5*-ben előforduló három cropped sister intron később elveszett az *EC38*-ból, valószínűbb, hogy ezek az intronok a két rokon szervezet szétválása után keletkeztek. Ez alátámasztja egy új sister stwintron és egy új 2-es típusú sister intron jelenlétét *EC38* két génjében, ami egy pozíció konzervált *CO27-5*-ben is jelenlévő sister stwintront is hordoz. Az *EC38* valószínűleg 770 aminosavból álló oligopeptid transzportert kódoló génjét egy HECc321A stwintron szakítja meg, ami megfelelője a *CO27-5*-ben található HCOc271A stwintronnak. Ezen felül a HECc321A stwintrontól 582 nukleotiddal 3' irányban egy új, cropped sister intront (HECc321B) tartalmaz. Hasonlóképpen egy valószínűleg 593 aminosav hosszúságú multicopper-oxidáz egy új sister stwintron szakít meg (HECc217A), ami 71 nukleotiddal 5' irányban található meg a *CO27-5*-ben (HCOc003A, HECc217B) is megtalálható stwintrontól.

Ezt követően a *CO27-5/EC38*-ban található mitokondriális szállító fehérje génjét (HCOc017A) is vizsgáltuk *Hypoxylaceae* és *Xylariaceae* fajokban. A két közelebbi rokon faj – *Hypoxylon sp. E7406B*- és *H. pulicicidum* – mellett távolabbi rokon fajokban – *Hypoxylon rubiginosum-ban* és 4 további *Daldinia-ban* (*D. eschscholzii*; *Daldinia sp. EC12*; *D. childiae*; *D. concentrica*) – is találtunk hasonló szekvenciával rendelkező

[D1,2] sister stwintronokat. A mellékletek között található 26. ábra szemlélteti a HCOc017A stwintronhoz homológ szekvenciák jelenlétét vagy hiányát rokon fajokban. 18 stwintront találtunk *D. childeae* (Dhch003A), 5-öt *H. rubiginosum* genomjában. *Daldinia* sp. EC12 esetében pedig RNA SRA adatokat gyűjtöttünk 5 feltételezett stwintronról. A *Daldinia* fajokban és *H. rubiginosum*-ban viszont cropped sister intronokat nem találtunk.

Hypoxylon sp. E7406B és *H. pulicicidum* genomjaiban 8 mindkét fajban jelen lévő stwintront találtunk. Egy sister stwintron csak az E7406B-ben (HE7c137A), egy másik sister stwintron pedig csak a *H. pulicicidum*-ban (Hpuc023A). Egy 2-es cropped sister intron mindkét fajban megtalálható – E7406B (HE7c057A) és *H. pulicicidum* (Hpuc021A) – ugyanabban a génben, ami a CO27-5 esetében a HCOc378A stwintront hordozza. Ez a gén kódolja az AzgA család egyik szállító fehérjéjét. Valójában az E7406B és/vagy *H. pulicicidum* 11 sister (stw)intronja közül egyik sem volt jelen CO27-5/EC38 fajokban. Ez erősen azt sugallja, hogy a CO27-5/EC38 taxontól való elválás után jelentek meg ezek a szekvenciák, bár lényegében ugyanabból a templátból keletkeztek.

A *Hypoxylaceae* családban 9 genomot vizsgáltunk meg, bennük csak [D1,2] típusú sister stwintronokat találtunk. A *Xylariales* fajokban azonban kimutattunk [D2,3] (51) (18) és egy [D5,6] (19) típusút is, ami a [D1,2] stwintronok duplikációjával és egy nagy intronként történő alternatív kivágódásával jöhetett létre. A stwintron kivágódása egy hatalmas kanonikus intronként történik, a külső intron donorjának G₁ nukleotidja visszamarad. Mindkét végén TIR régiók találhatóak elkerülve a kivágódáshoz szükséges elemeket. Ez megfelelő fiziológiai körülmények – másodlagos szárhurok, vagy hajtú alak – mellett bekövetkezhet (141). A

fiatalabb stwintronok RNAfold szerinti másodlagos szerkezete (HCOc004A, HCOc066A, HCOc070A és HCOc178A) hosszabban mutat szimmetriát, mint a 2-es cropped intronok, emellett SRA adatok is bizonyítják a *CO27-5* és az *EC38* esetében is, hogy a külső intron G₁-je visszamaradhat egy alternatív egylépéses splicing következtében. Továbbá *D. childiae* arilszulfatáz génjében található egy nagy kanonikus intron (Dhch003S), ami feltételezhetően egy [D1,2] stwintronból származik, a belső intron BP eleme eltűnt és a mérete kétszerese egy cropped intronéhoz képest.

5.3 Egyedi és ismétlődő stwintronok a *Hypoxylaceae* családban

5.3.1 Több, mint 100 [D1,2] stwintron felkutatása *Hypoxylon sp. CO27-5*-ben

Az előző eredmények függvényében tovább folytattuk a kutatásunkat a *Hypoxylon sp. CO27-5* genomjában. Az előzőekben talált nagy hasonlóságot mutató ismétlődő és többszörösen előforduló sister (multiple occurring; MO) [D1,2] típusú stwintronokat és kanonikus intronokat leginkább a *CO27-5/EC38* fajokban észleltük. Egy ősi sister stwintronból történő duplaszálú DNS törés következtében lezajló MMEJ javító mechanizmus hatására lehetséges intronizációs folyamatot is leírtuk, valamint bizonyítékot találtunk ezen sister megszakító szekvencia csoport proliferációjára is. Az első sister stwintront használva BLASTN segítségével könnyen rábukkantunk a többi – összesen 36 – sister stwintronra (MO). Azonban nem csak a sister [D1,2] stwintronok csoportja létezik a *CO27-5* genomjában, ugyanis a komparatív analízisük közepette két gén esetében is (HCOc091A és a HCOc121A) egyedülálló megszakító szekvenciákra (uniquely occurring; UO) is bukkantunk, amik nem

szerepeltek a korábbi BLASTN keresésünk találatai között (GenBank MW530491 és MW490721). Továbbá a humán polimeráz theta (POLQ) helikáz doménjének gomba ortológjában a DEXH-box DNS/RNS helikázt kódoló génjében is találtunk UO [D1,2] típusú szekvenciát (GenBank MW530496).

Az UO, nem-sister [D1,2] stwintronok azonosításához meg kellett vizsgálni a *CO27-5* genomját (107) és újabb kereső modelleket kellett kialakítani, hogy újabb stwintronokat találhassunk (5. ábra: *Keresőmodell A/B/C/D*). Elsődlegesen a **keresőmodell A** 282 jelöltet eredményezett, amiből 90 potenciális [D1,2] stwintront kaptunk. Ebből 59 nem szerepelt a sister stwintronok BLASTN találatai között. Ezeket az egyedülállóan előforduló stwintronokat a keresésből eredő sorszámaik alapján neveztük el. A 23 teljes sister stwintronból 21-et ez a keresőmodell felismert. A HCOc224-179 két kontigon átívelő stwintront egyik keresőmodellünk sem azonosította, továbbá egy rövidláncú dehidrogenáz/reduktáz nemrégiben pszeudogenizált génjében lévő sister stwintron (HCOc021A) sem került azonosításra, ugyanis a belső intronja nem kanonikus az 5'-donor végen lévő extra 4 nukleotid miatt (5'-GUACC₅UAUGU). Az *EC38* ortológ pszeudo génjében ez a 4 nukleotid hiányzik (5'-GUAUGU), így később ezt a stwintront azonosítottuk. 13 sheared sister stwintronból 3-at az elsődleges kereső modellünk nem észlelt, ugyanis a belső intronjuknak a donor és elágazási szekvenciájuk messzebb van egymástól (HCOc103A 131 nt; HCOc091A 82 nt), mint amit a keresőmodell A engedett (80 nt). Ezért a keresőmodellben a belső intron donor és BP elem közötti szekvenciahosszt 81-110 nukleotidra módosítottuk, így kaptuk meg a **keresőmodell B**-t (5. ábra). A keresőmodell B-vel 7 további [D1,2] UO szekvenciát találtunk,

amik elnevezése no-315-321 lett. Tovább növelve a távolságot nem kaptunk a HCOc103A-n kívül más eredményt.

Az eddig használt keresőmodell A/B sem találta meg a kimaradt sheared stwintront, a HCO263A-t, ugyanis a külső intronjának donor és BP eleme közötti távolság meghaladta az eddigi keresőmodellekben engedettet. A HCOc016B 5'-végét is felismerte a modell, ám a külső intron donor és lariat eleme közötti távolság valójában 8 nukleotiddal több a megengedettnél, a teljes stwintron hosszát pedig 58 nukleotiddal találtuk hosszabbnak kísérleti eredmények alapján. Ezek a megfigyelések arra készítettek bennünket, hogy a külső intron donorja és BP elemei közötti távolságot 25-110-ről 111-200 nukleotidra változtassuk ezzel megalkotva a **keresőmodell C-t** (5. ábra). Ez további 13 UO [D1,2] stwintron azonosítását eredményezte, amiknek a számozása no-302-314. A legnagyobb talált stwintron 313 nt hosszú volt (no-313), míg a legrövidebb 114 nt hosszú (no-061). Az elágazási pont és a hozzá tartozó 3'-akceptor (bármelyik U2 intron esetében) közötti távolság további növelése 4-20-ról 21-40 nukleotidra nem vezetett további [D1,2] stwintronok találatához.

A no-007 és a no-061 stwintronok előzetes összehasonlító elemzése azt mutatta, hogy ugyanabban a pozícióban található meg két paralóg génben, amik egy 243 és egy 250 aminosav hosszúságú rövidláncú dehidrogenázt/reduktázt kódolnak. Ezek ~54 %-ban azonosak egymással, valamint ezen kívül találtunk egy 3. paralóg gént is, ami az előzőekkel ~56 %- és ~59 %-ban hasonlít. Az utóbbi génben lévő stwintront (no-301) a keresőmodell A nem detektálta, ugyanis külső intronjának BP eleme kevésbé kanonikus 5'-C₁CTGAC, ami az eddigi keresőmodelljeinkkel inkompatibilis, ezért ismét módosítottunk a modellen. A külső intron 5'-D₁YTRAY lariat szekvenciáját 5'-C₁CTRAY-ra cseréltük ezzel

megalkotva a **keresőmodell D-t** (5. ábra). Ezzel találtunk egy újabb (no-300) [D1,2] stwintront, ami egy integrált membrán fehérje génjében található.

Az így összegyűjtött 81 db UO [D1,2] típusú stwintront – a sister stwintronok kivételével – háromféle módszerrel azonosítottuk:

1. BLASTN kereséseket végeztünk az NCBI SRA adatbázisában, amivel felállítottuk a splinter és a teljes stwintron kivágódása után maradt fuzionált szekvenciákat (SRX875229-34; (107)). Ezt mind a 117 esetben megtettük (MO (36) és UO (81) [D1,2] stwintronok), azonban ahol az adatbázisban nem szerepelt megfelelő SRA adat, ott a *CO27-5* közeli rokonját, az *EC38* adatait vettük figyelembe, mint ortológ génekben lévő stwintronokat (SRX872662-67; (107)).
2. Második megközelítésünk az RT-PCR által meghatározott köztes és teljes stwintron kivágódása után kapott transzkriptumok cDNS szekvenáltatása. A szekvenciákat GenBankba feltöltöttük (lásd: 36. oldal).
3. A stwintront tartalmazó gének 19 *Hypoxylaceae* fajban lévő ortológjait is felkutattuk TBLASTN kereséssel célzottan a WGS adatbázisokban. Komparatív elemzést követően manuálisan meghatároztuk a lehetséges intron-exon határokat, ugyanis az ortológ és paralóg génekben az intron pozíciók konzervatívak (142). Észleltünk néhány ritka esetet, amikor a donor régió 5'-GU helyett, kevésbé kanonikus 5'-GC található, ami előfordulhat belső vagy akár külső intronban is. Erre példa a no-71 ortológ *Entonaema liquescens* génjében található stwintron külső intron donor eleme.

Az összes 117 [D1,2] stwintron esetében megvizsgáltuk az ortológ és lehetséges paralóg géneket is. Három esetben – a no-007/ no-061/no-301; a

no-302/no-303; és a sheared sisterek közül a HCOc020A/HCOc263A – pozíciójukban konzervált stwintront tartalmazó paralógok. De más (nem pozíció konzervált) paralógokat is találtunk.

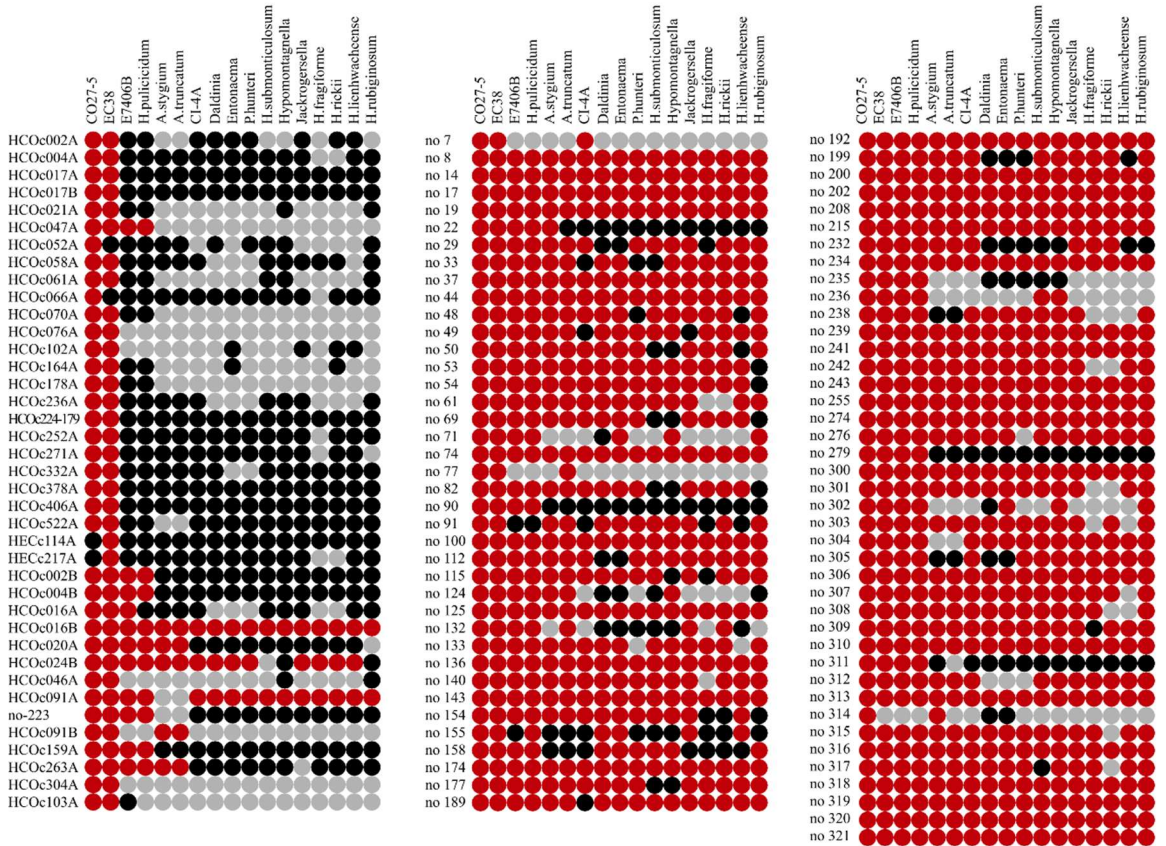
5.3.2 [D1,2] stwintronok előfordulása/megoszlása *Hypoxylaceae* családban

Jelen munkánkban bizonyítottuk, hogy több mint 1500 potenciális [D1,2] stwintron érintett a *Hypoxylaceae* családon belüli fajok között (*Sordariomycetes* osztály; *Xylariales* rend). Végre lehetőségünk adódott, hogy több, mint 100 ugyanazon típusú, egy fajon belül található stwintront statisztikai elemzés alá vegyünk, emellett eddigi kísérleteink alapján vizsgált nagy hasonlóságot mutató, terminális szimmetriával rendelkező [D1,2] sister stwintronjainkat összehasonlítsuk a jelenleg újonnan talált 81 [D1,2] egyedileg előforduló (egymáshoz sem funkcionálisan, sem szerkezetileg nem kapcsolódó) stwintronokkal.

Az első és lényeges különbség a sister- és nem-sister stwintronok között, hogy a sisterek csak *Hypoxylon sp. CO27-5*-ben és közeli rokonjában, a *Hypoxylon sp. EC38*-ban fordulnak elő, míg a nem-sister (UO) stwintronok nagy része (84 %) szinte az összes ortológ génben előfordul a *Hypoxylaceae* családon belüli vizsgált törzsek esetében. A stwintronok jelenlétét a 19. ábra piros pöttyei szemléltetik. Ez azt sugallja, hogy a 81 UO stwintron sokkal öregebb, mint a sister stwintronok, amik azután keletkeztek, hogy a *CO27-5/EC38* szétváltak a többi fajtól. A sheared sister stwintronok a két véglet között helyezkednek el. Közülük hat egy szűk taxonómiai kládra korlátozódik és a *CO27-5/EC38* és/vagy *Hypoxylon sp. E7406B/H. pulicicidum* fajok érintettek. Három sheared sister csak a *CO27-5/EC38* fajokban fordul elő, ugyanis más taxonokban nincs ortológjuk, ezzel szemben három másik sheared sister (HCOc016B,

HCOc024B és HCOc091A) konzervált pozíciókban helyezkedik el az egész családon belül.

Különös helyzetet figyeltünk meg a 2-es fázisú no-274 stwintront tartalmazó jól konzervált plazmamembrán fehérjét kódoló génben (monovalens kation: proton antiporter-1 családba tartozik), ami arra utal,



19. ábra 117 ortológ [D1,2] stwintron előfordulása ugyanazokban a génpozíciókban 17 Hypoxylaceae fajon belül. Piros pöttyök: az adott CO27-5-ben lévő stwintron előfordul más fajon belüli ortológ génjében, fekete pötty: a CO27-5-ben lévő stwintron nem fordul elő az ortológ gén azon pozíciójában, szürke pötty: az adott fajban hiányzik az ortológ gén (22).

hogy egy helyi mutáció eredményeképpen a stwintron típusa megváltozott [D1,2]-ről [D5,6]-ra. Három fajban – *H. rickii*, *H. fragiforme* és *H. lienhwacheense* – az eredetileg [D1,2] stwintron, amely a többi *Hypoxylaceae*, és más *Xylariales* fajokban is megtalálható, feltehetően

[D5,6] stwintronná alakult át. Négy nukleotid (5'-GTAA vagy 5'-GTGA) beékelődött közvetlenül a külső intron G₁ nukleotidja elé ezzel a következő felosztásokat eredményezve: 5'-GTAAG₅|T₆ vagy 5'-GTGAG₅|T₆. A folyamat során a belső intron és a határoló exonok sem voltak érintettek, ugyanis a stwintron fázisa nem változott. *H. fragiforme*-ban ez a folyamat a külső intron 5' végén a belső intron kapcsolódásánál történő 20 nukleotidos deléció eredménye lett (5'-TAG|T₂).

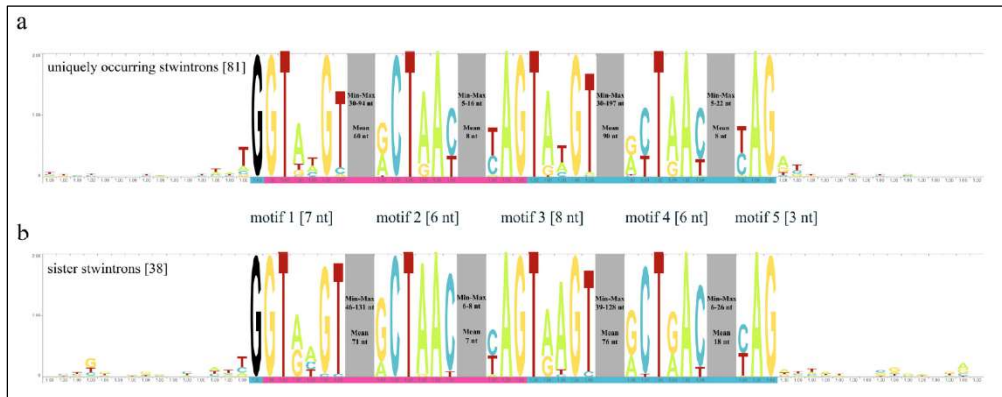
5.3.3 117 [D1,2] stwintron szekvencia elemzése

5.3.3.1 Intron-exon határok, szekvencia logo

Öt konzervált szekvencia motívumot hasonlítottunk össze az 5'- és 3' hasítási helyeken és elemeztük a 81 MO [D1,2] stwintron inszerciós helyeit. Megvizsgáltuk az exon-intron határokat úgy, hogy az exonok szekvenciáját 15 nukleotiddal kitöltük (20. ábra). A sisterek csoportját (korábban már elemeztünk az 5.2.4 bekezdésben) kontrollnak használtuk. Ahogy a sisterek esetében, az UO stwintronoknál sem találtunk szekvencia mintázatot, ami meghatározhatná az integrálódás helyeit. Nincs inszerciós mintázat a szomszédos exon régiókban sem.

Közülük az alternatívan kivágódó [D1,2]/[A2,3] stwintronok nincsenek elterjedve. A stwintronok előtti [AG] és a mögöttük elhelyezkedő [GT] szekvenciák kombinációja – kanonikus intronok kettős fúziós helyei lehetnek – sem volt megtalálható ebben a csoportban. Ezenkívül nem volt bizonyíték a stwintronok tandem hely duplikációjára valamely illesztési hely bevonásával (130). Két eset fordult elő, hogy egy [TAG] szekvencia közvetlenül egy 3' akceptoros TAG-gal végződő stwintron előtt áll, de [CAG] nem fordult elő közvetlenül a *Hypoxylon sp CO27-5* stwintronok

előtt. A [GT] közvetlenül a 3' akceptor után a 117 stwintron közül nyolcban fordult elő, a [GTA] pedig kettőben (no-279 és no-308) volt jelen.



20. ábra Szekvencia logo az exon-intron határokról és konszenzus szekvenciákról. a) 81 UO stwintron, b) 23 sister stwintron logo esetében (22)

Mind a 81 UO stwintron exon elvesztés-, vagy keletkezés nélkül integrálódott a genomba. Az elhelyezkedésük és fázisuk is konzerválódott a *Hypoxylaceae* fajokon belül, egyértelműen több 1-es fázisú stwintron van közöttük (~45 %). A két csoport stwintronjainak fázis profilja viszont hasonlóságot mutat.

5.3.3.2 Szekvencia hosszúságok vizsgálata

Egy hipotézis szerint, amit Collemare és munkatársai állítottak fel, a fonalas gombák legelterjedtebb, általános U2 intronjai (“regular spliceosomal introns” vagy RSI-k) introner-szerű elemekből származnak (azaz magas szekvenciahasonlósággal és stabil másodlagos szerkezettel rendelkező intronok). A terjedésük óta a megkettőzött intron szekvenciák a felismerhetetlenségig szétváltak. Az introner-szerű elemek degenerációja nem csak szekvencia megváltozásához vezetett, hanem a spiceoszómális U2 intronok hosszára történő lerövidüléshez is (117).

A stwintronok, és az őket alkotó intronok hosszát is megvizsgáltuk, miszerint az UO stwintronok szélesebb tartományba esnek, mint a sisterek. Ez nem meglepő, hiszen a sister stwintronok magas szintű szekvencia hasonlósággal rendelkeznek, és fiatalabbak a nem-sister stwintronoktól. Az UO stwintronok között 13 rövidebb bármelyik sister stwintronnál, viszont 10 hosszabb. A rövidebb UO stwintronok belső intronja zsugorodott, a hosszabbak esetében viszont a külső intron megnyúlt.

A stwintronok méretbeli különbsége leginkább az őket alkotó intronok 5'-donor és BP elemei közötti régióknál fontos. A lariat elem és a 3'-akceptor szekvenciák közötti távolság mindkét intron esetében elég kicsi. A belső intron lariat és akceptor közötti távolsága sister stwintronok közül 34 esetben (89%) 7 nukleotid, a nem-sister stwintronok között ez 49 esetben (60%) fordul elő. Az UO stwintronok között 11 esetben 6 nukleotid, 5 esetben 8 nukleotid volt ez a távolság. A külső intron lariat és akceptor elemei közötti távolság 6 nukleotid az UO stwintronok 62%-ában, míg 6 stwintron esetében ez 7 nukleotid és 3 másik stwintronnál pedig 5 nukleotid. A sister stwintronok külső intronjainak a BP elem és akceptor szekvenciája közötti különbség tipikusan 19 nt (25-ből 23 sister stwintron, és 13-ből 5 sheared sister stwintron esetében). 3 sheared sister stwintronnál ez a távolság 18 nt és 2 másik sheared sister stwintronnál pedig az UO stwintronokra is jellemző 6 nukleotidnyi volt a távolság.

Állatokban az U2 snRNP segítő faktor (U2AF protein) ismeri fel a pirimidin (CU) régió és a 3' vágási helyet, majd kapcsolatba lép az U2 snRNP-vel, hogy a lariat-hoz kapcsolódjon a spliceozóma összeszerelődésekor (143) (144). Mindkét alegység génje előfordul *Hypoxylon sp. CO27-5* gombában: U2AF65 [contig MDCL01000013, koordináták: c90609-887114] és U2AF35 [contig MDCL01000186,

koordináták: 104888-105517]. Általános megfigyelés alapján, ahol pirimidin régió van, az a lariat elágazási ponttól 5' irányban helyezkedik el, gyakrabban közelebb a donorhoz. Ez teljesen logikus, ugyanis a BP elem és az akceptor közötti rész az UO stwintronok nagyrésztében rövidebb, mint 8 nukleotid. A sister stwintronok külső intronjánál megfigyeltük, hogy szórványosan a pirimidin régió (CT/CU) hosszabb, mint 6 nukleotid. A sheared stwintronok két esetében (HCOc046A, HCOc263A) ez nagyobb, mint 7 nukleotid. Az UO stwintronok csoportjában a legtöbb külső intron tiszta pirimidin régióval rendelkezik a donor és a kanonikus BP szekvencia között. A no-071 és no-274 stwintronok külső intronjában a lariat és akceptor között azonosítottuk a pirimidin szakaszt (>7 nt). Ezek olyan UO stwintronok, amelyeknél ezen konszenzus szekvencia elemek közötti távolság lényegesen nagyobb (17 nt, illetve 15 nt), mint az átlagos méret.

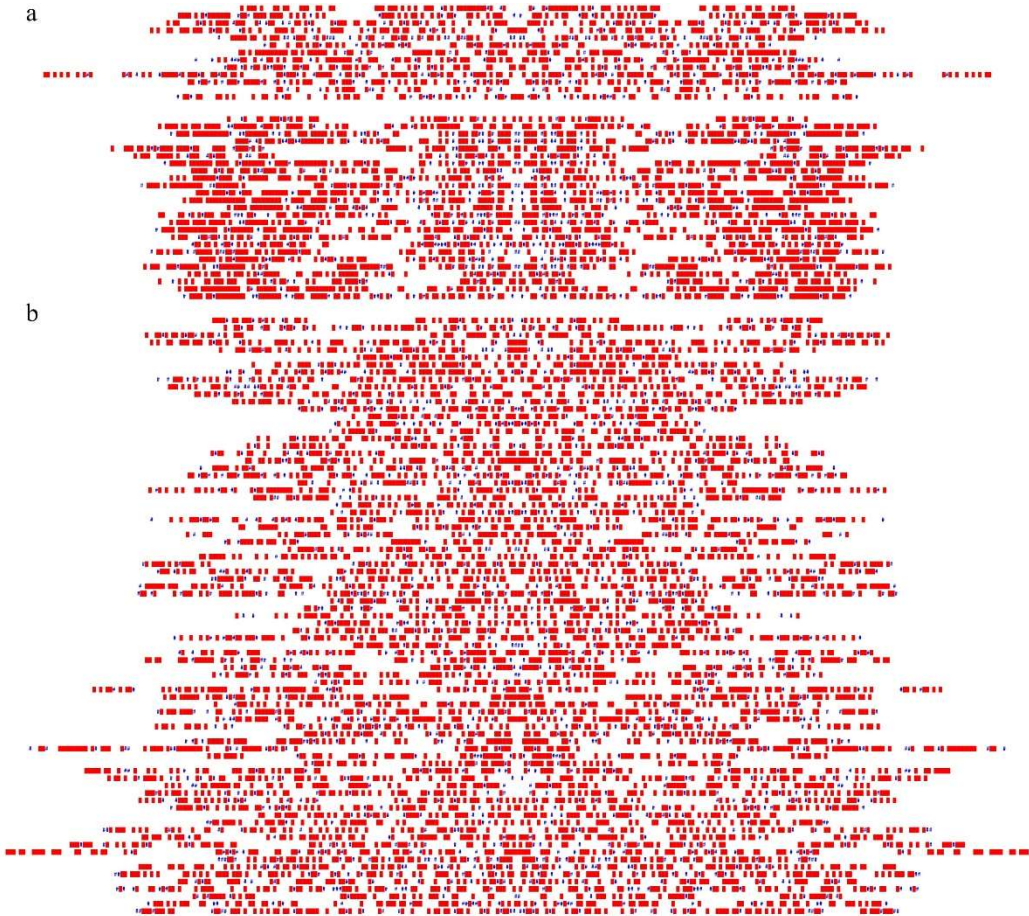
5.3.3.3 *A stwintronok AU tartalma*

Mind a 117 [D1,2] stwintron esetében elmondható, hogy az AU tartalom aránya GC-hez viszonyítva magasabb. A sister stwintronok csoportjának az % AU értéke lényegesen magasabb, átlagosan 631 %, mint a 81 UO stwintronból álló csoportnak, ahol az átlagérték 590 %. Az UO stwintronok közül csak 9 rendelkezik magasabb AU % értékkel, mint a sisterek átlagértéke. Az AU tartalom minden stwintronra külön kiszámításra került, majd azokat ezrelékre kerekítettük (%).

5.3.4 **Belső szimmetria bemutatása 117 [D1,2] stwintronban**

A kanonikus sister intronoknál és stwintronoknál rendkívüli szimmetriát figyeltünk meg a korábbiakban. A másodlagos szerkezetük előre jósolt alakjai hajtókanyarokat alkotnak, ahol a donor 5'-G₁ és az akceptor 3'-G₃ nukleotidok közel esnek egymáshoz és párosíthatlanok

maradnak. A belső szimmetria a megszakító szekvenciák saját reverz komplementerükkel végzett illesztésen alapszik. A MAFFT program által megalkotott eredmények adták a megfigyelésünk alapját, amelyet a 21. ábra prezentál.



21. ábra *A Hypoxylon sp. CO27-5* stwintronainak belső szimmetriáját ábrázoló séma. a) Legfelül a sheared sister stwintronok, majd a teljes hosszúságú sister stwintronok láthatóak, b) Az evolúciósan idősebb, 81 nem-sister stwintronok ábrázolása

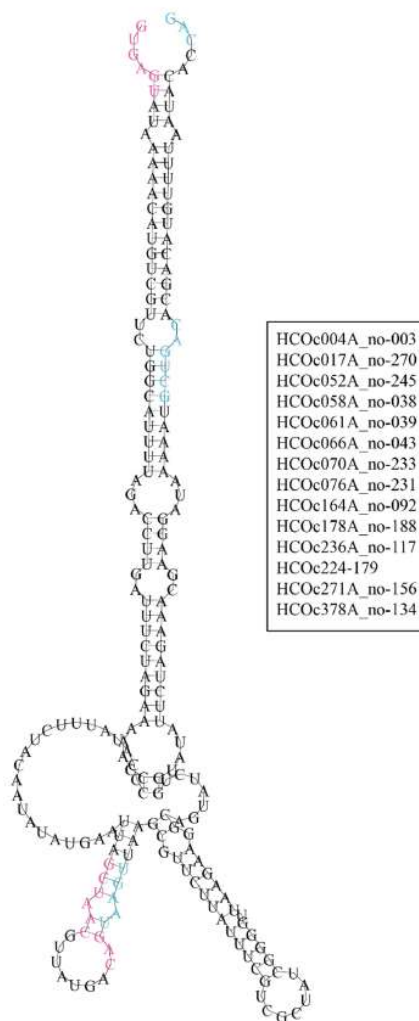
Az újonnan felfedezett 81 UO stwintronban is kivétel nélkül találtunk szimmetriai jellemzőket, melyeket több programmal is vizsgáltunk. Gyakran két szimmetriai középpel is rendelkezett egyazon stwintron, ezzel is azt mutatva, hogy többféle másodlagos szerkezetet is

felvehetnek. A stwintronok szekvenciájának saját reverz komplementerükkel végzett illesztéseit MAFFT programmal készítettük az E-INS-i modult használva. A pontozási mátrixok közül (200 PAM, 20 PAM, 1 PAM) azt választottuk a legmegfelelőbbnek, ahol minimalizálni tudtuk a réseket. A sister stwintronoknál a szimmetria a terminális részeken sokkal sűrűbb volt, aminek oka a láncvégi TIR szekvenciák. A MAFFT gyakran produkált hosszú kinyúlásokat a 3'- vagy 5' végeken, valamint hatalmas hézagokat tartalmazó igazításokat, amelyek nagy deléciós vagy inszerciós eseményekre utalnak a DNS-szinten. Két esetben (no-215 és no-319) olyan hosszú túlnyúlásokat észleltünk, amely során a szimmetriát mutató elem a stwintron külső intronjának felelt meg. Mindazonáltal legalább 30 esetben folytonos szimmetriát tapasztaltunk minimális hézagokkal. Ez a belső szimmetria viszont nem vesz részt az RNS másodlagos szerkezetének dinamikájában, valószínűleg ősi láncvégi ismétlődések maradványai okozhatják.

Ez a megfigyelésünk a *Hypoxylon sp. CO27-5* genomjában talált 117 [D1,2] stwintronról alátámasztja Collemare és munkatársainak publikációját (117), miszerint számos „reguláris spliceoszomális intron” (RSI), amely ma jelen van a fonalas gomba genomokban, valójában ősi introner elemekből, ismétlődő intron szekvenciákból származik, amelyekről azt jósolták, hogy másodlagos struktúrájuk egy jellegzetes hajtű alak, és képesek az intronok proliferációjára. De továbbra is kérdéses, hogy a *CO27-5* stwintronok belső szimmetriája és a másodlagos szerkezeteinek dinamikája mekkora mértékben és milyen szinten befolyásolhatják a splicing folyamatot és milyen esetekben válhatnak az alternatív kivágódás eseményére, ami által részt vehetnek az (stw)intronok terjedésében.

5.3.5 [D1,2] stwintronok másodlagos szerkezeteinek felépítése

Néhány évvel ezelőtt zebraadánióban (*Danio rerio*) a spliceoszómális intronoknak egy új osztályát fedezték fel, ahol megfigyelték, hogy a splicing mechanizmusát az intron másodlagos szerkezete befolyásolja (145) (146) (141). Az 5' donor végen lévő AC és a 3' terminálison lévő GU ismétlődések komplementerei egymásnak, ezáltal az egyszálú intron olyan hajtogatódásra képes, hogy a vágási régiói közel essenek egymáshoz, amivel egy speciális struktúrát alakít ki. Az RNAfold szoftver segítségével jósolták meg az optimális másodlagos szerkezeteit ezeknek a (AC)_m-(GU)_n típusú intronoknak. Állítólag a minimális szabad energiájuk (ΔG) kétszer is alacsonyabb, mint a zebraadánióban lévő hagyományos U2 típusú intronoknak. Az (AC)_m-(GU)_n intronok az U2 snRNP segéd faktor (U2AF protein) bevonása nélkül vágódnak ki, ami segítené a hagyományos intronok esetében a lariat elágazáshoz hívni az U2 snRNP-t. Ez a másodlagos szerkezet úgy írható le, mint egy meghosszabbított, bár tökéletlen hajtú, amely a G terminált párosítatlanul, de egymás közelében hagyja. A zebraadánióban talált cep97 gén 5. intronja (Accession number: NC_007112, koordináták: 13333-13503) egy jó példa erre (145). A konzervált elemek, mint az elágazási régió a kétszálú szakaszokban elfedve találhatóak. A zebraadánióban tapasztalt optimális RNS struktúrák hasonlóságot mutatnak a 2-es típusú cropped sister intronok és néhány sister stwintron RNAfold által elővetített másodlagos szerkezeteivel, különösen azokkal, amelyekben TIR (terminal inverted repeat) régiók találhatóak. A külső intron BP eleme részlegesen vagy teljesen takarásban van a spliceoszóma elől egy duplaszárú szár által, habár a TIR régióhoz komplementer dinukleotid ismétlődések (AC-GU) nincsenek sem sisterekben, sem a 81 UO stwintronok között.



22. ábra Egy konszenzus szerkezet stwintron optimális másodlagos szerkezet felépítése több szekvenciaillesztéséből. Belső intron konszenzus elemei magenta színnel, a külső intron konszenzus elemei türkizzel vannak jelölve.

Az RNAfold program szerint kigyűjtöttük mind a 117 [D1,2] stwintron minimum szabad energiáját (ΔG), hogy megbecsüljük, van-e különbség a sister (fiatalabb) és az UO (idősebb) stwintronok között. Amikor a kapott ΔG értékeket ábrázoltuk a növekvő méretre rendezett stwintronokhoz, a pontok szórtan helyezkedtek el. Ezért a sisterek mérettartományát alapul véve, 160-226 nt hosszúság között mind a két adatkészlet lineáris regresszióját alkalmaztuk, hogy meghatározzuk a trendeket az egyes stwintron csoportok esetében. A két illesztett vonal azt sugallta, hogy a minimális szabad energiák átlagosan 18–20%-kal

alacsonyabbak voltak a sister stwintronok esetében, mint az azonos mérettartományban lévő UO stwintronoknál, mivel sok egyedi kivétel is volt mindkét csoportban, ezek a ΔG értékbeli különbségek így kevésbé szembetűnők, mint a zebraadánióban lévő intron csoportok esetében, ahol a különbség 50–100% (145). A megjósolt másodlagos

szerkezetek megbízhatósága érdekében a 14 leghasonlóbb sister stwintronnal illesztést végeztünk. A másodlagos szerkezet vizsgálatát a ViennaRNA online webhely RNAalifold programjával modelleztük. A 22. ábra szemlélteti, a levezetett konszenzus szerkezet stwintron (205 nt) másodlagos szerkezetének optimális és takarékos megjósolt struktúráját (104). Ez egy szár-hurok struktúrát eredményezett, ahol a szár részt a láncvégi fordított ismétlődések alkotják. Ez a konszenzus szár 47 nt hosszú, 8 nt párosítatlan, két nem-kanonikus GU bázispárosodást tartalmaz. A duplaszálúsodott konszenzus szár elfedi a külső intron elágazási pontját. A konszenzus struktúra közepén két kis hajtű helyezkedik el egymás mellett. Az 5' irányban lévő hajtű tartalmazza a belső intron BP elemének 6 nukleotidját (5'-GCUAAC), ami párosodott a külső intron donorjának 6 nukleotidjával (5'-GUAAGU). Az elágazási A₅ és donor G₁ nukleotidja szomszédos bázispárokkal van elzárva. A belső vágási elemek egymást elzárják a spliceoszómális RNP elől. A struktúra nem tartalmaz 5-nél több egymás utáni pirimidint, de egy kisebb megszakítás nélküli pirimidin szakasz (5'-UUCUUAUUUC) van a 3' oldali hajtű régióban.

Ez az optimális konszenzus másodlagos struktúra az egy lépéses kivágódást preferálná, ami szemben áll az általunk megalkotott kétlépéses kivágódási szabállyal. Ebben a formában az egy lépéses kivágódás alternatív splicingként +1 frameshiftet okoz, ami csonka fehérje termékhez vezetne. Az így kivágódó RNS darabok az NMD (nonsense-mediated mRNS decay) (147) (148) (149) (150) folyamat által lebomlanak. Mindazonáltal találtunk erre az alternatív kivágódásra SRA adatokat NCBI-on belül sister stwintronokra és a nem sister stwintronok esetében is. Azonban úgy tűnik, hogy a kivágódáshoz szükséges kivágódási elemek az intron definíció alapján leírt módon párosítódnak és vágódnak ki, azaz a kétlépéses

kivágódás az uralkodó. Szelekciós nyomás hatására az idősebb stwintronok elterjedtebbek a fajok között és másodlagos szerkezetük nem konzervált, nem az optimális struktúrát mutatja. Ezt 9 stwintronon teszteltük (HCOc024B, no-53, no-54, no-82, no-143, no-208, no-223, no-249, no-301), melyek pozícióikban konzerváltak voltak a különböző fajokon belül. Nem találtunk bizonyítékot az optimális konszenzus másodlagos szerkezetre való konzerváltságukra sem *Hypoxylon sp. CO27-5*-ben, de a rokon fajokon belül sem.

6. ÖSSZEFOGLALÁS

A spliceoszómális intronok az eukariótákban található géneket megszakító szekvenciák. Kivágódásukat egy ribonukleoproteinekből álló komplex, a spliceoszóma katalizálja. A stwintronok (spliceoszómális ikerintronok) két spliceoszómális intronból épülnek fel, amelyek egymásba vannak ékelődve. Egy belső U2 intron beépül egy másik U2 intron valamely kanonikus kivágódási régiójába (donor, lariat vagy akceptor), ezzel ez a külső intron konszenzus régiója a spliceoszóma számára felismerhetetlenné válik, ezért a stwintronok kivágódása kétlépéses folyamat. Először a belső intron vágódik ki, mely során a külső intron megszakított konszenzus eleme helyreáll, majd ezt követően a külső intron is kivágódik.

Célkitűzéseimnek eleget téve elmondhatom, hogy a disszertációm alapját képező kísérlet sorozat sikerrel zárult, ugyanis mind a 6 tételre sikerült választ kapnunk.

Lipomyces élesztő fajok retikulum fehérjét kódoló génjeiben 3 új típusú stwintront, CIS elemet fedeztünk fel: [D4,5], [D6,7] és [D7,8]. Ezzel a megfigyeléssel nemcsak egy újfajta stwintron típust tudtunk kimutatni, amely nem a konszenzus elemek egyikébe ékelődött be, hanem egy **stwintronizációs folyamatot is felfedtünk**. Ezek az új megszakító szekvenciák részt vehettek egy stwintronizációs folyamatban, mely során az eredeti intron donorja (vagy egy része) megkettőződhetett, ezzel egy új vágási helyet kialakítva. A DSB mechanizmusnak köszönhetően, amit a cNHEJ folyamat kísér, létre jön egy új 5' vágási hely, ami a belső intron donorját fogja alkotni. A belső intron 3' végi konszenzus helyei pedig pontmutációk révén alakulhattak ki.

A. nidulans jól konzervált mitokondriális szállító fehérje génjéből kiindulva [D1,2] típusú stwintronokra bukkantunk *Hypoxylon sp. CO27-5*

és közeli rokonjainak ortológ génjeiben. A keresőmodellünket addig alakítottuk és szerkesztettük, amíg végül **több, mint 100 [D1,2] típusú stwintront találtunk a CO27-5 fajon belül és közeli rokonjaiban egyaránt**. Alaposabban megvizsgálva a stwintron halmazt, egy szembevető különbség alapján, két csoportra osztottuk őket: „sister” (többszörösen előforduló; MO) és „nem-sister” (egyedileg előforduló; UO) stwintronok csoportjára. Feltűnő volt a homológia mértéke 23 stwintron között, melyek így „sister” (sister) elnevezést kapták. Emellett találtunk még 13 [D1,2] stwintront, amik parciálisan, de magas szintű hasonlóságot mutatnak a sisterekhez, így „sheared sister” jelzöt kapták. További 12 kanonikus intronra is felfigyeltünk, amelyek szintén magas szintű homológiát mutatnak a sister stwintronokhoz. Két intron esetében, az 1-es típusú cropped sister intronok a sister stwintronok belső intronjaihoz hasonlítanak; viszont 10 intron esetében, a 2-es típusú cropped sister intronok a stwintron 5' végi belső intron szekvenciájának és a 3' végi külső intron szekvenciájának a fúziójaként véve mutat magas szintű hasonlóságot. Ez alapján a sister stwintronok/intronok csoportja egy **újfajta intronizációs folyamatot** követhet, ugyanis a stwintronokban található két palindrom szekvenciájának rekombinációja a stwintron középső részének kivágódását okozza, ezzel létrehozva egy középső részén palindromot (5'-TTTCTAGAAA) tartalmazó kanonikus intront (2-es típus). Egy 2-es típusú cropped intron (HCOc121A) esetében az intron inszerciója 3 nukleotid deléciójához köthető, ami alapján bemutattunk egy **proliferációt**. A sisterek önmaguk reverz komplementer szekvenciájukkal magas szintű hasonlóságot mutatnak, amit elősegít a terminális fordított ismétlődések régiója (terminal inverted repeat; TIR). Emellett minden sister szekvencia belső szimmetriát is mutat. Ezek a tényezők mind elősegítik az intron

beépülését, ugyanis ezek a tulajdonságok az introner-szerű elemekéivel hasonlóak. A magas szintű homológia adja a stabil másodlagos szerkezetet, mi szintén feltétele a folyamatnak, mely során a DSB, cNHEJ és reverz transzkripció folyamatok intron duplikációval kötött proliferációt okozhatnak.

A sister stwintronok elemzését követően elemeztük a 81 UO [D1,2] stwintront is. Lehetőségünk nyílt több, mint 100 ugyanazon típusú és fajon belül található stwintron statisztikai felmérésére, és a két stwintron csoport összehasonlítására. A stwintront tartalmazó géneket elemeztük és felkutattuk a közeli és távolabbi rokon fajok ortológ és paralóg génjeiket is. Ezzel több, mint 1500 stwintron volt érintett az összehasonlító elemzésben. A sister stwintronok fiatalabb szekvenciák, ezt bizonyítja, hogy a *CO27-5* mellett, csak nagyon közeli rokonjában, az *EC38*-ban fordulnak elő leginkább. Ezzel szemben a 81 UO stwintron távolabbi rokon fajokban is megjelenik. A stwintronok két csoportjában leginkább az egyes fázis dominál, inszerciós mintázatokat a határoló exonokban nem tapasztaltunk. A sister stwintronok külső intronjának BP és akceptor közötti része háromszor olyan hosszú (19 nt), mint az idősebb UO stwintronoknál (5-7 nt). Mind a 117 [D1,2] stwintron esetében elmondható, hogy az AU tartalom aránya GC-hez viszonyítva magasabb, a sister stwintronok csoportjának az % AU értéke viszont lényegesen magasabb, mint a 81 UO stwintronból álló csoportnak. Belső szimmetria ugyanúgy jellemzi mindkét csoportot, azonban a sister stwintronok a láncvégi TIR régiók miatt stabilabb másodlagos szerkezetet eredményeznek. Az RNAfold program szerint kigyűjtöttük mind a 117 [D1,2] stwintron minimum szabad energiáját (ΔG), majd összehasonlítottuk a két stwintron csoportot, ami szerint a sister stwintronok ΔG értéke átlagosan 18–20%-kal alacsonyabb,

mint az azonos mérettartományban lévő UO stwintronok értékei. Megfigyeléseink alapján néhány sister stwintron másodlagos szerkezete olyan elrendeződést mutat, ami lehetővé tehetné egy egylépéses alternatív splicing végbementelét, ami +1 nukleotidos eltolódást eredményezne, azonban az SRA adatok és kísérleti eredményeink azt mutatják, hogy többnyire a kétlépéses kivágódás a jellemző.

Nem csak a vizsgált komplex intron struktúrák újabb típusait sikerült felfedezni, de egy statisztikai felmérést is végeztünk. Elmondhatjuk, hogy a közelmúltban végzett kutatásaink és kísérleteink megtérültek, hiszen új dolgokat tapasztaltunk meg fonalas gombák és élesztők genomjában. A stwintronizáció, intronizáció és alternatív splicing eseményeivel megérthetjük az intronok proliferációját, a genomok evolúciós változását és persze a gének introntartalmának, valamint a szabályozásának lehetséges alakulását az idők folyamán. Továbbra is jó modellszervezetnek szolgálnak a gomba genomok a hasonló kísérletek elvégzésére. Biztos vagyok benne, hogy még sok dolog felfedezésre vár, azonban bízom benne, hogy jelenlegi kutatásommal hozzájárulok a további felfedezésre váró stwintronok funkcionális és szabályzó szerepének megértéséhez.

7. SUMMARY

Spliceosomal introns are gene-interrupting sequences found in eukaryotes. Their splicing is catalyzed by a complex of ribonucleoproteins, the spliceosome. Stwintrons (spliceosomal twin introns) consist of two spliceosomal introns that are nested in each other. An internal U2 intron interrupts one of the canonical splicing elements (donor, lariat or acceptor) of the external U2 intron, thereby making this consensus region of the external intron unrecognizable to the spliceosome, therefore the excision of stwintrons is a two-step process. First, the internal intron is excised, then the interrupted consensus element of the outer intron becomes intact and excisable.

Having fulfilled my objectives, I can say that the experimental series that formed the basis of my dissertation was successfully concluded, because we managed to get answers to all 6 goals:

1. Confirmation of stwintrons found in yeast species by molecular biological methods.
2. Revision and fine-tuning of our stwintron search model for [D1,2] type structures.
3. Search for genomes containing the same [D1,2] type stwintrons and verify them in laboratory conditions and by in silico analysis.
4. Analysis of the collected stwintron sequences in structure, sequentially, in symmetry and by examining the secondary structure.
5. After the data collection, reveal the possible stwintronization processes of the examined stwintrons.
6. Determination of factors affecting the proliferation of stwintrons.

In genes encoding reticulon-like proteins of *Lipomyces yeast* species, we discovered 3 new types of twintrons, CIS elements: [D4,5], [D6,7] and [D7,8]. With this observation, we did not only detect a new type of

stwintron, which was not interrupting consensus elements but we were also able to reveal a **stwintronization process**. These intervening sequences may have participated in a stwintronization process, when the donor (or a part) of the original intron could be duplicated and created a new splicing site. The DSB mechanism with the cNHEJ process form a new 5' splicing site that will become the donor of the internal intron. The consensus sites at the 3' end of the internal intron could have formed through point mutations.

In orthologous of the gene encodes a well-conserved mitochondrial transporter protein of *Aspergillus nidulans*, we found [D1,2] type stwintrons in *Hypoxylon sp. CO27-5*. We developed and edited our search model until finally **we found more than 100 [D1,2] type stwintrons both within the CO27-5 species and within its close relatives**. Taking a closer look at the stwintron set, based on a striking difference, we divided them into two groups: "sister" (multiple occurring; MO) and "non-sister" (uniquely occurring; UO) stwintrons. The degree of homology between 23 stwintrons, which we named sisters, was striking. In addition, we found 13 [D1,2] sheared sister stwintrons that have partially high level of similarity to the sisters. In two instances to be called type-1 cropped sister introns, the similarity extends to the whole internal intron; in 10 other instances, type-2 cropped sister introns consist of a fusion of 5' sequences of the internal intron with 3' sequences of the external intron. Based on this observation, the group of sister stwintrons/introns can also follow a new kind of stwintronization process, because the recombination of the two palindrome sequences in the stwintrons causes the middle part of the stwintron to be excised, thereby creating a canonical intron containing a palindrome (5'-TTTCTAGAAA) in the middle (type-2). In one type-2 cropped intron (HCOc121A) the insertion of the intron can be linked to the deletion of 3

nucleotides and we use this observation to demonstrate **proliferation of (stw)introns**. The sisters show a high level of similarity with their reverse complementary sequence, which is facilitated by the region of terminal inverted repeats (terminal inverted repeats; TIR). In addition, all sister sequences show internal symmetry. These factors all promote intron integration, as these properties are similar to those of intron-like elements. The high level of homology gives the stable secondary structure, which is also a condition for the process, during which the DSB, cNHEJ and reverse transcription processes can cause proliferation linked to intron duplication.

After the analysis of the sister twintrons, we also analyzed the 81 UO [D1,2] twintron. We had the opportunity to perform a statistical survey of more than 100 stwintrons of the same type in closely related species, and to compare the two stwintron groups. We analyzed all genes containing stwintrons and searched for their orthologous and paralogous genes in closely and more distantly related species. With this study, more than 1,500 twintrons were tested in the comparative analysis. The sister stwintrons are younger sequences, they occur mostly in its very close relative, *EC38*. In contrast, the 81 UO stwintron also appears in more distantly related species. Phase 1 dominates the two groups of twintrons, insertion patterns were not observed in the bordering exons. The part of the external intron of the sister stwintrons between BP and acceptor is three times as long (19 nt) as in the older UO stwintrons (5-7 nt). In the case of all 117 [D1,2] twintrons we can say that the ratio of AU content is higher compared to the ratio of GC, and the % AU of the group of sister stwintrons is significantly higher than that of the group of 81 UO stwintrons. Internal symmetry characterizes both groups in the same way, however, sister twintrons result in a more stable secondary structure due to the TIR regions. According to the RNAfold

program, we collected the minimum free energy (ΔG) of all 117 [D1,2] stwintrons, and then compared the two stwintron groups. The ΔG value of the sister stwintrons is on average 18–20% lower than the UO in the same size range values of stwintrons. Based on our observations, the secondary structure of some sister stwintrons shows an arrangement that could allow one-step alternative splicing that would result in a +1 frameshift. However, the SRA data and our experimental results show that two-step excision is mostly typical.

Finally, we managed not only to discover new types of the complex intron structures, but also to conduct a statistical survey. We can say that our recent research and experiments were successful. Through the events of twintronization, intronization and alternative splicing, we can understand the proliferation of introns, the evolutionary change of genomes and of course the possible evolution of the intron content and regulation of genes over time. I am sure that there is much more to be discovered, however, I am confident that my current research will contribute to the understanding of the functional and regulatory role of stwintrons that are yet to be discovered.

8. IRODALOMJEGYZÉK

1. *Molecular structure of nucleic acids.* **Watson, James D. és Crick, Francis H. C.** 1953. 04 25, Nature, 171. kötet, old.: 737-738.
2. *Genetical Implications of the Structure of Deoxyribonucleic Acid.* **Watson, James D. és Crick, Francis H. C.** 1953. 05 30, Nature, 171. kötet, old.: 964-967.
3. *An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA.* **Chow, Louise T., és mtsai.** 1977. 09 01, Cell, 12. kötet, old.: 1-8.
4. *Spliced segments at the 5' terminus of adenovirus 2 late mRNA.* **Berget, S. M., Moore, C. és Sharp, P. A.** 8, 1977. 08, Proc Natl Acad Sci U S A., 74. kötet, old.: 3171-3175.
5. *Why genes in pieces?* **Gilbert, Walter.** 1978. 02 01, Nature, 271. kötet, old.: 501.
6. *The Exon Theory of Genes.* **Gilbert, Walter.** 1987., Cold Spring Harb Symp Quant Biol., 52. kötet, old.: 901-905.
7. *The recent origins of introns.* **Palmer, Jeffrey D. és Logsdon, John M. Jr.** 4, 1991. 12, Current Opinion in Genetics & Development, 1. kötet, old.: 470-477.
8. *Origin of Genes.* **Gilbert, Walter, Souza, Sandro J. de és Long, Manyuan.** 15, 1997. 07 22, Proceedings of the National Academy of Sciences, 94. kötet, old.: 7698-7703.
9. *Self-splicing RNA and an RNA enzyme in Tetrahymena.* **Zaug, Arthur J. és Cech, Thomas R.** 4, 1987. 11, J Protozool, 34. kötet, old.: 416-417.
10. *Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena.* **Kruger, Kelly, és mtsai.** 1, 1982. 11, Cell, 31. kötet, old.: 147-157.
11. *Self-splicing of group I introns.* **Cech, Thomas R.** 1990., Annu. Rev. Biochem., 59. kötet, old.: 543-568.
12. *Self-splicing and enzymatic cleavage of RNA by a group I intervening sequence.* **Latham, John A., Zaug, Arthur J. és Cech, Thomas R.** 1990., Methods in Enzymology, 181. kötet, old.: 558-569.

13. *The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?* **Koonin, Eugene V.** 22, 2006. 08 14, *Biology Direct*, 1. kötet.
14. *Group II twintron: an intron within an intron in a chloroplast cytochrome b-559 gene.* **Copertino, Donald W. és Hallick, Richard B.** 2, 1991., *The EMBO Journal*, 10. kötet, old.: 433-442.
15. *Spliceosome twin introns in fungal nuclear transcripts.* **Flippi, Michel, és mtsai.** 2013. 01 19, *Fungal Genetics and Biology*, 57. kötet, old.: 48-57.
16. *Alternatively spliced, spliceosomal twin introns in *Helminthosporium solani*.* **Ág, Norbert, és mtsai.** 2015. 10 26, *Fungal Genetics and Biology*, 85. kötet, old.: 7-13.
17. *A mechanism for a single nucleotide intron shift.* **Fekete, Erzsébet, és mtsai.** 15, 2017. 06 08, *Nucleic Acids Research*, 45. kötet, old.: 9085-9092.
18. *Emergence and loss of spliceosomal twin introns.* **Flippi, Michel, és mtsai.** 7, 2017. 10 06, *Fungal Biology and Biotechnology*, 4. kötet.
19. *A spliceosomal twin intron (stwintron) participates in both exon skipping and evolutionary exon loss.* **Kavalecz, Napsugár, és mtsai.** 9940, 2019. 07 09, *Scientific Reports*, 9. kötet.
20. *Complex intron generation in the yeast genus *Lipomyces*.* **Ág, Norbert, és mtsai.** 6022, 2020. 04 07, *Scientific Reports*, 10. kötet.
21. *Internally Symmetrical Stwintrons and Related Canonical Introns in *Hypoxylaceae* Species.* **Fekete, Erzsébet, és mtsai.** 710, 2021. 08 29, *Journal of Fungi*, 7. kötet.
22. *Unique and Repeated Stwintrons (Spliceosomal Twin Introns) in the *Hypoxylaceae*.* **Fekete, Erzsébet, és mtsai.** 397, 2022. 04 13, *Journal of Fungi*, 8. kötet.
23. **Sharp, Philip A. és Roberts, Richard J.** Nobel Prize Outreach AB 2022. *Press release.* The Nobel Assembly at the Karolinska Institute : NobelPrize.org, 2022. 03 23.
24. *Group I introns: Moving in new directions.* **Nielsen, Henrik és Johansen, Steinar D.** 4, 2009., *RNA Biology*, 6. kötet, old.: 375-383.
25. *Nuclear group I introns in self-splicing and beyond.* **Hedberg, Annica és Johansen, Steinar D.** 17, 2013. 06 05, *Mobile DNA*, 4. kötet.

26. *Complete sequence of Euglena gracilis chloroplast DNA*. **Hallick, Richard B., és mtsai.** 15, 1993. 07 25, Nucleic Acids Research, 21. kötet, old.: 3537-3544.
27. *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*. **Christopher, D A és Hallick, R B.** 19, 1989., Nucleic Acids Res., 17. kötet, old.: 7591-608.
28. *The biology of intron gain and loss*. **Jeffares, Daniel C., Mourier, Tobias és Penny, David.** 1, 2006. 1, TRENDS in Genetics, 22. kötet.
29. *Intron evolution as a population-genetic process*. **Lynch, Michael.** 9, 2002. 04 30, PNAS, 99. kötet, old.: 6118-6123.
30. *Three distinct modes of intron dynamics in the evolution of eukaryotes*. **Liran, Carmel, és mtsai.** 7, 2007. 7, Genome Research, 17. kötet, old.: 1034-1044.
31. *Intron gain and loss in the evolution of the conserved eukaryotic recombination machinery*. **Hartung, Frank, Blattner, Frank R. és Puchta, Holger.** 23, 2002., Nucleic Acids REsearch, 30. kötet, old.: 5175-5181.
32. *Intron sliding in conserved gene families*. **Rogozin, Igor B., Lyons-Weiler, James és Koonin, Eugene V.** 10, 2000. 10 1, TRENDS in Genetics, 16. kötet, old.: 430-432.
33. *The Spliceosome*. **Lamond, Angus I.** 9, 1993., BioEssays, 15. kötet, old.: 595-603.
34. *Intron splicing: U12 spliceosomal introns not so 'minor' after all*. **Fast, Naomi M.** 14, 2021. 06 26, Current Biology, 31. kötet, old.: R891–R917.
35. *Spliceosome Structure and Function*. **Will, Cindy L. és Lührmann, Reinhard.** 3, 2011., Cold Spring Harb Perspect Biol.
36. *Introns and Splicing Elements of Five Diverse Fungi*. **Kupfer, Doris M., és mtsai.** 5, 2004. 10 3, Eukaryotic Cell, 3. kötet, old.: 1088–1100.
37. *A day in the life of the spliceosome*. **Matera, Gregory A. és Wang, Zefeng.** 2, 2014. 02, Nat Rev Mol Cell Biol, 15. kötet, old.: 108-121.
38. *The spliceosome: a ribozyme at heart?* **Valadkhan, Saba.** 2007., Journal of Biological Chemistry, 388. kötet, old.: 693–697.

39. *The spliceosome: a self-organized macromolecular machine in the nucleus?* **Rino, José és Carmo-Fonseca, Maria.** 8, 2009. 07 17, Trends in Cell Biology, 19. kötet, old.: 375-384.
40. *Alternative Isoform Regulation in Human Tissue Transcriptomes.* **Wang, Eric T., és mtsai.** 7221, 2008. 11 27, Nature, 456. kötet, old.: 470–476.
41. *Annotation of Alternatively Spliced Proteins and Transcripts with Protein-Folding Algorithms and Isoform-Level Functional Networks.* **Li, Hongdong, és mtsai.** 2017., Methods Mol Biol, 1558. kötet, old.: 415–436.
42. *Protein Diversity from Alternative Splicing: A Challenge for Bioinformatics and Post-Genome Biology.* **Black, Douglas L.** 2000., Cell, 103. kötet, old.: 367–370.
43. *Alternative splicing in ascomycetes.* **Kempken, Frank.** 97, 2013., Applied Microbiology and Biotechnology, old.: 4235-4241.
44. *How did alternative splicing evolve?* **Ast, Gil.** 10, 2004., Nat Rev Genet, 5. kötet, old.: 773-82.
45. *Comparison of intron-dependent and intron-independent gene expression.* **Buchman, A. R. és Berg, P.** 10, 1988., Mol Cell Biol, 8. kötet, old.: 4395-4405.
46. *Alternative splicing and nonsense-mediated decay modulate expression of important regulatory genes in Arabidopsis.* **Kalyana, Maria, és mtsai.** 6, 2012., Nucleic Acids Research, 40. kötet, old.: 2454–2469.
47. *Splicing promotes rapid and efficient mRNA export in mammalian cells.* **Valencia, Patricia, Dias, Anusha P. és Reed, Robin.** 9, 2008. 03 04, PNAS, 105. kötet, old.: 3386–3391.
48. *Longer First Introns Are a General Property of Eukaryotic Gene Structure.* **Bradnam, Keith R. és Korf, Ian.** 8, 2008., PLoS One, 3. kötet, old.: e3093.
49. *Proto-genes and de novo gene birth.* **Carvunis, Anne-Ruxandra, és mtsai.** 7407, 2012., Nature, 487. kötet, old.: 370-4.
50. *Introns: The Functional Benefits of Introns in Genomes.* **Jo, Bong-Seok és Choi, Sun Shim.** 4, 2015., Genomics & Informatics, 13. kötet, old.: 112-118.

51. *Convergent evolution of twintron-like configurations: One is never enough.* **Hafez, Mohamed és Hausner, Georg.** 12, 2015., RNA Biol, 12. kötet, old.: 1275-1288.
52. *A mixed group II/group III twintron in the Euglena gracilis chloroplast ribosomal protein S3 gene: evidence for intron insertion during gene evolution.* **Copertino, Donald W., Christopher, David A. és Hallick, Richard B.** 23, 1991. 12 11, Nucleic Acids Research, 19. kötet, old.: 6491-6497.
53. *Alternative splicing of the Euglena gracilis chloroplast roaA transcript.* **Jenkins, K P, Hong, L és Hallick, R B.** 6, 1995., RNA, 1. kötet, old.: 624-633. PMID: 7489521.
54. *Two new group-II twintrons in the Euglena gracilis chloroplast are absent in basally branching Euglena species.* **Thompson, M D, és mtsai.** 1, 1997., Curr. Genet, 31. kötet, old.: 89-95.
55. *Comparison of psbK operon organization and group III intron content in chloroplast genomes of 12 Euglenoid species.* **Doetsch, N A, és mtsai.** 5, 2001., Mol Gen Genet, 264. kötet, old.: 682-690.
56. *A maturase-encoding group III twintron is conserved in deeply rooted euglenoid species: are group III introns the chicken or the egg?* **Doetsch, N A, Thompson, M D és Hallick, R B.** 1, 1998., Mol Biol Evol, 15. kötet, old.: 76-86.
57. *Chloroplast group III twintron excision utilizing multiple 5'- and 3'-splice sites.* **Copertino, Donald W. , Shigeoka, Shigeru és Hallick, Richard B.** 13, 1992., The EMBO Journal, 11. kötet, old.: 5041-5050.
58. *A complex twintron is excised as four individual introns.* **Drager, Robert G. és Hallick, Richard B.** 10, 1993., Nucleic Acids Research, 21. kötet, old.: 2389-2394.
59. *Complete sequence of Euglena gracilis chloroplast DNA.* **Hallick, Richard B., és mtsai.** 15, 1993., Nucleic Acids Research, 21. kötet, old.: 3537-3544.
60. *A group III intron is formed from domains of two individual group II introns.* **Hong, Ling és Hallick, Richard B.** 13, 1994., GENES & DEVELOPMENT, 8. kötet.

61. *Twintrons are not unique to the Euglena chloroplast genome: structure and evolution of a plastome cpn60 gene from a cryptomonad.* **Maier, Uwe-G, és mtsai.** 1, 1995., Mol Gen Genet, 246. kötet, old.: 128-131.
62. *Lateral transfer of introns in the cryptophyte plastid genome.* **Khan, Hameed és Archibald, John M.** 9, 2008., Nucleic Acid Res., 36. kötet, old.: 3043-3053.
63. *Transposable group II introns in fission and budding yeast. Site-specific genomic instabilities and formation of group II IVS plDNAs.* **Schmidt, W M, és mtsai.** 2, 1994., J Mol Biol, 243. kötet, old.: 157-166.
64. *Evidence for transitional stages in the evolution of euglenid group II introns and twintrons in the Monomorpha aenigmatica plastid genome.* **Pombert, Jean-François, és mtsai.** 12, 2012., PLoS One, 7. kötet, old.: e53433.
65. *Recent mobility of plastid encoded group II introns and twintrons in five strains of the unicellular red alga Porphyridium.* **Perrineau, Marie-Mathilde, és mtsai.** 2015., PeerJ, 3. kötet, old.: e1017.
66. *Analysis of the Mitochondrial Genome in Hypomyces aurantius Reveals a Novel Twintron Complex in Fungi.* **Deng, Youjin, és mtsai.** 7, 2016., Int J Mol Sci, 17. kötet, old.: 1049.
67. *The mitochondrial genome of Endoconidiophora resinifera is intron rich.* **Zubaer, Abdullah, Wai, Alvan és Hausner, Georg.** 17591, 2018., Scientific Reports, 8. kötet.
68. *Comparison of the Mitochondrial Genome Sequences of Six Annulohyphomyces stygium Isolates Suggests Short Fragment Insertions as a Potential Factor Leading to Larger Genomic Size.* **Deng, Youjin, és mtsai.** 2079, 2018., Front Microbiol, 9. kötet.
69. *Intraspecific Mitochondrial DNA Comparison of Mycopathogen Mycogone pernicioso Provides Insight Into Mitochondrial Transfer RNA Introns.* **Deng, Youjin, és mtsai.** 4, 2021., Phytopathology, 111. kötet, old.: 639-648.
70. *Chloroplast genome expansion by intron multiplication in the basal psychrophilic euglenoid Eutreptiella pomquetensis.* **Dabbagh, Nadja, és mtsai.** 2017., PeerJ, 5. kötet, old.: e3725.

71. *In vivo* mobility of a group I twintron in nuclear ribosomal DNA of the myxomycete *Didymium iridis*. **Johansen, Steinar, és mtsai.** 4, 1997 : ismeretlen szerző, Mol. Microbiol, 24. kötet, old.: 737-745.
72. *The mtDNA rns gene landscape in the Ophiostomatales and other fungal taxa: twintrons, introns, and intron-encoded proteins.* **Hafez, Mohamed, és mtsai.** 2013., Fungal Genet Biol, 53. kötet, old.: 71-83.
73. *A homing endonuclease with a switch: characterization of a twintron encoded homing endonuclease.* **Guha, Tuhin Kumar és Hausner, Georg.** 2014., Fungal Genet Biol, 65. kötet, old.: 57-68.
74. *A functional twintron, 'zombie' twintrons and a hypermobile group II intron invading itself in plant mitochondria.* **Zumkeller, Simon, Gerke, Philipp és Knoop, Volker.** 5, 2020., Nucleic Acid Res, 48. kötet, old.: 2661-2675.
75. *Retroids in archaea: phylogeny and lateral origins.* **Rest, Joshua S. és Mindell, David P.** 7, 2003., Mol Biol Evol, 20. kötet, old.: 1134-1142.
76. *Sequential splicing of a group II twintron in the marine cyanobacterium Trichodesmium.* **Pfreundt, Ulrike és Hess, Wolfgang R.** 16829, 2015., Scientific Reports, 5. kötet.
77. *Structural Organization of S516 Group I Introns in Myxomycetes.* **Furulund , Betty M N, és mtsai.** 6, 2022., Genes (Basel), 13. kötet, old.: 944.
78. *Organellar Introns in Fungi, Algae, and Plants.* **Mukhopadhyay, Jigeesha és Hausner, Georg.** 8, 2021., Cells, 10. kötet, old.: 2001.
79. *An intronic enhancer regulates splicing of the twintron of Drosophila melanogaster prospero pre-mRNA by two different spliceosomes.* **Scamborova , Petra, Wong, Anthony és Steitz, Joan A.** 5, 2004., Mol Cell Biol, 24. kötet, old.: 1855-1869.
80. *Surprisingly high number of Twintrons in vertebrates.* **Janice, Jessin, Jąkaliski, Marcin és Makalowski, Wojciech.** 4, 2013., Biol Direct, 8. kötet.
81. *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.* **Brinkman, Kerry K., Mishra, Prakhar és Fraser, Nigel W.** 1, 2013., J Neurovirol., 19. kötet, old.: 102-108.

82. *Nested introns in an intron: evidence of multi-step splicing in a large intron of the human dystrophin pre-mRNA.* **Suzuki, Hitoshi, és mtsai.** 6, 2013., FEBS Lett., 587. kötet, old.: 555-561.
83. *Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions.* **Hatton, A R, Subramaniam, V és Lopez, A J.** 6, 1998., Mol Cell, 2. kötet, old.: 787-796.
84. *Genome-wide identification of zero nucleotide recursive splicing in Drosophila.* **Duff, Michael O, és mtsai.** 7552, 2015., Nature, 521. kötet, old.: 376-379.
85. *Subdivision of large introns in Drosophila by recursive splicing at nonexonic elements.* **Burnette, James M, és mtsai.** 2, 2005., Genetics, 170. kötet, old.: 661-674.
86. *Recursive splicing in long vertebrate genes.* **Sibley, Christopher R, és mtsai.** 7552, 2015., Nature, 521. kötet, old.: 371-375.
87. *Intrasplicing coordinates alternative first exons with alternative splicing in the protein 4.1R gene.* **Parra, Marilyn K, és mtsai.** 1, 2008., EMBO J, 27. kötet, old.: 122-131.
88. *Alternatively spliced, spliceosomal twin introns in Helminthosporium solani.* **Ág, Norbert, és mtsai.** 2015., Fungal Genetics and Biology, 85. kötet, old.: 7-13.
89. *The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences.* **Stothard, P.** 6, 2000., Biotechniques, 28. kötet, old.: 1102-1104.
90. *Basic local alignment search tool.* **Altschul, S F, és mtsai.** 3, 1990., J Mol Biol, 215. kötet.
91. *The Pfam protein families database in 2019.* **El-Gebali, Sara, és mtsai.** D1, 2019., Nucleic Acids Res, 47. kötet, old.: D427-D432.
92. *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* **Katoh, Kazutaka és Standley, Daron M.** 4, 2013., Mol Biol Evol, 30. kötet, old.: 772-780.
93. *BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence*

alignments. **Criscuolo, Alexis és Gribaldo, Simonetta**. 210, 2010., *BMC Evol Biol*, 10. kötet.

94. *SMS: Smart Model Selection in PhyML*. **Lefort, Vincent, Longueville, Jean-Emmanuel és Gascuel, Olivier**. 9, 2017., *Mol Biol Evol*, 34. kötet, old.: 2422-2424.

95. *An improved general amino acid replacement matrix*. **Le, Si Quang és Gascuel, Olivier**. 7, 2008., *Mol Biol Evol*, 25. kötet, old.: 1307-1320.

96. *Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative*. **Anisimova, Maria és Gascuel, Olivier**. 4, 2006., *Syst Biol*, 55. kötet, old.: 539-552.

97. *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. **Sievers, Fabian, és mtsai**. 539, 2011., *Mol Syst Biol*, 7. kötet.

98. *Confidence limits on phylogenies: An approach using the bootstrap*. **Felsenstein, Joseph**. 4, 1985., *Evolution*, 39. kötet, old.: 783-791.

99. *The Vienna RNA websuite*. **Gruber, Andreas R., és mtsai**. 2008., *Nucleic Acids Res*, 36. kötet, old.: W70-4.

100. *Partition function and base pairing probabilities of RNA heterodimers*. **Bernhart, Stephan H., és mtsai**. 3, 2006., *Algorithms for Mol Biol*, 1. kötet.

101. *Skyalign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models*. **Wheeler, Travis J., Clements, Jody és Finn, Robert D**. 7, 2014., *BMC Bioinform.*, 15. kötet.

102. *Mfold web server for nucleic acid folding and hybridization prediction*. **Zuker, Michael**. 13, *Nucleic Acids Research* : ismeretlen szerző, 2003., 31. kötet, old.: 3406–3415.

103. *UNAFold: software for nucleic acid folding and hybridization*. **Markham, Nicholas R. és Zuker, Michael**. 2008., *Methods Mol Biol*, 453. kötet, old.: 3-31.

104. *RNAalifold: improved consensus structure prediction for RNA alignments*. **Bernhart, Stephan H., és mtsai**. 474, 2008., *BMC Bioinformatics*, 9. kötet.

105. *Tempo and mode of genome evolution in the budding yeast subphylum.* **Shen, Xing-Xing, és mtsai.** hely nélk. : 175, 2018., Cell, 6. kötet, old.: 1533–1545.
106. *Comparative genomics of biotechnologically important yeasts.* **Riley, Robert, és mtsai.** 35, 2016., PNAS, 113. kötet, old.: 9882-9887.
107. *Characterization of four endophytic fungi as potential consolidated bioprocessing hosts for conversion of lignocellulose into advanced biofuels.* **Wu, Weihua, és mtsai.** 6, 2017., Appl Microbiol Biotechnol, 101. kötet, old.: 2603-2618.
108. *Origin of introns by 'intronization' of exonic sequences.* **Irimia, Manuel, és mtsai.** 2008., Trends Genet, old.: 378-381. 8.
109. *Spliceosome twin introns in fungal nuclear transcripts.* **Flipphi, Michel, és mtsai.** 2013., Fungal Genet Biol, 57. kötet, old.: 48-57.
110. *Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes Micromonas.* **Worden, Alexandra Z., és mtsai.** 5924, 2009., Science, 324. kötet, old.: 268-272.
111. *Extensive, recent intron gains in Daphnia populations.* **Li, Wenli, és mtsai.** 5957, 2009., Science, 326. kötet, old.: 1260-1262.
112. *Intron Invasions Trace Algal Speciation and Reveal Nearly Identical Arctic and Antarctic Micromonas Populations.* **Simmons , Melinda P., és mtsai.** hely nélk. : 32, 2015., Mol Biol Evol, 9. kötet, old.: 2219-2235.
113. *Origin and Spread of Spliceosomal Introns: Insights from the Fungal Clade Zymoseptoria.* **Wu, Baojun, Macielog , Allison I. és Hao, Weilong.** 10, 2017., Genome Biol Evol, 9. kötet, old.: 2658-2667.
114. *Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate.* **Denoëud , France, és mtsai.** 6009, 2010., Science, 330. kötet, old.: 1381-1385.
115. *Evidence for extensive recent intron transposition in closely related fungi.* **Torriani , Stefano F. F., és mtsai.** 23, 2011., Curr Biol, 21. kötet, old.: 2017-2022.
116. *Birth of new spliceosomal introns in fungi by multiplication of introner-like elements.* **Van der Burgt, Ate, és mtsai.** hely nélk. : 22, 2012., Curr Biol, 13. kötet, old.: 1260-1265.

117. *At the origin of spliceosomal introns: Is multiplication of introner-like elements the main mechanism of intron gain in fungi?* **Collemare, Jérôme, van der Burgt, Ate és Wit, Pierre J G M de.** 2, 2013., *Commun Integr Biol*, 6. kötet, old.: e23147.
118. *The complex intron landscape and massive intron invasion in a picoeukaryote provides insights into intron evolution.* **Verhelst, Bram, Van de Peer, Yves és Rouzé, Pierre.** hely nélk. : 5, 2013., 12. kötet, old.: 2393-2401.
119. *Hitchhiking selection is driving intron gain in a pathogenic fungus.* **Brunner, Patrick C., és mtsai.** 7, 2014., *Mol Biol Evol*, 31. kötet, old.: 1741-1749.
120. *Characterization of newly gained introns in Daphnia populations.* **Li, Wenli, és mtsai.** 9, 2014., *Genome Biol Evol*, 6. kötet, old.: 2218-2234.
121. *Novel Introner-Like Elements in fungi Are Involved in Parallel Gains of Spliceosomal Introns.* **Collemare, Jérôme, és mtsai.** 6, 2015., *PLoS One*, 10. kötet, old.: e0129302.
122. *Microhomology-mediated end joining: a back-up survival mechanism or dedicated pathway?* **Sfeir, Agnel és Symington, Lorraine S.** 11, 2015., *Trend in Biochemical Sciences*, 40. kötet, old.: 701-714.
123. *Microhomology-mediated end joining: new players join the team.* **Wang, Hailong és Xu, Xingzhi.** 6, 2017., *Cell Biosci.*, 7. kötet.
124. *Polymerase θ -helicase efficiently unwinds DNA and RNA-DNA hybrids.* **Ozdemir, Ahmet Y., és mtsai.** 14, 2018., *J Biol Chem*, 293. kötet, old.: 5259-5269.
125. *Molecular basis of microhomology-mediated end-joining by purified full-length Pol θ .* **Black, Samuel J., és mtsai.** 4423, 2019., *Nat. Commun.*, 10. kötet.
126. *Templated Insertions: A Smoking Gun for Polymerase Theta-Mediated End Joining.* **Schimmel, Joost, és mtsai.** 35, *Genet.* : ismeretlen szerző, 2019., rend, old.: 632-644.
127. *Non-Coding RNA Analysis Using the Rfam Database.* **Kalvari, Ioanna, és mtsai.** 1, 2018., *Curr Protoc Bioinformatics*, 62. kötet, old.: e51.
128. *Transposable elements in filamentous fungi.* **Daboussi, Marie-Josée és Capy, Pierre.** 2003., *Annu Rev Microbiol*, 57. kötet, old.: 275-299.

129. *Identifying the mechanisms of intron gain: progress and trends.* **Yenerall, Paul és Zhou, Leming.** 29, 2012., Biol. Direct, 7. kötet.
130. *Mechanism for DNA transposons to generate introns on genomic scales.* **Huff, Jason T., Zilberman, Daniel és Roy, Scott W.** 538, 2016., Nature, old.: 533–536.
131. *Evolution of the U2 Spliceosome for Processing Numerous and Highly Diverse Non-canonical Introns in the Chordate *Fritillaria borealis*.* **Henriet, Simon, és mtsai.** 19, 2019., Curr Biol, 29. kötet, old.: 3193-3199.e4.
132. *Mystery of intron gain: new data and new models.* **Roy, Scott William és Irimia, Manuel.** 2, 2009., Trends Genet, 25. kötet, old.: 67-73.
133. *Nonhomologous DNA end-joining for repair of DNA double-strand breaks.* **Pannunzio, Nicholas R., Watanabe, Go és Lieber, Michael R.** 27, 2018., J Biol Chem, 293. kötet, old.: 10512-10523.
134. *RNA in DNA repair.* **Vågbø, Cathrine Broberg és Slupphaug, Geir.** 102927, 2020., DNA Repair (Amst), 95. kötet.
135. *The roles of RNA in DNA double-strand break repair.* **Bader, Aldo S., és mtsai.** 5, 2020., Br J Cancer, 122. kötet, old.: 613-623.
136. *Jack of all trades? The versatility of RNA in DNA double-strand break repair.* **Ketley, Ruth F. és Gullerova, Monika.** 5, 2020., Essays Biochem., 64. kötet, old.: 721-735.
137. *One, No One, and One Hundred Thousand: The Many Forms of Ribonucleotides in DNA.* **Nava, Giulia Maria, és mtsai.** 5, 2020., Int. J. Mol. Sci., 21. kötet, old.: 1706.
138. *Dual loss of human POLQ and LIG4 abolishes random integration.* **Saito, Shinta, Maeda, Ryo és Adachi, Noritaka.** 16112, 2017., Nature Communications, 8. kötet.
139. *Inactivation of Pol θ and C-NHEJ eliminates off-target integration of exogenous DNA.* **Zelensky, Alex N., és mtsai.** 66, 2017., Nature Communications, 8. kötet.
140. *Low dose ionizing radiation effects on the immune system.* **Lumniczky, Katalin, és mtsai.** 106212, 2021., Environ Int, 149. kötet.
141. *The effects of structure on pre-mRNA processing and stability.* **Soemedi, Rachel, és mtsai.** 2017., Methods, 125. kötet, old.: 36-44.

142. *The function of introns.* **Chorev, Michal és Carmel, Liran.** 55, 2012., Front Genet, 3. kötet.
143. *RNA Splicing by the Spliceosome.* **Wilkinson, Max E., Charenton, Clément és Nagai, Kiyoshi.** 2020., Annual Review of Biochemistry, 89. kötet, old.: 359-388.
144. *How Is Precursor Messenger RNA Spliced by the Spliceosome?* **Wan, Ruixue, és mtsai.** 1, 2020., Annual Review of Biochemistry, 89. kötet, old.: 333-358.
145. *RNA structure replaces the need for U2AF2 in splicing.* **Lin, Chien-Ling, és mtsai.** 2016., Genome Research, 26. kötet, old.: 12-23.
146. *RNA structure in splicing: An evolutionary perspective.* **Lin, Chien-Ling, Taggart, Allison J. és Fairbrother, William G.** 9, 2016., RNA Biology, 13. kötet, old.: 766-771.
147. *Beyond quality control: The role of nonsense-mediated mRNA decay (NMD) in regulating gene expression.* **Nasif, Sofia, Contu, Lara és Mühlemann, Oliver.** 2018., Semin Cell Dev Biol, 75. kötet, old.: 78-87.
148. *How to get away with nonsense: Mechanisms and consequences of escape from nonsense-mediated RNA decay.* **Dyle, Michael C., és mtsai.** 1, 2000., Wiley Interdiscip Rev RNA, 11. kötet, old.: e1560.
149. *Nonsense-mediated mRNA decay: an intricate machinery that shapes transcriptomes.* **Lykke-Andersen, Søren és Jensen, Torben Heick.** 11, 2015., Nat Rev Mol Cell Biol, 16. kötet, old.: 665-677.
150. *Nonsense-Mediated mRNA Decay Begins Where Translation Ends.* **Karousis, Evangelos D. és Mühlemann, Oliver.** 2, 2019., Cold Spring Harb Perspect Biol, 11. kötet, old.: a032862.
151. *The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations.* **Comeron, J. M., Williford, A. és Kliman, R. M.** 1, 2008., Heredity (Edinb), 100. kötet, old.: 19-31.
152. *The reticulons: a family of proteins with diverse functions.* **Yang, Yvonne S és Strittmatter, Stephen M.** hely nélk. : 8, 2007., Genome Biol., 12. kötet, old.: 234.

9. TUDOMÁNYOS TEVÉKENYSÉG

AZ ÉRTEKEZÉS ALAPJÁT KÉPEZŐ PUBLIKÁCIÓK

Ág N, Kavalecz N, **Pénzes F**, Karaffa L, Scazzocchio C, Flippi M, Fekete E. Complex intron generation in the yeast genus *Lipomyces*. *Scientific Reports* **2020**, 10, 6022

Fekete E, **Pénzes F**, Ág N, Scazzocchio C, Flippi M, Karaffa L. Internally Symmetrical Stwintrons and Related Canonical Introns in Hypoxylaceae Species. *Journal of Fungi* **2021**, 7, 710

Fekete E, **Pénzes F**, Ág N, Ág-Rác V, Sándor E, Scazzocchio C, Flippi M, Karaffa L. Unique and Repeated Stwintrons (Spliceosomal Twin Introns) in the Hypoxylaceae. *Journal of Fungi* **2022**, 8, 397

Összesített impakt faktor: 15,827

AZ ÉRTEKEZÉS TÉMÁJÁBAN ELHANGZOTT ELŐADÁSOK ÉS

ABSZTRAKTOK

FEKETE E, PÉNZES F, ÁG-RÁCZ ÁG, ÁG N, FLIPPHI M, KARAFFA L
Position-conserved sequence-unique stwintrons, recently gained sequence-similar “sister” stwintrons and super symmetrical “daughter” introns in *Hypoxylon sp. CO27-5* (2022)

A Magyar Mikrobiológiai Társaság 2022. évi Nagygyűlése és a XV. Fermentációs Kollokvium, Aranyhomok szálló, Kecskemét
54 p. pp. 17-17., 1 p. Absztrakt

FLIPPHI M, ÁG N, ÁG-RÁCZ V, PÉNZES F, KARAFFA L, FEKETE E,
New sets of sequence-similar [D1,2] sister stwintrons in the genomes of two *Xylariaceae* species, *Xylaria sp. BCC_1067* and *Nemania abortiva* (2022)

A Magyar Mikrobiológiai Társaság 2022. évi Nagygyűlése és a XV. Fermentációs Kollokvium, Aranyhomok szálló, Kecskemét
54 p. pp. 18-18., 1 p. Absztrakt

ÁG-RÁCZ V, ÁG N, PÉNZES F, KARAFFA L, FLIPPHI M, FEKETE E
Identification and analysis of dozens of uniquely occurring [D1,2] stwintrons in the genome of the wood soft-rot fungus *Xylaria longipes* (2022)

A Magyar Mikrobiológiai Társaság 2022. évi Nagygyűlése és a XV. Fermentációs Kollokvium, Aranyhomok szálló, Kecskemét
54 p. pp. 3-3., 1 p. Absztrakt

PÉNZES F, MICHEL, FLIPPHI ; NORBERT, ÁG ; LEVENTE, KARAFFA ;
ERZSÉBET, FEKETE

Dissemination of intervening sequences of common origin in the genome / transcriptome of *Hypoxylon sp.* (2020)

A Magyar Mikrobiológiai Társaság 2020. évi Nagygyűlése és a XIV. Fermentációs Kollokvium, Budapest
28-28., 1 p. Absztrakt

PÉNZES F, ÁG N, KARAFFA L, FLIPPHI M, FEKETE E

Roles for stwintrons (spliceosomal twin introns) in alternative splicing events in *Aspergillus* (2020)

A Magyar Mikrobiológiai Társaság 2020. évi Nagygyűlése és a XIV. Fermentációs Kollokvium, Budapest
pp. 27-28., 2 p. Absztrakt

PÉNZES F

Fungális spliceoszómális iker intronok (stwintronok) szerepe alternatív splicing folyamatokban (2020)

A Magyar Mikrobiológiai Társaság 2020. évi Nagygyűlése és a XIV. Fermentációs Kollokvium, Aranyhomok Szálló, Kecskemét
előadás

PÉNZES F

Azonos eredetű komplex intronok vizsgálata *Hypoxylon* fajokban (2020)

A Magyar Mikrobiológiai Társaság 2020. évi Nagygyűlése és a XIV. Fermentációs Kollokvium, Aranyhomok Szálló, Kecskemét
előadás

Ág N, Kavalecz N, **PéNZes F**, Karaffa L, Flipphi M, Fekete E

Stwintron (spliceosomal twin intron) diversification: three types of [D] stwintron evolved at the same intron position in *Lipomyces* species (2019)

Szabó, Dóra; Kocsis, Béla; Horváth, Andrea Acta Microbiologica et Immunologica Hungarica Volume 66 Supplement 1, 209 p. pp. 115-116., 2 p.

Abstracts of the 18th International Congress of the Hungarian Society for Microbiology, Budapest
Akadémiai Kiadó

AZ ÉRTÉKEZÉS TÉMÁJÁVAL KAPCSOLATOS POSZTEREK

ÁG-RÁCZ V, ÁG N, PÉNZES F, KARAFFA L, FLIPPHI M, FEKETE E
[D1,2] típusú iker-intronok azonosítása és vizsgálata *Xylaria longipes*
gomba genomában (2022)

A Magyar Mikrobiológiai Társaság 2022. évi Nagygyűlése és a XV.
Fermentációs Kollokvium, Aranyhomok szálló, Kecskemét, FKP-1,
poszter

PÉNZES F, ÁG N, KARAFFA L, ÁG-RÁCZ V, FLIPPHI M, FEKETE E
Comparative and statistical analysis of 100 stwintrons found in an
Hypoxyton genome (2021)

6th Central European Forum for Microbiology [CEFORM] Aranyhomok
szálló, Kecskemét, MPP-31, poszter

ÁG N, KAVALÉCZ N, PÉNZES F, KARAFFA L, SCAZZOCCHIO C, FLIPPHI M,
FEKETE E

Formation of a new intron within an extant intron: how can
stwintronisation happen? (2020)

15th European Conference of Fungal Genetics, Róma, Olaszország, B2-
04, poszter

ÁG N, KAVALÉCZ N, PÉNZES F, KARAFFA L, FLIPPHI M, FEKETE E
Stwintron (spliceosomal twin intron) diversification: three types of [D]
stwintron evolved at the same intron position in *Lipomyces* species
(2019)

A Magyar Mikrobiológiai Társaság XVIII. nemzetközi nagygyűlése,
Budapest
MPP-8, poszter

10. KÖSZÖNETNYILVÁNÍTÁS

Elsősorban köszönetet szeretnék mondani **prof. Dr. Fekete Erzsébet** egyetemi tanár és témavezetőm kitartó munkájáért, aki négy éven át egyengette utam és bármilyen problémával fordulhattam hozzá.

Köszönettel tartozom **prof. Dr. Karaffa Levente** egyetemi tanár, az MTA doktora és tanszékvezetőnek, aki lehetőséget biztosított számomra, hogy a tanszéken elvégezhessem a doktori képzésemhez szükséges kutatásaim.

Hálámat fejezem ki a stwintron kutató csoport minden egyes régi és jelenlegi tagjának, hogy szakmai tudásukkal hozzásegítettek a kutató munkám sikerre viteléhez. Köszönöm **Dr. Flippi Michel** holland kutatónak, aki jótanácsokkal látott el, és **Dr. Ág Norbertnek**, aki pedig az ábrák készítésével segítette munkámat. Köszönöm már ledoktorált volt PhD társamnak és barátomnak, **Dr. Kavalecz Napsugárnak**, hogy nem csak a munkával kapcsolatos ügyekkel, de barátként is fordulhattam hozzá.

Hatalmas köszönet illeti PhD társaimat, **Márton Alexandrát** és **Bíró Vivient**, akik mindvégig segítettek és motiváltak nem csak munkatársként, de barátként is számíthattam rájuk a nehezebb időkben is.

Köszönöm **Fekete Zoltánnak**, tanszéki mérnökünknek, hogy segítségemre volt a műszaki ügyek intézésében. Valamint köszönetet szeretnék mondani minden volt, és jelenlegi tanszéki dolgozónak és szakdolgozóknak, akik bármilyen formában hozzájárultak a disszertációm elkészüléséhez.

Végezetül, de nem utolsó sorban köszönöm a szüleimnek, barátaimnak és férjemnek **Szabó Szilárdnak**, hogy mindvégig kitartottak mellettem és mindenben támogattak, még akkor is, ha éppen egy nehezebb időszakomban találtak.

Köszönöm!

11. MELLÉKLETEK

11.1 Genom adatbázisok és SRA adatok

3. táblázat Fajokhoz tartozó WGS adatbázisok és SRA adatok

Faj	WGS Master Accession (90)	SRA Accession
<i>Hypoxylon sp. CO27-5</i>	MDCL000000000 (107)	SRX875229-SRX875234
<i>Hypoxylon sp. EC38</i>	MDCK000000000 (107)	SRX872662-SRX872667
<i>Hypoxylon sp. E7406B</i>	JYCQ000000000	
<i>Hypoxylon pulicicidum</i> <i>ATCC 74245</i>	PDUJ000000000 CADCW000000000	
<i>Hypoxylon sp. CI-4A</i>	MDGY000000000	
<i>Annulohypoxylon stygium</i> <i>MG137</i>	PYLT000000000 QLPL000000000	
<i>Hypoxylon rubiginosum</i> <i>MUCL 52887</i>	CADCXA000000000	
<i>Daldinia sp. EC12</i>	MDGZ000000000	SRX872671-SRX872676
<i>Daldinia eschscholzii</i> <i>IFB-TL01</i>	AKGB000000000	
<i>Daldinia childiae</i> <i>JS-1345</i>	VYXO000000000	
<i>Daldinia concentrica</i> <i>CBS 113277</i>	CADCSW000000000	
<i>Entonaema liquescens</i> <i>ATCC 46302</i>	CADCSX000000000	
<i>Hypomontagnella monticulosa</i> <i>MUCL 54604</i>	CADCWR000000000	
<i>Hypoxylon fragiforme</i> <i>MUCL 51264</i>	CADCWU000000000	
<i>Hypomontagnella submonticulosa</i> (<i>Hypoxylon submonticulosum</i>)	CADCWV000000000	

<i>Hypoxylon lienhwacheense MFLUCC 14-1231</i>	CADCWW000000000	
<i>Annulohypoxylon truncatum CBS 140778</i>	CADCSV000000000	
<i>Pyrenopolyporus hunteri MUCL 49339</i>	CADCXC000000000	
<i>Hypoxylon rickii MUCL 53309</i>	CADCWY000000000	
<i>Jackrogsella multififormis CBS 119016</i>	CADCXD000000000	
<i>Lipomyces starkeyi</i>	LSGR00000000.120	MN689085- MN689088
<i>Lipomyces mesembrius</i>	PPJS00000000.219	
<i>Lipomyces arxii</i>	PPJT00000000.219	
<i>Lipomyces kononenkoae</i>	PPJW00000000.119	MN689089- MN689090
<i>Lipomyces lipofer</i>	PPJX00000000.119	MN689083-MN689084
<i>Lipomyces doorenjongii</i>	PPJU00000000.119	
<i>Lipomyces suomiensis</i>	PPJQ00000000.219	MN689081-MN689082
<i>Lipomyces oligophaga</i>	PPJR00000000.219	
<i>Lipomyces japonicus</i>	PPJV00000000.119	

11.2 Felhasznált primerek

11.2.1 *Lipomyces* fajokhoz használt primerek

4. táblázat Splicing intermedierek (splinter) kimutatásához használt primerek a *Lipomyces* fajokban

Fajok	Primer	Szekvencia (5'-)
<i>Lipomyces lipofer</i> NCAIM Y.00351	L.lipofer_F2	CTGCCGATTGATCCAACGTAC
<i>Lipomyces lipofer</i> NCAIM Y.00351	L.lipofer_R2	ACAGTTAGTATTAGAAGTTAC
<i>Lipomyces suomiensis</i> CBS 7251	L.suomi_F1	TTCCACGGTCTGATCCTACT
<i>Lipomyces suomiensis</i> CBS 7251	L.suomi_R1	ATCAGACGGAGCTATAGCAT
<i>Lipomyces starkeyi</i> CCY 33-1-1	L.starkeyi_R1	ATGATCCAGTCAGCCCTCATC
<i>Lipomyces starkeyi</i> CCY 33-1-1	L.starkeyi_seq_F1	CACAACACGAACCCAAAACATC
<i>Lipomyces kononenkoae</i> CBS 2514	L.kono_F1	GTCTGCCTGTAGATCGTATC
<i>Lipomyces kononenkoae</i>	L.kono_R2	CGTCAGCTCCTTATTCGGAT

CBS 2514		
----------	--	--

5. táblázat Felhasznált primerek a cDNS/gDNS szekvenáláshoz a *Lipomyces* fajokban

Fajok	Primer	Szekvencia (5'-)
<i>Lipomyces lipofer</i> NCAIM Y.00351	L.lipofer_F2	CTGCCGATTGATCCAACGTAC
<i>Lipomyces lipofer</i> NCAIM Y.00351	L.lipofer_R3	CACCAGATACGTCAATCATCC
<i>Lipomyces suomiensis</i> CBS 7251	L.suomi_F1	TTCCACGGTCTGATCCTACT
<i>Lipomyces suomiensis</i> CBS 7251	L.suomi_R3	GTGTATTTTCATGTCGTTACGC
<i>Lipomyces starkeyi</i> CCY 33-1-1	L.starkeyi_seq_F1	CACAACACGAACCCAAAACATC
<i>Lipomyces starkeyi</i> CCY 33-1-1	L.starkeyi_seq_R1	ATACCGCAACGACACCATCATC
<i>Lipomyces kononenkoae</i> CBS 2514	L.kono_F1	GTCTGCCTGTAGATCGTATC
<i>Lipomyces kononenkoae</i> CBS 2514	L.kono_R1	GGTTAGCATTAGGTATGGCT

11.2.2 *Hypoxylon* sp. CO27-5 sister (stw)intronokhoz használt primerek

6. táblázat Splicing intermedier kimutatására alkalmazott primerek a sister (stw)intronokhoz

<i>Hypoxylon</i> sp. CO27-5 gének	Primer	Szekvencia (5'-)
c004A	c004A_splinter_F	ATGGCCAAAACGAACAATCC
c004A	c004A_splinter_R	GTGTATTAAGACATGTCGTG
c017A	c017A_splinter_F	GACTGCAGATTCAGTACTCG
c017A	c017A_splinter_R	AGTATTAAAACATGTCGCGTC
c017B	c017B_splinter_F	GTACTIONGGGATATCAGGATC
c017B	c017B_splinter_R	TAAAACACGTCGTGTCAGTAC
c052A	c052A_splinter_F	CATTTCCCTACCCCTGGAGAAG
c052A	c052A_splinter_R	TAAAACATGTCGTATTAGCAT
c061A	c061A_splinter_F	GACGTCGAGGTTTACCTAAC
c061A	c061A_splinter_R	GTATTAAAACATATCGTGTTA
c043A	c043A_splinter_F	GAAATATTGCGACAGACATC
c043A	c043A_splinter_R	GTATTAGAACATGTCGTGTC
c070A	c070A_splinter_F	CAATACCATTATGTGGCGATG
c070A	c070A_splinter_R	GTGTTAGCATTTCATCCTTC
c076A	c076A_splinter_F	GCATGGTCCGATAATGAAAAA
c076A	c076A_splinter_R	CCTAATCCCGATAACGACGA

c102A	c102A_splinter_F	ATGGTATTTACTGTAATATAT
c102A	c102A_splinter_R	GTGTATTAAAACATGTCGTGTC
c164A	c164A_splinter_F	GTTCACTACATCGAAGACCAG
c164A	c164A_splinter_R	GTAGTGTC AATATCTATAGTC
c178A	c178A_splinter_F	AAAAACATGTCATTCTGGCATT
c178A	c178A_splinter_R	TCGTCCCAACCTAAACATCC
duf636	duf636_splinter_F	CCTAGGTAGCTATATGCTGT
duf636	duf636_splinter_R	GCATGTTGGACAGACTGATG
c236A	c236A_splinter_F	GCTCTCCACAGTCGAGTCTCA
c236A	c236A_splinter_R	TAACCTTGTTAGCATTTTAACAC
c252A	c252A_splinter_F	CAGGTCAAGAGGCTTCATCAC
c252A	c252A_splinter_R	AGGATATTACAACATATCGTG
c271A	c271A_splinter_F	GCACTCGTACGGTCTTCTCC
c271A	c271A_splinter_R	GCGTCGTGTCAGCATTTTTTA
c332A	c332A_splinter_F	GCATATTTTGCGCCACTTCT
c332A	c332A_splinter_R	CATTCTTGGCCTTTGCTTCT
c378A	c378A_splinter_F	AGGTTTGCTGTAGATGAGAG
c378A	c378A_splinter_R	ATTAAAACATGTCGTGTCAG
c406A	c406A_splinter_F	ATCCTCAATGCGACGAGCGG
c406A	c406A_splinter_R	GCATGTCGTGTCAGCATTGT
c522A	c522A_splinter_F	ATGAAGTGGTATACTCTACTT
c522A	c522A_splinter_R	TCCTGTGTCAGCAATTTTGACTTC
c004B	c004B_splinter_F	AAGGACTGTGCGCCATGAT
c004B	c004B_splinter_R	TGTATGAAACCGGTCATATCAGT
c016A	c016A_splinter_F	GAGACCTGCCGGAACGTTGT
c016A	c016A_splinter_R	GTTTACCTTGGACCGTTTGAG
c016B	c016B_splinter_F	GTCATTGATCTGACCTCCAT
c016B	c016B_splinter_R	TGTTAGCTCTCTCCAGCTGA
c020A	c020A_splinter_F	GATCCCATGACACGAACAGG
c020A	c020A_splinter_R	CCGTGTCAGTACTTCTAACTCC
c024B	c024B_splinter_F	GTCGCGATTAAGCTGCAGAAG
c024B	c024B_splinter_R	TCATTGAGCCCTTAATCCGTA
c046A	c046A_splinter_F	CAGAAGCATTCAAAAAGGAAC
c046A	c046A_splinter_R	GGAAAACATCTTGTGTCAGC
c091A	c091A_splinter_F2	CGCCTCAGCGGAGCAGATCTA
c091A	c091A_splinter_R2	GTTGTCAGTAATCTTATAAC
c091A	c091A_splinter_F1	ACGGCGAGCATTCCATCGTTC
c091A	c091A_splinter_R1	AGGTATGTATTAGTGTGCGGTG
c103A	c103A_splinter_F	ATCCTGCGGATGCATCTAAC
c103A	c103A_splinter_R	TATACCAGGCGGTGCTTCT
c159A	c159A_splinter_F	CGATACTAGGCTTCATAAATAC
c159A	c159A_splinter_R	ATACCAGAGTATATCATGTCA
c263A	c263A_splinter_F	TCGCTTTCCCAGACTAAGA
c263A	c263A_splinter_R	TAGGGGTGACCATTCCACAT
c304A	c304A_splinter_F	GAGATACATTGTCCCACCAAC
c304A	c304A_splinter_R	CAGCAATCAGCCTAGCCATG
helQ	helQ_splinter_F	GCACATCTATAGCACATTCT

helQ	helQ splinter R	CTAAATGTTAGCTTACCTGA
------	-----------------	----------------------

7. táblázat Felhasznált primerek a gDNS/cDNS szekvenáláshoz a *sister (stw)* intronokhoz

<i>Hypoxyton sp.</i> CO27-5 gének	Primer	Szekvencia (5'-)
c047B (I)	c047B seq F	ACCGGCGCTAATTTCTAAT
c047B (I)	c047B seq R	GCATCGTATCCCACGCTATT
c153A (II)	c153A seq F	GCTAGCAACAGAAAAGCAAC
c153A (II)	c153A seq R	AGTTGCATGTGTGGTTATGA
c229A (I)	c229A seq F	TGACCGAACAACAGCAACAG
c229A (I)	c229A seq R	GTCTATCGGCATACCTCGCT
c105A (II)	c105A seq F	TTACTAGCCGGCGGATAAGG
c105A (II)	c105A seq R	ATTACATTGCCCCGGATCCA
c164B (II)	c164B seq F	TATGCCCAGAGAGCCATTGT
c164B (II)	c164B seq R	TCACCCGAATTGTCTCTTGA
c343A (II)	c343A seq F	AGAGGGGCAAACGTCATCTA
c343A (II)	c343A seq R	CGCTGAGTACATGCCTACCT
c024A (II)	c024A seq F	TCGAGAGTCGCGCCGGAATC
c024A (II)	c024A seq R	TAAGTCTGCCAGGAGATCGC
c147A (II)	c147A seq F	ATCCTCAAGCCGCCTCATAA
c147A (II)	c147A seq R	GCATCACAACATAATTGCACGA
c121A (II)	c121A seq F	AAGGAGAAGGAACCAGAGCC
c121A (II)	c121A seq R	CGCAACGCCTCTCATCTAAC
c004A	c004A seq F	CATCGTTCTTATGTAATTAT
c004A	c004A seq R	AGAGCTGTAATAGCTATCCT
c017A	c017A seq F	GTACCCCAATGTAGTATACG
c017A	c017A seq R	CAGATCGCATAACTCGATGA
c017B	c017B seq F	GAACAGTGTCTTCGTCAATCT
c017B	c017B seq R	GTTACCGCAAGATTGTTAGAC
c021A	c021A seq F	ATGATACAGAATCTCTAGTCG
c021A	c021A seq R	ACATGTTGCTTCCACTGTTTC
c052A	c052A seq F	CATTTCTACCCTTGGAGAAG
c052A	c052A seq R	AGGCCATCATACTCATCCATC
c061A	c061A seq F	GACGTCGAGGTTTACCTAAC
c061A	c061A seq R	ATTTACGGTATATACCATTTC
c066A	c066A seq F	CTTGCTAGAGACACTCTCTA
c066A	c066A seq R	TCCCTAGGTTCCAAGGAGAC
c070A	c070A seq F	GATGAGGTGGTGCAGTTGCG
c070A	c070A seq R	TCAGAAACTCATCGACTACAG
c076A	c076A seq F	CGACGTCAGCATCCCTACTA
c076A	c076A seq R	AGGTTTCCCACGTCAAACAG
c102A	c102A seq F	CCTAGGTATTTACACACCATG
c102A	c102A seq R	TCCCTCTCATGTAGCTCAAG
c164A	c164A seq F	CAATTCTTTACGATATAGATA
c164A	c164A seq R	CATTATTTAGACATTGCAGCAG

c178A	c178A seq F	ACAGTCAACTTATAAAAAGCTC
c178A	c178A seq R	CTAACTGTAACCTCCTATCGG
c236A	c236A seq F	GCTCTCCACAGTCGAGTCTCA
c236A	c236A seq R	CTAGCGACGTAATATAACCGA
duf636	duf636 seq F	GACTTACTACTCGGAACCTCC
duf636	duf636 seq R	GTACTTGGATAAAAACCTTT
c252A	c252A seq F	CAGGTCAAGAGGCTTCATCAC
c252A	c252A seq R	TGCTCCAGCGTATAATAGATG
c271A	c271A seq F	GCTCTGCGTATGTTGTGGAA
c271A	c271A seq R	AGAGAATGGGGAGAAGCACA
c332A	c332A seq F	GCATATTTTTGCGCCACTTCT
c332A	c332A seq R	TTCTAACACAGCGGGAGTCA
c378A	c378A seq F	AGGTTTGCTGTAGATGAGAG
c378A	c378A seq R	AAGTACCACACAGAGATATT
c406A	c406A seq F	GAGATTGCACCGGCTCTCAC
c406A	c406A seq R	CTTGCCGCAAAACTACCAAC
c522A	c522A seq F	GAAGACGTTCCCTAGTGCTA
c522A	c522A seq R	GACATATCTCTCAGGACGGA
c004B	c004B seq F	AAGGACTGTGCGCCATGAT
c004B	c004B seq R	CAAGAACTCCTAGGCCATCG
c016A	c016A seq F	CATCGGATAAAGCTCGAGGCCCT
c016A	c016A seq R	CACGGGCAGCACCCAATTTAAC
c016B	c016B seq F	ACTTGCTATGTCCAAGCTTGT
c016B	c016B seq R	TACATTCTAGATTACGTATC
c020A	c020A seq F	TCCAACTCCACCTCGTCTTC
c020A	c020A seq R	AATGTTGGTAGCGATGAGGC
c024B	c024B seq F	TCATTGATCTTCGCTCCGTA
c024B	c024B seq R	TAGATGTATCTCTAGATCTA
c046A	c046A seq F	CAGAAGCATTCCAAAAAGGAAC
c046A	c046A seq R	CTGGACAAAACACCATTGCTC
c091A	c091A seq F	GTAAACATCTTGTGTCAAAAC
c091A	c091A seq R	TAGGTACGTTAATTATATCAT
c103A	c103A seq F	GCATGACTGAAAGCCTCACA
c103A	c103A seq R	CGCAGGCAGAAAATGGACTA
c159A	c159A seq F	GTTCGTCTTTCGATAGTTCGT
c159A	c159A seq R	GTGCAAGGTGAACACTTGAAC
c263A	c263A seq F	CAACGGTTCTAGCCTGAGGA
c263A	c263A seq R	ATTCGCCCTTTTGAGATGAA
c304A	c304A seq F	GAGAACTTAAATAGTTCGATAT
c304A	c304A seq R	CGACGTATCTCTAGAATATTTTC
helQ	helQ seq F	GCATCTTGCCAGCGTCTATC
helQ	helQ seq R	CACAAGGCAAGGTACAGAAG

11.2.3 *Hypoxylon sp. CO27-5* egyedi/ismétlődő stwintronjaihoz felhasznált primerek

8. táblázat *Splicing intermedier* kimutatására használt primerek egyedi/ismétlődő stwintronokhoz

<i>Hypoxylon sp. CO27-5</i> gén	Primer	Szekvencia (5'-)
No 8	No8_splinterF	TGGCGAACATCTTGGTTGCT
No 8	No8_splinterR	AATTCGCCCCCTCGAATAAT
No 37	No37_splinterF	GACCGCTGATCCTTTTCGTC
No 37	No37_splinterR	AGCAGTGAAATGAGGTAGGACA
No 50	No50_splinterF	GGCAAGAGACACGGATCAAG
No 50	No50_splinterR	AGCAAGACTAGCACAGGAGA
No 69	No69_splinterF	ACCCATCGTCCTTCTCTCG
No 69	No69_splinterR	GCCTAGGAAAGAAATGAAGGGG
No 71	No71_splinterF	AGACCTTGTGCTGGATCGTT
No 71	No71_splinterR	GGGTAGAAAAGATGCAACGTCA
No 77	No77_splinterF	ACTTTCGTTTCGAGGGGTTTT
No 77	No77_splinterR	TTGACCACGAGAACAAGAGC
No 82	No82_splinterF	TCAAGTTGGATGCGGTGATC
No 82	No82_splinterR	TGCAAGACGGGAAGTAGAGT
No 90	No90_splinterF	TATGTTATCTGCGGTTACAC
No 90	No90_splinterR	CAGAGTATATCATGTCAGTAT
No 100	No100_splinterF	ATCGGAGTTCCTGTTTGCT
No 100	No100_splinterR	GACCCTAAAACCTCCATGTCGG
No 115	No115_splinterF	GGAGGCCGAGCTCAAAGATA
No 115	No115_splinterR	GCCAAAGTTGAGAAGAAGGGA
No 132	No132_splinterF	TTGCGAAAGCCGTCGAGCTC
No 132	No132_splinterR	CTGGACATCAGCATTTGGAGTTC
No 133	No133_splinterF	CGACTGACCAATTAAGGCC
No 133	No133_splinterR	GTATGATTAGCAACCCTCAT
No 140	No140_splinterF	GGGGAATGCTCACCATTGTT
No 140	No140_splinterR	ATTAGCGAAGAACGCGAAAG
No 155	No155_splinterF	CGCATCGACATTCCGGAATG
No 155	No155_splinterR	TTCGTGCCAAGATGCGTATG
No 174	No174_splinterF	TCCAACCTTCTACCCCGTCAC
No 174	No174_splinterR	GCTCCTCCTTCTCTTTGGCT
No 177	No177_splinterF	CTCTTGGTCAGCCCAGGTAA

No 177	No177_splinterR	CCGAATGAAAATCGGTGGTA
No 189	No189_splinterF	AAGATTCGAGCTTGGGTCCA
No 189	No189_splinterR	CGCCTTTGATGATTTGACATGT
No 208	No208_splinterF	CTGGTGGTTATGGCTGGAGA
No 208	No208_splinterR	AACCCAAATCATACTCCCCTT
No 215	No215_splinterF	CATAAACGCCAGATCCGCC
No 215	No215_splinterR	TTCCGATTAAGCCCCTGTGT
No 243	No243_splinterF	TGAAAGCGCTCAATGATCCG
No 243	No243_splinterR	TTCGTTAGTAAGCGTCTCATGTG
No 279	No279_splinterF	TCCCAGTTCTTCTGACGGAC
No 279	No279_splinterR	GAAAGCTCGTAGTTGTCCCC
No 301	No301_splinterF	CGCCTCGGTTGTTCAGAAA
No 301	No301_splinterR	GCCGTTTAGTCAGGTTTGCT
No 302	No302_splinterF	AGGCAGCCAGGATTAATGTC
No 302	No302_splinterR	TGGGAAAGAGGAAGCGTAGA
No 303	No303_splinterF	CCCTACCGAACCAGATAGTGT
No 303	No303_splinterR	AGATCAACACTACGGGAGGA
No 306	No306_splinterF	CCCGCTTTCTAGTTCCTAATC
No 306	No306_splinterR	GAGACGCAGGTGTTTGGATC
No 311	No311_splinterF	AACTACGGTACCCTGGTTGG
No 311	No311_splinterR	TTTTCCCTTCACGGCGCA

9. táblázat Felhasznált primerek a gDNS/cDNS szekvenáláshoz az egyedi/ismétlődő stwintronokhoz

<i>Hypoxyton sp. CO27-5 gén</i>	Primer	Szekvencia (5'-)
No 8	No8_seqF1	CGACCATTCGATATTCACGA
No 8	No8_seqR1	AATGCTAGTCGAGGCACGAG
No 8	No8_seqF2	AGATCAGTTCGCGACAATCC
No 8	No8_seqR2	AGAAACCAATGCCGACAGAC
No 37	No37_seqF1	GACCGCTGATCCTTTTCGTC
No 37	No37_seqR1	TTCCCAAATCCGAACCTCCA
No 37	No37_seqF2	CCTGAAGAGAGTGTCATACCG
No 37	No37_seqR2	AGATAGCTCAGGGTTTCGCAT
No 50	No50_seqF	GGACGCTGGACCAATTCTTC
No 50	No50_seqR	CAGCGGAGATCAGGTACCTA
No 61	No61_seqF	CGACACTTTGTGGACGGATA

No 61	No61_seqR	AGACCCAGCATCCTTAGCAA
No 69	No69_seqF1	CGAATTGGCTTCTTCGTAC
No 69	No69_seqR1	TGCCGAAGAAATGGGGAATG
No 69	No69_seqF2	ACCCATCGTCCTTCTCTCG
No 69	No69_seqR2	TCAAATGTACAACCACTGGCT
No 77	No77_seqF1	TTGGAACGAGGTACCAACAA
No 77	No77_seqR1	CTCGATATCCTGGGCCATAA
No 77	No77_seqF2	CGACGAGTCGTTGCTAAACA
No 77	No77_seqR2	CGTAGAAGCCTCTGGGACTG
No 82	No82_seqF	TTTTCTGCCCAGTTCTCA
No 82	No82_seqR	AACAGGTCTTGGCAAGTGC
No 90	No90_seqF1	CCATGTAAGTTAATCAATCAG
No 90	No90_seqR1	TGTGTAACCGCAGATAACATA
No 90	No90_seqF2	TATGTTATCTGCGGTTACACA
No 90	No90_seqR2	CCGGCTAATAGTGCAAGGTGA
No 100	No100_seqF	AGGTAGCAGCCTGATCACGT
No 100	No100_seqR	TCTAACATGCCCCGAGTCGAGC
No 132	No132_seqF1	CTCCCAATTTTCGGAAAGACGT
No 132	No132_seqR1	GTACCTATGCAAAGTAATTAC
No 132	No132_seqF2	GTATGCGTTATTGGGTGGCC
No 132	No132_seqR2	ACCTTCGAGCTCTATGGTT
No 133	No133_seqF	ATCCGAGCGATCTCGCCGATC
No 133	No133_seqR	ATAGTTTGAAGCAACTGGGAG
No 140	No140_seqF1	GTGCTGTTTTGCGAAGGTCT
No 140	No140_seqR1	CAGCAGAGCCGACACATAAG
No 140	No140_seqF2	GCCTCCAGATTCCGAAC
No 140	No140_seqR2	AAGCAAAGATACCTCCCGTA
No 155	No155_seqF1	AGCTCAACATCCCATCACCA
No 155	No155_seqR1	TGTGTAAGTCTGCGTCGAGA
No 155	No155_seqF2	GAGGAGGACAAGGAACCACA
No 155	No155_seqR2	CCGCACACTCCTGATAAAGC
No 177	No177_seqF	CTCTTGGTCAGCCCAGGTAA
No 177	No177_seqR	GCGTTAGTTACGATGCGAAAT
No 208	No208_seqF1	TGCAGAAGATCAGAGGTCGT
No 208	No208_seqR1	CCGTCGGTCATAGGATGTCT
No 208	No208_seqF2	ATGTCTCGGTACTTTTGC GG

No 208	No208_seqR2	TTCATGTATCGCGTTTGGGC
No 215	No215_seqF	CATAAACGCCAGATCCGCC
No 215	No215_seqR	TGGCTTCCTCAACGGTACAT
No 243	No243_seqF	AATACCTCATTCTGCTGGGC
No 243	No243_seqR	CGCTCCTTACGTTCTGTCTCT
No 279	No279_seqF	CCATTCAAGCTGCCCAATCA
No 279	No279_seqR	TCCTACAGCGCTAAACCACT
No 303	No303_seqF1	CCCTACCGAACCCAGATAGTGT
No 303	No303_seqR1	AGTGGCGGTGACAGAATAT
No 303	No303_seqF2	GAGGAATTCGTTGGTGCTCC
No 303	No303_seqR2	TCATCCGTGGAAAGTAGAGCA
No 304	No304_seqF	GGCAACGAAACATAAGTGACCT
No 304	No304_seqR	AGGTGCCTCTATATGTGGGG
No 306	No306_seqF	CCCCTTTCTAGTTCCTAATC
No 306	No306_seqR	AGAAATGCTTTTATCGCCGAC
No 307	No307_seqF1	TCGCTGAAATGCTCCCAGAT
No 307	No307_seqR1	TTAGAACCCCGTACCAGCTG
No 307	No307_seqF2	ATCGGAAATCAGGGCCATCA
No 307	No307_seqR2	TCAAAGAGAGGAGGCCGAAA
No 309	No309_seqF	TTCGCGGGCACATATCATTG
No 309	No309_seqR	AGAGGTTCAAAGGCAGACGA
No 311	No311_seqF	AACTACGGTACCCTGGTTGG
No 311	No311_seqR	TCACAGTGTGGAATTGGCG

11.3 Sister (stw)intronok függelékei

11.3.1 Teljes sister stwintronok (23)

>HC0c002A

GAATATGGACTCTATggtatgttcaggtactcagtgcttagaaagacctcaaaa**tttctaaaat**at
 aggaacaa**gctaacatcataacagtaagt**aaagaattctaataatcgccgctgtccagattaaaaag
 gcgatacttggtacatattgaaaacaaaaggtaagat**gctgac**acgacatggtgcgatacac**tag**
 ACCACGTTTTCTGGT

>HC0c004A

ACTGGGGTACTACAggtgagtatcaaaacacgtcgtcctggcattttagtccttaa**tttctaaag**
 aaaaaaaaaaccctaaatttctacaacataggaataa**actaacattatgaa**tagtaagttaaccggt
 cctatttcggttgcatcggggttaagagaaatgtctagtac**attccagaaa**ccaggataaaaaatg
 ctgacacgacatgtcttaataacac**tag**TCGAGGTGAAAGTCA

>HC0c017A [mitochondrial carrier gene]

GTTTTCTGGATTACGGgtgagttataaaaacatgtcgttctggcatttttagatattgatttctagaa
aaaaccccccgtttctacaatatatgaatcagctaacggtataacagtaagttaagcgttcttatt
tcggtgctatcgggattaagaaggcagtagctggatgttctataaaacgaaggataaaaatgctga
cgcgacatgttttaatacatcagAGTTACCCAAGCGAC

>HC0c017B [other gene]

AGAGCGGATTCCTCTgtaagtgaaaaatctgccgttttgggtgtttatacttcaatttcttgaaa
gatcaccaaattcgctaacggtatgacagtaagtcaagtgcttctgttctcgtcgcacatcgggggta
agaaggcaatatctggtatattctagaaacgacgtataaagtactgacacgacgtgttttaatagA
TGTTGGAGATGAAC

>HC0c021A

TTTCTAAAGATGTTTCgttacctatgtattaagatatatcatggttaggatttcggcctttgatttct
agaagaaccctatgacatagaaactagctaacggtatagcagtgagtaaatctttttttttttt
ttttcaatactatgaaacttaaaaaaggcaatttccggcgactataagaacgaaggtaaaaaatag
actgacataatatgtttcaacacactagGGTCCATATGAAGCT

>HC0c047A

GCCCTGGGTTGCAATggtacgtacaaaaacttcctgtggagggttttaaaccttgacgtctgaaa
aatgtcgaataataactaattttataacagtgagataaatgttcctctgttcctcttttaacgcc
tggcatagcctggaagagaaggtaaaaaaaacgaaggagagaaaaagctgaacacgatatgt
tttaataccttagGTTACAGCAAGTTGT

>HC0c052A

CCATCTGGCTACTCGgtgagttataaaaacatctcgttctggcatttttagacctttgatttctag
aagactccccatatttccagtgatatagaattagctaacggtatgacagtaagttaagcgttctta
tctcgtcgtatcggggtatgaaggtagatctggtatattctagaaatgaaggataaaaatgct
aatagacatgttttaatacattagACTATCGCGCGCAGC

>HC0c058A

CCGTACACGAACAATggtatgtatgaagacataaccgtcttggcgtatttagaccttgatttctagaa
agaccctcaggttttcaaatatagggattagctaacggttaggacagtaagtaagcgttgctcgtt
tcgctcatcattgaggttaaaaaggaaatgtctggtatattccggaagcaatggtcgaaaaacgctg
acacgacatatttaatacacacagATCATTGGGTCCAG

>HC0c061A

GCTATTGGTTCGGATggtgagcacaaaaacatgtcgttctgacattttataccttaatttctacaa
acccccctatatttttacaacatataaattagctaacggtatgacagtaagttatgcgttcttatt
tctttactatcagaggtcagaaggtagatttttagtatatttttagacacgaaggataaaaatgctaa
cacgatatgttttaatacacacagGGCAGTACTGGTGGGC

>HC0c066A

TATGCTCTGGTGTGAgtgagttataaaaacatgtcgttctggcatttttaaacctaaatttctaga
aaatctcccattgttttacaatatatgaattagctaacattataacagtaagttaagcgttcttatt
ttcgtcgtatcgggattaagaaggcagtagctggatattctagaaacgaaggatgaaaatgctg
acacgacatgttctaatacattagCCTGCGCCTGTCTAC

>HC0c070A

CTTTGTTGCTCTGGGgtaagttataaaaatagtcgttccagcatttttagaccttgatttctagaa
aacccccctatttctacgatataatgaattagctaacattatgacagtaagcgaagcgttcttatttc
gtcgtatccgggttaaggaggcagatgtctggtatattctacaaacgaaggatgaaaatgctaac
cggaatgttttaatacacacagCCGGGGTGGCTTTGA

>HC0c076A

ACGCTAGTGAAGCCCggtgagtataaaaacatgtcgttctagcatttttagaccttgatttctagag
aaaccctatatactttacaatatatgaattagctaacggtataatagtaagtaagcgttcctatt
tcgtcgttatcgggattaggaaggcagtagctggatattccagaaacgaaggataaaaatgctaa
cacgacatggttttaatacaccagTGTGCAGAAGGTGG

>HC0c102A

CATTTAATAGTCAGTggtgagtataaaaacatgtcatttctggcatttttaggccttgtttctagaaa
accctcatatcttacaatatatgaattagctaacggtatgatagtaagc taagcgctcttatgt
cgttgctatcagggttagggaggaaggataaaaatgctgacacgacatggttttaatacac cagTTC
AAGAAGCTACAA

>HC0c164A

TAAGGCAACTACAACggtatgtattaggtagatatatggtgtggggattccggcccttgcttcc
tagaaaagccttcatttctctatgatgcaggaacca gctaatttttatgacagtgagt taaaacac
ttttatgtgctgcgctgaactgtgaagagaatatcaagtatatcctagaaatgaagactatag
atattgacactacatggttttaacactttagGAAAGGGAGCGATGT

>HC0c178A

CCTAGACCACCCTTCggtgagtataaaaacatgtcatttctggcatttttaacattgatttctagaa
aaccctcatatcttaccgtatatgaattagctgacggttgtagtaagtaagcgttcttatt
tcatcgtaccgggttaagaaggcgtatctggatattctagaacgaaggataaaaatgctgaca
cgacattttttaatacac cagGGATATCAGCGAGCAT

>HC0c224-179 [split over contigs 224 and 179] [GenBank Accession
number MW477887]

ACGGACAGCGCCTCTggtgagtataaaaacatgtcgttctggcatttttagaccttgatttctagaa
gaccccgaccccccccccaacatatttctacaagatatgaattagctaacggtatgtagtaagt
taaagcttcttatctcgtcgtatcgggttaagaaggcagtagctcttatattctagaaacgaag
gataaaaatgctaaccacgacatggttttaatacac cagTCTGTCCAACATGCG

>HC0c236A

CGCGCTTAGCATCTTggtgagtatacgaatatgtcgtttttgtcacttttagaccttgatttctaga
aagatcctcaaatgtcttcaatgtaggaattagctaacgctcgtgacagtaagtgaaagcgttctat
ttcgtcgccatcgggttaagaaaacaatgctaggtatatcttagaaacgaaggttataaatgcta
acaagttatggttttaattacac cagCGCGCTTAGCATCTT

>HC0c252A

AGGGTTGGCGTCCCAGgttacgtataaaactatgccatttctggcatttttagatcctggtcctaccta
gaggtttagctaacgtaagacagtaagtgatcgtttttatttcgttattgtcaggttaacaag
gaaatggttggtttattctagaatgaagtattgaaaag gctaaccacgataggttgaatatcc tag
GGATGGAGGGATGCAA

>HC0c271A

ATTGGATACGTAATTggtgagtataaaaactatgtcgttttggcgttttagacatttggtttctgga
aagaccatcaaatctcaatataaaaatcagctaacggtatgacagtgagtaaatgttcttatttc
gtcgcgatcgggttaagaaggcaatgcttggtataattctagaacgcacaggtataaaaatgctgac
acgacgcgttttaacacgttagTTATATGCCGCATTG

>HC0c332A

GGTGGATGCTACAGTggtatgtataaaaatagtcgtttggcatcttgagactggttttcataaa
atatatatgaattagctaacactgggacagtaagtaaacggttctcatttcggtgctatcaggtcc
aagaaggcaatgtctagtatattctagaagcaaggccaagaatgctgacacgacatggttctaata
cac cagAATCGAAGTGAAGT

>HC0c378A

GAAAGGGTGCATATTggtgagtacaaaagcatgtcattctggcatttttagaccttgatctctaaga
aaaccactatatttctataatatattaattggctaacggttatgacagtaagttaagcgttcttatt
tcgtcgtatcgcggttaagaaagtagtgtttgggtatattctagaacgaaggataaaatgctgac
acgacatgttttaatacacagTCTTTACCGAACTGA

>HC0c406A

TGGCGCAAGACTGTTggtgagtataaaaagacgtcattttggcatttcataccagattcccacaa
tataggaattcgctaacactatagaagtaagttgagcgtttttatctcgtcgtcatcggggttaag
aaggcaatgcctacctagtatattctagaatgatgggtaacaatgctgacacgacatgctttaac
acactagTTACATGCCATCGAC

>HC0c522A

CGACGCCCCGACATggtatgtataaaacatggttgccttggatttccaagcattgatctctggaaa
gactcgtaaatttccatagtttaggaattaagctaacgttgggtagtgagttaagaattatcattt
tatcgccatcgagtttaagagcgcgacgacctggtaagcttagaacaagaagtcaaaattgctga
caggatacattttaatacactagAAAAGGTGGCATGAC

2 sister stwintrons unique to Hypoxylon sp. EC38

>HEC0c114A

GGTGAGTATAAAAACATGTCGTTCCGACATTTTAGACCTTGATTTCTAGAAAATCCCATATTTCTA
CAATATATGAATTAACCTAACGTTATGACAGTAAGTCAAGCATCCTTATTTTCGTCGCTATCGGCCTT
AAGAAGGCAGTATCTAGTATATTCTAGAAACGAAGGATAAAAATGCTGACACGACATGTTTTAATA
CACCAG

>HEC0c217A

GGTGAGTATAAAGACATGTCGTTCTGGCATTTTAGATCTTGATTTCTGCAAAAATCCCCATGTTTC
TACAATATATGAACCACTAACGTTATAACAGTAAGTTAAGCGTTCTTATTTTCGTCGCTATCGGGG
TTAAGAAGGCAGTACCTGGTATATTCTAGAAACGAAGGATAAAAATGCTGACACGACATGTTTTAA
TACATCAG

11.3.2 Sheared sister stwintronok (13)

>HC0c103A

GGTATGTATCAAAAATATGTGGATTCCAGGATTTTGAAAATTTGGTTTTTACAAGAATGTTTTTGA
TTGGTTTTTACAAGAATGTTTTTGAATTTGGTTTTTACAAGAATTTTCAAGTATAGCTATAATATAG
TGGAAAAGCTAACGTTAGAACTAGTATGTTCAAACACCCCTTATTTTCGTTGCTACCAAGCCTAGAAAGCA
CCGCCTGGTATATTCTAGAACACGAATCTTAAAATTAGGTACTAATACTATAGTTGTGTCTAAAAA
AAAAAG

>HC0c002B

GGTACGTATAAAAACGGCAAGAGATATTCATATTTTCCAGTCCCCAGATTTTTCGAACATCCTTCAAAT
TTTTCACAACGCAATACTAACATTGTGAAGTGAGTAGAAAATATCGTTGCCTTACTATCAAATTTG
CGAGGGCAACGATCTGTTCCCTCCATATGAGAATTTTACAATGCTAACATGACATATTTTGATACA
TCAG

>HC0c004B no-280

GGTATGTAGGAAGACGCGTCTCGTCGTTATTCTAACTCCGACTGTCAAATGACTTCTAAATTTCTG
TATAGTAGTAACTAACCTTCTAGTAGTGAGTCGAATATCCTCGTTTTATCCTCCCCAACTTAAGA
TGGAAATTTCCAGTATATTCTAGAAATAGATGTTATAATACTGATATGACGCGTTTCATACACAG

>HC0c016A

GGTAAGTATCCCTACATAACCTGTCCGTCTCAGCTTATAGTTCTTGATCCCCCTTCTGTTTCAGTA
GTCAATCAAACGTTGACTAACATTCCTTAGTACGTACAGGATCTTCTTCTTCTCAATCATTTTT

CTATAACAATTAAATGCCTAGGAGACCTCAAACGGTCCAAGGTAAACGTTGTTAGATTACTAAC
GCGACATGATGTAATATGACAG

>HC0c016B

GGTGGCTTCATTTCTTTATCTCCATTCATCCTTGCTCATACTTCATTTATATGCATGCATACGCTA
ACGATAGGACAGTAAGTAACCAAGATTTCCTAACCTCATCTTCAGGCTAGATCTCCTTCCTGGAG
AAACCTACGAGAGGCTCCTTTCTCTAACCTCTATAAAGCGGCCGGGAGATCAAAAATTTCTTCGCGT
TCAGCTGGAGAGAGCTAACAAATAGTAG

>HC0c020A

GGTGGCTGTAAGACCATCTCGTATCTTAATTCTTGTCCTTGACTATTTAATAATTTTAAGATTTCGT
ATAGGATAGTAATAACTATACGGTAGTGAGTGAACATTTGTTTATTAACACTGAATCTAGAAAA
GAAAAGCAACTTCTAGTGATAGGAAAGGAGTTAGAAGTACTGACACGGTATGTTTCAATAAAAACAG

>HC0c024B

GGTATGTAGTAGATGGCCTCAATTTCCGATTCCATACCCCCTTTTTATCTCTTCTACTCTTCCG
TACTAACCTTGTGGATAGTGAGTGACTAATTCACCTTCTACCTATTCTACGGATTAAGGGCTCAATG
AACGATTTGAGATACTAACGTATTGCAG

>HC0c046A

GGTATGTACGAAAACATCTCGTTTGAGCATTTCAAAGCCCCTTTTCTAGAATAATCTTCGAACTTC
TCGTACATCCATTGCTAACCTCCCACTAGTGAGTTATAACACTCTAAGACGTTTCGAGATGATGGTGA
TATAATGCTGACACAAGATGTTTTCTCCACCAG

>HC0c091A

GGTAAGTATTAGAGCAAGGGATCATATTTCTGAACCCTTCAGCCCCATTTTCATATTTGCCTACCC
GATTGACTGTCCGTAGGTAATTAGCTAACCTCCTCATAGTATGTGAACTATTTCTCCGTTCCAC
GTTCTATTACTATAGTGAGCTATGAAATAAGTTATAAGATTACTGACGAACAATTTATCCATGTT
AG

>HC0c091B_no-222

GGTACGTAGTGTCTGTTCGAGATTTCAAGAGCTATCTTTGAAACGGTCTCCAATTTCTTTGGTAT
AGATAAAAAGCTAACCTCCTGGCAGTGAGTAAAATGATCCTATTCCATTTGCACGATATTTATTCCG
TATGTACTAAAAGTAAATGACGAGATACTGACGTGCCATAATATACTAG

>HC0c159A_no-090

GGTACGTATGAAGATATGCCATGTTACAGTTTGAACCTCTAGCCCCCTTGAAGATCTCCTAGTTTA
TTCAATACTTCAAGCTAACGTTACGGCAGTAAGTGAACACATTCATTTCTCCGCTGCTCAATTTAG
AACAGCAACGTCTAGCGTTTTAGAGAAAAGAGATGATGGCATACTGACATGATATACTCTGGTATT
AAG

>HC0c263A

GGTACGTCTAATCCCATATCGTGTGATATCGAATGATCTCCGTGTTCTATAATACAGTACTAACG
CTCAGACAGTGAGTAACATCGAGTTCTTTATTTCCCTGTGACCTGTGCGTGGGCCAGCGTTTC
CAAGTTAGGAGGAATGTGAATGGTCAACCCTAAAACATGCCTAAAACCTCCCGAACCTTGAAA
AAATCATCACTAACCTTCTCTACAG

>HC0c304A_no-129

GGTACGTATAGATGCATTTCTAGAGTATCTTTCAATTTCTCTCACACGCAAACGCTAACATTTTCA
AGTAAGTTGAATAATTTCTGCTATGCCCAAGCTTGCTACGGATCCCCTAGATGCACTCGAGCTAGG
GGAATATGATGGTAGCCCTTCATGGCTAGGCTGATGCTGACACAACACGCTTTAATACAAATAG

11.3.3 Sister kanonikus (cropped) intronok (11)

HCOc047B és HCOc229A 1-es típusú cropped sister intronok. A 3' végi plusz guanin (G) és timin (T) magenta színnel van jelölve.

```
>HCOc024A (2)
TGCGGACCATGCTCTgtcagtagataaaaaacatgtcgttctggcatttttagacctcaatttctagaaa
cgaaaaataaaaatgctaacacgacatgttttaatacGGGCCGAATGCAGG

>HCOc047B (1)
AGTATTCGAGGTTTGTgtgagtagataaaaaacatgtcgttctggcatttttagacctcaatttctagaaa
aatcgccggtattttctattaaaaaatatgaattaactaacggtataaacagGTTGCATAGGGGCAT

>HCOc096A (2)
CGATGATGTCCTAATgtgagtagataaaaaacatgtcgttcttagcatttttagaccttaatttctagaaa
cgaaagataaaaatgctgacatgatatgttttaatacacgtagCGCGGCTATCAGACA

>HCOc100A (2)
CAGGAGAGTTTTAACgtgagtagataaaaaacatgtcgttctggcatttttagaccttaatttctagaaa
cgaagtataaaaatgctgacacgacatgttttaataatacctagTCGAACGGAAACCAG

>HCOc105A (2)
CGAGACGATGCAGTGTgtgagtagataaaaaacatgtcgttctggcatttttagacctgttttctagaaa
cgaaagataaaaatgctgacgacacggtttttatacacgtagAAACTAGACCGCGAG

>HCOc121A (2)
TCGCATCAGCTACTGTgtgagtagataaaaaacatgtcattctggcatatttagaccttagtttctagaaa
cgaaagataaaaacgctgacacgacatgtttttatacacgtagACCAAGCCTACGCAC

>HCOc147A (2)
CACCTCGCTTCCGTgtgagtagataaaaaacatgtcattaaggaatttttagaccttaatttctagaaa
cgaaagataaaaatgctgacacgacatattttttatacacgtagACCACGGACGCGTAC

>HCOc153A (2)
ACATGCCAATTGAAGgtgagtagataaaaaacatgtcgttctggcatttttagaccttaatttctagaaa
cgaaagataaaaatgctgacacgacatgtttttatacacgtagcagCTCTACGACGAAGCC

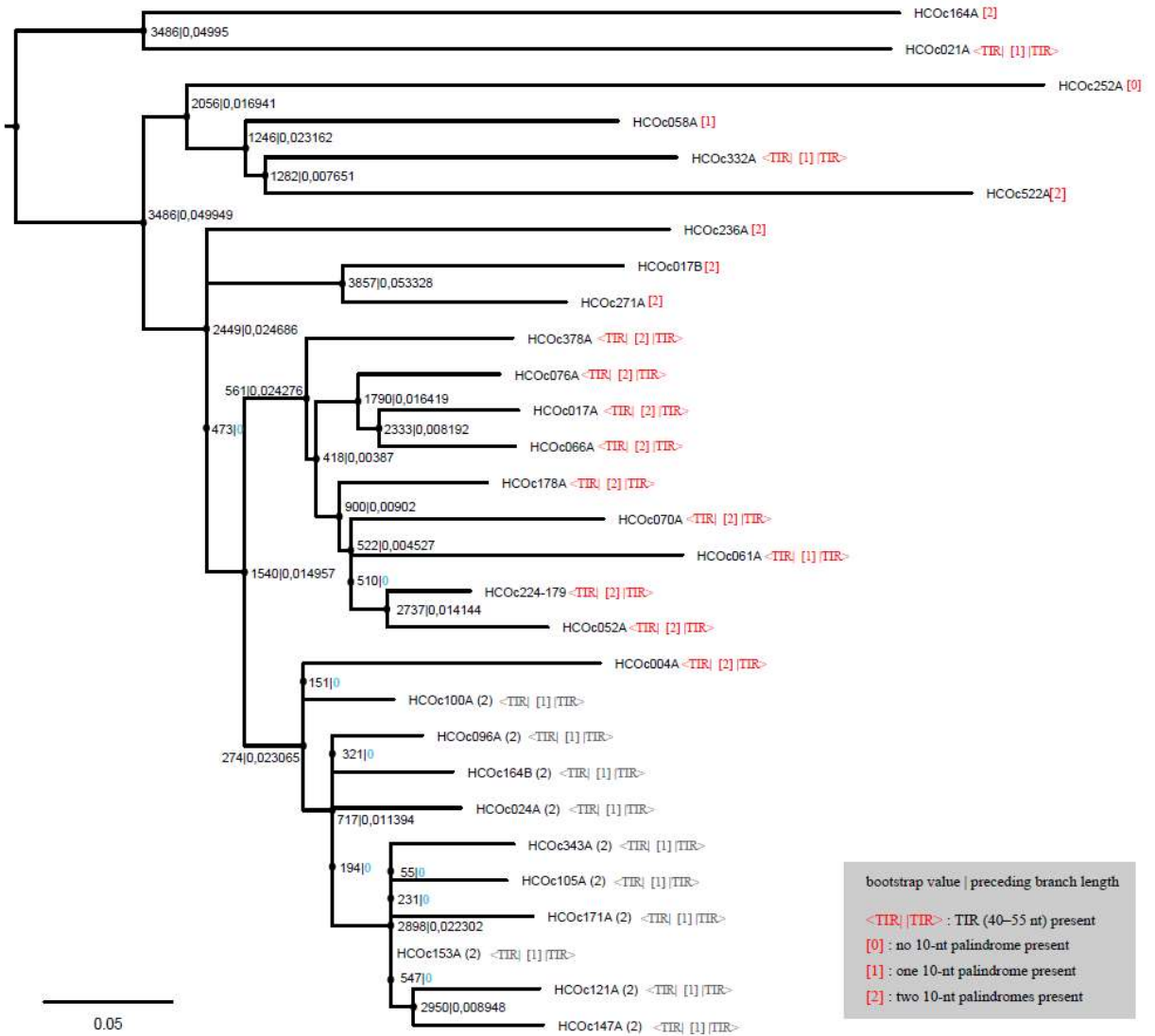
>HCOc164B (2)
GAGACACGTCTATCgtgagtagataaaaaacatgtcgttctggcatttttagaccttaatttctagaaa
cgaaagataaacagctgacacgacatgttttaataatacctagCACTAGGACCGTGT

>HCOc171A (2)
CTCGGCTATATATTGTgtgagtagataaaaaacatgtcgttctggcatcttagaccccaatttctagaaa
cgaaagataaaaatgctgacacgacatgtttttatacacgtagGGACACTCCAATCG

>HCOc229A (1)
TGGATATTGCCTTTGTgtatgtatgaaaacgtgacattttagacatttagatttgaatctgaaaagat
ccttaaaagtctattataaccagacctagcctaacggtatgacagACGTCTTTTCGCTAG

>HCOc343A (2)
TGTTTTCGCCGGTTgtgagtagataaaaaacatgtcgttctggcatttttagactttaatttctagaaa
gaaagattaaaatgctgacacgacatgtttttataatacctagagATGGATCGTGGTACG
```

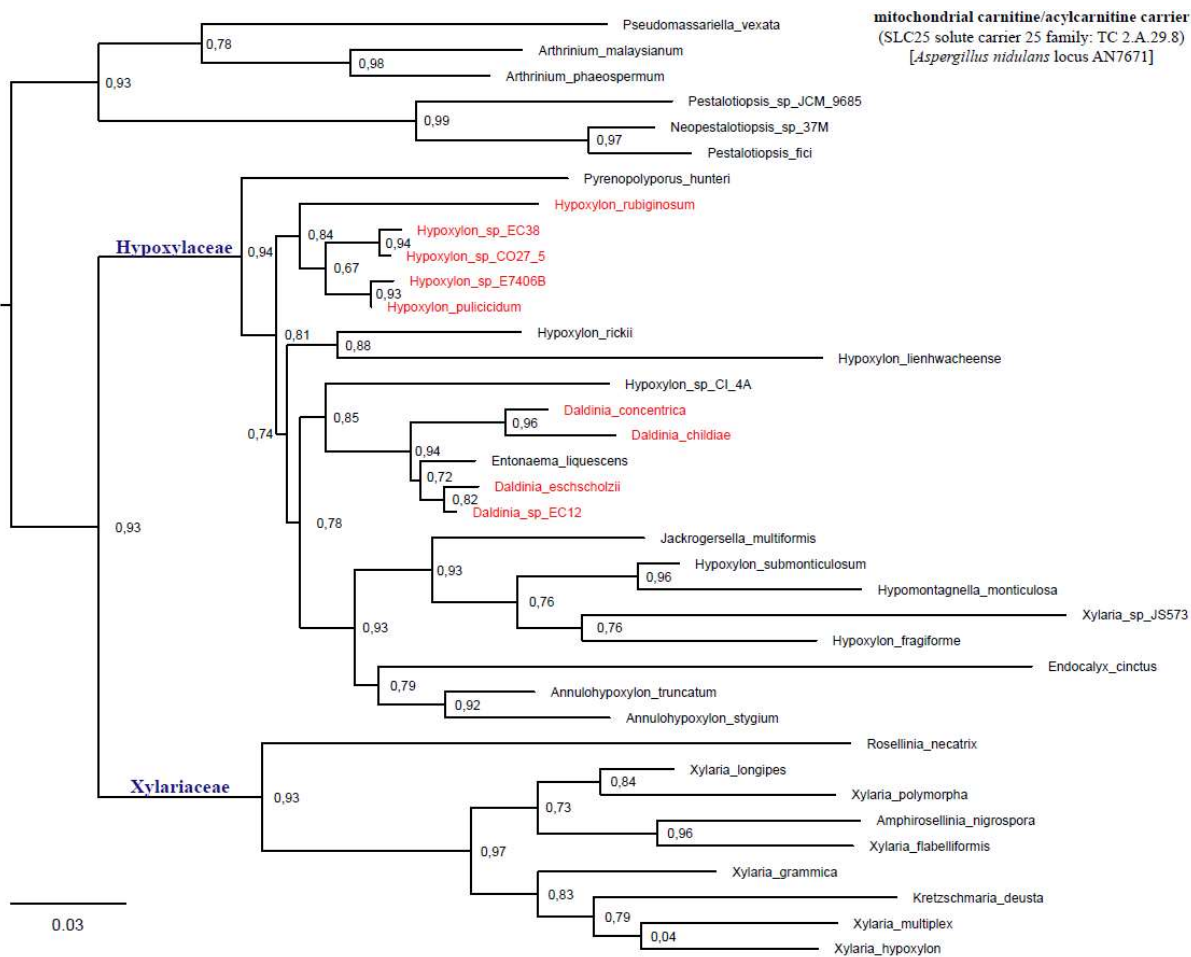
11.3.4 *Hypoxylon sp. CO27-5* sister (stw)intronok filogenetikai törzsfái és szimmetriai tulajdonságai



24. ábra Maximum likelihood filogenetikai törzsfá. 19 paralóg sister stwintron és 10 2-es típusú cropped sister intronl *Hypoxylon sp. CO27-5* gombában.

sister (stw)intron	length repeat palindrome incl.	no. identities	palindrome(s)	% ID inverted repeat	minimum free energy for dimerisation [Kcal/mol]
HCDc153A (2)	51	39	TTTCTAGAAA	~ 76.5 %	-79.8
HCDc105A (2)	51	37	TTTCTAGAAA	~ 72.5 %	-73.2
HCDc100A (2)	49	35	TTTCTAGAAA	~ 71.4 %	-68.2
HCDc164B (2)	49	32	tctctagaaa	~ 65.3 %	-53.8
HCDc096A (2)	49	36	TTTCTAGAAA	~ 73.5 %	-59.4
HCDc343A (2)	47	35	TTTCTAGAAA	~ 74.5 %	-75.8
HCDc024A (2)	46	34	TTTCTAGAAA	~ 73.9 %	-60.8
HCDc147A (2)	51	36	TTTCTAGAAA	~ 70.6 %	-47.8
HCDc121A (2)	47	34	TTTCTAGAAA	~ 72.3 %	-54.6
HCDc171A (2)	51	35	ttcttagaaa	~ 68.6 %	-71.0
HCDc004A	49	32	tttctaaga ttctagaaa	~ 65.3 %	-106.8
HCDc017A	47	32	TTTCTAGAAA gttctaàaaa	~ 68.1 %	-122.2
HCDc052A	44	33	tttctagaag TTTCTAGAAA	75 %	-124.8
HCDc066A	48	35	TTTCTAGAAA TTTCTAGAAA	~ 72.9 %	-118.4
HCDc070A	47	33	TTTCTAGAAA ttcttagaaa	~ 70.2 %	-112.4
HCDc076A	49	37	tttctagaga ttctagaaa	~ 75.5 %	-134.6
HCDc178A	49	34	TTTCTAGAAA ttcttagaa=	~ 69.4 %	-96.5
HCDc224-179	47	37	tttctagaag TTTCTAGAAA	~ 78.7 %	-136.8
HCDc378A	50	34	tctctaagaaa TTTCTAGAAA	68 %	-109.6
HECc034A (2) (Hypoxyton EC38)	49	36	TTTCTAGAAA	~ 73.5 %	-56.4
HECc057A (Hypoxyton EC38)	47	34	attctagaag ttcttagaaa	~ 72.3 %	-119.2
HECc321B (2) (Hypoxyton EC38)	51	41	TTTCTAGAAA	~ 80.4 %	-93.2
HECc217A (Hypoxyton EC38)	47	32	tttctgaaa TTTCTAGAAA	~ 68.1 %	-126.1
Dchc003A (D. childiae)	49	33	tttctagaac ttcttagaat	~ 67.3 %	-109.4

25. ábra Sister (stw)intronok szimmetriai tulajdonságához gyűjtött adatok.



26. ábra A CO27-5 stwintonjaihoz hasonló szekvenciák jelenléte vagy hiánya Hypoxylaceae fajokban