



1949

**A hasadó élesztőgombák (*Schizosaccharomyces*) komparatív  
genomikai vizsgálata**

Egyetemi doktori (PhD) értekezés

Ács-Szabó Lajos

Témavezető

Gálné Dr. Miklós Ida

Tanszékvezető egyetemi docens

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

Debrecen, 2023

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

*Debrecen, 2023.03.10.*

.....  
*a jelölt aláírása*

*Tanúsítom, hogy **Ács-Szabó Lajos** doktorjelölt **2011-2014** között a fent megnevezett Doktori Iskola **Fermentációs biotechnológiai és biomérnöki (biológia) doktori** programjának keretében irányításommal végezte munkáját. Az értekezésben foglalt eredményekhez a jelölt önálló alkotó tevékenységével meghatározóan hozzájárult. Nyilatkozom továbbá arról, hogy a tézisekben leírt eredmények nem képezik más PhD disszertáció részét.*

*Az értekezés elfogadását javasolom.*

*Debrecen, 2023.03.10*

.....  
*a témavezető aláírása*

# A hasadó élesztőgombák (*Schizosaccharomyces*) komparatív genomikai vizsgálata

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

Írta: **Ács-Szabó Lajos** okleveles Biológus

Készült a Debreceni Egyetem **Juhász-Nagy Pál Doktori Iskolája**  
(**Fermentációs biotechnológiai és biomérnöki (biológia) doktori programja**) keretében

Témavezető: Gálné Dr. Miklós Ida

A doktori szigorlati bizottság:

elnök:	Dr. Batta Gyula	.....
tagok:	Dr. Miskei Márton	.....
	Dr. Gácsér Attila	.....

A doktori szigorlat időpontja: 2015.11.12.

Az értekezés bírálói:

	Dr. Sveiczzer Ákos	.....
	Dr. Benkő Zsigmond	.....

A bírálóbizottság:

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

Az értekezés védésének időpontja: .....

## Tartalomjegyzék

1. Bevezetés.....	9
2. Irodalmi áttekintés.....	11
2.1. A komparatív genomika jelentősége .....	11
2.2. Az eukarióta genom szerveződése.....	14
2.2.1. A genomszerveződés fogalmának értelmezése.....	14
2.2.2. Korai bizonyítékok a random génsorrend cáfolatára .....	14
2.2.3. Az eukarióta génsorrendet befolyásoló tényezők .....	16
2.2.4. A gombagenomok génsorrendjére ható tényezők.....	18
2.3. A szekvencia és struktúra változások kapcsolata .....	20
2.4. A hasadó élesztőgombák bemutatása .....	22
2.4.1. A hasadó élesztőgombák, mint modellszervezetek .....	22
2.4.2. A hasadó élesztőgombák filogenezeise és genom szerveződése.....	27
3. Célkitűzések .....	33
4. Anyagok és módszerek .....	34
4.1. A doktori munka során felhasznált törzsek .....	34
4.2. A doktori munka során felhasznált tápközegek.....	34
4.3. A doktori munka során alkalmazott kísérleti módszerek .....	35
4.3.1 A sejtek tenyésztési körülményei .....	35
4.3.2. Genomi DNS izolálása.....	35
4.3.3. Polimeráz láncreakció (PCR).....	36
4.3.4. Agaróz gélelektroforézis.....	38
4.3.5. Szekvenálás.....	39
4.3.6. Pulzáló erőterű gél elektroforézis .....	39
4.4. A doktori munka során alkalmazott bioinformatikai módszerek .....	40
4.4.1. Programok, algoritmusok, adatbázisok.....	41
4.4.2. Szekvencia adatok.....	43
4.4.3. Genom szekvencia illesztések .....	44
4.4.4. BLAST analízis és szekvencia összehasonlítások .....	46
4.4.5. Szinténia analízis .....	46

4.4.6. Kolineáris blokkok manuális feltárása a hasadó élesztő genomokban .....	47
4.4.7. Kromoszómális átrendeződések feltárása .....	48
4.4.8. Átrendeződések evolúciós modellezése.....	48
4.4.9. Filogenetika .....	50
4.4.10. Génsorrendet meghatározó tényezők feltárása .....	50
4.4.10.1. Funkcionális csoportosulás .....	51
4.4.10.2. Ko-expresszió.....	51
4.4.10.3. Esszenciális gének lokális denzitása .....	52
4.4.10.4. Rec12 hasítóhelyek .....	52
4.4.11. Evolúciós ráta és intront tartalmazó gének vizsgálata .....	52
4.4.12. Normalizáció, randomizáció és statisztika.....	52
5. Eredmények.....	54
5.1. A hasadó élesztőgombák genom evolúciója I.: a <i>S. cryophilus</i> szuperkontigok összerendezése és komparatív analízise .....	54
5.1.1. A <i>S. cryophilus</i> szuperkontigok összerendezése szinténikus kapcsolatok segítségével.....	54
5.1.1.1. A szuperkontigok rendezése a Mauve programmal .....	54
5.1.1.2. A pericentromerikus és szubtelomerikus szuperkontigok azonosítása .....	56
5.1.1.3. A <i>S. cryophilus</i> szuperkontigok sorrendjének meghatározása ..	58
5.1.2. A <i>S. cryophilus</i> szuperkontigok összeszerelésének validálása .....	60
5.1.2.1. Szomszédos szuperkontigok ellenőrzése PCR-rel .....	60
5.1.2.2. A <i>S. cryophilus</i> kariogramja.....	61
5.1.2.3. A rDNS-ek lokalizációjának meghatározása .....	62
5.1.3. A <i>S. cryophilus</i> genom összevetése közeli és távolabbi rokon fajokkal .....	64
5.1.3.1. Kromoszóma-átrendeződések típusai és dinamikája a hasadó élesztőkben.....	64
5.1.3.2. Kromoszóma-átrendeződések összevetése más fajcsoporttal ...	65
5.2. A hasadó élesztőgombák genom evolúciója II.: a stabil hasadó élesztő genomok genetikai és evolúciós háttere .....	70

5.2.1. Strukturális és szekvencia evolúció összehasonlítása a hasadó élesztőgombákban.....	70
5.2.2. Genom konzerváltság és közös szegmensek .....	71
5.2.2.1. Genom konzerváltság nukleotid szinten .....	73
5.2.2.2. Az illesztések statisztikai értékelése .....	75
5.2.2.3. A közös szegmensek vizualizációja .....	77
5.2.3. Ősi kolineáris blokkok azonosítása.....	78
5.2.4. Az ősi kolineáris blokkok fennmaradása és a természetes szelekció kapcsolata.....	80
5.2.4.1. A neutrális evolúció szimulációja .....	81
5.2.4.2. Szimuláció az Artificial Life Framework szimulátorral .....	81
5.2.5. Az ősi kolineáris blokkok lehetséges eredete .....	83
5.2.5.1. Teljes genom illesztések egyéb Taphrinomycotina fajokkal ...	83
5.2.5.2. Gén szintű kolinearitás vizsgálat különböző gombafajokban random blokkok k segítségével.....	85
5.2.6. Az ősi kolineáris blokkok és az azokon kívüli régiók összehasonlítása .....	88
5.2.6.1. A protein szekvenciák evolúciós rátájának összehasonlítása....	89
5.2.6.2. A gének intron vesztese, intron nyerése.....	91
5.2.7. Funkcionális csoportosulás lehetősége az ősi kolineáris blokkokban .....	92
5.2.7.1. GO kategóriák az ősi kolineáris blokkokban .....	92
5.2.7.2. A szomszédos gének ko-expressziós rátájának vizsgálata.....	96
5.2.7.3. Rec12 hasítóhelyek gyakorisága a különböző GO kategóriájú gének mellett .....	96
5.2.7.4. Esszenciális gének denzitása a különböző GO kategóriába tartozó gének között .....	98
5.2.7.5 Rec12 hasítóhelyek és esszenciális gének kapcsolata.....	98
6. Eredmények megbeszélése.....	100
6.1. A <i>S. cryophilus</i> szuperkontigok összerendezése és evolúciós analízise .....	100
6.1.1. A <i>S. cryophilus</i> szuperkontigok összerendezése szinténikus kapcsolatok segítségével.....	100

6.1.2. A <i>S. cryophilus</i> szuperkontigok összeszerelésének validálása molekuláris módszerekkel .....	102
6.1.3. A <i>S. cryophilus</i> genom összevetése közeli és távoli rokon fajokkal .....	103
6.2. A stabil hasadó élesztőgomba genomok genetikai és evolúciós háttere .....	105
6.2.1. Strukturális és szekvencia evolúció összehasonlítása a hasadó élesztőkben.....	105
6.2.2. Genom konzerváltság és ősi kolineáris blokkok azonosítása a hasadó élesztő genomokban.....	106
6.2.3. Az ősi kolineáris blokkok fennmaradása és a természetes szelekció kapcsolata.....	107
6.2.4. Az ősi kolineáris blokkok lehetséges eredete .....	107
6.2.5. Az ősi kolineáris blokkok konzerváltsága szekvencia szinten .....	108
6.2.6. A hasadó élesztőgombák stabil génsorrendjének lehetséges okai .	109
7. Összefoglalás.....	112
8. Summary .....	116
9. Köszönetnyilvánítás .....	119
10. Irodalomjegyzék.....	120
11. Függelék I.....	145
12. Függelék II. ....	148

## Rövidítések listája és rövid magyarázatuk

**aLCB:** „ancient locally collinear blocks” – manuálisan beazonosított lokálisan kolineáris blokkok, amelyek legalább 5 génből álltak azonos sorrendben és egymáshoz képest azonos irányultságban a vizsgált hasadó élesztógomba fajok mindegyikében.

**CCC:** „chromosome conformation capture” – a kromoszóma sejten belüli konformációjának megállapítására irányuló molekuláris biológiai módszer.

**DSB:** „double strand break” – dupla szálú DNS törés.

**EDTA:** etilén-diamin-tetraecetsav

**GO:** „gene ontology” – gének csoportosítása a biológiai funkcióik szerint (jelen esetben).

**IGD:** „intergenic distance” – intergénikus távolság, két kódoló gén közötti fizikai távolság nukleotidokban meghatározva.

**LCB:** „locally collinear blocks” - lokálisan kolineáris blokkok. Szekvencia hasonlóság alapján beazonosított közös genom szegmensek, amelyek megtalálhatók voltak a vizsgált fajok genom szekvenciáiban.

**MCD:** „multi chromosomal distance” – multi kromoszómális disztancia. Egy érték, amely megmutatja, minimum hány átrendeződési esemény szükséges ahhoz, hogy egy vizsgált genom hasonló kromoszóma struktúrával és génsorrenddel rendelkezzen, mint a referencia genom.

**NCB:** „non-collinear blocks” – az aLCB-ken kívül eső genomi régiókat jelöli.

**SNP:** „single nucleotide polymorphism” – nukleotid pontmutáció.

**TBE:** Trisz-(hidroximetil)-amino-metán - bórsav - EDTA

**TE:** Trisz-(hidroximetil)-amino-metán - EDTA

**tf:** transzkripció faktor

**UTR:** „untranslated region” – az mRNS-ek azon részei, amelyek átíródnak a transzkripció során, de translációra már nem kerül sor az esetükben.

**WGD:** „whole genome duplication” – teljes genom duplikáció.

## 1. Bevezetés

A feltevés, miszerint az élőlények változ(hat)nak az idők során, már az ókori gondolkodók körében is megfogalmazódott. Ennek ellenére az első alaposan kidolgozott teória csak a 19. században jelent meg Jean-Baptiste Lamarck (1744-1829) tolmácsolásában. A hosszú kihagyás után felpezsdült a fajok változásával, leszármazásával foglalkozó tudomány, gondoljunk csak Charles Darwin (1809-1882) és Alfred Russel Wallace (1823-1913) munkásságára. Ezt követően a természetes szelekció alternatívái, a mendeligenetika újra felfedezése és a populációgenetika integrálása vezetett az evolúció biológia, mint önálló tudományterület megszületéséhez. Igazi forradalmát viszont az 1950-es évek után kezdte meg a molekuláris genetika térhódításával. A hangsúly a fenotípusról a genotípusra került, a változások „egysége” a gén lett. Később, a DNS-szekvenálás és a modern filogenetika (és filogenomika) újabb távlatokat nyitottak a tudományterület előtt.

Ma már tudjuk, hogy a változások három fő szinten játszódnak le: a fenotípus, a kariotípus és a DNS szekvencia szintjén. Míg a fenotípus változásainak általában az adaptáció a fő mozgatórugója, addig a DNS szekvencia szintjén számos neutrális változás is bekövetkezik, mint például a szinoním mutációk. Talán a kariotípus esetében a legárnyaltabb a kép, hiszen sokszor egy fajon belül is eltérő kromoszómaszámot és mintázatot figyelhetünk meg. Az egy fajon belüli eltérő kromoszómaszámok leggyakrabban a növények között fordulnak elő, bár az esetükben a poliploiditás következménye az eltérő kromoszómaszám. Azonban mind a gombák, mind az állatok között ismertek olyan fajok, amelyek eltérő számú kromoszómával rendelkező geográfiailag elkülönülő populációkból állnak.

Manapság a teljes genom szekvenciák összehasonlítása lehetőséget ad a változások és kapcsolatuk együttes vizsgálatára. Az ilyen típusú (komparatív genomikai) elemzések vezettek például a teljes genom duplikációk felfedezéséhez, különböző molekuláris órák kalibrálásához vagy a random

génsorrend cáfolatához. Ezekből a kutatásokból természetesen nem csak teoretikus következtetések vonhatók le, számos gyakorlati „hasznuk” is van. Például egy-egy járvány esetén a komparatív genomika segít előre jelezni bizonyos változásokat, amelyek megkönnyítik a védekezést. Emellett a rokon fajok genomjainak összehasonlítása jelentősen megkönnyítheti a megfelelő transzgén integrációt is, legyen az kutatási vagy gazdasági célú.

Jelen doktori munka a gombagenomok változásainak bizonyos aspektusait kívánja bemutatni, középpontban a hasadó élesztő (*Schizosaccharomyces*) genomokkal. A hasadó élesztőgombák az Ascomycota törzs egyik alapi elágazásához tartoznak, így sok közös tulajdonságot őriznek a Metazoa csoport tagjaival, úgymint a relatíve nagy kromoszóma méretek, a hasadásos típusú sejtosztódás, a sejtciklus-kontroll, metabolikus útvonalak, RNS-interferencia és még sorolhatnánk. Másrészt egyediek, hiszen nagyon konzerváltak a genomjaik géntartalom szempontjából, míg a szekvenciáik gyorsan változnak. A genusz legismertebb tagja a régóta közkedvelt modell organizmusnak számító *Schizosaccharomyces pombe*. Jelenlegi tudásunk szerint a nemzetséget öt faj képezi, amelyből háromnak közel kromoszóma-szintű összeszerelési állapotban van a genomja (*Schizosaccharomyces pombe*, *Schizosaccharomyces octosporus* és *Schizosaccharomyces japonicus*). Egy további fajt megszekvenáltak ugyan, de nem szerelték teljesen össze a genomját (*Schizosaccharomyces cryophilus*), míg a legutóbb leírt faj (*Schizosaccharomyces osmophilus*) elérhető genom szekvenciával nem rendelkezik.

## 2. Irodalmi áttekintés

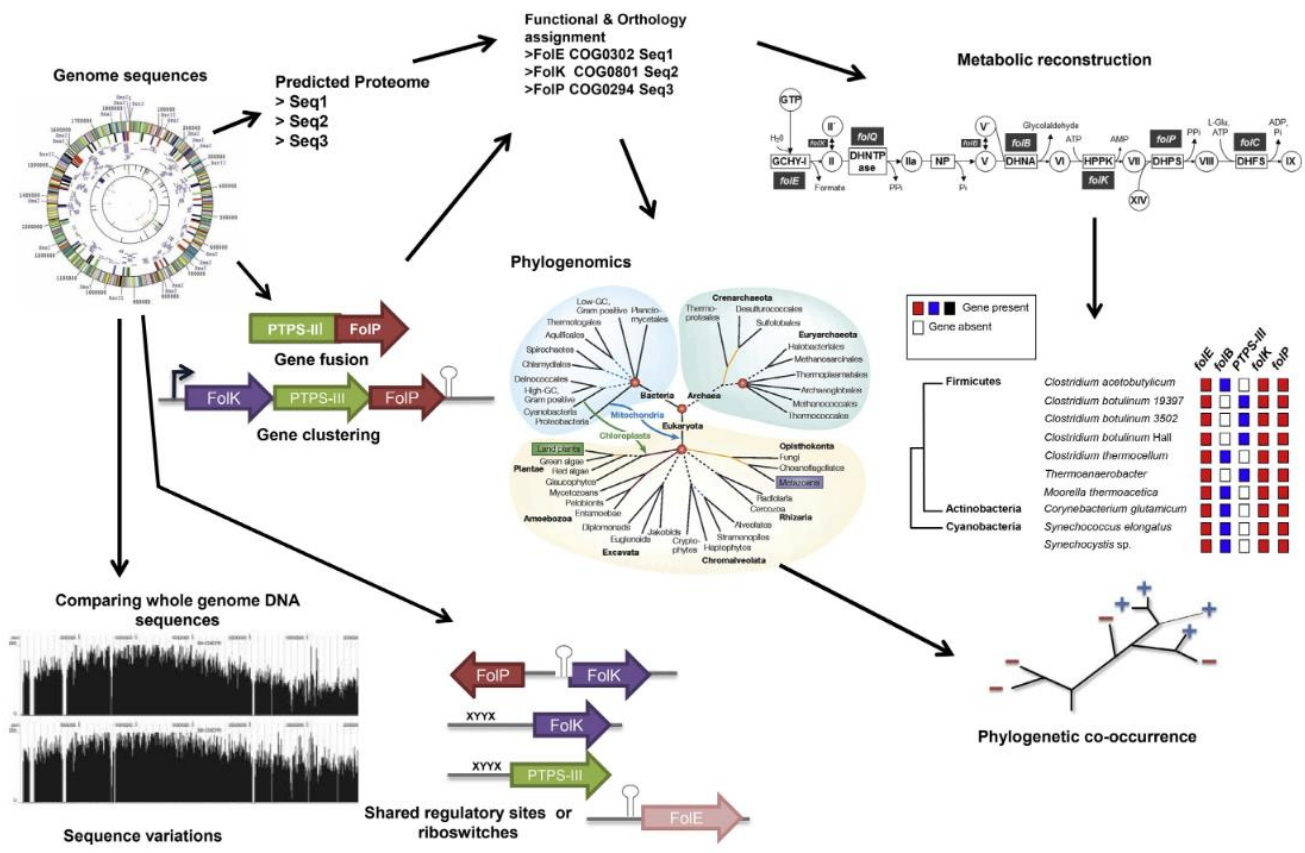
### 2.1. A komparatív genomika jelentősége

Watson & Crick (1953) úttörő munkásságát követően a kutatók világszerte nagy energiát fektettek a különböző élőlények DNS szekvenciáinak meghatározásába. Bár a Sanger típusú szekvenálás (Sanger és mtsi, 1977) egy ma is közkedvelt és széles körben alkalmazott szekvenálási módszer, az igazi áttörést a molekuláris biológia tudományterületén a második generációs szekvenálási technológiák (Illumina, Roche454, Ion Torrent) hozták el a 2000-es évek elején. Gondoljunk csak az 1990-ben indult Humán Genom Projektre, aminek befejezéséhez 10 évre volt szükség a Sanger típusú szekvenálást alkalmazva (Lander és mtsi., 2001), ellenben egy ilyen kaliberű feladatot a második generációs technológiák mindössze néhány napra redukálnak. Ennek köszönhetően manapság mindennapos dolognak számít egy-egy újabb genomszekvencia publikálása. Azonban precízen összeszerelni egy genomszekvenciát csupán a második generációs technológiákra alapozva mind a mai napig komoly kihívásnak számít (Salzber & Yorke, 2005; Sims és mtsi., 2009; Bradnam és mtsi., 2013). E „probléma” kiküszöbölését segíthetik elő a harmadik generációs szekvenálási technikák (Pacific Bioscience, Oxford Nanopore), amelyek segítségével nagy méretű DNS szakaszok bázissorrendjei is egy menetben megfejthetők. Viszont a második és harmadik generációs technológiák együttes alkalmazása még mindig jelentős kiadásokkal jár, mindemellett a megfelelő informatikai háttér hiánya és a kísérletes validációk időigényessége miatt, számos esetben a genomok úgynevezett „draft”-ként vagyis csak részlegesen összeszerelt állapotban kerülnek közlésre. Az NCBI Genome List adatbázisának 2022. június 14-i állapota szerint (<https://www.ncbi.nlm.nih.gov/genome/browse/#!/eukaryotes/>) 23101 eukarióta genom szekvenciája érhető el, amiből mindössze 226 van teljesen összeszerelve. Habár a részlegesen összerakott szekvenciák is lehetőséget biztosítanak az adatgyűjtésre és számos vizsgálat elvégzésére, az alapos és széles körű

komparatív genomikai vizsgálatokhoz elengedhetetlenül szükségesek lennének gondosan összeszerelt genomszekvenciák.

A komparatív genomika leegyszerűsítve tulajdonképpen két vagy több genom szekvenciájából származó biológiai információk összehasonlítását jelenti (Crécy-Lagard & Hanson, 2018). Természetesen az összehasonlítás maga is nagyon sokféleképpen történhet és a levonható következtetések is széles spektrumon mozoghatnak (1. ábra). A vizsgálatok irányulhatnak az egyedekre (törzsekre) jellemző pontmutációk (SNP-k) feltárására, közös ortológ szekvenciák azonosítására, génfunkciók prediktálására, kromoszóma struktúra variánsok meghatározására, metabolikus útvonalak rekonstruálására vagy filogenomikai vizsgálatokra is, sok egyéb aplikáció mellett.

A komparatív genomikai vizsgálatoknak köszönhetően ismerhettünk meg számos evolúciós szempontból nagyhatású jelenséget, például a génátrendeződések (gene shuffling) bakteriális metabolizmusra gyakorolt hatását (Tatusov és mtsi., 1996), vagy a teljes genom duplikációt a pékésztő esetében (Goffeau és mtsi., 1996; Wolfe, 2015). Megismerhettük a kromoszóma átrendeződések és szekvencia mutációk fajképződésre gyakorolt hatását (Seoighe és mtsi., 2000; Sharakhov és mtsi., 2002; Haerty és mtsi., 2007; Hillier és mtsi., 2007; Dujon, 2010; Hane és mtsi., 2011; Maguire és mtsi., 2013). Összehasonlító munkák sora vezetett a felismeréshez, miszerint bár az eukarióta operonok nem általánosak (Blumenthal, 2004), de a gének sorrendje a kromoszómákon egyáltalán nem véletlenszerű (Hurst és mtsi., 2004; Dávila-López és mtsi., 2010; Zakharov, 2010). A komparatív genomikai megközelítések segítettek a tumorképződésben meghatározó, illetve szerepet játszó gének azonosításában is (Pickeral és mtsi., 2000; Deakin & Belov, 2012; Dornburg és mtsi., 2021). Az egyre több elérhetővé váló genomszekvencia egyedülálló lehetőséget biztosít olyan jelenségek vizsgálatára, amelyekre korábban egyáltalán nem volt lehetőség, ezáltal jobban megismerhető a genomok szerveződése, működése.



1. ábra. A komparatív genomika alkalmazásának néhány lehetősége a teljesség igénye nélkül (Crécy-Lagard & Hanson, 2018).

## **2.2. Az eukarióta genom szerveződése**

### 2.2.1. A genomszerveződés fogalmának értelmezése

A genomok szerveződése kétféle módon értelmezhető: egyrészt a genetikai elemek elhelyezkedése, sorrendje és egymáshoz viszonyított irányultsága szerint, másrészt a kromoszómák sejtmagon belüli elhelyezkedése és egymással való kapcsolata szerint. Az utóbbi terület vizsgálata viszonylag gyerekcipőben jár, függetlenül attól, hogy a téma kutatásához szükséges alapvető technika, a „chromosome conformation capture” (CCC vagy 3C) már 2002-ben megalapozásra került (Dekker és mtsi., 2002). A terület igazán Lieberman-Aiden és mtsi. (2009) munkásságának köszönhetően kezdte meg a fejlődést, amikor is kifejlesztették a “Hi-C” technikát, ami a CCC technikát a második generációs szekvenálási technikákkal kombinálta. A genomok 3D-s szerveződésének vizsgálata kiemelt fontosságú a megfigyelt struktúrát kialakító molekuláris mechanizmusok megismerésében, továbbá a genetikai, epigenetikai reguláció megértésében (Fudenberg & Mirny, 2012; Gong és mtsi., 2015; Franke és mtsi., 2016; Rodríguez-Carballo és mtsi., 2017; Kraft és mtsi., 2019; Belokopytova & Fishman, 2021).

Jelen doktori dolgozatban a genomszerveződést az eukarióta génsorrend aspektusából szeretnénk bemutatni. Maga a gén elhelyezkedése is kétféleképpen közelíthető meg: melyik kromoszómán lokalizálódik egy adott gén, vagy egy adott kromoszómán hol helyezkedik el az a bizonyos gén (Hurst és mtsi., 2004). Mivel számos feltételezhetően ortológ gén a különböző fajokban különböző kromoszómákon helyezkedik el, ezért a gének egymáshoz viszonyított sorrendjét, illetve a sorrendet befolyásoló tényezőket vesszük górcső alá.

### 2.2.2. Korai bizonyítékok a random génsorrend cáfolatára

Mivel az eukarióta élőlények szomszédos génjei sok esetben nem mutatnak funkcionális kapcsolatot, sőt egy adott genomban megtalálható gén feltételezett ortológja egy másik élőlény genomjában akár teljesen eltérő

génkörnyezetben is előfordulhat, elterjedt a nézet miszerint az eukarióta génsorrend véletlenszerű. Ez a teória azonban jóval a rutinszerű genomszekvenálások elterjedése előtt megdőlt (Hurst és mtsi., 2004). Számos kísérleti eredmény bizonyította, hogy egyes gének gyakrabban fordulnak elő egymás mellett vagy egymás közelében, mint azt a feltételezett véletlenszerű elrendeződés indokolta volna. Az első példák egyike az ortológ gének klaszterei voltak a házi egér és az ember kromoszómáin (Lundin, 1993). Egy másik korán megismert példa a „Homeobox” (HOX) gének klasztereződéseire kötődik (Ruddle és mtsi., 1994). Habár ezt a jelenséget a tandem duplikációk rovására írták, már sejthető volt, hogy nem lehet teljesen random az eukarióta gének elhelyezkedése. Az első teljesen megszekvenált eukarióta organizmus, a *Saccharomyces cerevisiae* (Goffeau és mtsi., 1996) esetében bebizonyosodott, hogy a sejtciklus adott szakaszaihoz szükséges gének egymás szomszédságában helyezkednek el a kromoszómákon (Cho és mtsi., 1998), sőt számos génpár vagy triplet ko-expressziót mutat (Cohen és mtsi., 2000; Kruglyak & Tang, 2000). Mindemellett az esszenciális gének alacsony rekombinációs rátájú genomi régiókban lokalizálódnak (Hurst és mtsi., 2002). Számos más élőlény esetében is bebizonyosodott, hogy a hasonló expressziót mutató gének egymás szomszédságában/egymás közelében helyezkednek el: például a *Drosophila melanogaster* (Boutanaev és mtsi., 2002; Spellman és mtsi., 2002; Kalmykova és mtsi., 2005), a *Plasmodium falciparum* (Florens és mtsi., 2002) vagy az *Arabidopsis thaliana* (Birnbaum és mtsi., 2003; Williams & Bowles, 2004) esetében. De a humán genom „háztartási”, magasan expresszáló vagy szövet-specifikus génjei is hajlamosak a klasztereződésre (Caron és mtsi., 2001; Lercher és mtsi., 2002; Mégy és mtsi., 2003). Kiderült, hogy a *Caenorhabditis elegans* génjeinek 15%-a operon-szerű struktúrát mutat és együtt expresszálódnak (Blumenthal és mtsi., 2002; Lercher és mtsi., 2003), valamint a génjeinek sorrendje nagyon hasonlít a rokon fajnak számító *C. briggsae* génsorrendjéhez (Hillier és mtsi., 2007). Egy későbbi tanulmány bemutatta, hogy az előző

mondatokban felsorolt élőlények mindegyikének genomjában hajlamosak az egy metabolikus útvonalhoz tartozó gének is klasztereződni, bár nem egyforma gyakorisággal (Lee & Sonnhammer, 2004).

### 2.2.3. Az eukarióta génsorrendet befolyásoló tényezők

Ahogy egyre nőtt az elérhető genomszekvenciák száma, széles körben elfogadottá vált a nézet, miszerint az eukarióták génsorrendje sem véletlenszerű. Mivel a magasabb rendű élőlényeknél a prokariótáknál megismert operon szerkezettel nem volt magyarázható bizonyos gének együttes előfordulása, felmerült a kérdés, hogy milyen okok állhatnak a megfigyelt jelenségek hátterében. A darwini teória szerint minden változásnak az adaptáció a mozgatórugója (Darwin, 1859), azonban a genomban bekövetkező változások esetében nem mutatható ki minden esetben egyértelműen a természetes szelekció hatása (Kimura, 1983; Hurst, 2009; Nei és mtsi., 2010; Palazzo & Kejiou, 2022).

A korábbi fejezetben már említésre került több faj esetében is a ko-expresszió, mint a konzervált génsorrend fenntartásának egyik lehetséges oka, viszont a megítélése nem egyértelmű. Két vagy több gén együttes (egy idejű) expressziójának több különböző oka is lehet: fizikai átfedés a gének között, tandem duplikációk, közös promóter (bidirekcionális promóter) szekvenciával rendelkeznek, közös a ko-regulációjuk transzkripciós faktorok által vagy egészen egyszerűen transzkripciós zaj a tapasztalt jelenség (Michalak, 2008; Hurst, 2009). Ha a ko-expresszió nem is minden esetben, de maga az expresszió közvetett szelekciós hatás alatt kell, hogy álljon, amit a „pozíció effektus” és „dózis hatás” jól szemléltet. Azt már korán felismerték, hogy ugyanaz a gén eltérő aktivitást mutat, ha megváltozik a kromoszómális helyzete (például egy transzlokáció miatt), ami a fenotípusban is megjelenik (Sturtevant, 1925; Weiler & Wakimoto, 1995). Közismert tény, hogy egy gén expressziós szintje a teljes elcsendesüléstől akár 20-szorosára is változhat a kromoszómális kontextustól

függően, bizonyos esetekben egy ártalmatlan gén letállissá is válhat (Kleinjan & van Heyningen, 1998; Akhtar és mtsi., 2013; Berloco és mtsi., 2014; Chen & Zhang, 2016; Tsujino és mtsi., 2019; Goodwin és mtsi., 2019). A „dózis hatás” az a jelenség, amikor egy több interakciós partnerből álló protein komplex egyik kódoló génjének megváltozik a kópiaszáma a többi partnerhez képest és ez hatással van a celluláris folyamatokra. Habár a biológiai rendszerek képesek a génkópia változást tolerálni egy bizonyos szintig, számos tanulmány bizonyítja, hogy a géndózsist megváltoztató genommodosulások az érintett protein-komplexek számának nem kívánatos csökkenéséhez vezetnek, bizonyos esetekben végzetes kimenetelűek is lehetnek (Papp és mtsi., 2003; Veitia, 2003; Teichman & Veitia, 2004; Cooper és mtsi., 2011; Rice & McLysaght, 2017). Fontos azonban megemlíteni, hogy populáció szinten a DNS szekvenciában bekövetkező pontmutációk mellett a legvariábilisabb elem a génkópiaszám (Conrad és mtsi., 2010; Dolatabadian és mtsi., 2017).

Számos tanulmány következtetései alapján megállapítható, hogy a ko-expresszió, ko-reguláció, ezáltal a funkció szerinti csoportosulás hatással lehet az eukarióta génsorrend konzerváltságára (Singer és mtsi., 2005; Sémon & Duret, 2006; Mezey és mtsi., 2008; Michalak, 2008; Dávila López és mtsi., 2010). Azonban a hosszú távon megőrzött génsorrendre az említett jelenségek nem minden esetben adnak magyarázatot, hiszen maguk az egy folyamathoz köthető géncsoportok is hajlamosak az átrendeződésre, hogy újabb funkcionális csoportosulásokat alkossanak (Poyatos & Hurst, 2006; Al-Shahrour és mtsi., 2010).

A következő alfejezetekben a gombagenomok tulajdonságaira szeretnénk koncentrálni, mivel a nagy egyedszámmal és rövid generációs idővel rendelkező élőlények genomjára hatékonyabban képes hatni a szelekció (pékélesztő vs ember) (Hurst, 2009).

#### 2.2.4. A gombagenomok génsorrendjére ható tényezők

A gombagenomok kromoszómastruktúrájára, génsorrendjére számos esemény lehet befolyással. Génvesztések, gén- és szegmensduplikációk, horizontális géntranszferek, inverziók, transzpozíciók, transzlokációk, kromoszómafúziók és –hasadások, bizonyos esetekben a teljes genomot érintő duplikációk aszimmetrikus génvesztésekkel karöltve, továbbá hibridizációk következhetnek be, amelyek gyakorisága jelentősen eltér a különböző fajok között (Dujon, 2010; Hane és mtsi., 2011; Scannell és mtsi., 2011; Karanyicz és mtsi., 2017; Rajeh és mtsi., 2018). Ezek a jelenségek legyenek akár előnyösek, akár hátrányosak, a fajképződés fő mozgatórugóinak számítanak (Fischer és mtsi., 2006; Hane és mtsi., 2011; Rajeh és mtsi., 2018). A genomok plasztikusságában is számottevő különbség tapasztalható a különböző gombafajok között. Bizonyos élőlények fajon belül is jelentős diverzitást mutatnak (pl. *Candida*-k, *Saccharomyces*-ek, *Fusarium*-ok), míg mások esetében a fajok között is csak limitált számú és típusú változások figyelhetők meg (pl. *Microsporidia*-k, *Schizosaccharomyces*-ek, *Pneumocystis*-ek) (Keeling & Slamovits, 2004; Corradi és mtsi., 2007; Rhind és mtsi. 2011; Ma és mtsi., 2016; Waalwijk és mtsi., 2017; Legras és mtsi., 2018; Todd és mtsi., 2019; Giannakou és mtsi., 2020; Mba és mtsi., 2022). A genomokat érintő átrendeződések gyakorisága és a fajgazdagság nagy valószínűséggel összefügg: az Ascomycota törzs Pezizomycotina altörzse hatalmas fajgazdagsággal rendelkezik, szemben a Taphrinomycotina altörzsszel, amely csupán néhány tucat fajt számlál a jelenlegi tudásunk szerint, és az előbbi csoportban rengetek átrendeződés érintette a genomokat (Rajeh és mtsi., 2018).

A génsorrendre ható tényezőket legalaposabban az élesztők esetében vizsgálták, hiszen általában kisméretű, kompakt és jól szerveződő genommal rendelkeznek, amelyekben talán könnyebben tetten érhető a természetes szelekció aktivitása. Hurst és munkatársai (2002) bemutatták a *S. cerevisiae* és *C. albicans* génsorrendjének vizsgálatával, hogy a már említett ko-expresszió

mellett az intergénikus távolság (a távolság két gén kódoló régiója között) is meghatározó lehet a génsorrend fenntartásának szempontjából. Pál & Hurst (2003) ismertették, hogy a génsorrend változásai korrelálnak a különböző genomi régiók rekombinációs rátájával, továbbá az esszenciális gének alacsony rekombinációs rátájú genomszegmensekben lokalizálódnak. Mások munkái alapján leszögezhető, hogy az egy metabolikus útvonalhoz tartozó vagy stabil protein komplexeket kódoló gének is hajlamosak csoportosulni, szelekciós hatás alatt állni (Teichman & Veitia, 2004; Wong & Wolfe, 2005, Poyatos & Hurst, 2006). Az is bebizonyosodott, hogy viszonylag nagy távolságban lévő gének is alkothatnak ko-expressziós klasztereket, amely kromatin-szintű regulációból származhat (Lercher & Hurst, 2006). Poyatos & Hurst 2007-es munkájukban együttesen vizsgálták az előzőekben ismertetett jelenségeket (metabolikus kapcsolatok, ko-expresszió, ko-reguláció, intergénikus távolság (IGD), esszenciális gének lokális denzitása, rekombinációs ráta és protein-protein interakciós távolság) és arra a következtetésre jutottak, hogy az IGD és ko-expresszió a két legmeghatározóbb tényező a génsorrend megőrzésében. Érdekes módon az esszenciális gének lokális denzitása lett a harmadik legfontosabb paraméter, pedig e gének mellett sok esetben átlagos vagy éppen az átlagtól nagyobb az IGD. Másik érdekesség, hogy a ko-expresszió és ko-reguláció sem mindig jár kéz a kézben, legalábbis kiderült, hogy a megfigyelt ko-expresszióknak nem mindig a közös transzkripciós faktorok (tf-ok) általi cisz-reguláció az oka (Poyatos & Hurst, 2007; Tsai és mtsi., 2007). Egy másik tanulmány bemutatta, hogy még a divergens irányú génpárok (olyan génpárok, amelyek egymással ellentétes irányban transzkriptálódnak:  $\leftarrow\rightarrow$ ), amelyek közös tf kötő hellyel rendelkeznek sem mindig expresszálódnak együtt, a ko-expressziójuk körülményfüggő (Su és mtsi., 2010). Sugino & Innan (2012) megállapította, hogy az újonnan kialakuló génpárok esetében számottevően kevesebbnek bizonyulnak a divergens irányú génpárok, tehát a szelekció az IGD vesztes és ko-reguláció lehetősége ellen dolgozik. Bizonyos konvergens

génpárok (olyan génpárok, amelyek egymás irányába transzkriptálódnak:  $\rightarrow\leftarrow$ ) esetében kiderült, hogy közös transzkriptumot képezhetnek mRNS duplex formájában, amelyek a 3'-végüknél kapcsolódnak és szerepük lehet a poszttranszlációs regulációban (Gilet és mtsi., 2020). Az olyan génpárokról, amelyek nemcsak a kromoszómán közvetlen szomszédok, hanem egy metabolikus útvonalban is „szomszédosak” bebizonyosodott, hogy az általuk kódolt enzimek félkész termékei számos esetben toxikusak (McGary és mtsi., 2013). Ebből következően a megfelelő reguláció miatt is célszerű, ha az ilyen típusú gének egymás mellett helyezkednek el. Több tucat gombagenomot megvizsgálva be is igazolódott, hogy az ilyen típusú génpárok valóban szelekciós hatás alatt állnak (McGary és mtsi., 2013).

Azonban a véletlenszerűen bekövetkező külső hatások és az élesztő saját genomregulációja mellett számolnunk kell a transzpozonok génsorrendre gyakorolt hatásával is. A transzpozonok aktivitása egyaránt lehet pozitív vagy negatív kimenetelű: elősegíthetik az adaptációt, de káros mutációkat is okozhatnak (Gabriel és mtsi., 2006; Chang és mtsi., 2013; Michel és mtsi., 2017).

### **2.3. A szekvencia és struktúra változások kapcsolata**

Általánosan igaz az a megfigyelés, miszerint minél nagyobb a szekvenciabeli változásokra épülő filogenetikai távolság két genom között, annál nagyobb a kromoszómákat érintő struktúrabeli változások mértéke is (Rolland & Dujon, 2011; Drillon & Fischer, 2011). Ebből következően a két típusú változás erős korrelációt mutat számos esetben: például egyes gerincesek, fonálféreg, ízeltlábúak és élesztők esetében (Burt és mtsi., 1999; Coghlan & Wolfe, 2002; Sharakhov és mtsi., 2002; Vakirlis és mtsi., 2016). Azonban számos kivétel létezik, amikor is feltételezhetően közös eredetű genomok hasonló mértékű szekvenciaevolúció mellett eltérő mértékű átrendeződési dinamikát mutatnak és fordítva. Ez arra utal, hogy a különböző változások

„molekuláris órái” nem egyformán ketyegnek és nem minden esetben függnék egymástól (Dujon, 2010; Ho és mtsi., 2016; Luo & Ho, 2018).

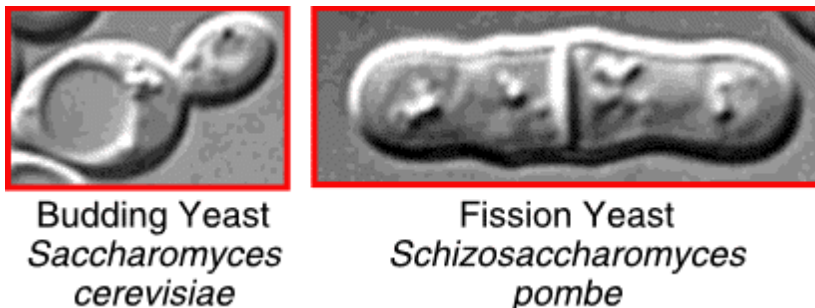
Érdekes példái ennek a jelenségnek az *Encephalitozoon* (Microsporidia) genomok. A szekvenciáik dinamikus változásaihoz képest átrendeződések sokkal kisebb számban fordulnak elő a genomjaikban (Stechmann, 2004). Ezzel szemben egyes *Fusarium* törzsekre jellemző, hogy egész kromoszómákat vagy kromoszóma-régiókat veszítenek (Williams és mtsi., 2016). Három *Aspergillus* faj komparatív genomikai összehasonlításánál is kiderült, hogy bár az *A. fumigatus* és *A. oryzae* közel egyforma mértékű szekvencia divergenciát mutat az *A. nidulans*-hoz képest, az *A. oryzae* genomját kétszer annyi átrendeződés érte, mint az *A. fumigatus* genomot (Galagan és mtsi., 2005). A *Drosophila melanogaster* és *D. yakuba* közeli rokonoknak számítanak, bár kereszteződni nem tudnak. A *D. melanogaster* vonalon számottevően kevesebb átrendeződés következett be, mint a *D. yakuba* esetében (Ranz és mtsi., 2007). Érdekes példa a csirke, egér és ember esete is. Az egér és az ember protein szekvencia szinten 88%-os átlagos egyezést mutatnak, míg az ember-csirke 75,3%-ost, de meglepő módon az utóbbi páros génsorrendjüket tekintve sokkal jobban hasonlít egymásra (Pevzner & Tesler, 2003; International Chicken Genome Sequencing Consortium, 2005).

A hasadó élesztők nemzetsége egy rendkívül érdekes csoport, mivel a genomjaik géntartalmához képest meglehetősen gyors a szekvenciájuk evolúciója (Rhind és mtsi., 2011; Rajeh és mtsi., 2018). Például a *Saccharomyces* vagy *Kluyveromyces* nemzetség tagjai kisebb mértékű szekvencia divergencia mellett géntartalmukat tekintve nagyobb mértékben térnek el egymástól, mint a hasadó élesztők (Rhind és mtsi., 2011).

## 2.4. A hasadó élesztőgombák bemutatása

### 2.4.1. A hasadó élesztőgombák, mint modellszervezetek

A hasadó élesztők (*Schizosaccharomyces*) nemzetségét jelenlegi tudásunk szerint 5 faj és számos változat alkotja. A legismertebb tag a *S. pombe* (Lindner, 1893), amely régóta közkedvelt modellszervezete a molekuláris folyamatok vizsgálatának (Hoffman és mtsi., 2015). A nemzetség egyedüli dimorf tagja a *S. japonicus* (Yukawa & Maki, 1931), amely kutatása még gyerekcipőben jár ugyan, de egyre több labor kezdi modellként használni. A további három közeli rokon faj a *S. cryophilus* (Helston és mtsi., 2010), *S. octosporus* (Beijerinck, 1894) és a *S. osmophilus* (Brysch-Herzberg és mtsi., 2019), amelyek inkább a *S. pombe*-re hasonlítanak, de számos egyedi tulajdonsággal is rendelkeznek. E fajok mellett előfordul még számos változat (pl. *S. maledivorans*, *S. liquefaciens*, *S. sloofiae*, *S. versatilis*, *S. longobardus*) is, amelyek megítélésében még ma sincs konszenzus (Brysch-Herzberg, 2019). A hasadásos típusú sejtosztódás az egyik legalapvetőbb tulajdonság, amely elkülöníti a hasadó élesztőket az összes többi élesztőtől (2. ábra).



**2. ábra.** A legismertebb sarjadzó élesztő (*S. cerevisiae*) és hasadó élesztő (*S. pombe*) sejtmorfológiája és osztódása. Forrás:

<https://joelhuberman.net/HubermanLabArchives/BIR572/YeastGeneticsLectureNotes.html> (letöltés dátuma: 2022.07.16)

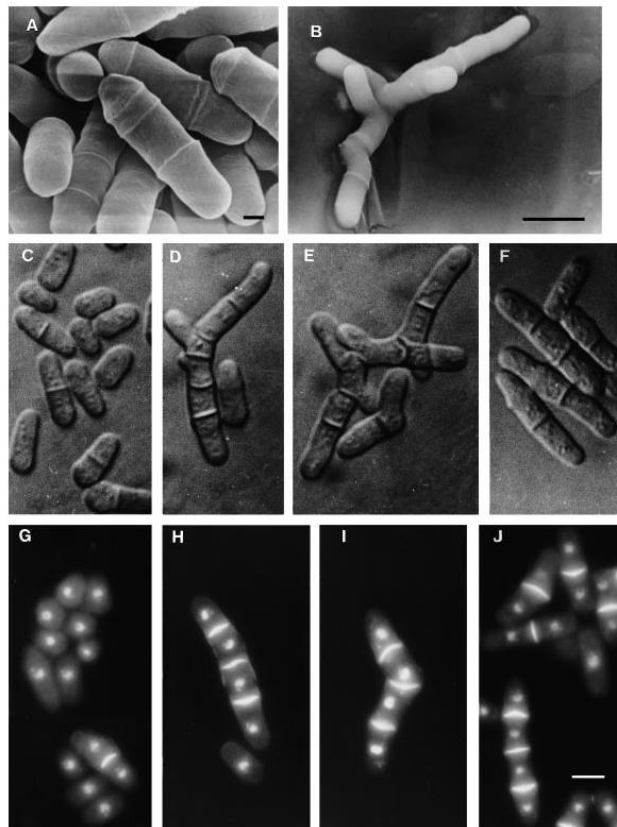
A két leginkább kutatott fajra (*S. pombe* és *S. japonicus*) általánosan jellemző, hasonlóan más élesztő modellszervezetekhez (pl. *S. cerevisiae*) a könnyű laboratóriumi tenyésztetőség, a rövid generációs idő, a fenntartható haploid állapot. Emellett, széles körű molekuláris eszköztár (genetikai transzformáció, plazmid vektorok) is rendelkezésre áll e fajoknál. A legújabb molekuláris technikákat is adaptálták a hasadó élesztőkre, úgymint az RNS szekvenálást, a CRISPR-Cas9 rendszert, a Hi-C és Chip-Seq technikákat (Rhind és mtsi., 2011; Jacobs és mtsi., 2014; Murray és mtsi., 2016; Matsuda és mtsi., 2017; Rodríguez-López és mtsi., 2017; Zhang és mtsi., 2018, Tong és mtsi., 2019). Ezáltal, valamint az eukarióta sejtfelepítésüknek köszönhetően, számos konzervatív molekuláris folyamat megismerésére nyílt (és nyílik) lehetőség általuk. Nagy előnye továbbá a hasadó élesztőknek, hogy jónéhány tulajdonságukban eltérnek a sarjadzó élesztőktől, így ugyanazon folyamatok változatainak megismerését teszik lehetővé. Ráadásul számos tekintetben jobban hasonlítanak a magasabbrendű eukariótákra, mint más élesztőkre. Ezek a különbségek teszik őket megkerülhetetlen modellszervezetekké.

A továbbiakban a *S. pombe* és a *S. cerevisiae* közötti néhány különbséget mutatjuk be. A *S. cerevisiae*-ből számos vad típusú törzs létezik, amelyeket használnak is a különböző laboratóriumokban és rendre előfordul, hogy ugyanaz a mutáció némileg eltérő fenotípust eredményez a különböző törzsekben (Elion és mtsi., 1991; Ralser és mtsi., 2012). Habár *S. pombe*-ből is több izolátum áll rendelkezésre, de szinte a világ minden táján ugyanazokat a laboratóriumi törzseket használják, amelyek egy izolátumból származnak (Fantes & Hoffman, 2016). Ezek a törzsek pedig az Urs Leupold által izolált 968 h<sup>90</sup> homotallikus és a belőle származó 972 h<sup>-</sup> és 975 h<sup>+</sup> heterotallikus törzsek. A *S. cerevisiae*-vel szemben ez jelentősen megkönnyíti az eltérő laborok *S. pombe*-vel kapcsolatos eredményeinek összevetését (Hoffman és mtsi., 2015). Emellett Rhind és mtsi. (2011), továbbá Hoffman és mtsi. (2015) munkái további különbségekre is rávilágítanak.

Bár a két élesztő genomja hasonló méretű, a *S. cerevisiae*-nek általában 16 kromoszómája van haploid állapotban (~12,1 Mb), de a diploid állapotot részesíti előnyben, míg a *S. pombe* 3 nagyméretű kromoszómával rendelkezik (~12,5 Mb) és a haploid állapot a természetes számára. Ez utóbbi egyszerűbbé teszi a génmutációk kialakítását, ezáltal a mutáns fenotípus vizsgálatát. Eltérnek a centromer-régiójuk szerveződésében is, mivel a hasadó élesztők nagy, repetitív régiókat tartalmazó centromerekkel rendelkeznek. Ezzel összefüggésben a *Schizosaccharomyces*-ekben megtalálható és aktív maradt az RNS interferenciához szükséges génekészlet is. Mindemellett a „spliceosome”-komponensek is jelen vannak a hasadó élesztő genomokban, bár egyértelmű bizonyíték alternatív „splicing”-ra legjobb tudásunk szerint egyelőre nincs. Számos hasadó élesztő protein mind szekvenciájában, mind szerkezetében is jobban hasonlít a Metazoa élőlények proteinjeihez, mint a sarjadzó élesztők proteinjeihez. A hasadó élesztő modellek segítségével könnyebben megérthetők a magasabb szerveződésű élőlények hasonló folyamatai, mint például a heterokromatikus hiszton metiláció folyamatai, a sejtciklus-kontroll, a citokinezis és számos szignál-transzdukciós útvonal, mivel a hasadó élesztők e folyamatokban is sokkal nagyobb hasonlóságot mutatnak a Metazoa élőlényekkel, mint a sarjadzó élesztők (Vyas és mtsi., 2021).

A kutatási eredmények közül kiemelkedő az a hasadó élesztőkkel kapcsolatos eredmény, amely Paul Nurse és csapata munkásságához köthető, aki a sejtciklus kutatásaiért megosztott orvosi Nobel-díjban részesült 2001-ben (Beach és mtsi., 1982; Lee & Nurse, 1987). A sejtciklus kutatások mellett a kromoszóma dinamika (replikáció, rekombináció, szegregáció), a gén-expresszió és az epigenetika területén is esszenciális modellt váltak a hasadó élesztők (Hirano és mtsi., 1986; Samejima és mtsi., 1993; Watanabe & Nurse, 1999; Mata & Bähler, 2003; Kitajima és mtsi., 2004; Forsburg & Rhind, 2006; Mojardín és mtsi., 2013; Allshire & Ekwall, 2015).

A tanszékünk sejtciklus kutatásokban végzett munkássága sem megkerülhető, hiszen a tanszék munkatársai izoláltak először sejtszeparációban sérült mutánsokat (3. ábra). E mutánsok segítségével számos kulcsfontosságú transzkripció faktor és enzimet kódoló gén szerepét sikerült tisztázni (Sipiczki és mtsi., 1993; Ribár és mtsi., 1997; Grallert és mtsi., 1999; Sipiczki & Bozsik, 2000; Zilahi és mtsi., 2000; Szilagyi és mtsi., 2002, 2005; Martín-Cuadrado és mtsi., 2003; Alonso-Nunez és mtsi., 2005; Lee és mtsi., 2005; Sipiczki, 2007; Miklos és mtsi., 2008; Batta és mtsi., 2009). Azonban az említett téma csak egy része a tanszék hasadó élesztőkkel végzett kutatási profiljának, viszont a többi téma és publikáció felsorolásától a terjedelmi korlátok miatt eltekintünk.



**3. ábra.** *S. pombe* L972 vad típusa (A, C, G) és különböző sejtszeparációban sérül mutansok (B, D-F, H-J) fotói. Forrás: Grallert és mtsi., 1999.

Megkerülhetetlen azonban a hasadó élesztők egyetlen dimorf fájának, a *S. japonicus*-nak az említése, ami egy feltörekvőben lévő modellszervezetnek számít (Klar, 2013). Mivel e faj jelentősen különbözik a többi hasadó élesztőtől, kiváló modellpárost alkotnak a *S. pombe*-vel. Hiszen a két faj párhuzamos használatával lehetőség nyílik ugyanazokat a jelenségeket egy élesztő és egy dimorf modellben tesztelni (Tanaka & Hirata, 1982; Alfa & Hyams, 1990; Bullerwell és mtsi., 2003; Linder & Gustaffson, 2008; Xu és mtsi., 2012; Gu és mtsi., 2012, 2015; Balazs és mtsi., 2012; Guo és mtsi., 2015; Domizio és mtsi., 2017; Princová és mtsi., 2019; Makarova és mtsi., 2020). A *S. japonicus* képes valódi invazív hifákat képezni különböző környezeti hatásokra reagálva, ami a dimorfizmus kutatás nem-patogén alapmodellévé teheti a fajt (4. ábra) (Sipiczki és mtsi., 1998a,b; Enczi és mtsi., 2007; Furuya & Niki, 2012; Papp és mtsi., 2014; Nozaki és mtsi., 2018; Kinnaer és mtsi., 2019; Gómez-Gil és mtsi., 2019; Papp és mtsi., 2021a,b).

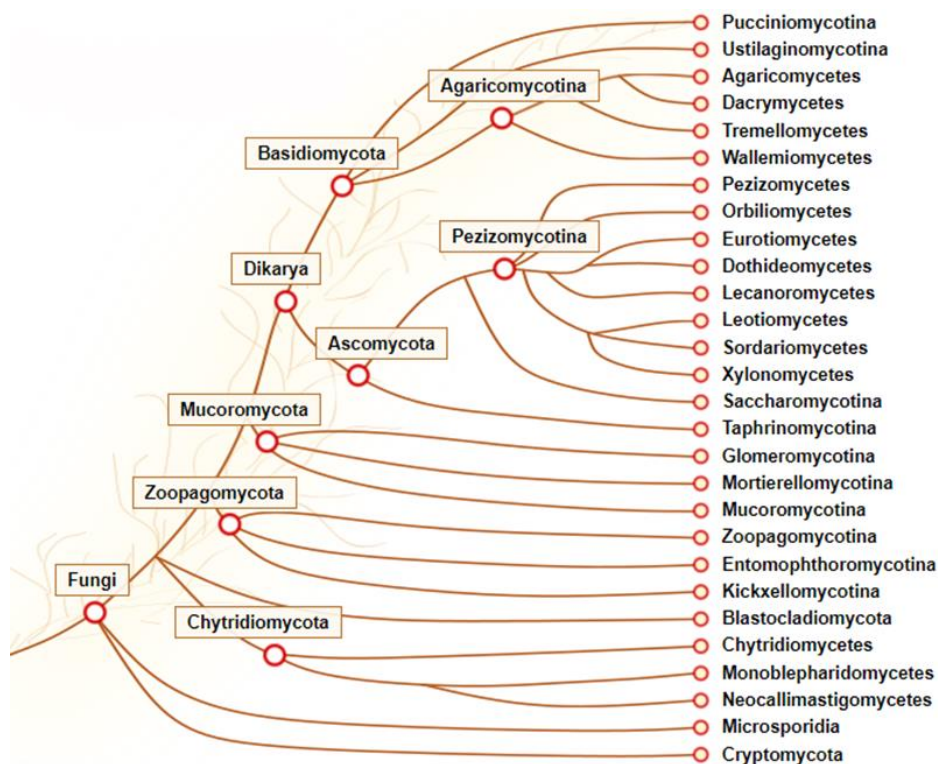


**4. ábra.** *S. japonicus* 7-1 vad típus élesztő állapotú sejtjei (A), egy ismeretlen sejtszeparációban sérült mutáns törzs sejtjei (B) és a vad típus invazív hifaképzése szilárd tápközegen (C). Forrás: Ács-Szabó Lajos

A hasadó élesztőket leginkább az alapkutatásokban hasznosítják, de néhány alkalmazott területen is felhasználják őket (pl. borászat) (Minnaar és mtsi., 2017; Benito 2019; Millarini és mtsi., 2020).

#### 2.4.2. A hasadó élesztőgombák filogenezeise és genom szerveződése

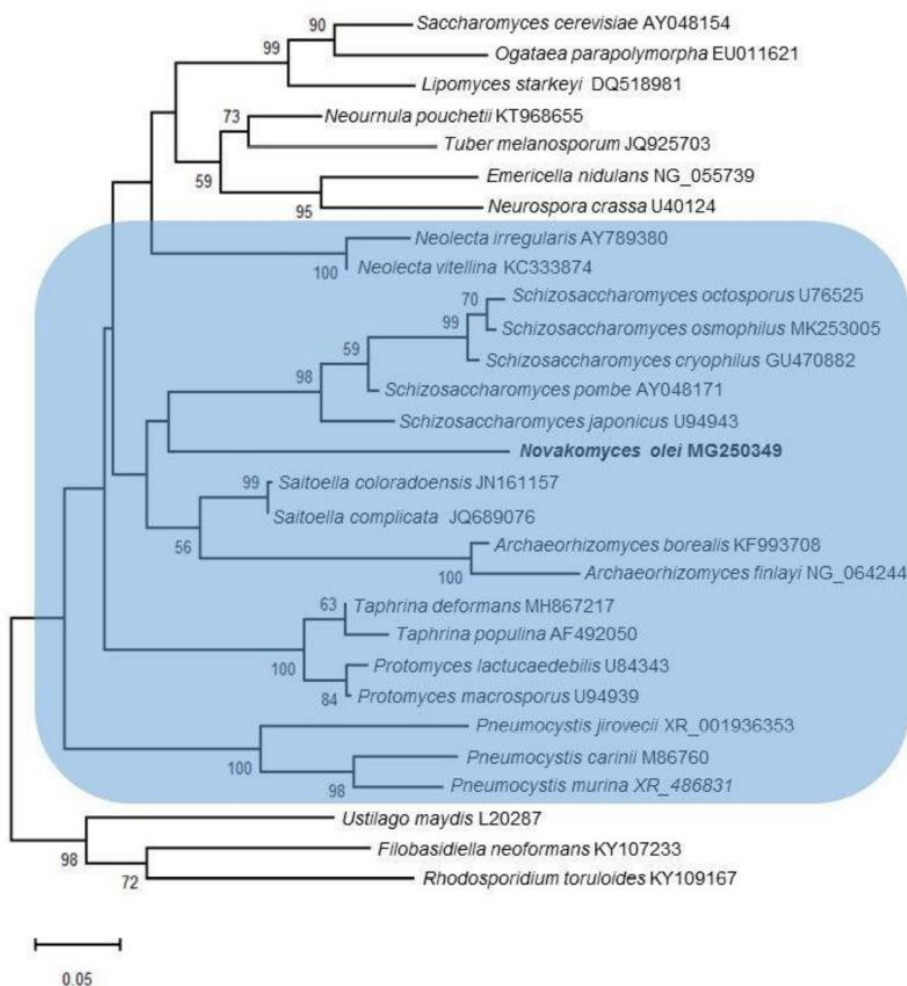
A hasadó élesztők rendszertanilag az Ascomycota törzsbe és a Taphrinomycotina (Archiascomycotina) altörzsbe tartoznak (5. ábra).



**5. ábra.** A gombák filogenezeise. A Taphrinomycotina altörzs az Ascomycota alapi ágához tartozik. Forrás: <https://mycocosm.jgi.doe.gov/mycocosm/home> (letöltés dátuma: 2022.07.16.)

Ez az altörzs rendkívül változatos gombanemzetségeket foglal magában (*Archaeoryzomyces*, *Novakomyces*, *Pneumocystis*, *Taphrina*, *Protomyces*, *Saitoella*, *Neolecta* és *Schizosaccharomyces*), aminek következtében felmerült a kérdés, hogy vajon a Taphrinomycotina altörzs nem egy filogenetikai műtermék-

e („long branch attraction” jelenség) (Sipiczki, 2000; Liu és mtsi., 2009)? Azonban a jelenleg rendelkezésre álló bizonyítékok és filogenetikai elemzések támogatják a Taphrinomycotina altörzs monofiletikus egységét (6. ábra) (Liu és mtsi., 2009; Rhind és mtsi., 2011; Rajeh és mtsi., 2018; Shen és mtsi., 2020; Li és mtsi., 2021; Čadež és mtsi., 2021).



**6. ábra.** A Taphrinomycotina altörzs főbb nemzetségeinek és azok néhány fájának filogenezise. A kézzel kiemelt fajok tartoznak a szóban forgó altörzsbe, a csoport legújabb leírt faja, a *Novakomyces olei* vastag betűvel kiemelve szerepel. Forrás: Čadež és mtsi., 2021

Azonban nemcsak az altörzs változatos, hanem maguk a hasadó élesztő fajok is. Olyannyira, hogy a múltban több alkalommal is felmerült a gondolat, miszerint külön nemzetségekbe lenne célszerű rendezni a fajokat: *S. pombe* → *Schizosaccharomyces*, *S. japonicus* → *Hasegawa*, *S. octosporus* → *Octosporomyces* (Kudriawzew, 1960; Yamada & Banno, 1987). Ezeket a változtatásokat viszont a genomszekvenciák vizsgálata szükségteenné tette, hiszen rendkívül nagy mértékű a hasonlóság a géntartalmukat tekintve a nemzetség tagjai között (Rhind és mtsi., 2011).

A *S. pombe* volt a hatodik eukarióta élőlény, amelyet megszekvenáltak teljes egészében és már ekkor egyértelművé vált, hogy drasztikusan eltérnek a hasadó élesztők a sarjadzó élesztőktől (Wood és mtsi., 2002). A további hasadó élesztő fajok genomszekvenciái majd' egy évtizeddel később váltak elérhetővé, amikor is a *S. japonicus* és *S. octosporus* genomok közel kromoszóma szintű összeszerelési állapotba kerültek, viszont a *S. cryophilus*-t csak nagyobb kontigokba rendezték (Rhind és mtsi., 2011). Jelen doktori dolgozat készítése ideje alatt a *S. osmophilus* faj genomszekvenciája még nem volt elérhető. Kiderült, hogy a hasadó élesztők általában 3 nagy méretű kromoszómával rendelkeznek, és a kódoló génjeik száma pedig 5000 körül mozog (*S. pombe* 5114 kódoló gén, *S. japonicus* 4886 kódoló gén) (Rhind és mtsi., 2011; Pombase; JaponicusDB). A hasadó élesztő fajok legalább 4218 közös feltételezett ortológ génnel rendelkeznek és az intronjaik közül is 2901 (~80%) megtalálható mind a 4 fajban, azonos pozícióban. Ez bizonyítja, hogy a hasadó élesztők mind a géntartalmukat, mind a génszerkezetüket tekintve rendkívül konzerválnak tekinthetők (Rhind és mtsi., 2011). Mindezek mellett a fajok génsorrendje is meglehetősen hasonlóságot mutat, hiszen a genomok közös géntartalmának jelentős része szinténikus blokkoka alkot (7. ábra).



**7. ábra.** A hasadó élesztők szinténia térképei. A legfelső sorban a *S. pombe*, a középső sorban a *S. octosporus*, míg az alsó sorban a *S. japonicus* látható referenciaként. Az egyszínű középvonalak a referencia faj kromoszómáit jelképezik, a középvonal feletti és alatti színes téglalapok pedig a másik két faj szinténikus régióit hivatottak bemutatni. A ciklámen szín az I. kromoszómát, a kék a II. kromoszómát, a sárga a III. kromoszómát jelöli. Forrás: Rhind és mtsi., 2011

Azonban a fajok szekvencia-evolúciója meglehetősen gyors: protein szekvencia szinten a *S. japonicus* ~55%-os átlagos aminosav-azonosságot mutat a másik három faj szekvenciáival, de a két legközelebbi rokonnak számító *S. octosporus* és *S. cryophilus* között is „csak” 85%-os azonosság figyelhető meg (Rhind és mtsi., 2011). Ezek az azonosságok megfeleltethetők az ember és a kacsacsőrű emlős, illetve az ember és kutya szekvenciabeli eltéréseinek (Sipiczki, 2000; Rhind és mtsi., 2011). További érdekesség, hogy más élesztő genuszokhoz (pl. *Saccharomyces* és *Kluyveromyces*) képest a hasadó élesztők nagyobb mértékű géntartalom konzerváltság mellett számottevően gyorsabb szekvencia-evolúciót mutatnak (Rhind és mtsi., 2011). Sőt, egy viszonylag friss tanulmány adataiból arra lehet következtetni, hogy nem is ismert az Ascomycota gombák között olyan genusz, amely hasonló mértékű genom konzerváltság mellett még csak közelítőleg is olyan mértékű szekvencia evolúciós rátát mutasson, mint a hasadó élesztők (Rajeh és mtsi., 2018).

Számos tanulmány foglalkozott a törzsek szekvencia és kariotípus változásainak dinamikájával, illetve a változások hatásával is. A korábbiakhoz képest érdekes eredmény, hogy a különböző geográfiai eredetű vad típusú *S. pombe* törzsek szekvencia szinten rendkívül nagymértékű hasonlóságot mutatnak (95-100%), viszont kariotípusukban eltérhetnek egymástól (Rhind és mtsi., 2011; Brown és mtsi., 2011; Jeffares és mtsi., 2015). Fawcett és mtsi. (2014) górcső alá vették számos *S. pombe* törzs teljes genomszekvenciáját és arra a megállapításra jutottak, hogy a törzsek szekvenciái nagy valószínűséggel erős szelekciós hatás alatt állhatnak, mivel még az intergénikus régiókban is nagyfokú hasonlóság volt megfigyelhető. Felvetették, hogy ezek a nem kódoló szekvenciák is funkcióval rendelkezhetnek, amely felvetés a későbbiekben bizonyításra is került (Grech és mtsi., 2019). Jeffares és mtsi. (2015) vizsgálatai alapján az átlagos diverzitás a *S. pombe* izolátumokon belül  $3 \times 10^{-3}$  „SNP/site” volt, ami körülbelül 3 SNP-t jelent kilobázisonként. Vizsgálataikban arra is rámutattak, hogy tisztító szelekció alatt főleg az exon, intron és UTR szekvenciák állnak. Ezzel összhangban megfigyelték, hogy a fenotípusos varianciáért inkább az inszerciók/deléciónak a felelősek, mintsem az SNP-k (Jeffares és mtsi., 2015). Avelar és mtsi. (2013) bebizonyították, hogy már egy inverzió vagy egy transzlokáció is képes nemcsak a változást érintő kromoszóma, hanem a teljes genom génexpressziós mintázatát megváltoztatni. Egyértelművé vált, hogy bár bizonyos kromoszóma strukturális változások jelenthetnek előnyt egyes környezeti körülmények között, de az általuk okozott reprodukív izoláció végső soron nagyobb evolúciós hátrányt jelent (Avelar és mtsi., 2013; Zanders és mtsi., 2014; Naumov és mtsi., 2015; Jeffares és mtsi., 2017). Tusso és mtsi. (2019) tanulmányukban megállapították, hogy a ma megfigyelhető biodiverzitás a *S. pombe* természetes izolátumai között, két korai divergens vonalra vezethető vissza (*S. pombe* és *S. kambucha* kládok). Harnqvist és mtsi. (2021) megvizsgálták jelentős számú paramétert, ami a nemzetközi szakirodalom alapján hozzájárulhat a szekvenciákat érintő szelekciós

nyomáshoz. Három fő variábilist találtak, amelyek sok esetben meghatározóak voltak az evolúciós ráták szempontjából: a protein-protein interakciós hálózatban betöltött hely, az aminosav kompozíció és a celluláris lokalizáció (citoplazma, organelumok).

### 3. Célkitűzések

Annak érdekében, hogy megérthessük a hasadó élesztőgombák stabil genomszerveződésének genetikai és evolúciós hátterét, különböző komparatív vizsgálatoknak kívántuk alávetni a genusz azon négy faját, amelyek rendelkeznek ismert genom szekvenciával.

Mivel a *S. cryophilus* megszekvenált genomjának összeszerelése csak részlegesen történ meg (Rhind és mtsi., 2011), célul tűztük ki a faj nagy szuperkontigjainak összeszerelését, majd az összeszerelés validálását különböző molekuláris biológiai módszereket alkalmazva. Az összeállított genom szekvencia birtokában lehetőségünk volt összehasonlítani a kromoszómák átrendeződési dinamikáit közeli és távoli rokon fajok esetében. Ezt követően meg kívántuk vizsgálni a hasadó élesztők szekvencia és kromoszóma struktúra evolúciójának kapcsolatát, a fajok közötti DNS szintű genomkonzerváltság mértékét. Lehetőségünk nyílt a fajok legközelebbi közös ősetől származó genom-szegmensek felderítésére, majd a beazonosított szegmensek tulajdonságainak vizsgálatára. Végül megkíséreltük feltárni a konzervált régiók fennmaradásának néhány lehetséges okát, azok egymásra hatását.

## 4. Anyagok és módszerek

### 4.1. A doktori munka során felhasznált törzsek

A doktori tanulmányaim során kísérleti munkában felhasznált élesztő törzsek listáját az 1. táblázat tartalmazza.

#### 1. táblázat. Felhasznált törzsek listája.

Fajnév	Leírás	Forrás
<i>Schizosaccharomyces pombe</i>	Vad típus L-972 h <sup>r</sup>	U. Leopold
<i>Schizosaccharomyces cryophilus</i>	Vad típus CBS11777	CBS
<i>Saccharomyces cerevisiae</i>	Vad típus S288C	ATCC

### 4.2. A doktori munka során felhasznált tápközegek

Az élesztő sejtek tenyésztésére használt komplex tápközegek és összetevőik listáját a 2. táblázat tartalmazza.

#### 2. táblázat. Felhasznált tápközegek listája.

Megnevezés	Összetétel		Leírás
YEL (1000 ml)	Glükóz	30,0 g	Komplex tápfolyadék
	Élesztőkivonat	10,0 g	
YEA (1000 ml)	YEL	1000 ml	Komplex táptalaj
	Poragar	20,0 g	
YPL (1000 ml)	Glükóz	30,0 g	Komplex tápfolyadék
	Élesztőkivonat	10,0 g	
	Kazein pepton	10,0 g	
YPA (1000ml)	YPL	1000 ml	Komplex táptalaj
	Poragar	20,0 g	

### 4.3. A doktori munka során alkalmazott kísérleti módszerek

#### 4.3.1 A sejtek tenyésztési körülményei

A *S. cerevisiae*, *S. pombe* és *S. cryophilus* élesztőket YEA vagy YPA tartalmú Petri-csészéken tartottunk fenn, illetve a kísérletekhez 100-200 ml YEL vagy YPL tápfolyadékban tenyésztettük őket 30°C-on, illetve a *S. cryophilus* 25°C-on, mivel ez utóbbi faj nem kedveli a magasabb hőmérsékletet (Helston és mtsi., 2010).

#### 4.3.2. Genomi DNS izolálása

A 3. táblázatban látható a genomi DNS (gDNS) izolálásához használt oldatok és anyagok listája.

#### 3. táblázat. A gDNS izolálásához felhasznált oldatok listája.

Oldatok	Összetétel		Felhasználás
0,5 M EDTA (pH=8,0)	EDTA	25,0 g	TE, TBE
	NaOH	186,1 g	
	Millipore MilliQ víz	1000 ml	
10x TE	0,5 M EDTA	2 ml	gDNS izolálása
	1 M TRIS	10 ml	
	Millipore MilliQ víz	889 ml	
„Lysis” puffer	1 M TRIS	0,5 ml	gDNS izolálása
	0,5 M EDTA	0,1 ml	
	5 M NaCl	1 µl	
	Triton X	1 µl	
	10 %-os SDS	1 ml	
	Millipore MilliQ víz	48,398 ml	

A *S. cryophilus* sejtek tenyésztése 100 ml YEL vagy YPL tápoldatban egy éjszakán át 25°C-on történt. Másnap reggel 50 ml-es falcon csövekben

lecentrifugáltuk (4000 rpm, 5 perc), majd vízzel átmostuk a sejteket. Újabb centrifugálás (4000 rpm, 5 perc) és vizes mosás után 2 ml MilliQ vízzel felszuszpendáltuk a sejtüledéket, és az elegyből 1 ml-t eppendorf csőbe vittünk át. Ismételt centrifugálás (2500 rpm, 5 perc) és a felülúszó eltávolítása után hozzáadtunk 0,2 ml „lysis” puffert, 0,2 ml fenol:kloroform:izo-amilalkoholt (25:24:1; Sigma P2069) és ~0,3 g üvegyöngyöt (425-600 µm, Sigma G8772-10G). Az eppendorf cső parafilmmel történő lezárását követően 4 percen keresztül vortexeltünk, azután hozzáadtunk 0,2 ml 1x TE-t, végül lecentrifugáltuk (2500 rpm, 5 perc) az elegyet. A tiszta felső fázist új eppendorf csőbe vittük át és hozzáadtunk 1 ml abszolút etanolt, majd -20°C-on tároltuk 10 percig. Centrifugálás (2500 rpm, 5 perc) és 70%-os etanolos mosás után hozzáadtunk 0,4 ml 1x TE-t és 3 µl 10 mg/ml-es RNáz enzimet (Sigma, R5000) és 37°C-on inkubáltuk 15-20 percig. Ezt követően hozzámértünk 40 µl 3 M nátrium-acetátot (Sigma, S7899) és 1 ml abszolút etanolt (Spektrum 3D, 3.05032). -20°C-on 10 percig történő tárolás után ismét lecentrifugáltuk (10000 rpm, 10 perc) és a felülúszó alapos eltávolítása után szárítottuk a mintát. Legvégül 50 µl 1x TE-ben vettük fel az üledéket és későbbi felhasználásig -20°C-ra helyeztük. A minták tisztaságát gélelektroforézissel és nanodrop-pal (Biosystems ACTGene UVS-99) ellenőriztük, illetve mértük.

#### 4.3.3. Polimeráz láncreakció (PCR)

A doktori munka során felhasznált PCR primerek listáját a 4. táblázat tartalmazza.

#### **4. táblázat. A felhasznált PCR primerek listája.**

<b>Primer gyűjteményi száma</b>	<b>Primer neve</b>	<b>Szekvencia (5'-3')</b>	<b>Leírás</b>
664.	Sc2 Forw_A	gggacctttgg gtttcaagt	Az Sc2* és Sc6* összetartozásának validálására
665.	Sc6 Rev_A	gegtgcctc cttatcaat	Az Sc2* és Sc6* összetartozásának validálására
666.	Sc9 Forw_B	tagtttatggcc gccacagt	Az Sc9* és Sc1* összetartozásának validálására

667.	Sc1 Rev_B	ccgtctgcttct cagtttg	Az Sc9* és Sc1* összetartozásának validálására
668.	Sc5 Forw_C	gcttcaagctg ccacatttt	Az Sc5* és Sc8* összetartozásának validálására
669.	Sc8-Rew-C	gcgatctcttta gcatttcca	Az Sc5* és Sc8* összetartozásának validálására
794.	Sc6endF	ggaaataccttt tggcgact	Az Sc6* és Sc8* összetartozásának validálására
795.	Sc8startR	ggtctaagggg gcagattta	Az Sc6* és Sc8* összetartozásának validálására
49.	NL1	gcataatcaataa gcggaggaaa ag	Riboszómális DNS lokalizálása
50.	NL4	ggtccgtgtttc aagacgg	Riboszómális DNS lokalizálása
926.	SPOG_04999	tgttggtgtgat gagcagc	Riboszómális DNS lokalizálása
929.	SPOG_00001	agctcctgtaac gattgctac	Riboszómális DNS lokalizálása
930.	SPOG_00002	gatggatcact ggacaggtg	Riboszómális DNS lokalizálása
928.	18S rDNA1	tcattacggcg gtcctagaa	Riboszómális DNS lokalizálása
927.	18S rDNA2	ttctaggaccg ccgtaatga	Riboszómális DNS lokalizálása

\* Sc1, Sc2, Sc5, Sc6, Sc8 és Sc9: *S. cryophilus* szuperkontigok

A PCR reakciókhoz DreamTaq DNS polimerázt (Thermo Scientific EP0711) használtunk. Az 5. táblázat tartalmazza a PCR reakció összetételét, a 6. táblázat pedig a használt programokat hivatott bemutatni.

### 5. táblázat. A PCR reakcióelegy összetétele.

Összetevők	Térfogat (50 µl végtérfogat esetén)
<i>S. cryophilus</i> gDNS	1 µl (~100 ng)
DreamTaq Green puffer 5×	10 µl
Thermo Scientific dNTP mix 10 mM (R0191)	2 µl
Forward primer (5 µM)	2 µl
Reverse primer (5 µM)	2 µl
DreamTaq DNS polimeráz (5U/µl)	0,5 µl
MilliQ víz	32,5 µl

**6. táblázat.** A PCR protokoll bemutatása.

Hőmérséklet	Idő	Ciklus
95°C	3 perc	
95°C	0,5 perc	25-30x
Primer TM-2°C	0,5 perc	
72°C	0,5-3 perc	
72°C	10 perc	
4-10°C	∞	

4.3.4. Agaróz gélelektroforézis

Az izolált gDNS és amplifikált PCR termékek ellenőrzését gélelektroforézissel végeztük. A munka során 1%-os agaróz (BioRad 162-3102) gélt használtunk, 1x TBE pufferben, amelybe kiöntés előtt 8 µl ethidium-bromidot (10 mg/ml törzsoldatból, Sigma E8751-5G) tettünk. A futtatást 30-60 percen át 120 Volton végeztük. A gélben a kisebb molekulák a gél ellenállását könnyebben legyőzve gyorsabban mozognak, így hosszabb utat tesznek meg. A DNS helyzetét UV transzilluminátor (BioDoc-It Imaging System) segítségével tettük láthatóvá. A DNS mintánk méretének meghatározásához 1 kb méretű markert (Thermo Scientific, SM0332) használtunk, amelyet a mintánkkal párhuzamosan futtattunk. A gélelektroforézishez felhasznált oldatok összetételét a 7. táblázat tartalmazza.

**7. táblázat.** A gélelektroforézishez felhasznált oldatok listája.

Oldatok	Összetétel		Felhasználás
10x TBE puffer	TRIS-HCl	109,0 g	Gélelektroforézis
	Bórsav	55,6 g	
	0,5 M EDTA	40 µl	
	Millipore MilliQ víz	1000 ml	
1%-os agaróz	Agaróz	4,0 g	Gélelektroforézis
	1x TE	400 ml	

#### 4.3.5. Szekvenálás

Egyes PCR termékek esetében szükséges volt, hogy szekvenálásnak is alávéssük a mintákat. A szekvenálandó DNS fragmentumokat Na-acetátos kicsapással tisztítottuk, szükség esetén gélből visszaizoláltuk. Előbbi módszer keretében 0,1 térfogat Na-acetát és 2,5 térfogat abszolút etanol hozzáadásával a PCR termékeket kicsaptuk, majd 1x TE-ben vagy MilliQ vízben újra felvettük. Minden esetben tisztított PCR amplifikátumot küldtünk szekvenálásra a felszaporításhoz használt primerekkel egyetemben. A szekvenáltatásokat a GeneArt és a Microsynth AG szolgáltatókkal végeztettük.

#### 4.3.6. Pulzáló erőterű gél elektroforézis

A törzsek elektrokariogramját BioRad CHEF DR III (BioRad 170-3695) pulzáló erőterű gélelektroforézis rendszerrel készítettük el. A hasadó élesztő minták agaróz dugóba ágyazását a következő módon hajtottuk végre. A kiindulási sejtenyészeteket 30°C-on, illetve 25°C-on 200 ml YEL-ben vagy YPL-ben inkubáltuk egy-két napon át. A sejt-szuszpenziót centrifugáltuk 4000 rpm-en 5 percig, majd a leülepitett sejteket a tápfolyadék eltávolítását követően 4 ml 50 mM EDTA (pH=8) oldattal mostuk, és ismét centrifugáltuk a korábbi paraméterekkel. Felvettük a mintákat 5 ml W pufferben (8. táblázat) és szonikáltuk 2 percen keresztül. Ezt követően a sejt-szuszpenzióhoz adtunk 20 mg Lysing és 50 mg Novozyme enzimet, továbbá 50 µl β-merkaptóetanolt és 2 órán át inkubáltuk 37°C-on. Újabb centrifugálás és felülúszó eltávolítás után felvettük a leemésztett sejtfalú sejteket 500 µl S pufferben (8. táblázat). Ezzel párhuzamosan 1,5%-os Low Melting (Sigma A 9414) agarózt készítettünk szintén 500 µl S pufferben és összekevertük azokat a sejtekkel. Dermedést követően az elkészült dugókat 0,75 mg Proteinase-K (Sigma P 2308) enzimet tartalmazó „Lysis” pufferben (8. táblázat) inkubáltuk 1 éjszakán át 37°C-on. Másnap a folyadék eltávolítása után 1x TE oldatban áztattuk a dugókat 20 percig 37°C-on, majd szobahőmérsékleten is. Ekkor vagy felhasználtuk a dugókat

azonnal, vagy 0,5 M EDTA oldatban 4°C-on tároltuk őket későbbi felhasználás céljából. A felhasznált oldatok összetételét a 8. táblázat tartalmazza.

Az agaróz dugó készítése *S. cerevisiae* esetén Nguyen & Gaillardin (1997) leírása alapján történt.

**8. táblázat.** A PFGE-hez felhasznált oldatok listája.

Oldat	Összetétel	Felhasználás
S puffer	1,2M szorbit	Agaróz dugó készítése
	50mM Na-citrát	
	50mM Na <sub>2</sub> HPO <sub>4</sub>	
W puffer	1,2M szorbit	Agaróz dugó készítése
	50mM Na-citrát	
	50mM Na <sub>2</sub> HPO <sub>4</sub>	
	40mM EDTA (pH=8)	
Lysis puffer	10mM TRIS-HCl (pH=8)	Agaróz dugó készítése
	0,45M EDTA (pH=8)	
	1% Na-Lauril-Szarkozin	

Az elektroforetikus elválasztást 14°C-os 0,5x TBE pufferben végeztük 1%-os agaróz (BioRad 162-3102) gélben, 50 V-on (1,5 V/cm) a következő futtatási paraméterekkel: (1) 48 h 40 min; (2) 70 h 50 min; és (3) 24 h 55 min (Naumov és mtsi, 2015).

Az elválasztást követően a gélt 5 µl/ml ethidium-bromidot tartalmazó szűrt (Millipore) vízben festettük, steril desztillált vízben mostuk, majd UV fény segítségével fényképeztük.

**4.4. A doktori munka során alkalmazott bioinformatikai módszerek**

A doktori munka során használt bioinformatikai módszerek esetében nagyrészt online dolgoztunk, hiszen többnyire nagy teljesítményű szerverek állnak a bioinformatikai platformok mögött és ezáltal felgyorsíthatók bizonyos munkafolyamatok. A dolgozatban a programoknál és algoritmusoknál felsorolt

paraméterek esetében az eredeti angol kifejezéseket használtuk, mivel sok esetben nehézkes ezek megfelelő magyar nyelvre történő fordítása.

#### 4.4.1. Programok, algoritmusok, adatbázisok

A felhasznált interneten keresztül elérhető adatbázisok, programok és algoritmusok, továbbá az asztali számítógépen futtatott programok listáját a 9. táblázat tartalmazza. A felhasználásuk részletes bemutatása a későbbi fejezetekben történik.

#### 9. táblázat. A munka során felhasznált programok, algoritmusok listája.

Online programok/algoritmusok	Felhasználás	Elérhetőség
PSA	Páronkénti szekvencia illesztés	<a href="http://www.ebi.ac.uk/Tools/psa/">http://www.ebi.ac.uk/Tools/psa/</a>
Clustal Omega	Többszörös szekvencia illesztés	<a href="http://www.ebi.ac.uk/Tools/msa/clustalo/">http://www.ebi.ac.uk/Tools/msa/clustalo/</a>
MUSCLE	Többszörös szekvencia illesztés	<a href="http://www.ebi.ac.uk/Tools/msa/muscle/">http://www.ebi.ac.uk/Tools/msa/muscle/</a>
MAFFT v. 7.221	Többszörös szekvencia illesztés	<a href="http://mafft.cbrc.jp/alignment/server/">http://mafft.cbrc.jp/alignment/server/</a>
GBLOCKS	Illesztések felülvizsgálata	<a href="http://www.phylogeny.fr/one_task.cgi?task_type=gblocks">http://www.phylogeny.fr/one_task.cgi?task_type=gblocks</a>
Neighbor Joining (NJ)	Filogenetika	<a href="http://mafft.cbrc.jp/alignment/server/phylogeny.html">http://mafft.cbrc.jp/alignment/server/phylogeny.html</a>
Phylogeny.fr	Filogenetika	<a href="http://www.phylogeny.fr/alacarte.cgi">http://www.phylogeny.fr/alacarte.cgi</a>
YASS Genomic Similarity Search Tool	DNS dot plot	<a href="http://bioinfo.lifl.fr/yass/yass.php">http://bioinfo.lifl.fr/yass/yass.php</a>
OrthoclusterDB	Szinténia vizualizációja	<a href="http://genome.sfu.ca/cgi-bin/orthoclusterdb/runortho.cgi">http://genome.sfu.ca/cgi-bin/orthoclusterdb/runortho.cgi</a>
Genome Synteny Viewer (GSV) (megszűnt)	Szinténia vizualizációja	<a href="http://cas-bioinfo.cas.unt.edu/gsv/homepage.php">http://cas-bioinfo.cas.unt.edu/gsv/homepage.php</a>

SimpleSynteny	Szinténia vizualizációja	<a href="https://www.dveltri.com/simplesynteny/">https://www.dveltri.com/simplesynteny/</a>
GRIMM v2.01	Átrendeződések becslése	<a href="http://grimm.ucsd.edu/cgi-bin/grimm.cgi">http://grimm.ucsd.edu/cgi-bin/grimm.cgi</a>
Random Integer Generator	Statisztika	<a href="https://www.random.org/integers/">https://www.random.org/integers/</a>
<b>Számítógépen futtatott programok/algorithmusok</b>	<b>Felhasználás</b>	<b>Elérhetőség</b>
SnapGene Viewer	Szekvencia kezelés	<a href="http://www.snapgene.com/products/snapgene_viewer/">http://www.snapgene.com/products/snapgene_viewer/</a>
FigTree v1.4.2	Filogenetikai fa megjelenítése	<a href="http://tree.bio.ed.ac.uk/software/figtree/">http://tree.bio.ed.ac.uk/software/figtree/</a>
Archaeopteryx	Filogenetikai fa megjelenítése	<a href="https://sites.google.com/site/cmzmasek/home/software/archaeopteryx">https://sites.google.com/site/cmzmasek/home/software/archaeopteryx</a>
Mauve	Teljes genom illesztés	<a href="http://darlinglab.org/mauve/mauve.html">http://darlinglab.org/mauve/mauve.html</a>
Artificial Life Framework (ALF)	Átrendeződések szimulációja	<a href="http://alfsim.org/#index">http://alfsim.org/#index</a>
Past4	Statisztika	<a href="https://past.en.lo4d.com/windows">https://past.en.lo4d.com/windows</a>
Microsoft Excel	Statisztika/adatelemzés	
Parser (Python szkript)	Adatelemzés	Fejlesztő: Tajti Ákos
Genetic Modelling (Python szkript)	Átrendeződések modellezése	Fejlesztő: Dr. Papp László Attila
<b>Online adatbázisok</b>	<b>Felhasználás</b>	<b>Elérhetőség</b>
Broad Institute (megszűnt)	BLASTp, szekvencia letöltés, ortológ ellenőrzés	<a href="https://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/schizosaccharomyces-genomes-project">https://www.broadinstitute.org/scientific-community/science/projects/fungal-genome-initiative/schizosaccharomyces-genomes-project</a>
Pombase	BLASTp, szekvencia letöltés, ortológ ellenőrzés, GO kategóriák, esszenciális gének	<a href="https://www.pombase.org/">https://www.pombase.org/</a>
NCBI: National Center for Biotechnology Information	BLASTp, szekvencia letöltés	<a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a>

EnsemblFungi	BLASTp, szekvencia letöltés, ortológ ellenőrzés	<a href="http://fungi.ensembl.org/index.html">http://fungi.ensembl.org/index.html</a>
JGI Genome Portal	BLASTp, szekvencia letöltés	<a href="http://genome.jgi.doe.gov/">http://genome.jgi.doe.gov/</a>
Saccharomyces Database	BLASTp, szekvencia letöltés, ortológ kuráció	<a href="https://www.yeastgenome.org/">https://www.yeastgenome.org/</a>
Saccharomyces sensu stricto	BLASTp, szekvencia letöltés, ortológ kuráció	<a href="http://sss.genetics.wisc.edu/cgi-bin/s3.cgi">http://sss.genetics.wisc.edu/cgi-bin/s3.cgi</a>

#### 4.4.2. Szekvencia adatok

A doktori munka során használt szekvenciák és adataik a 10. táblázatban láthatók.

**10. táblázat.** A munka során felhasznált genomszekvenciák listája.

Fajnév	Törzs	Kromoszóma szám	Genom méret (Mb)	Hivatkozás
<i>Schizosaccaromyces japonicus</i>	yFS275	3	11,7332	Rhind és mtsi., 2011
<i>S. pombe</i>	972 h-	3	12,5913	Rhind és mtsi., 2011
<i>S. octosporus</i>	yFS286	3	11,6345	Rhind és mtsi., 2011
<i>S. cryophilus</i>	OY26	3	11,5548	Rhind és mtsi., 2011
<i>Pneumocystis murina</i>	B123	17	7,50134	Ma és mtsi., 2016
<i>Saitoella complicata</i>	NRRLY-17804	Draft*	14,1362	Riley és mtsi., 2016
<i>Taphrina deformans</i>	PYCC 5710	Draft*	13,7735	Cissé és mtsi., 2013
<i>Protomyces lactucae-debilis</i>	12-1054	Draft*	12,93	Mondo és mtsi., 2017
<i>Neolecta irregularis</i>	DAH-3	Draft*	14,18	Nguyen és mtsi., 2017
<i>Neurospora crassa</i>	OR74A	7	41,1024	Galagan és mtsi., 2003

<i>Botrytis cinerea</i>	B05.10	16	42,6301	Amselem és mtsi., 2011
<i>Aspergillus nidulans</i>	FGSC A4	8	29,8283	Galagan és mtsi., 2005
<i>Yarrowia lipolytica</i>	CLIB122	6	20,5509	Dujon és mtsi., 2004
<i>Debaryomyces hansenii</i>	CBS76	8	12,1819	Dujon és mtsi., 2004; Sacerdot és mtsi., 2008
<i>Meyerozyma guilliermondii</i>	ATCC 6260	Draft*	10,61	Butler és mtsi., 2009
<i>Saccharomyces cerevisiae</i>	S288C	16	12,1571	Liti és mtsi. 2009
<i>Saccharomyces uvarum</i>	CBS7001	16	11,4775	Scannell és mtsi., 2011
<i>Naumovozyma castelli</i>	CBS 4309	10	11,2195	Gordon és mtsi., 2011
<i>Cryptococcus gattii</i>	R265	14	18,3748	D'Souza és mtsi., 2011; Farrer és mtsi., 2015
<i>Ustilago maydis</i>	521	23	19,6644	Kämper és mtsi., 2006
<i>Rhizopus oryzae</i>	99-880	Draft*	39,0638	Ma és mtsi., 2009

\* Draft: a kifejezés arra utal, hogy e fajok esetében csak részleges genom összeszerelés történt, a szekvenciák csak fragmentálva érhetők el (adott esetben sok száz kontig).

A szekvenciákat minden esetben letöltöttük „fasta” vagy egyszerű szöveg (.txt) formátumban a megfelelő adatbázisokból. Az annotált kromoszóma fájlokat a SnapGene Viewer programban kezeltük.

#### 4.4.3. Genom szekvencia illesztések

Páronkénti és többszörös teljes genom illesztéseket az online YASS (Noé & Kucherov, 2005) és az asztali gépen futtatott Mauve (Darling és mtsi., 2010) programokkal végeztünk.

A YASS segítségével DNS szekvenciák illeszthetők páronként. Eredményként egy interaktív „dot plot” ábrázolást kapunk, ahol az egyik szekvencia az X-tengelyhez, a másik pedig az Y-tengelyhez van hozzárendelve. Ha az illesztett két szekvencia identikus, akkor a program azt folytonos középvonallal jelöli. Minél inkább eltér a két szekvencia, annál inkább eltérnek a pontok a középvonaltól, míg az illeszthetetlen szakaszok koordinátái üresen maradnak, úgynevezett „gap”-ek alakulnak ki. Ezzel a módszerrel gyorsan felmérhető a genom szekvenciák globális hasonlósága, továbbá láthatóvá válnak a kromoszómákat érintő átrendeződések (transzlokációk, inverziók) is. Az elkészült „dot plot” ábra letölthető nagy felbontásban a hozzá tartozó statisztikával együtt, továbbá a lokális illesztések elérhetőek fasta és tabuláris formátumban is. Esetünkben szükséges volt a letöltött genom szekvenciák összefűzése (konkatenálása), amelyeket aztán „txt” formátumban mentettünk el. A szekvenciák online felületre történő feltöltését követően, a következő illesztési paramétereket állítottuk be: „E value”: 1.0E-30 ( $1 \times 10^{-30}$ ); „X-drop”: 50; „window range”: 100-200000; „window incr.”: 2x; „hit criterion”: double. A többi paraméter esetében az alapértelmezett beállításokkal dolgoztunk.

A Mauve teljes genom szekvenciák illesztésére alkalmas program. Ellentétben a YASS-szal, a Mauve algoritmus képes akár több genom egyszerre történő illesztésére is, ami lehetővé teszi a genomszekvenciák közös régióinak feltárását. Emellett képes kiszűrni az ismétlődő szekvenciákat is, amelyek gyakran megnehezítik a pontos illesztéseket. A program lokálisan kolineáris blokkokat (LCBs – locally collinear blocks) keres, figyelembe véve a genomokat ért változásokat (transzlokációk, inverziók, duplikációk, deléciók), majd ezeket a rövidebb szakaszokat próbálja összefűzni a felhasználó által megadott paraméterek szerint. Ezáltal a végeredmény közvetett módon felhasználható a genomokat potenciálisan ért változások jellemzésére és számszerűsítésére, továbbá a szinténikus kapcsolatok felfedésére is. Munkánk során a „progressiveMauve” algoritmust használtuk a „use seed families” opció

bekapcsolásával, továbbá a „minimum LCB weight” értékét 40-re állítottuk be. A többi paraméter esetén az alapértelmezett beállításokat használtuk. Maximum négy genomot illesztettünk egyszerre, mivel a program hardverigénye a genomok számának növekedésével exponenciálisan nő.

#### 4.4.4. BLAST analízis és szekvencia összehasonlítások

Feltételezett ortológok keresésére protein BLAST módszert alkalmaztunk a 9. táblázatban felsorolt adatbázisokban, azok alapértelmezett keresési beállításait alkalmazva. Az értékelésnél figyelembe vettük a kapott „E” értékeket, a protein szekvenciákban található domének jelenlétét vagy hiányát, továbbá a proteinek kódoló gének kromoszómális környezetét (szomszédos gének). A releváns találatokat reciprok BLAST-tal ellenőriztük. Valamint ortológok keresésénél felhasználtuk a Fungal Orthogroups Repository (megszűnt) (Wapinski és mtsi, 2007) és a *Saccharomyces* Sensu Stricto Resources (<http://sss.genetics.wisc.edu/cgi-bin/s3.cgi>) (Scannell és mtsi, 2011) adatait is.

Páronkénti szekvencia összehasonlításokra (PSA) Needleman-Wunsch (Needleman and Wunsch, 1970) algoritmust használtunk (9. táblázat).

Több szekvencia egyidejű illesztésére progresszív és iteratív módszereket alkalmaztunk, így Clustal Omega (Sievers és mtsi, 2011), MUSCLE (Edgar, 2004) és MAFFT (Katoh & Standley, 2013) algoritmusokat egyaránt (9. táblázat).

#### 4.4.5. Szinténia analízis

A különböző fajok közötti szinténikus kapcsolatokról a YASS és Mauve által generált illesztésekből vontunk le általános következtetéseket. Vizualizációhoz a Genome Synteny Viewer GSV (Revanna és mtsi., 2011), SimpleSynteny (Veltri és mtsi., 2016) és OrthoClusterDB (Ng és mtsi., 2009) online platformokat használtuk (9. táblázat). Előbbi kettőt rövidebb szakaszok, kromoszómaméretlek bemutatására, míg az utóbbit teljes genomok

szinténiájának megjelenítésére alkalmaztuk. Mindhárom esetben a manuális BLAST analízisekből származó génlistákat használtuk fel.

A GSV futtatásához készítettünk egy egyszerű szövegfájlt, amely tartalmazta a páronkénti ortológokat listaszerűen és azok kromoszómális koordinátáit.

A SimpleSynteny használatához szükséges volt feltölteni a vizsgálandó genomi régiók nyers szekvenciáit fasta formátumban, továbbá egy referencia génlistát, amely alapján beazonosításra kerülnek a feltételezett ortológ szekvenciák.

Az OrthoClusterDB felületére feltöltöttük egy-egy fájlban az összes genom génsorrendre vonatkozó adatait, úgymint a gének azonosítóját, a vonatkozó kromoszóma számát, a gének lokalizációjának koordinátáit és végül a gének irányultságát. Ezt követően feltöltöttünk egy olyan fájlt is, amely tartalmazta az összes feltételezett ortológot megfelelően párosítva. A futtatási paraméterek a következők voltak: „order and strandedness”: -r -s, “synteny block size lower bound”: 2, “upper bound”: 2000. A többi paraméter esetén az alapértelmezett beállításokat használtuk.

#### 4.4.6. Kolineáris blokkok manuális feltárása a hasadó élesztő genomokban

A magas evolúciós rátával rendelkező protein szekvenciákat sok esetben nehézkes BLAST alapú keresésekkel azonosítani, és ez az állítás hatványozottan igaz a génszekvenciák esetében is. Emiatt a BLAST analízisek és a teljes genom illesztések eltéveszthetik vagy be sem képesek azonosítani a szekvenciájukat tekintve gyorsan változó, de génsorrend szempontjából konzervált szegmenseket.

Ennek kiküszöbölése végett, továbbá a hasadó élesztő genomok legkonzerváltabb régióinak feltárása érdekében olyan kolineáris blokkokat kerestünk manuálisan, amelyek legalább 5 gént tartalmaztak és e gének sorrendje és irányultsága mind a 4 hasadó élesztő fajban megegyezett. Ezekre a

blokkokra ezt követően aLCB-ként hivatkozunk (aLCB – ancestral locally collinear blocks), mivel ezeket a legközelebbi közös őstől örökölhették a szóban forgó fajok. Ha már csak egy fajban is eltérést tapasztaltunk, a továbbiakban azt a régiót nem kezeltük aLCB-ként.

#### 4.4.7. Kromoszómális átrendeződések feltárása

A generált teljes genom illesztésekből levonhatók bizonyos következtetések a kromoszómákat ért átrendeződések tekintetében, azonban azokat számszerűsíteni szemrevételezés alapján az esetek nagy többségében lehetetlen. Emiatt szükségszerűvé válik erre alkalmas algoritmusok használata.

Az átrendeződések típusainak és számának becsléséhez a GRIMM v2.01 (Tesler, 2002) online elérhető algoritmust használtuk. Első lépésként szükséges a genomok közös kolineáris blokkjainak feltárása. Ez történhet a Mauve segítségével, vagy manuálisan is. Ezután ezek a blokkok kapnak egy számot, amely számok a referenciaként használt genomban sorba rendezve találhatóak (1,2,3,4...). A többi genomban a számok a referencia genom permutációiként (pl.: 2,3,-1,4...) jelennek meg. Ezek összevetésével az algoritmus megállapít egy lehetséges eseménysorozatot, amelynek segítségével egyik genom átranzformálható a másikba (MCD – multi kromoszómális disztancia).

Munkánk során két vagy több genom vonatkozó adatait vetettük alá GRIMM analízisnek. Két genom összevetése esetén az algoritmus generált egy optimális átrendeződési esemény sorozatot (inverzió, transzlokáció, fúzió és hasadás), míg több genom analízise során egy MCD disztancia mátrixot kaptunk eredményként, ahol az értékek az átrendeződések számát jelölték.

#### 4.4.8. Átrendeződések evolúciós modellezése

Annak érdekében, hogy megállapíthassuk a beazonosított kolineáris blokkok esetében, hogy pusztán csak egy „ősi” génsorrend véletlenszerű maradványai vagy a fenntartásukban szerepet játszhat a természetes szelekció is, különböző „in silico” tesztek végeztünk.

Elsőként egy neutrális modellt teszteltünk, amikor is meghatározott számú átrendeződést (transzformációt) hajtottunk végre mindenféle szelekciós hatás nélkül. Ennek érdekében létrehoztunk egy 5000 génből álló referencia genomot, amelyben a géneket pozitív egész számok jelölték (1,2,3... 5000). Ebben a fiktív genomban hajtottunk végre meghatározott számú átrendezéseket (inverziókat), random pozíciókban. Az átrendeződési értékeket a hasadó élesztő fajtárok becsült MCD értékei alapján adtuk meg (*S. pombe*: 590; *S. cryophilus*: 592; *S. octosporus*: 598). Egyszerűen fogalmazva, egy fiktív *S. japonicus* genomot transzformáltunk *S. pombe*, *S. cryophilus* és *S. octosporus* genomokká. A modellezéshez egy általunk készített Python szkriptet használtunk fel ([https://github.com/Laci01/Laci01/tree/Schizosaccharomyces\\_synthetic](https://github.com/Laci01/Laci01/tree/Schizosaccharomyces_synthetic)) (9. táblázat).

Egy még kifinomultabb elemzés érdekében használtuk az Artificial Life Framework (ALF) modellező csomagot is (Dalquen és mtsi., 2012). Az ALF egy általunk megadott filogenetikai fa mentén „evolválatja” a genomokat, számos előre beállított paraméter figyelembe vételével. Az evolúciós referencia egységként a „szubsztitúció/site” volt megadva. Ebben az esetben is egy 5000 génből (proteinből) álló referencia genomot adtunk meg. Csak transzlokációkat és inverziókat engedtünk bekövetkezni, mivel a kromoszóma fúzió és hasadás nem jellemző a hasadó élesztőkre. Ezek rátái 0,13-ra lettek beállítva és maximum 300 gént érinhettek. Az inverz transzlokációk aránya 0,5 volt.

Mindkét modell esetében 100-100 független adatsort generáltunk, és az így létrehozott fiktív genomokban ugyanolyan aLCB-ket kerestünk, mint amiket megfigyelhettünk a valós genomokban. Az aLCB-k azonosításához egy Parser nevű Python szkriptet használtunk (9. táblázat).

Mivel a modellezések részletes leírásai és a paraméterek kalkulációjának bemutatásai terjedelmüket tekintve jelentősen felborítanák jelen dolgozat egységét, a leírások online elérhetőségét adjuk meg: Supplementary File S1 <https://www.mdpi.com/article/10.3390/jof7100864/s1>.

#### 4.4.9. Filogenetika

A fajok közötti filogenetikai kapcsolatok feltárására, továbbá válogatott szekvenciák alkalmasságának tesztelésére filogenetikai fákat készítettünk. Ezekhez a fajok ITS (internal transcribed spacer) régióit, vagy bizonyos protein szekvenciák összefűzött változatait használtuk. Az analízisekhez az online Phylogeny.fr (Dereeper és mtsi, 2008), MAFFT+NJ (Kato & Standley, 2013) és PhyML 3.0 (Guindon és mtsi, 2010) oldalakat/algorithmusokat használtuk.

A Phylogeny.fr oldalon az „A la Carte” analízis választásával létrehoztunk egy saját munkafolyamatot, ahol MUSCLE (Edgar, 2004) algoritmust használtunk az illesztésekhez, GBLOCKS (Castresana, 2000) programot a rosszul illeszhető szakaszok eltávolításához, BIONJ és PhyML algoritmusokat a filogenetikához, majd az elkészült fákat Newick formátumban letöltöttük és a FigTree v1.4.2 programmal jelenítettük meg.

A MAFFT online felületén az illesztésekhez a MAFFT E-INS-i algoritmusát választottuk, a filogenetikai analízishez a honlapon rendelkezésre álló Neighbor Joining algoritmust használtuk. Az elkészült fát az Archaeopteryx (Han & Zmasek, 2009) programmal jelenítettük meg, majd Newick formátumban mentettük. A fát a FigTree v1.4.2 programmal szerkesztettük.

A PhyML 3.0 oldalára feltöltöttük a MUSCLE vagy MAFFT algoritmusokkal illesztett szekvencia fájlokat PHYLIP formátumban. A megfelelő szubsztitúciós modell kiválasztásához a SMS (Lefort és mtsi, 2017) programot használtuk. A további PhyML paramétereket az online algoritmussal becsültettük.

A fák topológiájának megbízhatóságának tesztelésére aLRT (Anisimova & Gascuel, 2006) vagy Bootstrap analízist végeztünk 100 ismétléssel.

#### 4.4.10. Génsorrendet meghatározó tényezők feltárása

Első lépésként felderítettük a szakirodalom által említett, génsorrend konzerváltságát meghatározó, illetve azt elősegítő tényezőket. A paraméterek

között vannak fizikai és genetikai paraméterek egyaránt. Amely paraméter mind a négy hasadó élesztő faj esetében vizsgálható volt, azokat fajonként vizsgáltuk, más esetben a *S. pombe*-re vonatkozó adatokat tekintettük a többi faj esetében is mérvadónak. Így vizsgáltuk a ko-expresszió, esszenciális gének lokális denzitása, Rec12 hasítóhelyek és funkcionális csoportosulás tényezőket. Ezek részletes bemutatása a következő alfejezetekben történik.

Az összes analízishez a *Schizosaccharomyces* fajok NCBI adatbázisából letöltött annotált kromoszóma fájljait használtuk a következő akcessziós számokkal: CU329670, CU329671 és CU329672 - *S. pombe*, KE651166, KE651167 és KE651168 - *S. japonicus*, KE503206, KE503207 és KE503208 - *S. octosporus*, KE546988, KE546989, KE546990, KE546991, KE546992, KE546993, KE546994, KE546995 és KE546996 - *S. cryophilus*. A fájlakat a SnapGene Viewer programba importáltuk és ott dolgoztunk velük tovább.

#### 4.4.10.1. Funkcionális csoportosulás

Annak megállapítása érdekében, hogy létezik-e funkcionális kapcsolat a kolineáris blokkokat alkotó gének termékei között, megvizsgáltuk a gének „GO” (Gene Ontology) besorolását. Az adatsort a PomBase adatbázisából töltöttük le (<http://www.pombase.org/downloads/go-associations>) (Lock és mtsi, 2018). Ha egy gén több GO kategóriában is szerepelt, abban az esetben mindegyik kategóriában számításba vettük.

#### 4.4.10.2. Ko-expresszió

A szomszédos gének ko-expressziójának vizsgálatához Koch és mtsi (2012) *S. pombe*-re vonatkozó adatsorait használtuk, amely 5063 génről szolgáltat információt. Az adatok 9 független expressziós vizsgálatból származtak. Az analízis során meghatároztuk a gének egymáshoz viszonyított ko-expressziós rátájának átlagértékét. Majd ezeket az értékeket összehasonlítottuk a kolineáris blokkokon belül, illetve azokon kívül elhelyezkedő régiók esetében is.

#### 4.4.10.3. Esszenciális gének lokális denzitása

Az esszenciális géneknek tekintettük azokat a géneket, amelyek deléciója letálisnak bizonyul az élőlény számára. Első lépésként a Pombase adatbázisból (<https://www.pombase.org/downloads/phenotype-annotations>) kigyűjtöttük az összes esszenciális gént (4899 génről rendelkezünk adattal, ebből 1415 volt esszenciális), majd vizsgáltuk a lokális denzitásukat kolineáris blokkokon belül és azokon kívül is a különböző GO kategóriák viszonyában.

#### 4.4.10.4. *Rec12* hasítóhelyek

A *Rec12* hasítóhelyek lokalizációjának vizsgálatához Fowler és mtsi (2014) adatsorait használtuk fel (603 vágó hely). A megadott genomi koordináták alapján megkerestük a pontos elhelyezkedésüket és viszonyítottuk a korábban azonosított kolineáris blokkok helyzetéhez a *S. pombe* genomjában.

#### 4.4.11. Evolúciós ráta és intront tartalmazó gének vizsgálata

A hasadó élesztő proteinek evolúciós rátáinak vizsgálatához Rhind és mtsi (2011) adatsorait használtuk, amely tartalmazta 4220 1:1:1:1 ortológ szekvencia evolúciós rátáját. Az intront vesztett, illetve nyert gének listáját Zhu & Niu (2013) munkájából vettük, amelyben 2108 intronnal rendelkező ortológ gént vizsgáltak.

#### 4.4.12. Normalizáció, randomizáció és statisztika

Az adatsoraink normalizálásához meghatároztuk a vonatkozó adatok átlagát, majd a kapott értékkel osztottuk az egyedi adatokat (Fischer és mtsi, 2006).

Randomizációs teszteket Microsoft Excel 2016 program segítségével végeztünk a következő módon: random számsorozatok felhasználásával véletlenszerűsítettük a gének kromoszómális sorrendjét és vetettük össze a

kapott adatokat az eredeti sorrend adataival. A valószínűségi értéket ( $p$ ) a következő képlet alapján számítottuk ki:

$$p = \frac{r + 1}{n + 1}$$

ahol  $r$  jelölte azt az eset számot, amikor a teszt érték nagyobb vagy megegyező volt a valós értéknél, illetve értékkel,  $n$  pedig a randomizációk számát adta meg (Hurst és mtsi, 2004). Valódi random számsorozatok a Random.org (<https://www.random.org/integers/>) weboldal segítségével generáltunk, amely értékek az atmoszférikus zörejéből származtak.

Statisztikai tesztek és ábrákat a Past4 program (Hammer és mtsi., 2001) segítségével végeztünk, illetve készítettünk. A szignifikancia-szint ( $p$ ) minden teszt esetében 0,05 volt. Annak tesztelésére, hogy az adatsoraink normál eloszlású számtani populációból származnak-e, Shapiro-Wilk és Anderson-Darling próbákat végeztünk. E tesztek eredményei alapján használtunk paraméteres vagy nem paraméteres tesztek a továbbiakban. Páros adatsorok egyenlő eloszlásának vizsgálatára Kolmogorov-Smirnov tesztet végeztünk. Egymintás  $t$ -próbát alkalmaztunk abban az esetben, ha egy számadatról szeretnénk volna megállapítani, hogy beletarthat-e egy számtani tesztpopulációba. Az adatsorok átlagértékeinek összehasonlításához normál eloszlás, illetve nem nagymértékben eltérő variancia esetén kétmintás  $t$ -próbát végeztünk, nem normál eloszlás esetén nem paraméteres Mann-Whitney U, és páros Wilcoxon tesztet használtunk. Több adatsor együttes összehasonlításához normál eloszlás és nem szignifikánsan eltérő varianciák (Levene teszt) esetén egy szempontú ANOVA tesztet vagy ismételt méréses ANOVA (RM-ANOVA) tesztet végeztünk, nem normál eloszlás esetén Kruskal-Wallis tesztet használtunk, amelyet „post-hoc” analízisként páronkénti Mann-Whitney U vagy Dunn teszt követett. A „post-hoc” tesztek esetében Bonferroni korekciót is végeztünk. Az értékek korrelációjának megállapításához lineáris Pearson vagy Spearman tesztet végeztünk.

## 5. Eredmények

### 5.1. A hasadó élesztőgombák genom evolúciója I.: a *S. cryophilus* szuperkontigok összerendezése és komparatív analízise

Habár a ma ismert öt hasadó élesztőgomba fajból négy genomszekvenciája ismerté vált (Wood és mtsi., 2002; Rhind és mtsi., 2011), a *S. cryophilus* szekvenciáinak összeszerelése nem történt meg. Márpedig egy alapos és kiterjedt evolúciós analízis kivitelezhetetlen megfelelően összeillesztett genomszekvenciák nélkül. Ezért első lépésként e faj nagy méretű (Sc1-Sc9) szuperkontigjainak összerendezését kívántuk megvalósítani.

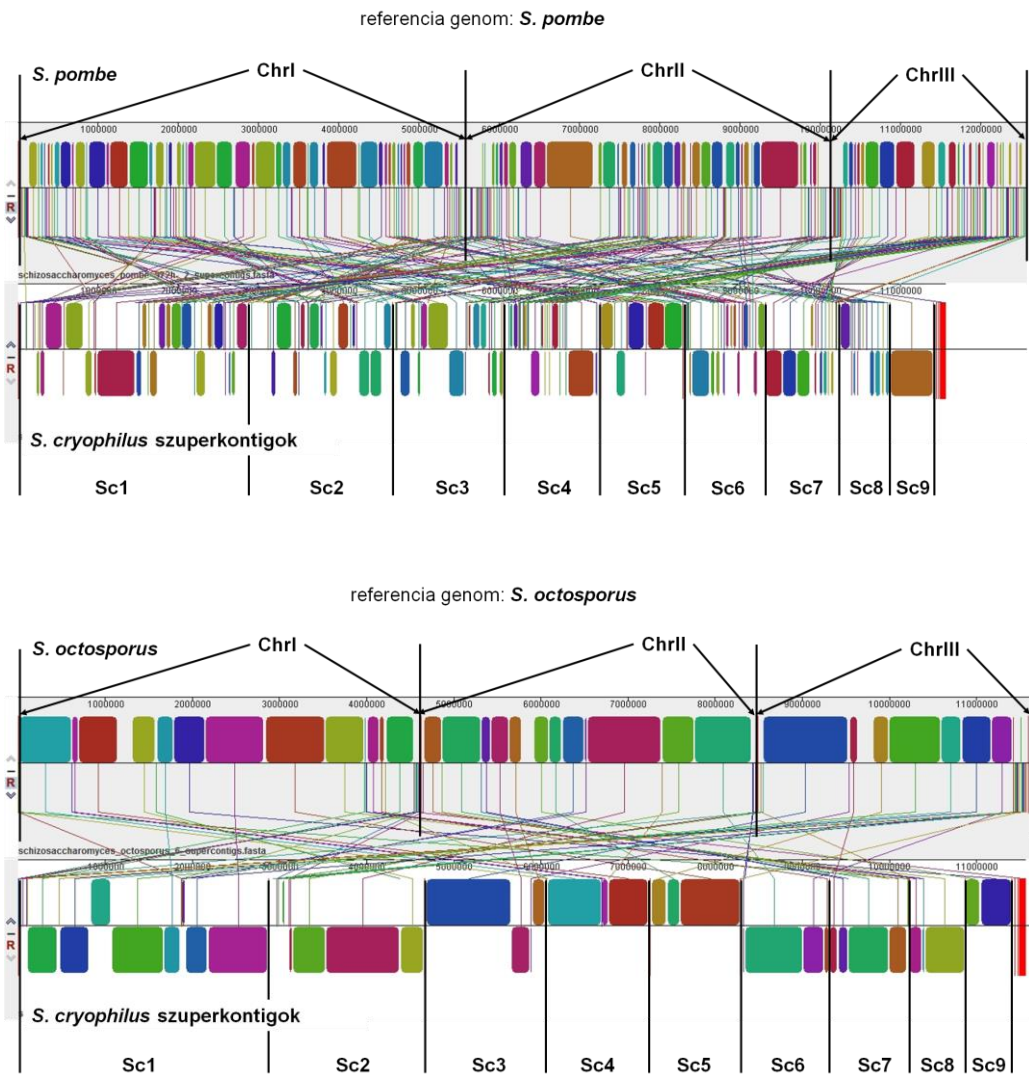
#### 5.1.1. A *S. cryophilus* szuperkontigok összerendezése szinténikus kapcsolatok segítségével

Munkánk során feltételeztük, hogy az adatbázisba feltöltött szuperkontigok egyenként megfelelően lettek összeillesztve, ezért mi magunk nem szekvenáltuk újra a genomot. Az összerendezést a közeli rokon fajok közötti szinténikus kapcsolatok segítségével kiviteleztük.

##### *5.1.1.1. A szuperkontigok rendezése a Mauve programmal*

Mivel korábbi munkák rávilágítottak, hogy a géntartalom és génsorrend meglepően konzervált a hasadó élesztőkben (Helston és mtsi., 2010; Rhind és mtsi., 2011), ezért úgy gondoltuk, hogy lokálisan kolineáris blokkok (LCB-k) feltárással sikerülhet a *S. cryophilus* szuperkontigjainak egymáshoz viszonyított sorrendjét megállapítani, ezáltal a szekvenciákat kromoszómákká rendezni.

Ennek érdekében teljes genom szekvencia illesztéseket végeztünk a Mauve program segítségével a *S. pombe* és *S. octosporus* genomokat használva referenciaként. Az első illesztések során a *S. cryophilus* esetében az adatbázisban elérhető szuperkontig sorrendet (Sc1-Sc9) használtuk (8. ábra).



**8. ábra.** Teljes genom szekvencia illesztések a Mauve program segítségével.

Az illesztések során a *S. pombe* és *S. octosporus* referencia genomokhoz hasonlítottuk a *S. cryophilus* szuperkontigjait az alapértelmezett sorrendet használva. Az azonos színű és vonallal összekötött téglalapok lokálisan kolineáris régiókat (LCB-eket) jelölnek. A középvonal alatti LCB-k fordított irányultságban helyezkednek el a referencia genom vonatkozó szakaszához képest.

Az illesztések egyértelműen jelezték, hogy a *S. cryophilus* génsorrendjét tekintve is sokkal jobban hasonlít a *S. octosporus*-hoz, mint a *S. pombe*-hez, mivel az illesztett genomok kevésbé fragmentált képet mutattak.

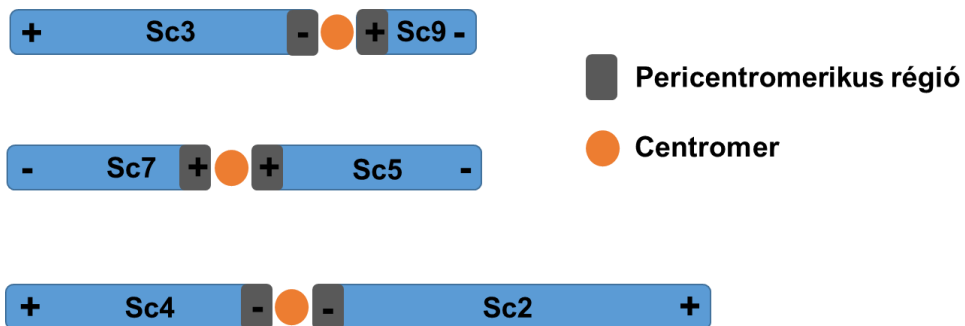
Mivel a Mauve program képes automatikusan újra rendezni a szuperkontigok sorrendjét a kromoszóma szinténia alapján a referenciaként megadott genomok szekvenciája mentén, ezért újabb illesztéseket készítettünk. Sajnos az így kapott szuperkontig sorrendek teljesen eltértek a két referencia genom esetében (Sc5,3,2,6,8,9,1,7,4 – *S. pombe* referencia és Sc4,9,5,8,7,2,6,3,1 – *S. octosporus* referencia) (Függelék 1. ábra). Így arra a megállapításra jutottunk, hogy a *S. cryophilus* fajnál sem határozható meg pontosan a szuperkontigok sorrendje kizárólag a globális szinteniára támaszkodva.

#### 5.1.1.2. *A pericentromerikus és szubtelomerikus szuperkontigok azonosítása*

Köztudott, hogy közeli rokon fajok genomjai rendelkeznek bizonyos közös karakterisztikával, ami segítségünkre lehet a *S. cryophilus* szuperkontigok sorrendjének meghatározásában. Például egy Saccharomycotina altörzsben végzett analízis rávilágított arra, hogy egy faj szubtelomerikus régióban lokalizálódó génjei hajlamosak a közeli rokon fajok esetében is a szubtelomerikus régióban elhelyezkedni, a kromoszómákat ért számos átrendeződés ellenére is (Fabre és mtsi, 2005). Ez még hatványozottabban igaz a pericentromerikus génekre, mivel a centromer környéki átrendeződések kifejezetten előnytelenek, illetve veszélyesek lehetnek az adott élőlény számára (Sasaki és mtsi, 2010; Fowler és msi, 2014).

Ezen ismeretek birtokában először a *S. cryophilus* pericentromerikus régióit kívántuk azonosítani. Mivel Rhind és mtsi (2011) munkájukban bemutatták, hogy a *S. pombe* és *S. octosporus* pericentromerikus géntartalmukat és génsorrendjüket tekintve szinte teljesen azonosak, így valószínűleg a *S. cryophilus* esetében is hasonló jelenséget tapasztalhatunk. Ezért azonosítottuk a *S. pombe* és *S. octosporus* pericentromerikus génjeit (10-13 gént a centromer

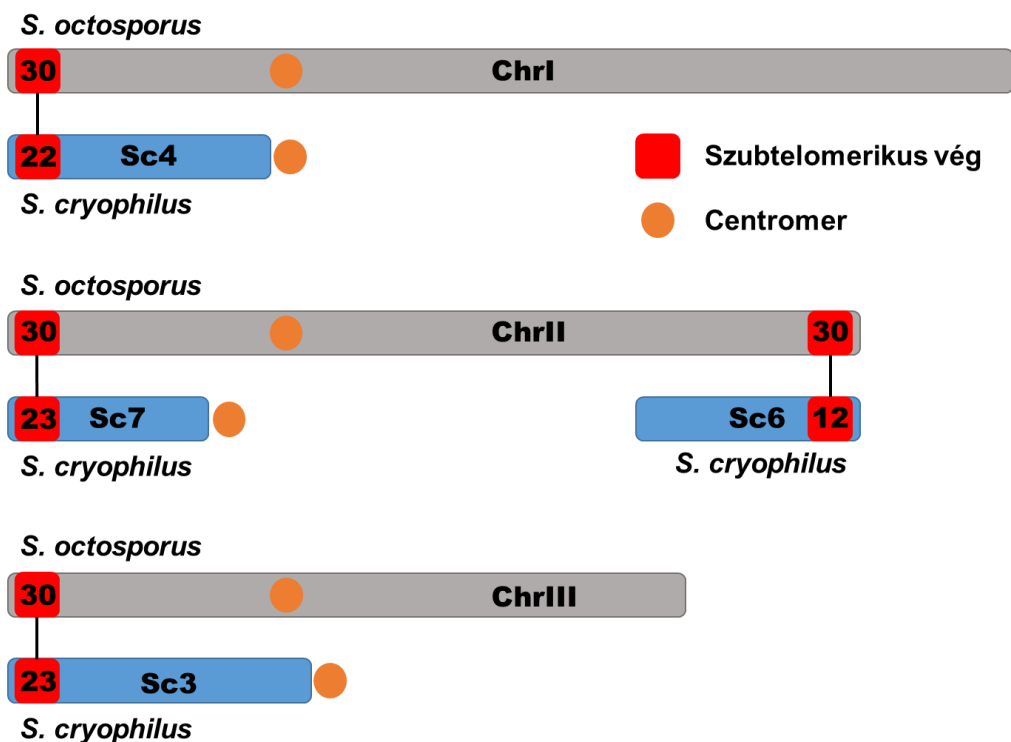
mindkét oldaláról) és azok protein szekvenciáinak segítségével BLASTp alapú keresést hajtottunk végre a *S. cryophilus* szekvenciáiban (Függelék 1. táblázat). Eredményeink megerősítették, hogy a pericentromerikus génsorrend közel azonos mind a három fajban. Mivel ezek a szakaszok a *S. cryophilus* esetében bizonyos szuperkontigok szélein helyezkedtek el, ezért feltételeztük, hogy ezek a szuperkontigok közvetlenül a centromerek mellett lokalizálódhatnak. Ezért az Sc4-Sc2; Sc3-Sc9; Sc7-Sc5 párokat pericentromerikus kontigoknak tekintettük (9. ábra). Elméletünket erősíti, hogy ezekben a régiókban jelentős számú tRNS szekvenciát is találtunk, amelyek szintén hajlamosak a centromer környékén lokalizálódni (Kuhn és mtsi, 1991).



**9. ábra.** A *S. cryophilus* feltételezett pericentromerikus szuperkontigjai. A „+” és „-” jelölés a szuperkontigok irányultságára vonatkozik: a „+” jellel kezdődő szuperkontigok irányultsága megegyezik az adatbázisból letölthető szekvencia irányultságával, míg a „-” jellel kezdődő fordított irányultságú.

A szubtelomerikus kontig végek beazonosításához is hasonló forgatókönyvet követtünk. A *S. octosporus* szubtelomerikus génjeinek felderítését követően (30 gén minden kromoszómavégről), azok protein szekvenciáinak segítségével BLASTp keresést alkalmazva megkerestük a *S. cryophilus* feltételezett ortológjait (Függelék 2. táblázat). Mint ahogyan az várható is volt, nagyfokú génsorrend konzerváltságot nem lehetett minden

vonatkozó régióban találni, mindazonáltal négy szubtelomerikus szuperkontig véget mégis képesek voltunk azonosítani (Függelék 2. táblázat, Függelék 2. ábra). A továbbiakban a *S. cryophilus* Sc3, Sc4, Sc6 és Sc7 vonatkozó végeit szubtelomerikus végekként kezeltük (10. ábra).



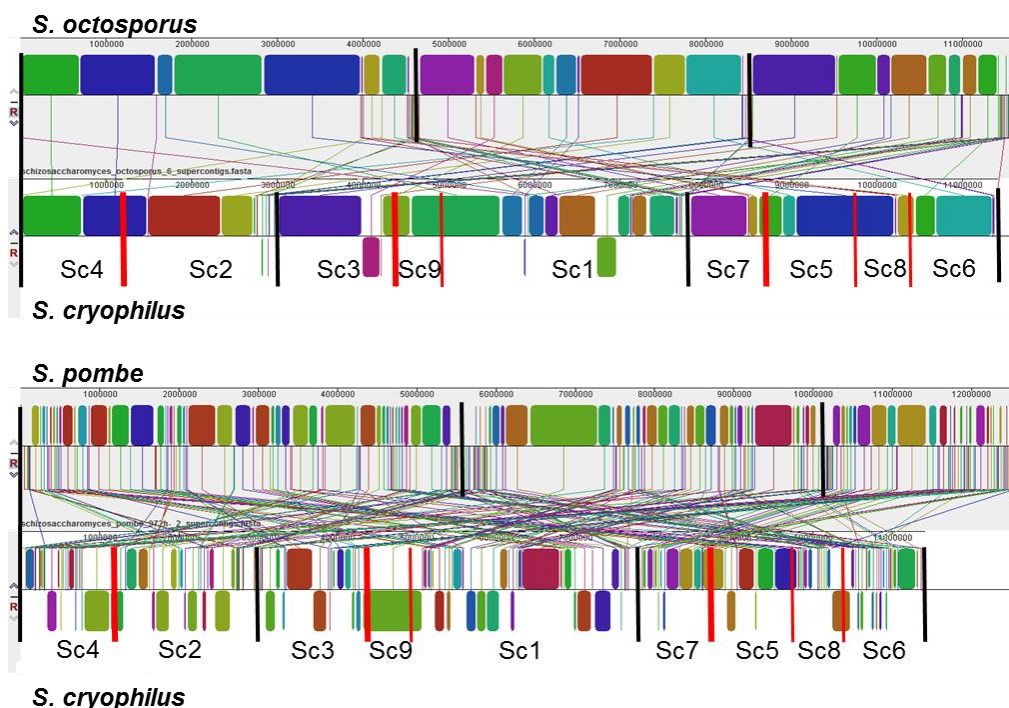
**10. ábra.** A *S. cryophilus* feltételezett szubtelomerikus régiókat tartalmazó szuperkontigjai az *S. octosporus* referenciához hasonlítva. A piros négyzetekbe írt számok a feltételezett közös ortológok számát jelölik.

### 5.1.1.3. A *S. cryophilus* szuperkontigok sorrendjének meghatározása

Az előző alfejezetek eredményeinek megfelelően újra rendeztük a *S. cryophilus* szuperkontigjait, figyelembe véve a kontigok orientációját is. Mivel három pericentromerikus véggel rendelkező szuperkontig egyben szubtelomerikusnak is bizonyult másik végüket tekintve, így ezek képezték a leendő kromoszómák széleit. Majd melljük rendeztük a megfelelő pericentromerikus szakasszal rendelkező párjukat. A fennmaradó

szuperkontigok pozícionálása a fajtársak kromoszóma méreteinek figyelembe vételével történt.

Ezután újra illesztettük a genomokat a Mauve programmal, hogy tesztelhesük a *S. cryophilus* hipotetikus kromoszómáit (11. ábra).



**11. ábra.** A *S. cryophilus* „új” szuperkontig-sorrendjének illesztése a *S. octosporus* és *S. pombe* referencia genomokhoz. A fekete függőleges vonalak a feltételezett kromoszóma-határokat, míg a piros függőleges vonalak a belső szuperkontig határokat jelölik. A vastagabb piros vonalak a centromer pozíciót hivatottak bemutatni.

Az illesztéseket bemutató ábrákon jól megfigyelhető, hogy a *S. cryophilus* pericentromerikusnak és szubtelomerikusnak tekintett kontig végein lokalizálódó LCB-k a többi faj esetében is hasonló módon helyezkednek el. A jelenség a *S. octosporus*-szal történő összehasonlítás esetében hangsúlyos különösen.

Az új illesztések felvetettek további lehetséges szuperkontig párosításokat, például: Sc9-Sc1; Sc5-Sc8; Sc8-Sc6, amelyeken nagy kiterjedésű átfedő LCB-k voltak megfigyelhetők (11. ábra). Mindezek fényében a következő szuperkontig sorrendet javasoljuk: Sc4-Sc2; Sc3-Sc9-Sc1; Sc7-Sc5-Sc8-Sc6 (12. ábra A).

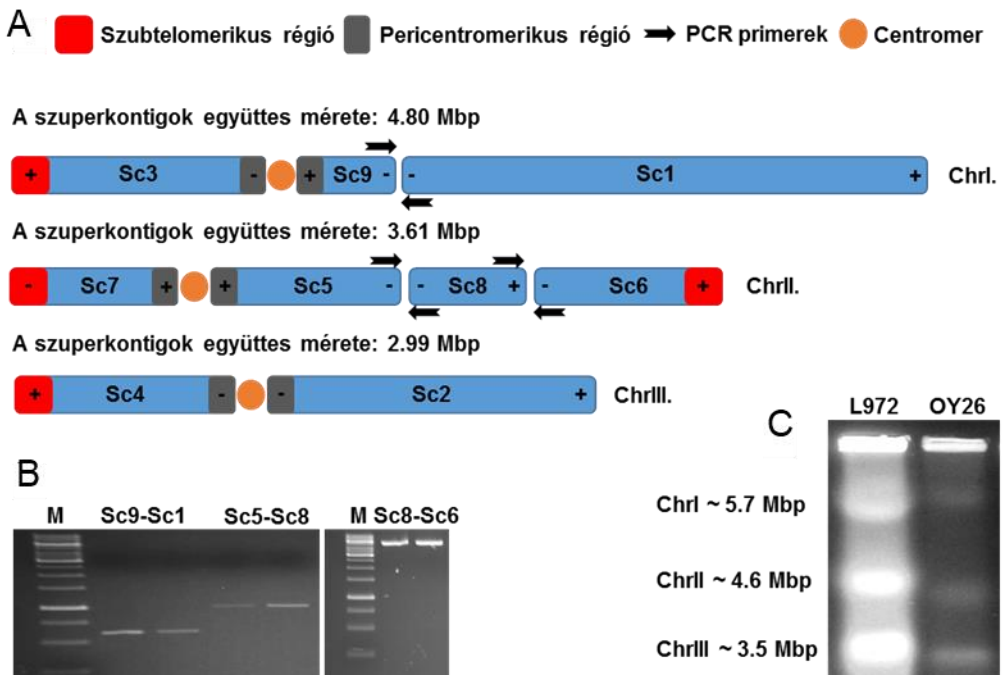
#### 5.1.2. A *S. cryophilus* szuperkontigok összeszerelésének validálása

Annak érdekében, hogy a bioinformatikai eredményeink alapján összeállított hipotetikus szuperkontig sorrendet validálhassuk, különböző laboratóriumi vizsgálatokat végeztünk, amelyeket a következő alfejezetekben ismertetünk.

##### *5.1.2.1. Szomszédos szuperkontigok ellenőrzése PCR-rel*

A nem pericentromerikus és nem szubtelomerikus szuperkontig végek esetében lehetőségünk volt a kontigok szomszédosságát PCR reakcióval ellenőrizni. Ezért megvizsgáltuk a Sc9-Sc1; Sc5-Sc8; Sc8-Sc6 párok vonatkozó végeit és a szekvenciákban rövid (400-500 bp hosszúságú) átfedéseket találtunk. Bár a Mauve illesztések a fent említett szuperkontig sorrendet támasztották alá, de a vizsgálatok során felmerült a Sc2-Sc6 szomszédosság lehetősége is. Ezért ezt az eshetőséget is számításba vettük a PCR-es ellenőrzéseknél. A PCR reakcióhoz szükséges primereket (4. táblázat) úgy terveztük meg, hogy az átfedő szakaszokon kívüli egyedi szekvenciákra essenek.

A négy PCR reakció közül három (Sc9-Sc1; Sc5-Sc8; Sc8-Sc6) bizonyult sikeresnek (12. ábra B), ami alátámasztotta a bioinformatikai eredményeinket.



**12. ábra.** A *S. cryophilus* feltételezett szuperkontig sorrendjének validálása. **A)** A hipotetikus kromoszómák bemutatása a pericentromerikus, szubtelomerikus és PCR validált szakaszok jelölésével. **B)** Bizonyos szuperkontigok összetartozásának validálása PCR reakció segítségével. A validáláshoz használt primerpárokat a 4. táblázat tartalmazza. **C)** A *S. cryophilus* (OY26) kariogramja a *S. pombe*-hez (L972) hasonlítva.

#### 5.1.2.2. A *S. cryophilus* kariogramja

További ellenőrzések végett megvizsgáltuk a *S. cryophilus* kariogramját PFGE módszerrel. Standardként a *S. pombe* 972 h- és a *S. cerevisiae* S288C törzseit használtuk.

Megerősítettük, hogy a *S. cryophilus* is három jól elkülöníthető kromoszómával rendelkezik (Rhind és mtsi, 2011), amelyek méretüket tekintve eltérnek a *S. pombe* kromoszóma méreteitől (12. ábra C). A *S. cryophilus* kromoszómaikat a méreteik szerint csoportosítottuk, így a legnagyobb kromoszóma a ChrI, a középső a ChrII és a legkisebb a ChrIII. elnevezést kapta.

A 12. ábra C-n látható kariogrammok alapján megfigyelhető, hogy a *S. cryophilus* ChrI-e nagyobb méretű, mint a *S. pombe* ChrI-e (~5,7 Mbp), a ChrII és ChrIII viszont kisebbek, mint a *S. pombe* vonatkozó kromoszómái (~4,6 Mbp és ~3,5 Mbp).

Rhind és mtsi (2011) becslése alapján a *S. cryophilus* genom mérete körülbelül 11,55 Mbp-t tesz ki. Az általunk összeállított 9 szuperkontig együttes mérete 11,4 Mbp. A kettő különbsége abból fakad, hogy a szuperkontigok sorrendjének megállapítása során a kis méretű kontigok (Sc10-Sc34, összesen 150610 bp) lokalizációjának meghatározása nem történt meg, mivel újabb szekvenálások nélkül az kivitelezhetetlen. Köztudott, hogy az egyedi kromoszómák méreteinek pontos meghatározása igen nehézkes egy PFGE analízis alapján, különösen, ha nem rendelkezünk megfelelő felbontású méret standardokkal. Viszont a faj kariogramja a korábbi becslésektől is nagyobb kromoszóma méreteket sugall. Ennek háttérében az állhat, hogy jelenleg nem rendelkezünk információval a *S. cryophilus* kromoszómák centromereinek, telomereinek hosszúságáról, továbbá a rDNS ismétlődések elhelyezkedéséről, illetve azok kiterjedéséről sem.

#### 5.1.2.3. A rDNS-ek lokalizációjának meghatározása

Mivel az rDNS ismétlődések (18S-5,8S-28S) jelenléte egy kromoszómán nagymértékben befolyásolhatja annak hosszát, ezért célszerű volt meghatározni a *S. cryophilus* kromoszómáin történő elhelyezkedésüket.

A *S. pombe* esetében két rDNS klaszter azonosítható, amelyek a ChrIII végein helyezkednek el és együttes méretük jóval meghaladja az 1 Mbp-t (Pasero & Marilley, 1993; Wood és mtsi., 2002; Brown és mtsi., 2011). A *S. octosporus* annotált kromoszóma fájljaiban (Rhind és mtsi, 2011) a II. és III. kromoszómák szélein egyaránt találtunk rDNS szekvenciákat. Mivel a *S. cryophilus* jobban hasonlít a *S. octosporus*-ra, mint a *S. pombe*-re, feltételeztük, hogy a *S. cryophilus* is két külön kromoszómán található rDNS klaszterrel rendelkezik.

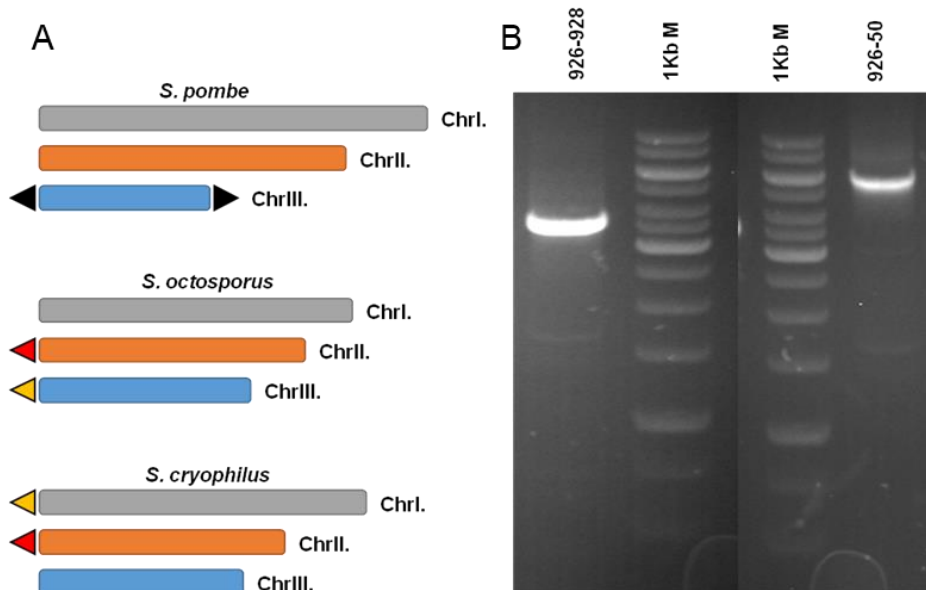
Ezért első lépésként rDNS szekvenciákat kerestünk az adatbázisból letöltött annotált *S. cryophilus* szuperkontig fájlokban. Sajnos csak a Sc7 szubtelomerikus régiójában sikerült rDNS-t találnunk, a többi szuperkontigon nem. Hipotézisünk szerint a Sc7 a II. kromoszóma részét képezi.

Ezt követően megvizsgáltuk a *S. octosporus* rDNS ismétlődései mellett lokalizálódó géneket, illetve LCB-ket. A *S. octosporus* II. és III. kromoszómáinak vonatkozó végei és a *S. cryophilus* Sc7 és Sc3 szinténikusnak bizonyultak (10. ábra, Függelék 2. táblázat). Ezek alapján elképzelhető, hogy a *S. cryophilus* Sc3 is tartalmaz rDNS klasztert, ami az I. kromoszóma részét képezi (13. ábra).

Elméletünk ellenőrzése céljából megpróbáltuk PCR reakciókkal bizonyítani a rDNS klaszterek jelenlétét. Ehhez a 28S rDNS alegység (D1/D2 domén) specifikus NL1-NL4 primereket (O'Donnell, 1993), illetve az általunk tervezett 18S alegység specifikus primereket használtuk különböző génspecifikus primerekkel párosítva. Ez utóbbiak olyan génekhez lettek tervezve, amelyek a feltételezéseink szerint az rDNS klaszterekhez közel helyezkednek el. Sajnos csak az Sc7 esetében voltak sikeresek a PCR reakciók (13. ábra), a Sc3 esetében nem sikerült bizonyítanunk az rDNS ismétlődések jelenlétét.

Ennek egyik oka az lehet, hogy az általunk kiválasztott gének túlságosan távol helyezkednek el az rDNS klaszterektől. A másik lehetőség az, hogy nem ott vannak az rDNS-ek, ahol feltételeztük. Viszont a *S. cryophilus* I. kromoszómájának rendkívül nagy mérete és a *S. octosporus*-sal való szinténikus kapcsolata azt a feltevést erősíti, hogy a szóban forgó kromoszómán lokalizálódik a *S. cryophilus* második rDNS klasztere.

Mindezeket figyelembe véve, úgy gondoljuk, hogy az általunk felállított szuperkontig sorrend helyes és sikerült a *S. cryophilus* szekvenciákat közel kromoszóma szintű egységekké rendeznünk.



**13. ábra.** Az rDNS klaszterek potenciális lokalizációja a *S. cryophilus* kromoszómáin. **A)** Az rDNS klasztereket színes háromszögek jelölik a hasadó élesztők kromoszómáin, az azonos színek szinténikus klasztereket jelölnek. **B)** Az rDNS klaszterek jelenlétének PCR-es validálása a Sc7 esetében. Az ábrán látható számok a PCR-hez alkalmazott primerek számait jelölik.

### 5.1.3. A *S. cryophilus* genom összevetése közeli és távolabbi rokon fajokkal

A következő alfejezetekben összevetettük az immáron összerendezett *S. cryophilus* genom szekvenciáit közeli, illetve távolabbi rokon fajokkal.

#### *5.1.3.1. Kromoszóma-átrendeződések típusai és dinamikája a hasadó élesztőkben*

A korábbiakban megállapított kontig sorrend és szinténikus kapcsolatok (11. ábra) lehetővé tették, hogy a *S. cryophilus* is bevonjuk abba a vizsgálatba, amely felderítheti, hogy milyen típusú kromoszóma átrendeződések formálták a hasadó élesztő fajok genomjait. Ezek feltárásához a YASS és Mauve illesztéseket használtuk fel, illetve a GRIMM v2.01 (Tesler, 2002) algoritmussal becsültünk meg lehetséges átrendeződési eseménysorokat.

A *S. cryophilus* és *S. octosporus* genomok vonatkozásában a leggyakrabban megfigyelhető változás az interkromoszómális transzlokáció volt (11. ábra és 5.1.3.2 fejezet). Meglepő módon, hosszabb szakaszokat érintő inverziók csak ritkán fordultak elő. Hogy megerősíthessük, illetve számszerűsíthessük ezeket a változásokat, kigyűjtöttük a két genom 20 000 bp-tól nagyobb méretű LCB-it és GRIMM analízisnek vetettük alá őket: egy optimális átrendeződési forgatókönyv szerint 7 interkromoszómális transzlokáció és 4 inverzió történhetett a *S. cryophilus* és *S. octosporus* genomok között.

A *S. cryophilus* és *S. pombe* összevetése során szintén gyakorinak bizonyultak az interkromoszómális transzlokációk, viszont jelentősen megnőtt a bekövetkezett inverziók száma is (11. ábra és 5.1.3.2 fejezet). Szintén a 20 000 bp-tól nagyobb LCB-k változásainak nyomon követésével 9 interkromoszómális transzlokációt és 54 inverziót azonosítottunk a GRIMM algoritmus segítségével. A *S. octosporus* és *S. pombe* összehasonlítása is hasonló tendenciát mutatott, szám szerint 14 interkromoszómális transzlokációt és 53 inverziót találtunk.

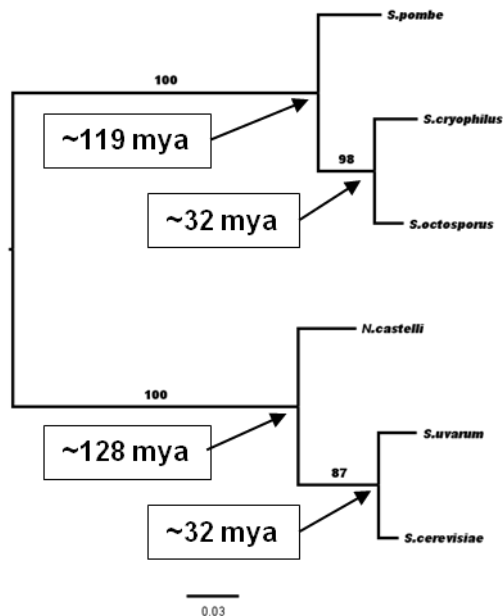
Egy ma megfigyelhető átrendeződés több különböző módon is kialakulhatott, például egy fordított irányultságú és megváltozott helyű LCB lehet egymást követő inverziók vagy egy inverzióval egybekötött intrakromoszómális transzlokáció következménye is. Mivel a GRIMM algoritmus a transzlokációkat és inverziókat két külön eseményként kezeli, ezért a fenti eredményeket fenntartással kezeltük, csak irányadóként használtuk fel az összehasonlításokhoz. Mindazonáltal az elemzések azt sugallták, hogy az összes interkromoszómális transzlokáció levezethető reciprok transzlokációkból.

### 5.1.3.2. Kromoszóma-átrendeződések összevetése más fajcsoporttal

Rhind és mtsi (2011) tanulmányukban bemutatták, hogy a géntartalom a hasadó élesztőkben sokkal konzerváltabb, mint a *Saccharomyces* vagy

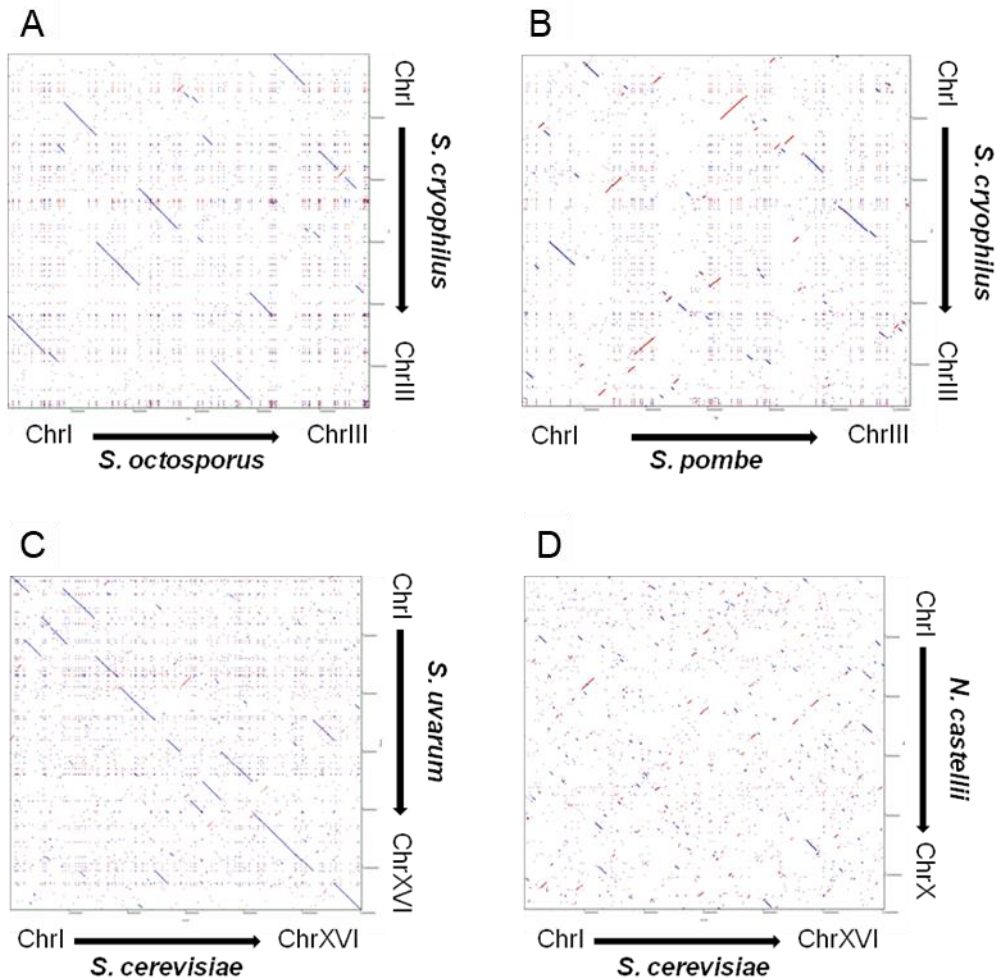
*Kluyveromyces* genuszokban, amelyek viszont sokkal kisebb divergenciát mutatnak aminosav szinten.

Annak érdekében, hogy a hasadó élesztő genomok evolúciós változásait összehasonlíthassuk más fajcsoportokkal, olyan alanyokat kerestünk, amelyek divergencia ideje körülbelül megegyezik a *Schizosaccharomyces*-ekkel (14. ábra). Rhind és mtsi adatai alapján a *S. cryophilus* és *S. octosporus* vonalak kb. 32 millió évvel ezelőtt váltak el egymástól, úgymint a *S. cerevisiae* és *S. uvarum*. A *S. cryophilus* és *S. pombe* vonalak kb. 119 millió évvel ezelőtt, a *S. cerevisiae* és *N. castellii* pedig kb. 128 millió évvel ezelőtt váltak szét (14. ábra). Mivel egyetlen módszer sem hibamentes és a különböző megközelítések jelentősen befolyásolhatják a végeredményt (Darling *et al*, 2010; Ghiurcuta & Moret, 2014), ezért a sarjadzó élesztő genomok vizsgálatához is a YASS, Mauve és GRIMM eszközöket használtuk és nem mások korábbi eredményeire hagyatkoztunk.



**14. ábra.** A vizsgált hasadó és sarjadzó élesztők filogenezeise és divergencia ideje. Mya: milliárd év.

Először teljes genom dot plot analíziseket végeztünk (15. ábra). Meglepő módon, ugyanazt a tendenciát figyeltük meg a kiválasztott sarjadzó élesztő fajoknál, mint a hasadó élesztőknél, miszerint az interkromoszómális transzlokációk a gyakoribbak, az inverziók csak a filogenetikai távolság növekedésével kerülnek túlsúlyba (15. ábra C,D).



**15. ábra.** Páronkénti teljes genom illesztések a YASS programot használva. A kék színű vonalak az azonos irányultságú régiókat, a pirosak az invertált irányultságú szakaszokat jelölik. A középvonaltól eltérő elhelyezkedésű szakaszok interkromoszómális transzlokációkat jelölnek. Nagy felbontású

képek a következő linken keresztül érhetők el: <https://static->

Ezt követően illesztettük a genomokat a Mauve program segítségével is. Ezek alapján úgy tűnt, hogy a sarjadzó élesztők esetében sokkal több kromoszómális átrendeződés következhetett be (Függelék 3. ábra).

A pontosabb összevethetőség kedvéért GRIMM analízisnek is alávetettük a genomokat, kétféle megközelítési módot használva. Egyik esetben felhasználtuk a Mauve által azonosított összes LCB-t, a másik esetben csak a 20 000 bp-tól hosszabb LCB-eket vettük figyelembe, ahogy korábban is.

A GRIMM által becsült MCD (multi kromoszómális disztancia) az első megközelítésnél 46 volt a *S. cryophilus* – *S. octosporus* vonatkozásában és 72 a *S. cerevisiae* – *S. uvarum* esetében (11. táblázat). Tehát, ha a genomok egészét figyelembe vettük, akkor valóban több átrendeződési esemény érinthette a sarjadzó élesztőket. Ha a szubtelomerikus régiók változásait nem vettük figyelembe (mert azok hajlamosak az átrendeződésre) és csak a nagy LCB-eket követtük nyomon, akkor a MCD értéke 11-re (*S. cryophilus* – *S. octosporus*) és 5-re módosult (*S. cerevisiae* – *S. uvarum*) (egy inverziót, egy ektopikus és három reciprok transzlokációt azonosítottunk a *Saccharomyces*-ekben, ahogy Fischer és mtsi (2000, 2001) is) (11. táblázat). Ezek alapján úgy tűnik, hogy több, nagyobb régiókat érintő kromoszómális átrendeződési esemény zajlott a *S. cryophilus* – *S. octosporus* genomokban, mint a hasonló divergencia idejű sarjadzó élesztők esetében.

Az összehasonlítás komplikáltabbnak bizonyult a *S. cryophilus* – *S. pombe* és *S. cerevisiae* – *N. castelli* esetében, mivel az utóbbi párosnál nagymértékű fragmentáltság volt megfigyelhető (15. ábra B, D). Viszont az elemzések hasonló eredménnyel zárultak e fajtárok esetében is, miszerint jelentősen lecsökkent az átrendeződések száma, ha csak a 20 000 bp-tól nagyobb LCB-eket vettük figyelembe (11. táblázat).

Ha az összes, illetve a 20 000 bp-tól hosszabb régiókat érintő átrendeződések arányait vizsgáltuk, akkor azt találtuk, hogy sokkal több nagy szakaszokat érintő változás érhet a hasadó élesztő genomokat, mint a sarjadzó élesztőket, függetlenül attól, hogy kromoszómánként vagy megabázisonként vizsgáltuk az eseményeket (11. táblázat).

**11. táblázat.** A hasadó és sarjadzó élesztők összehasonlítása a genomjaikat ért átrendeződések becsült száma alapján.

		Multi kromoszómális disztancia (MCD)						
		Teljes genom		Kromoszóma		Megabázis		
	Krom. szám	Össz. vált.	Nagy vált.	Össz. vált.	Nagy vált.	Össz. vált.	Nagy vált.	Nagy vált./ Össz. Vált
<i>So - Scry</i>	3 – 3	46	11	15,33	3,67	3,97	0,95	24,00%
<i>Sp - Scry</i>	3 – 3	150	63	50	21	12,5	5,25	42,00%
<i>Su - Scer</i>	16 – 16	72	5	4,5	0,31	6,1	0,42	7,00%
<i>Nc - Scer</i>	10 – 16	607	102	37,94	6,38	51,88	8,72	17,00%

*So*: *S. octosporus*

*Scry*: *S. cryophilus*

*Sp*: *S. pombe*

*Su*: *Saccharomyces uvarum*

*Scer*: *Saccharomyces cerevisiae*

*Nc*: *Naumovozyma castelli*

Krom. szám: kromoszóma szám

Össz. vált.: a GRIMM algoritmus segítségével megbecsült összes változások száma

Nagy vált.: a 20000 bp-tól nagyobb szakaszokat érintő változások száma

## 5.2. A hasadó élesztőgombák genom evolúciója II.: a stabil hasadó élesztő genomok genetikai és evolúciós háttere

Ahogy az már korábban említettük, több tanulmány is közölte, miszerint a hasadó élesztő genomok géntartalma és génsorrendje meglepően konzervált az aminosav-szekvenciáik divergenciájához képest (Helston és mtsi, 2010; Rhind és mtsi, 2011). Ebből az a felvetés következett, hogy vagy a *Schizosaccharomyces*-ek genomstruktúrája igen konzervált vagy a szekvenciáik evolválódnak rendkívüli sebességgel (Rhind és mtsi, 2011). A következő fejezetekben arra a kérdésre kerestük a választ, hogy mutat-e korrelációt a hasadó élesztők strukturális és szekvencia evolúciója.

### 5.2.1. Strukturális és szekvencia evolúció összehasonlítása a hasadó élesztőgombákban

Első lépésként össze kívántuk hasonlítani a hasadó élesztő fajok szekvenciális és strukturális evolúciójának dinamikáját. Mivel a fajcsoport filogenetikailag legkorábbinak (~220 millió év) számító leágazása a *S. japonicus*, ezért az összehasonlítások során e fajt használtuk referenciaként.

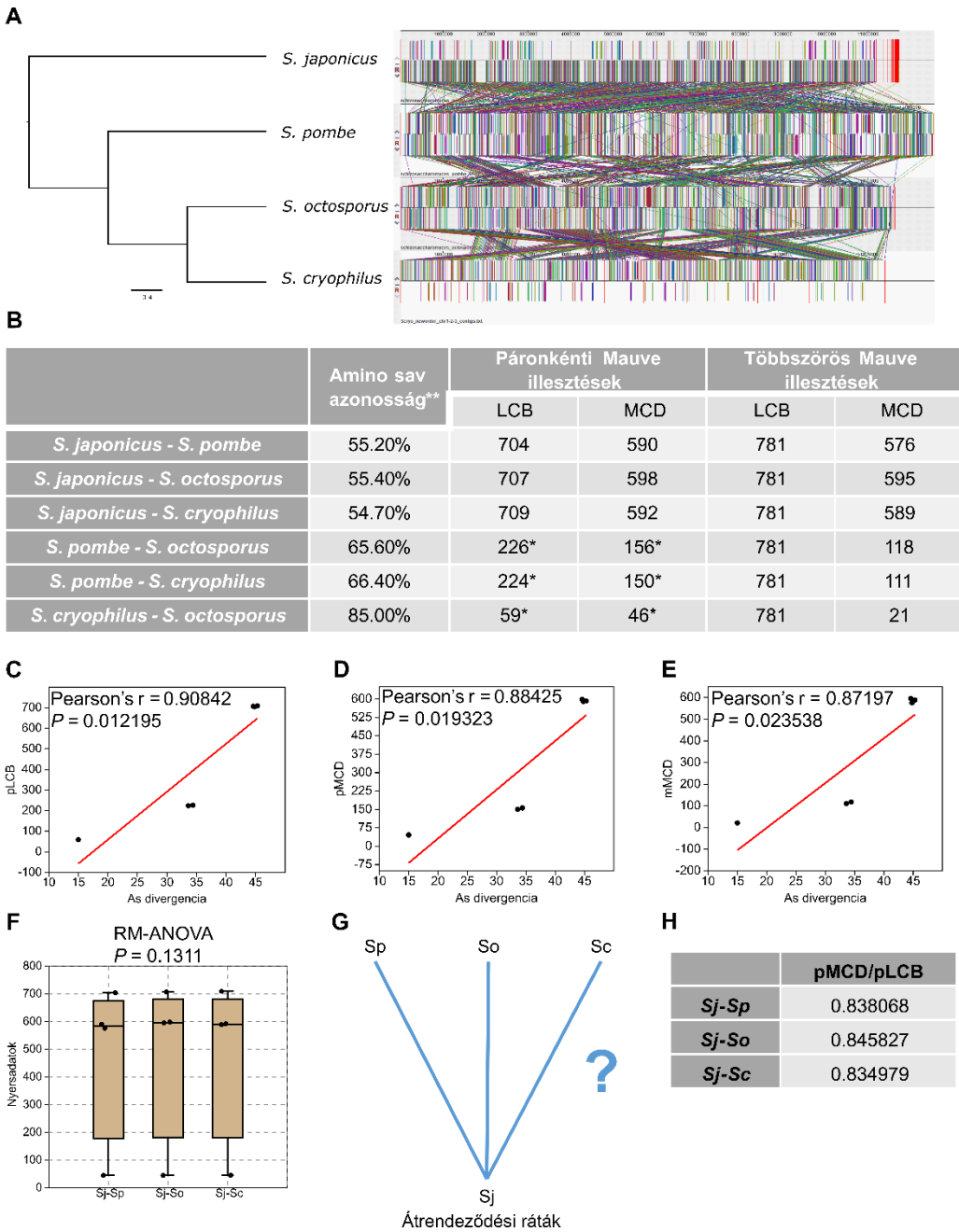
A fajok strukturális változásainak megbecslése érdekében páronkénti, illetve többszörös teljes genom illesztéseket készítettünk a Mauve program segítségével (16. ábra A és Függelék 4. ábra), amelyek ugyan konzervált, de sok átrendeződést sejtető képet festettek a vizsgált genomokról. A nagy számú átrendeződést megerősítették a feltárt LCB-vel végzett GRIMM analízisek is (16. ábra B), amelyek során megbecsültük az MCD-k számát. A szekvencia evolúció vizsgálatához a Rhind és mtsi által megállapított páronkénti aminosav azonosság értékeit vettük figyelembe (16. ábra B). Annak érdekében, hogy a vizsgált értékek kapcsolatait feltárhassuk. Pearson-féle lineáris korrelációs analíziseket végeztünk (16. ábra C-E). Ennek keretében megállapítottuk a következő értékek korrelációját: páronkénti LCB (pLCB) – aminosav divergencia (asdiv=1-aminosav azonosság); páronkénti MCD (pMDC) - asdiv

és multi MCD (mMCD) - asdiv (16. ábra C-E). Az adatok alapján kijelenthetjük, hogy e fajcsoport esetében a szekvencia és strukturális evolúció erős pozitív korrelációt mutat.

Viszont az adatokból az is kiderült, hogy a változások dinamikája is hasonló a fajok között, mivel mind a szekvencia, mind a strukturális változások egybevágóan a fajok filogenetikai kapcsolataival (16. ábra A). Tehát ha a *S. japonicus*-hoz viszonyítjuk a többi faj vonatkozó értékeit, akkor megfigyelhető, hogy az asdiv, az LCB-k száma és a MCD értékek is hasonlóak, illetve nincsenek szignifikáns eltérések (RM-ANOVA-teszt,  $P = 0,1311$ ) (16. ábra F). Továbbá az egy LCB-re jutó változások száma is nagyon hasonló a fajok között: *S. japonicus* – *S. pombe*: 0,838; *S. japonicus* – *S. octosporus*: 0,846; *S. japonicus* – *S. cryophilus*: 0,835 (16. ábra G, H).

### 5.2.2. Genom konzerváltság és közös szegmensek

Mivel a hasadó élesztők strukturális és szekvencia evolúciójának összehasonlítása azt mutatja, hogy közel egyforma mértékű divergencia figyelhető meg a fajok között a *S. japonicus* szemszögéből, meg kívántuk vizsgálni a genomjaik konzerváltságát nukleotid szinten is. A hasadó élesztő genomok konzerváltságának meghatározásához és közös, kolineáris régióinak feltárásához a Mauve, illetve YASS páronkénti illesztéseket és azok adatait használtuk fel. A kapott adatokat statisztikai értékelésnek is alávetettük.



**16. ábra.** A hasadó élesztők strukturális és szekvencia evolúciója erős pozitív korrelációt mutat. **A)** Többszörös teljes genom illesztés a négy *Schizosaccharomyces* faj között a *S. japonicus*-t használva referenciaként. Az ábrán látható színes téglalapok és vonalak közös kolineáris szegmenseket (LCB-eket) jelölnek a fajok között. **B)** A táblázat a Rhind és mtsi (2011) által

megállapított aminosav azonosságokat és a különböző teljes genom illesztésekből származó LCB-k számát, továbbá a becsült MCD értékeket hivatott bemutatni. **C-E)** Az említett adatok páronkénti korrelációs vizsgálata. As divergencia: 1-aminosav azonosság; pLCB: páronkénti LCB értékek; pMCD: páronkénti MCD értékek; mMCD: többszörös MCD értékek. Az ábrákon látható fekete pöttyök a fájárokat jelölik, a piros vonalak a regressziós egyenest jelölik. Az összes vizsgált érték pozitív korrelációt mutat. **F)** Az értékek variancia analízise a *S. japonicus*-t használva referenciaként. Az értékek nem bizonyultak szignifikánsan eltérőnek. **G, H)** Az átrendeződési ráták azt sejtetik, hogy a többi hasadó élesztő közel egyforma divergenciát mutat az *S. japonicus*-hoz képest.

#### 5.2.2.1. Genom konzerváltság nukleotid szinten

A genom szintű nukleotid konzerváltság meghatározásához páronkénti teljes genom dot-plot illesztéseket készítettünk a YASS programmal a következő fájárok esetében: *S. japonicus* – *S. pombe*; *S. japonicus* – *S. octosporus*; *S. japonicus* – *S. cryophilus* (Függelék 3-8 táblázatok). Két esetet vizsgáltunk: feltártuk a legkonzerváltabb régiókat (E érték = 0) és vizsgáltuk az illesztéseket egy kevésbé szigorú beállítás mellett is (E érték  $\leq 1 \times 10^{-30}$ ). Az illesztések adatait exportáltuk, majd a legszigorúbb illesztések (E érték = 0) listájából kiszűrtük az 1000 nukleotidnál rövidebb illesztéseket, továbbá a nem szinténikus ismétlődő szekvenciákat (pl. 5S rDNS-ek, tRNS-ek, sok kópiás gének), ezáltal is elkerülve a genom konzerváltság túlbecsülését. Az eredményeket összefoglaltuk a 12. táblázatban. Az illesztések számának és méretének jelentős csökkenése (a többi fájár korábbi adataihoz képest) egybevághat a *S. japonicus* filogenetikai helyzetével és a fajok aminosav-szekvencia divergenciájával. Így az ismert *S. japonicus* genom 39,2%-a, 36,1%-a és 35,6%-a (E  $\leq 1 \times 10^{-30}$ ), illetve 6,8%-a, 6,4%-a és 6,6%-a (E = 0) volt illeszthető a többi fajhoz.

**12. táblázat.** A YASS teljes genom illesztések számszerűsített adatai. nt: nukleotid.

<b><i>S. japonicus</i> – <i>S. pombe</i> (szűrés: <math>E \leq 1 \times 10^{-30}</math>)</b>					
	Talált illesztések száma	Méret (nt)	Átlag (nt)	Min (nt)	Max (nt)
<i>S. japonicus</i>	4922	4385225	891	122	7089
<i>S. pombe</i>	4922	4388365	892	124	7095
<b><i>S. japonicus</i> – <i>S. octosporus</i> (szűrés: <math>E \leq 1 \times 10^{-30}</math>)</b>					
	Talált illesztések száma	Méret (nt)	Átlag (nt)	Min (nt)	Max (nt)
<i>S. japonicus</i>	4549	4041397	888	119	7288
<i>S. octosporus</i>	4549	4043653	889	121	7261
<b><i>S. japonicus</i> – <i>S. cryophilus</i> (szűrés: <math>E \leq 1 \times 10^{-30}</math>)</b>					
	Talált illesztések száma	Méret (nt)	Átlag (nt)	Min (nt)	Max (nt)
<i>S. japonicus</i>	4504	3999159	888	116	7098
<i>S. cryophilus</i>	4504	4000482	888	118	7092

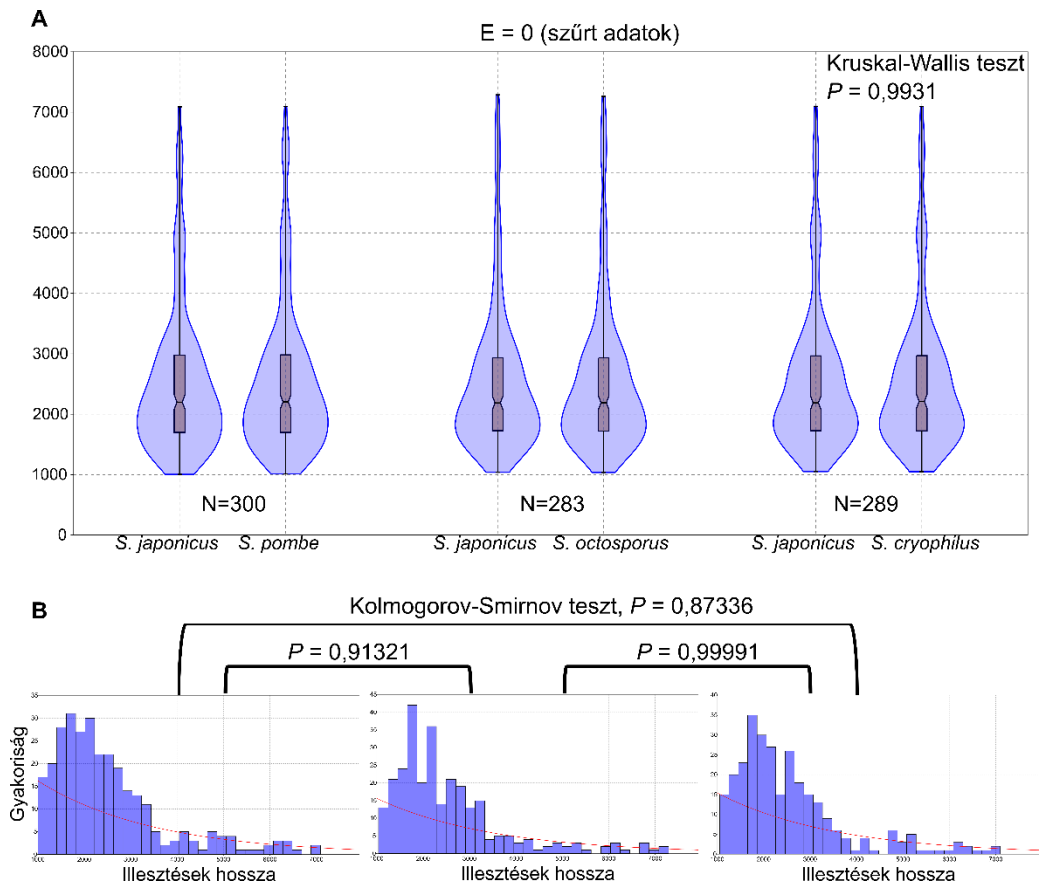
<b><i>S. japonicus</i> – <i>S. pombe</i> (szűrés: <math>E = 0</math>)</b>					
	Talált illesztések száma	Méret (nt)	Átlag (nt)	Min (nt)	Max (nt)
<i>S. japonicus</i>	300	759046	2530	1004	7089
<i>S. pombe</i>	300	760717	2536	1011	7095
<b><i>S. japonicus</i> – <i>S. octosporus</i> (szűrés: <math>E = 0</math>)</b>					
	Talált illesztések száma	Méret (nt)	Átlag (nt)	Min (nt)	Max (nt)
<i>S. japonicus</i>	283	721186	2548	1038	7288
<i>S. octosporus</i>	283	721024	2548	1036	7261
<b><i>S. japonicus</i> – <i>S. cryophilus</i> (szűrés: <math>E = 0</math>)</b>					
	Talált illesztések száma	Méret (nt)	Átlag (nt)	Min (nt)	Max (nt)
<i>S. japonicus</i>	289	739252	2558	1051	7098
<i>S. cryophilus</i>	289	739027	2557	1049	7092

#### 5.2.2.2. Az illesztések statisztikai értékelése

Mivel mind az illesztések száma, mind a kiterjedésük mértéke hasonló volt a fajok között, ezért kíváncsiak voltunk, hogy van-e szignifikáns eltérés az illesztések mértékében. Ennek vizsgálatához az illesztések egyéni méreteit használtuk fel. Kiindulási hipotézisünk az volt, hogy ha a fajpárok közötti illesztések méretükben és azok számában jelentősen eltérnek, akkor ennek egy variancia-analízis, illetve páros teszt során is meg kell mutatkoznia.

Annak érdekében, hogy kiválaszthassuk a célra legalkalmasabb próbákat, megvizsgáltuk a számsorok eloszlását, mind az  $E = 0$ , mind  $E \leq 1 \times 10^{-30}$  valószínűségű illesztéseket figyelembe véve. Mivel az értékek nem bizonyultak normál eloszlásúnak (Anderson-Darling és Shapiro-Wilk tesztek,  $P < 0,05$ ), ezért az egytényezős ANOVA teszt helyett, a nem paraméteres Kruskal-Wallis tesztet alkalmaztuk az adatsorok összehasonlítására. Az  $E = 0$  értékű illesztések varianciája nem tért el szignifikánsan (Kruskal-Wallis teszt,  $P = 0,9931$ ) (17. ábra A). További megerősítésként összehasonlítottuk az adatok páronkénti eloszlásait is, és ezek sem bizonyultak szignifikánsan eltérőnek (17. ábra B). A kevésbé szigorú illesztések is hasonló tendenciát mutattak (Kruskal-Wallis teszt,  $P = 0,7430$ ) (Függelék 5. ábra).

Mivel az eltérések nem bizonyultak szignifikánsoknak, arra következtethetünk, hogy a négy faj közötti genom konzerváltság mértékei sem térnek el egymástól jelentősen az *S. japonicus* perspektívájából. Ez felveti annak a lehetőségét, hogy számos esetben a genomok illeszthető részeinek ugyanazon szegmenseit kerülték el az átrendeződések, tehát ugyanazon régiók maradtak változatlanok.

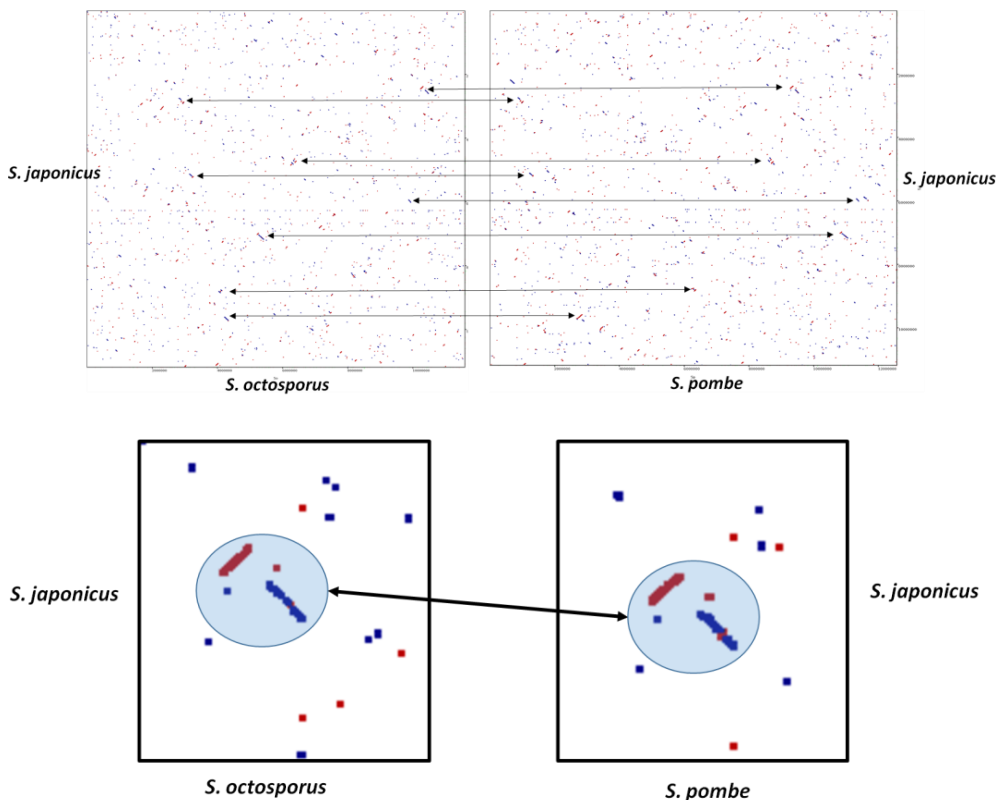


**17. ábra.** A hasadó élesztő genomok közös konzervált szegmensei egyforma eloszlást mutatnak. **A)** Az ábra a páronkénti teljes genom illesztésekből származó egyedi szegmensek hosszának eloszlását hivatott bemutatni. A hegedűplotok a kernel denzitást szemléltetik, a boxplotok az értéktartományok középső két kvartilisét (25% - 75%) mutatják be. A vízszintes vonalak a boxploton belül a mediánt jelölik, a becspések a mediánok a 95%-os valószínűségi intervallumát adják meg, míg a talpak a minimum és maximum értékekre vonatkoznak.  $N$  = mintaszám. Az Y-tengely az egyéni illesztések méreteit (nukleotidban megadva) jelöli. Szignifikáns eltérés nem volt megfigyelhető az adatsorok varianciájában. **B)** A hisztogramok a *S. japonicus* genomjának legkonzerváltabb szegmenseit mutatják be méret és gyakoriság szerint, a különböző páronkénti illesztések esetében (a sorrend azonos az ábra

A részével). A piros vonalak az adatok alapján becsült exponenciális eloszlást mutatják be. Az adatsorok eloszlása sem bizonyult szignifikánsan eltérőnek.

### 5.2.2.3. A közös szegmensek vizualizációja

Az előző felvetést, miszerint számos közös, átrendeződés-mentes szegmens található a fajok genomjaiban, a páronkénti Mauve (Függelék 4. ábra) és YASS illesztések vizuálisan is alátámasztják (18. ábra és Függelék 6. ábra).



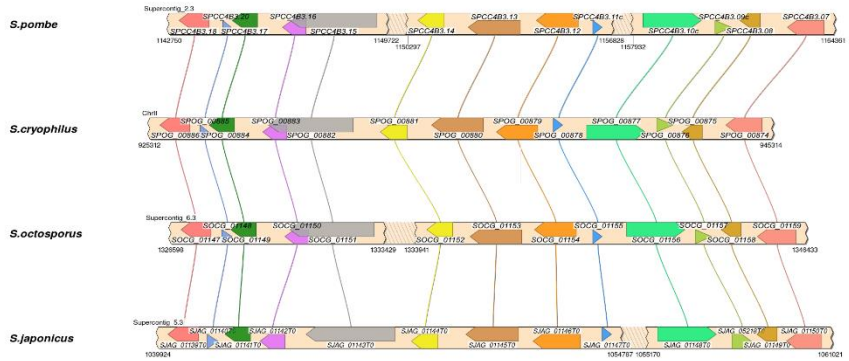
**18. ábra.** Számos esetben ugyanazok a genomszegmensek maradtak érintetlenek a hasadó élesztő genomok esetében. Az ábra felső része két páronkénti teljes genom illesztést mutat be *S. japonicus* – *S. octosporus* és *S. japonicus* – *S. pombe* fajpárok között. A nyilakkal jelölt részek az *S. japonicus* faj genomjának azonos szegmenseire mutatnak. Az ábra alsó része egy kiemelt és felnyagyított szegmenst mutat be.

### 5.2.3. Ősi kolineáris blokkok azonosítása

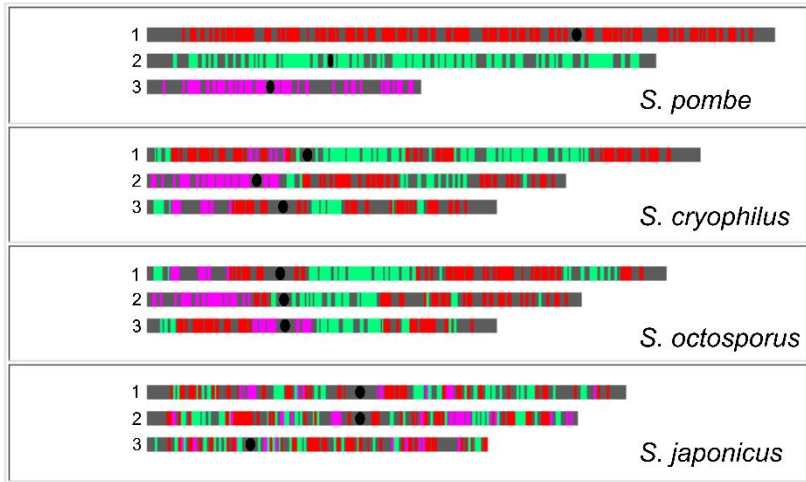
A korábbi fejezetek eredményei arra utaltak, hogy számos közös, változatlan szegmens maradhatott fenn a hasadó élesztők genomjában. Viszont a fajok közötti jelentős mértékű szekvencia divergencia miatt, a DNS alapú összehasonlítás mellett célszerű a genomokat a gének által kódolt proteinszekvenciáik felhasználásával is összevetni.

Ezt megvalósítandó, BLASTp keresést hajtottunk végre a *S. japonicus* genomban a *S. pombe* szekvenciáit használva referenciaként, mivel ez utóbbi faj rendelkezik a legjobban annotált genom szekvenciával. A feltételezett közös ortológok azonosítását követően kiterjesztettük a vizsgálatot a *S. octosporus*-ra és *S. cryophilus*-ra is. A kapott találatokat táblázatokba rendeztük, oly módon, hogy a *S. pombe* kromoszómális génsorrendjéhez illesztettük a többi faj vonatkozó találatait, azok lokalizációs adataival egyetemben (Függelék 9. táblázat). Az így kialakított adatbázisban olyan lokálisan kolineáris blokkokat kerestünk, amelyek legalább 5 génből álltak és a gének egymáshoz viszonyított sorrendje és irányultsága teljesen megegyezett (19. ábra A és Függelék 10. táblázat). E szigorú kritériumoknak megfelelő LCB-kről feltételezzük, hogy közvetlenül a hasadó élesztők legközelebbi közös őseitől származhatnak, ezért ezentúl aLCB-kként (ősi lokálisan kolineáris blokkokként) hivatkozunk rájuk. Ha már csak egy fajban is eltérést tapasztaltunk – inszerciót, deléciót vagy csak megváltozott irányultságot – akkor azt a továbbiakban nem tekintettük aLCB-nek. A kromoszómák szubtelomerikus régióiban nem voltak jellemzők az aLCB-k, viszont 266 blokkot találtunk a kromoszómák egyéb részein 7,73 átlagos gén tartalommal (19. ábra B). Ezek az aLCB-k összesen 2055 gént tartalmaztak, amely a genomok teljes ismert géntartalmának 40-42%-ának felel meg, a kiterjedésük pedig a teljes genomok 37-38%-ának (19. ábra C).

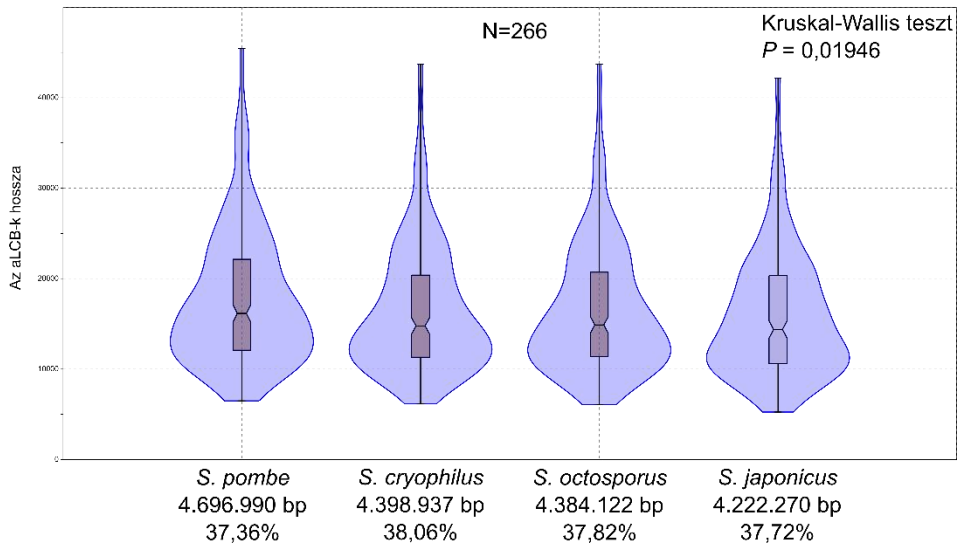
**A**



**B**



**C**



**19. ábra.** A hasadó élesztők ősi kolineáris blokkjai (aLCB-k), amelyeket vélhetően a legközelebbi közös őstől örökölhettek. **A)** Az ábra egy tipikus aLCB-t mutat be. A feltételezett ortológ gének mind a négy faj esetében azonos sorrendben és irányultságban helyezkednek el. **B)** Az aLCB-k kromoszómális elhelyezkedése. Jelen esetben a *S. pombe* génsorrendjéhez hasonlítottuk a többi faj génsorrendjét. A piros, zöld és lila téglalapok jelölik az aLCB-ket, a szürke színű szegmensek a genomok egyéb régiói. A fekete ellipszisek a centromer régiókat jelölik. **C)** Az aLCB-k hosszának megoszlása méret szerint a 4 hasadó élesztő faj genomjában. A *S. pombe* és *S. japonicus* értékei szignifikánsan eltérnek egymástól az intergénikus régiók eltérő méretei miatt.

Az aLCB-k hosszainak összehasonlításakor szignifikáns különbséget figyelhettünk meg a *S. pombe* és *S. japonicus* adatai között (19. ábra C). Ennek hátterében a fajok eltérő méretű intergénikus régiói (a kódoló gének közötti régiók) állnak. Fontos azonban megjegyeznünk, hogy a fenti értékek jelentősen módosultak volna, ha nem 5 génből álló aLCB-ket keresünk, hanem csak 4 vagy 3 génből állókat szelektálunk. Meggyőződésünk szerint ezek a „relikviák” nagyban hozzájárulhatnak, hogy megérthessük a hasadó élesztők genomevolúcióját, továbbá a stabil genom-struktúrájuk genetikai hátterét.

#### 5.2.4. Az ősi kolineáris blokkok fennmaradása és a természetes szelekció kapcsolata

Meggyőződünk arról, hogy a hasadó élesztők nagyfokú genom konzerváltságot mutatnak a viszonylag nagy filogenetikai távolság ellenére is, továbbá, hogy számos átrendeződés érte a kromoszómaikat. Így felmerült a kérdés, hogy a ma megfigyelhető konzervált génsorrend egyszerűen csak a legközelebbi közös ős génsorrendjének maradványa vagy pedig ezek az aLCB-k erős szelekciós hatás alatt is állnak? A következő alfejezetekben erre a kérdésre kerestük a választ, különböző szimulációkat használva.

#### *5.2.4.1. A neutrális evolúció szimulációja*

Amennyiben a hasadó élesztők genomjaiban megfigyelhető aLCB-k csupán csak az ősi génsorrend maradványai és nem állnak szelekciós nyomás alatt, akkor egy neutrális (minden restriktiótól mentes) szimuláció során hasonló eredményeket kell kapnunk, mint ami a valóságban is tapasztalható. Ezt az elméletet tesztelendő, egy tanszékünkön fejlesztett Python szkript segítségével lefuttattunk 100 független szimulációt. E szimulációk során, a korábban megbecsült kromoszóma-átrendeződések számait alapul véve transzformáltuk a kiindulási genomot további három utódgenomot létrehozva. Ezekben az utódgenomokban olyan aLCB-eket kerestünk, amelyeket a valós genomokban figyelhattunk meg (13. táblázat A). A neutrális szimulációk jelentősen eltértek a valós adatoktól minden vizsgált paramétert esetében (13. táblázat B).

#### *5.2.4.2. Szimuláció az Artificial Life Framework szimulátorral*

Az előző alfejezetben ismertetett szimulációk során semmilyen szelekciós tényezőt sem vettünk figyelembe, a változások minden utódgenomban egymástól függetlenül következhetnek be. A jelenség további vizsgálatának céljából egy szofisztikáltabb megközelítést is alkalmaztunk, amikor is az Artificial Life Framework (ALF) szimulátort alkalmaztuk (Dalquen és mtsi, 2012). A szimuláció során az általunk készített hasadó élesztő specifikus filogenetikai fa mentén evolváztattuk a genomokat. Az ALF segítségével lefuttatott 100 független szimuláció elemzése során a korábbiaktól eltérő eredményeket figyelhattunk meg (13. táblázat A,B). A létrehozott genomokban hasonló számú aLCB-eket találtunk, mint a valós genomokban. Ez megerősített minket abban, hogy az átrendeződési rátákat megfelelően becsültük meg. Viszont a többi vizsgált paramétert tekintve az ALF szimuláció eredményei is jelentősen eltértek a valóságban megfigyelhető adatoktól (13. táblázat A,B). A szimulációk eredményei azt sejtetik, hogy az aLCB-k nem csupán az ősi génsorrend maradványai, hanem szelekciós nyomás alatt is állhatnak.

**13. táblázat.** A valós és szimulációs adatok összehasonlítása az aLCB paraméterekre vonatkozóan. **A)** A táblázat ismerteti a két eltérő megközelítésű szimuláció során tapasztalt adatokat a valós adatok fényében. N: a független szimulációk száma, Min: a megfigyelt minimum értékeket jelöli, Max: a megfigyelt legmagasabb értékeket jelöli, Átlag: a 100 független szimulációra vonatkozó átlagos értéket ismerteti az adott paraméterre vonatkozóan. **B)** A valós és szimulációs adatok statisztikai összehasonlítása.

(A)	aLCB-k száma			Gének száma az aLCB-kben			Átlagos gén szám az aLCB-kben		
Valós adatok	266			2055			7,73		
	Min.	Max.	Átlag	Min.	Max.	Átlag	Min.	Max.	Átlag
Random evolúció (N=100)	5	23	13,12	33	140	77,75	5,09	7,4	5,93
ALF evolúció (N=100)	214	289	254,54	1387	1930	1656,44	6,15	6,9	6,5

(B)	aLCB-k száma		Gének száma az aLCB-kben		Átlagos gén szám az aLCB-kben	
	Valós vs random	Valós vs ALF	Valós vs random	Valós vs ALF	Valós vs random	Valós vs ALF
Randomizációs <i>P</i>	0,00990099	0,188118812	0,00990099	0,00990099	0,00990099	0,00990099
Egymintás t-próba <i>P</i>	$1,53 \times 10^{-83}$	0,44274	$7,36 \times 10^{-95}$	0,00016125	$1,24 \times 10^{-5}$	$3,33 \times 10^{-14}$

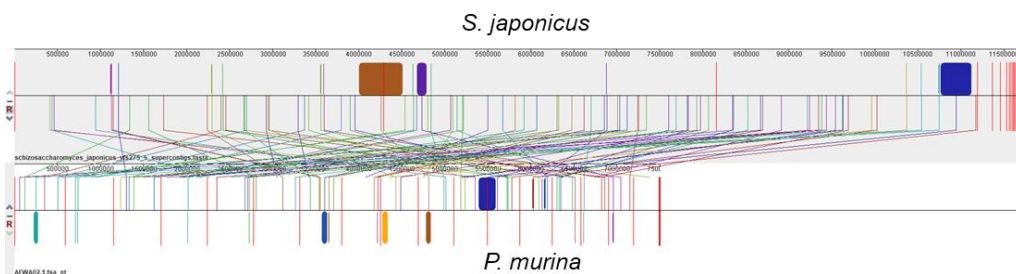
### 5.2.5. Az ősi kolineáris blokkok lehetséges eredete

Ha az aLCB-k fennmaradásában a természetes szelekció is szerepet játszhatott, akkor elképzelhető, hogy a megfigyelt génsorrendek egyéb fajokban is előfordulhatnak. Mivel számos más Taphrinomycotina altörzsbe tartozó faj szekvenciája is elérhetővé vált az évek során, lehetőségünk nyílt megvizsgálni ezt a felvetést.

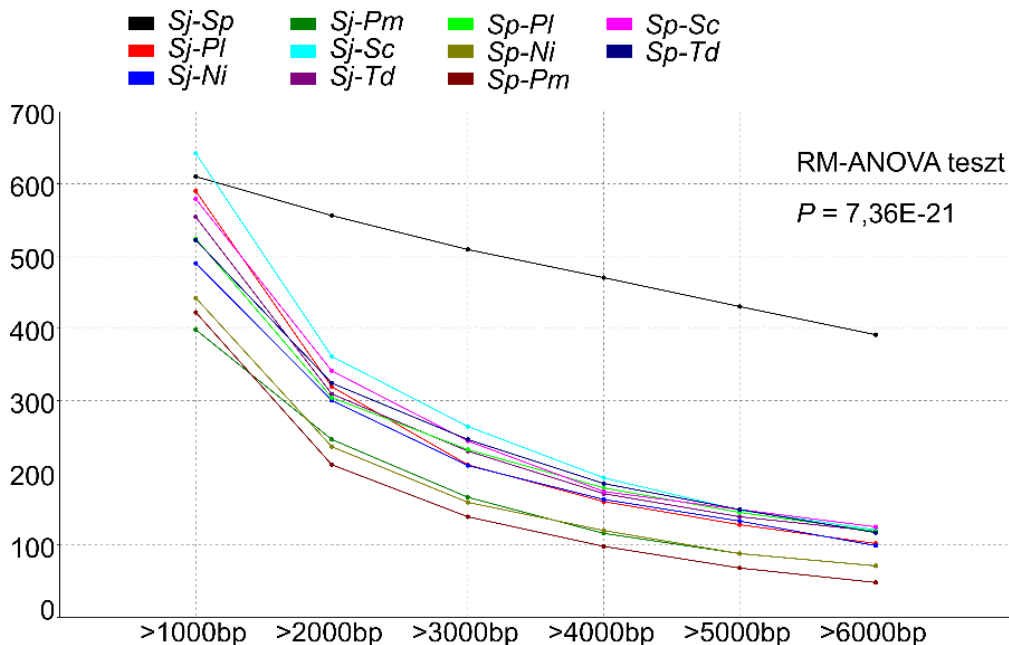
#### *5.2.5.1. Teljes genom illesztések egyéb Taphrinomycotina fajokkal*

Mivel a *Schizosaccharomyces* genusz a Taphrinomycotina altörzsbe tartozik, ezért megvizsgáltuk, hogy van-e, illetve milyen szintű a genomok közötti kolinearitás más Taphrinomycotina fajokhoz viszonyítva. Irodalmi adatokból tudhatjuk, hogy kiterjedt kolineáris régiók azonosítására nem számíthatunk, azonban kisebb kiterjedésű LCB-k létezése nem kizárható.

A hasadó élesztő klád legdivergensebb tagja, a *S. japonicus* mellett a *S. pombe* genomot használtuk referenciaként a páronkénti teljes genom illesztésekhez. A vizsgálatok során a következő fajok genomjainak felhasználásával készítettük az illesztéseket: *Pneumocystis murina* (Pm); *Taphrina deformans* (Td); *Protomyces lactucae-debilis* (Pl); *Saitoella complicata* (Sc) és *Neolecta irregularis* (Ni). Példaként a *S. japonicus*-*P. murina* illesztést mutatjuk be (20. ábra). Az illesztésekből egyértelműen látszik, hogy a genomok szerkezetüket tekintve is nagyon különböznek, azonban elképzelhető, hogy kis kiterjedésű, 2-3 génből álló kolineáris blokkok létezhetnek. Az analízisek során a Mauve program segítségével kigyűjtöttük, hogy milyen számú közös LCB-vel rendelkeznek a fájparok különböző LCB mérettartományokat használva (21. ábra). Azt tapasztalhattuk, hogy vizsgált fajok egyike sem rendelkezett releváns számú közös LCB-vel.



**20. ábra.** Egy példa ábra, amely a *S. japonicus* és *P. murina* genomok páronkénti illesztését mutatja be a Mauve program segítségével. Az ábra azt az állapotot tükrözi, amikor csak az 5000 nukleotiddal egyenlő vagy annál hosszabb LCB-k kerülnek jelölésre.



**21. ábra.** Az ábra a Mauve program segítségével illesztett fajpárok különböző mérettartományú közös LCB-ire vonatkozó gyakorisági adatokat hivatott bemutatni. Az Y-tengely az előfordulási gyakoriságot jelöli: hány darab LCB volt megfigyelhető az adott mérettartományban. Az X-tengelyen a különböző LCB mérettartományok láthatók bázispárban megadva. A *S. japonicus* és *S.*

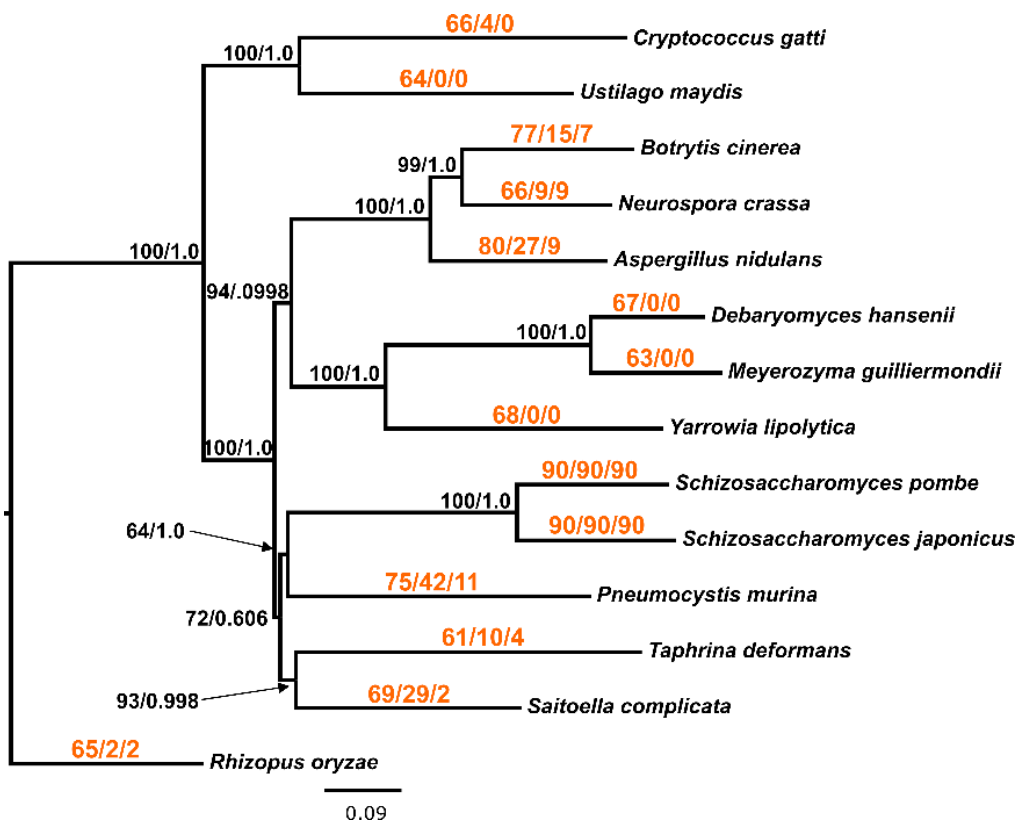
*pombe* fajpárok szignifikánsan eltérő számú közös LCB-vel rendelkeznek a többi fajpárhoz képest.

#### 5.2.5.2. Gén szintű kolinearitás vizsgálat különböző gombafajokban random blokkok k segítségével

Annak érdekében, hogy gén szinten is vizsgálhassuk a különböző genomok kolinearitását, néhány aLCB génsorrendjét összevetettük más fajok génsorrendjével. Egy generált random számsorozat segítségével kiválasztottunk 11 aLCB-t. E blokkok összesen 90 gént tartalmaztak, amelyek változatos funkcióval rendelkeztek (Függelék 11. táblázat).

Majd kiválasztottunk további 9 gombafajt a hasadó élesztőkön és egyéb Taphrinomycotina fajokon kívül, amelyek a következők voltak: *Yarrowia lipolytica*; *Debaryomyces hansenii*; *Meyerozyma guilliermondii*; *Aspergillus nidulans*; *Botrytis cinerea*; *Neurospora crassa*; *Cryptococcus gattii*; *Ustilago maydis*; *Rhizopus oryzae*. A Saccharomycotina altörzsből szándékosan választottunk protoploid fajokat, mivel a WGD eseményt átélte fajoknál fokozottabb a szinténia elvesztése (Fischer és mtsi., 2006).

A *S. japonicus* és *S. pombe* protein szekvenciáinak segítségével BLASTp keresést hajtottunk végre a vonatkozó fajok adatbázisaiban, hogy felderítsük a protein szekvenciák feltételezett ortológjait. A kapott eredményeket reciprok BLASTp kereséssel is ellenőriztük. Mint ahogyan azt vártuk is, változatos számú közös feltételezett ortológot sikerült azonosítanunk a fajok között, a számuk 61 és 80 között mozgott. Ezután megvizsgáltuk a szekvenciák lokalizációját a különböző genomokban és megszámoztuk azokat a géneket, amelyek egymás közelében lokalizálódtak, továbbá közvetlen szomszédok voltak. A kapott eredményeket az 22. ábrán látható filogenetikai fa mutatja be, amely a fajok között azonosított 26 közös ortológ aminosav-szekvenciájának összefűzésével készült (Függelék 12. táblázat).



**22. ábra.** 26 közös vélhetően ortológ aminosav-szekvencia összefüzéséből képzett filogenetikai fa. A narancs színű számok a következő adatokat mutatják be: az első szám a hasadó élesztőkkel közös ortológ gének számát mutatja be.

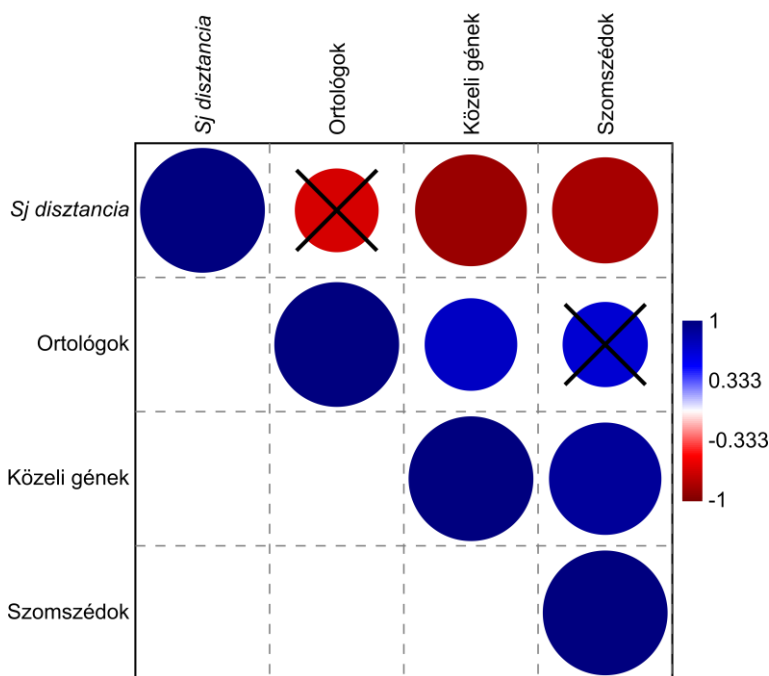
A második szám az egymás közelében lévő gének számát mutatja be, a harmadik pedig a közvetlen szomszédok számát jelöli. A fekete számok a

csomópontoknál az elágazások statisztikai megbízhatóságát jelölik:

„Neighbour joining bootstrap” (100 ismétlés) és „aLRT (approximate likelihood ratio teszt)”.

Az első megfigyelhető trend az, hogy a fajok között talált közös feltételezett ortológok száma nem tükrözi teljes mértékben a filogenetikai kapcsolatokat (Spearman  $r_s = -0,66079$ ,  $P = 0,060534$ ) (23. ábra). Hiszen elvárható lett volna, hogy minél nagyobb a filogenetikai távolság két faj között, annál kevesebb közös ortológ legyen azonosítható. Viszont az egymás közelében

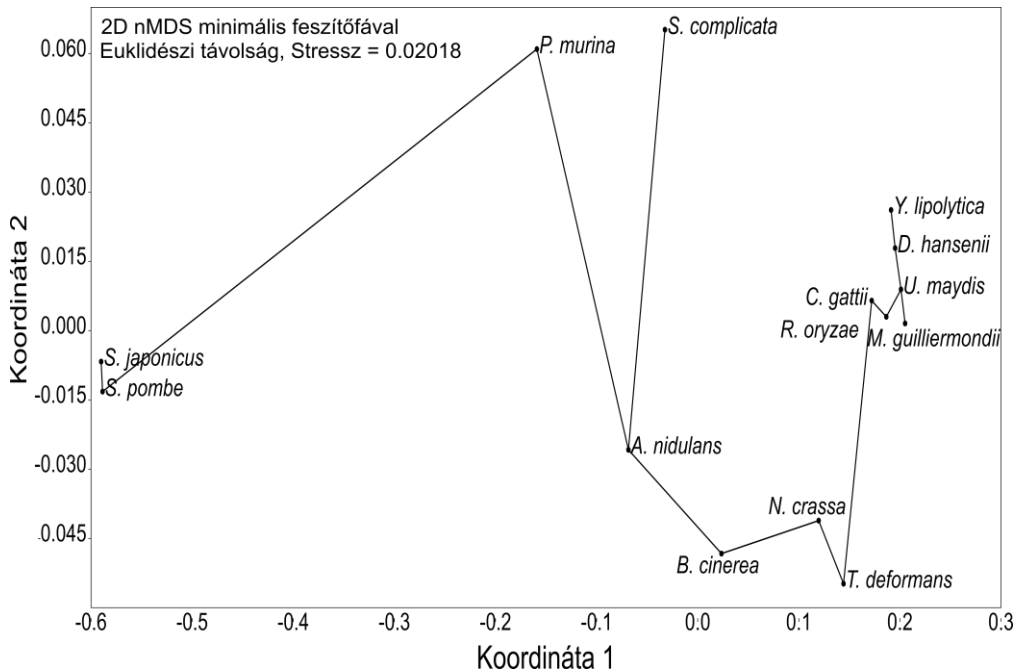
lokalizálódó gének száma és közvetlen szomszédos gének száma jobban egybevágh a filogenetikai kapcsolatokkal (Spearman  $r_s = -0,89217$  és  $-0,84583$ ,  $P = 0,000108$  és  $0,000827$ ) (23. ábra). Érdekes módon a szomszédos gének száma a *P. murina* faj és a fonalas fajok esetében a legmagasabb, a sarjadzó élesztők esetében pedig egy szomszédos génpárt sem sikerült azonosítanunk (22. ábra).



**23. ábra.** A vizsgált paraméterek páronkénti Spearman korrelációi. A kék szín pozitív, a piros negatív irányú korrelációt jelöl. A körök mérete arányos a korreláció mértékével, az áthúzott körök nem szignifikáns értékeket jelölnek. Az S<sub>j</sub> disztancia a *S. japonicus*-tól számított filogenetikai távolságot jelöli. A többi esetben a 22. ábrán látható adatok kerültek közvetlen felhasználásra.

Habár a teljes genomok géntartalmának csupán kis hányadával végeztük az összehasonlítást, mégis fontos megemlítenünk, hogy mind a közös ortológok számában, mind azok lokalizációjában a hasadó élesztők a fonalas gombákhoz (Pezizomycotina) hasonlítottak a *P. murina* mellett leginkább (24. ábra). Így nem zárhatjuk ki annak a lehetőségét, hogy a megfigyelt hasonlóság az ősi

fonalas fajok hagyatéka. Ugyanakkor meggyőződhattünk arról is, hogy az aLCB-k génsorrendje leginkább a hasadó élesztőkre jellemző tulajdonság.



**24. ábra.** 2 dimenziós skálázás minimális feszítőfával a fajok filogenetikai disztanciájából, a közös ortológok számából, a közeli és szomszédos gének adataiból kalkulálva. Az adatok alapján a hasadó élesztők jobban hasonlítanak a fonalas fajok bizonyos köréhez, mint például a sarjadzó élesztőkhöz.

#### 5.2.6. Az ősi kolineáris blokkok és az azokon kívüli régiók összehasonlítása

Köztudott, hogy a szekvenca- és strukturális evolúciót egymástól független molekuláris órák határozzák meg, így ha egy régió génsorrendjét tekintve változatlan maradt, az nem feltétlenül jelenti azt, hogy az alkotó gének szekvenciája sem változott jelentősen az idők során (Dujon, 2010). Viszont az átrendeződések általában mutagénnek bizonyulnak, mivel a DNS törése, majd a töréseket javító molekuláris mechanizmusok akár deléciókhöz, ezáltal információvesztéshez vezethetnek (Rodgers & McVey, 2016). Ha a DNS a töréspontok mentén megváltozik - még ha a változás maga nem is érinti egy gén kódoló régióját - a szabályzó szekvenciákban bekövetkező esetleges változások

a génexpresszió megváltozásához vezetnek. Ez pedig további szekvenciabeli mutációkhoz vezethet. Tehát elképzelhető, hogy az aLCB-ken belüli gének szekvenciája alacsonyabb evolúciós rátát mutat, mint azok a gének, amelyek ezeken a blokkokon kívül helyezkednek el.

A következő alfejezetekben összevetettük az aLCB-ket és a nem kolineáris blokkokat (NCB-ket) alkotó gének protein-szekvenciának evolúciós rátáját és a gének intron vesztését, nyerését.

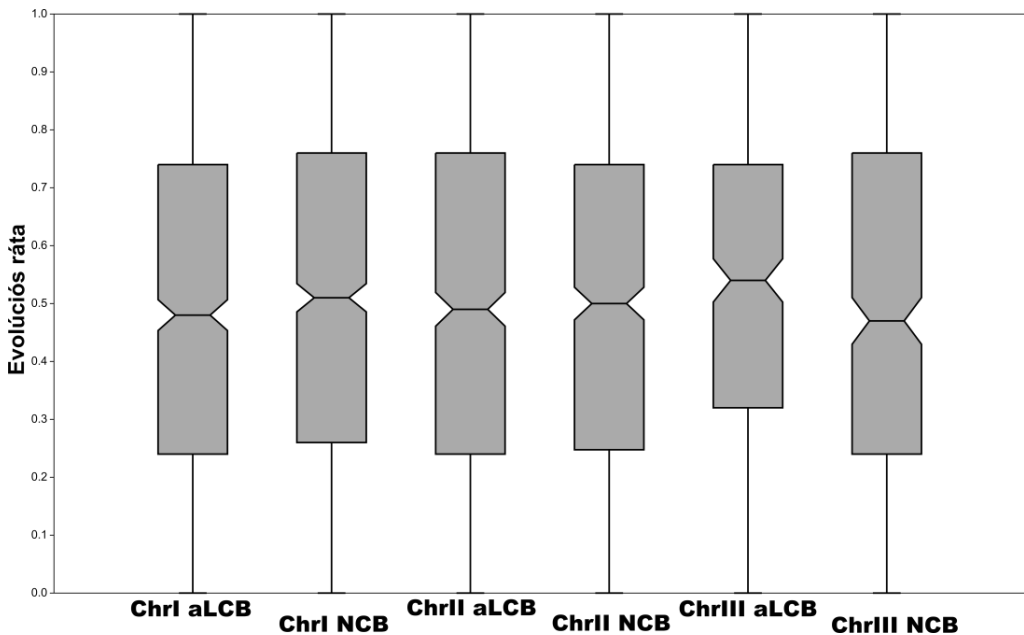
#### 5.2.6.1. A protein szekvenciák evolúciós rátájának összehasonlítása

Habár a proteinszekvenciák változásai nem tükrözik szükségszerűen a DNS-ben bekövetkezett változások összességét (szinonim mutációk, homoplázia), mégis kaphatunk egy általános képet a változásokról, különösen azok dinamikájáról. Kigyűjtöttük az aLCB-kben és NCB-kben lokalizálódó gének által kódolt protein szekvenciák evolúciós rátáit, amelyeket összehasonlítottunk kromoszómánként és a teljes genomra nézve is. Az analízishez Rhind és mtsi (2011) által megállapított 4220 1:1:1:1 feltételezett ortológ szekvencia evolúciós rátáit használtuk fel. Az eredményeket a 14. táblázat mutatja be.

**14. táblázat.** Az aLCB-kben és NCB-kben lokalizálódó gének által kódolt proteinszekvenciák evolúciós rátái.

	aLCB-k			NCB-k	
	Gének száma	Evolúciós ráta		Gének száma	Evolúciós ráta
<b>ChrI</b>	878	0,4923	<b>ChrI</b>	1047	0,5077
<b>ChrII</b>	799	0,4973	<b>ChrII</b>	766	0,4964
<b>ChrIII</b>	314	0,5237	<b>ChrIII</b>	416	0,4924
<b>Teljes genom</b>	1991	0,4993	<b>Teljes genom</b>	2229	0,5009

A táblázat adataiból kiderül, hogy bár nem rendelkezünk adattal az összes szekvenciáról, de megfigyelhető, hogy nincs jelentős különbség a különböző evolúciós rátával rendelkező szekvenciák eloszlásában (Mann-Whitney U teszt,  $P = 0,85212$ ). Tehát hozzávetőlegesen ugyanolyan arányban vannak jelen a gyorsan, illetve lassan evolválódó gének az aLCB-kben és NCB-kben. A jobb feloldás kedvéért, megvizsgáltuk a különböző evolúciós rátájú szekvenciák kromoszómánkénti megoszlását is (25. ábra). Habár az ábrán látható némi különbség a kromoszómánkénti megoszlások és medián értékek között, azonban ezek az eltérések sem bizonyultak szignifikánsnak (Kruskall-Wallis teszt,  $P = 0,5562$ ).



**25. ábra.** A hasadó élesztők proteinszekvenciáinak evolúciós rátái kromoszómánként az aLCB-kben és NCB-kben. A különbségek nem szignifikánsak (Kruskall-Wallis teszt,  $P = 0,5562$ ).

Az eredmények bemutatták, hogy az aLCB-kben lokalizálódó gének proteinszekvenciái nem bizonyultak konzerváltabbaknak, mint a genom többi részében elhelyezkedők.

### 5.2.6.2. A gének intron vesztese, intron nyerése

Mindezek után kíváncsiak voltunk, hogy megfigyelhető-e strukturális különbség az aLCB-kben és NCB-kben található gének között. Ennek érdekében megvizsgáltuk az intron nyert, illetve veszített ortológok megoszlását a blokkok között.

A Pombase adatbázisa szerint 2512 intronnal rendelkező gén létezik a *S. pombe* genomjában. Zhu és Niu 2013-as tanulmányukban 2963 1:1:1:1 ortológ szekvenciát vizsgált meg, amelyek közül 2018 intronnal rendelkező gént azonosítottak. 1775 konzervált és 808 egyedi intron pozíciót figyeltek meg.

Ezeket az adatokat felhasználva megvizsgáltuk a különböző gének megoszlását az aLCB-kben, illetve NCB-kben. A kapott eredményeket a 15. táblázat mutatja be.

**15. táblázat.** Az intront nyert és veszített gének megoszlása az aLCB-kben és NCB-kben. A zárójelben látható százalékok mutatják a változást elszenvedett gének arányát a blokkokban található összes gének számához viszonyítva.

	aLCB-k	NCB-k
<b>Intront nyert gének száma</b>	31 (1,51%)	7 (0,23%)
<b>Intront veszített gének száma</b>	232 (11,29%)	167 (5,35%)

Mivel Zhu és Niu csak olyan géneket szelektált, amelyek szinténikus blokkokban lokalizálódtak, ezért a legtöbb gén az aLCB-kben csoportosult, csak a gének kisebb hányada képezte az NCB-k részeit. Tehát a 15. táblázat adatai ennek következtében sugallhatják azt, hogy több változás érte az aLCB-k génjeit, mint az NCB-k génjeit. Arról viszont mindenképpen meggyőződhattünk, hogy a gének szerkezet változásai sem bizonyultak ritkébbnek az aLCB-kben, mint az NCB-kben.

Ezek alapján úgy gondoljuk, hogy bár az aLCB-k génsorrendje rendkívül konzervált, a szekvenciáik ugyanolyan mértékben változékonyak, mint a genomok egyéb részeiben csoportosulók.

### 5.2.7. Funkcionális csoportosulás lehetősége az ősi kolineáris blokkokban

Ezt követően kíváncsiak voltunk, hogy milyen biológiai funkcióval rendelkező gének alkotják az aLCB-ket, illetve megfigyelhető-e valamilyen rendező elv a gének csoportosulásában. Ehhez a Pombase adatbázisából letöltött biológiai folyamatokra vonatkozó 53 fő „GO” kategória adatsorait vettük alapul, amelyek 728 gén kivételével az összes *S. pombe* gén besorolását tartalmazzák. Ezt követően görcső alá vettük a szomszédos gének ko-expressziós értékeit is, majd megvizsgáltuk, hogy a különböző funkcióval rendelkező gének milyen gyakran lokalizálódtak Rec12 hasítóhelyek mellett, továbbá milyen arányú volt az esszenciális gének denzitása a kategóriákban. Mivel a többi hasadó élesztő fajról nem rendelkezünk hasonló információkkal, ezért a *S. pombe*-re vonatkozó adatokat tekintettük mérvadónak.

#### *5.2.7.1. GO kategóriák az ősi kolineáris blokkokban*

Korábban említettük, hogy az aLCB-kben csoportosuló gének száma 2055, ami a teljes géntartalom 39,97%-ának felel meg, viszont ha a GO annotációval nem rendelkező géneket nem vesszük figyelembe, akkor ez az érték 42,07%-ra emelkedik. Abból a feltevésből indultunk ki, hogy ha a különböző kategóriák génjei random módon oszlanak szét a genomban, akkor az egyes GO kategóriákba tartozó gének számának sem szabad meghaladnia ezt a bizonyos 42,07%-ot az aLCB-kben. Ezért megvizsgáltuk, hogy az 53 fő kategóriában vannak-e olyanok, amelyek génjei nagyobb számban lokalizálódnak az aLCB-kben, mint azokon kívül. Eredményeinket a 26. ábra és a Függelék 13. táblázata foglalja össze. 31 GO kategóriát találtunk, amelyek génjei nagyobb számban voltak jelen az aLCB-kben. Ez lehet a véletlen műve, vagy pedig valamilyen rendező elv hatása. Annak érdekében, hogy a véletlent kizárjuk, 50 független randomizációs tesztet végeztünk a 31 GO kategória elemeivel, majd az adatokat egymintás *t*-próbának is alávetettük (16. táblázat és

Függelék 14. táblázat). A tesztek követően 13 kategória számított szignifikánsnak, tehát az ő esetükben zárhattuk ki a véletlen hatását.

Függetlenül attól, hogy bizonyos GO kategóriák jelentős csoportosulást mutatnak az aLCB-kben, a kategóriákba tartozó gének diszperz elhelyezkedést mutatnak az aLCB-ken belül is. Az egy biológiai funkcióhoz tartozó gének nem egymás mellett vagy egymás közelében helyezkednek el, hanem elszórtan a kromoszómákon. Ebből következik, hogy a csoportosulások nem fakadhatnak a gének fizikai közelségéből, hanem más tényezők játszhatnak szerepet a csoportosulások kialakításában.



**16. táblázat.** A vizsgált GO kategóriák, amelyek szignifikáns mértékben csoportosultak az aLCB-kben. 14 GO kategória bizonyult szignifikánsnak a randomizációkat követően, de a „meiotic nuclear division” GO kategória kiesett ebből a körből az egymintás *t*-próbák eredménye alapján.

GO kategóriák	Összes gén a kategóriában	Gének száma az aLCB-kben	Gének száma az NCB-kben	Randomizáció <i>P</i> -érték	Egymintás <i>t</i> -próba <i>P</i> -érték
chromatin organization (GO:0006325)	275	128	147	0,019608	0,043888
meiotic nuclear division (GO:0140013)	142	67	75	0,039216	0,07286
mitochondrial gene expression (GO:0140053)	150	79	71	0,019608	0,010351
mitotic sister chromatid segregation (GO:0000070)	182	91	91	0,019608	0,005432
mRNA metabolic process (GO:0016071)	281	140	141	0,019608	0,001176
nucleocytoplasmic transport (GO:0006913)	109	64	45	0,019608	0,000021
protein catabolic process (GO:0030163)	233	114	119	0,019608	0,009971
protein modification by small protein conjugation or removal (GO:0070647)	185	89	96	0,019608	0,033979
protein-containing complex assembly (GO:0065003)	262	122	140	0,019608	0,030717
regulation of mitotic cell cycle phase transition (GO:0044772)	170	80	90	0,019608	0,040627
ribosome biogenesis (GO:0042254)	323	156	167	0,019608	0,004532
transcription, DNA-templated (GO:0006351)	459	205	254	0,039216	0,044262
tRNA metabolic process (GO:0006399)	166	93	73	0,019608	0,000877
vesicle-mediated transport (GO:0016192)	319	152	167	0,019608	0,000801

### 5.2.7.2. A szomszédos gének ko-expressziós rátájának vizsgálata

Köztudott, hogy a szomszédos gének ko-expressziója bár ritka, de annál erősebben ható tényező lehet a génsorrend megőrzésében (Poyatos & Hurst, 2006, 2007). Ezért megvizsgáltuk az aLCB-ken belül és azokon kívül lokalizálódó gének egymáshoz viszonyított átlagos ko-expressziós rátáját, amelyeket összehasonlítottunk a teljes genomra vonatkozó ko-expresszió mértékével is (17. táblázat).

Az aLCB-kben és NCB-kben lokalizálódó szomszédos gének átlagos ko-expressziós értékei között nincs szignifikáns különbség (Mann-Whitney U teszt,  $P = 0,86051$ ), de rátájuk magasabb, mint a teljes genomra vetített átlag (Kruskal-Wallis teszt,  $P = 3,878 \times 10^{-33}$ ) Tehát az aLCB-k génei nem expresszálódnak nagyobb mértékben együtt, mint a genom egyéb részein található gének.

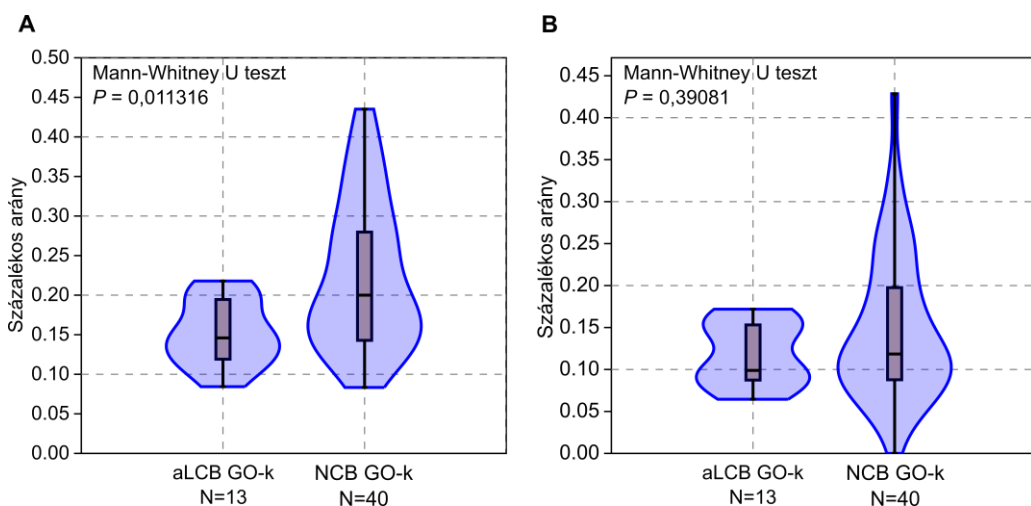
**17. táblázat.** A ko-expresszió átlagos értéke a genom különböző régióiban.

	Blokkok száma	Gének száma	Ko-ex. esetek száma	Ko-ex. átlag
<b>Teljes genom</b>	-	5063	12814452	0,0287
<b>aLCB</b>	266	2055	7895	0,0408
<b>NCB</b>	232	1922	8175	0,0394

### 5.2.7.3. *Rec12* hasítóhelyek gyakorisága a különböző GO kategóriájú gének mellett

Pál és Hurst 2003-as tanulmányukban bebizonyították, hogy a *S. cerevisiae* esetében a génsorrend és rekombinációs ráta ko-evolúciót mutat, és ehhez kapcsolódóan bemutatták, hogy az esszenciális gének alacsony rekombinációs rátával rendelkező régiókban lokalizálódnak a genomban. Ezért mi is megvizsgáltuk, hogy esetünkben állhatnak-e hasonló okok a megfigyelt csoportosulások hátterében.

A hasadó élesztő Rec12 egy topoizomeráz-szerű protein (Spo11 a *S. cerevisiae*-ben), amely programozott módon dupla-szálú DNS töréseket (DSB-eket) hoz létre, a meiotikus rekombinációt elősegítendő (Fowler és mtsi, 2014). Ezért megvizsgáltuk, hogy a különböző GO kategóriák génjei mellett milyen arányban fordulnak elő kísérletesen igazolt Rec12 hasítóhelyek (27. ábra és Függelék 15. táblázat).

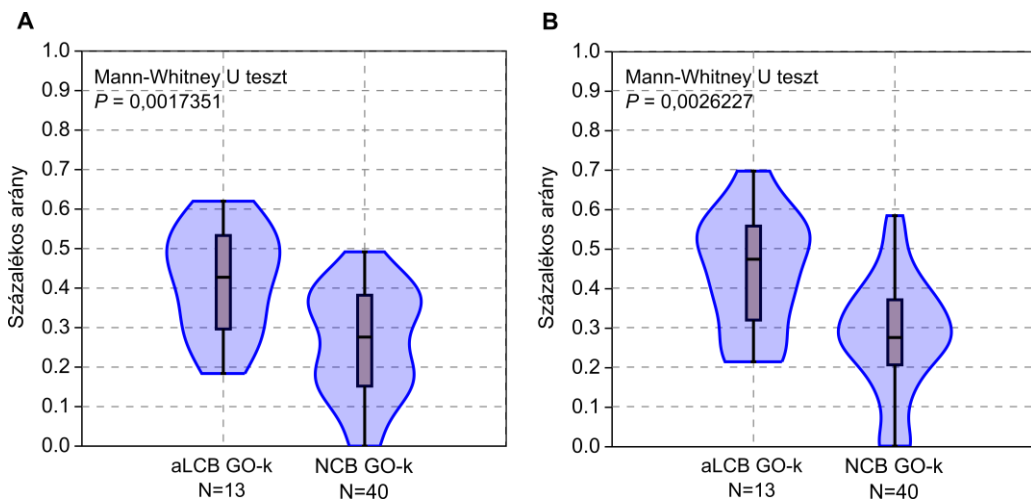


**27. ábra.** Rec12 hasítóhelyek megoszlása a különböző GO kategóriákba tartozó gének mentén. **A)** A hasítóhelyek megoszlása a szignifikánsan aLCB-kben csoportosuló 13 és NCB-kben csoportosuló 40 GO kategória génjei mentén. Az előbbi csoportban jelentősen alacsonyabb a Rec12 hasítóhelyek száma. **B)** Ha csak az aLCB-ken belül lokalizálódó gének esetében vizsgáltuk a hasítóhelyek számát, akkor a különbség nem bizonyult szignifikánsnak.

Megfigyeltük, hogy a kategóriákban 8% és 44% közötti volt azon gének aránya, amelyek mellett hasítóhelyek voltak. Találtunk számos csoportot, ahol alacsony volt a Rec12 hasítóhelyek száma, viszont a 13 aLCB-kben csoportosuló GO kategória átlaga (15%) jóval alacsonyabb volt, mint a többi kategória átlaga (23%) (27. ábra A). Viszont ha a hasítóhelyek gyakoriságát csak az aLCB-ken belül lokalizálódó gének esetében vizsgáltuk, a különbség nem bizonyult szignifikánsnak (27. ábra B).

#### 5.2.7.4. Esszenciális gének denzitása a különböző GO kategóriába tartozó gének között

Ezután megvizsgáltuk az esszenciális gének megoszlását a GO kategóriákon belül (28. ábra és Függelék 16. táblázat).



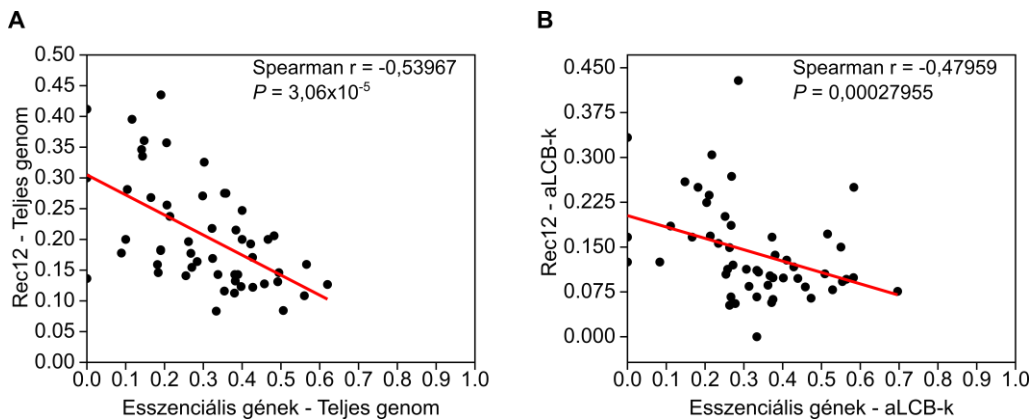
**28. ábra.** Az esszenciális gének lokális denzitása a különböző GO kategóriákban. **A)** Esszenciális gének denzitása a teljes genomra vonatkoztatva és **B)** csak az aLCB-kre vonatkoztatva. Mind a két esetben számottevően több esszenciális gén lokalizálódik a szignifikáns GO kategóriákon belül.

Számos csoportot találtunk, amelyek nagy számú esszenciális génnel rendelkeztek. A 13 aLCB-ben csoportosuló GO kategóriákban megfigyelhető esszenciális gének átlagos aránya jóval magasabb volt (42%), mint az egyéb GO kategóriákban (26%) (28. ábra A) és a különbség megmaradt abban az esetben is, ha csak az aLCB-kben lokalizálódó géneket vizsgáltuk (44% és 28%) (28. ábra B).

#### 5.2.7.5 Rec12 hasítóhelyek és esszenciális gének kapcsolata

Annak érdekében, hogy megállapíthassuk létezik-e valamiféle kapcsolat a Rec12 hasítóhelyek és esszenciális gének lokalizációja között, megvizsgáltuk, hogy milyen gyakran fordulnak elő egymás mellett. Kiderült, hogy az

esszenciális gének kis hányada mellett található csak Rec12 hasítóhely (14%), illetve a hasítóhelyek mellett lokalizálódó gének 17%-a esszenciális csak. Ebből talán arra következtethetünk, hogy a Rec12 endonukleáz hajlamos elkerülni az esszenciális géneket az *S. pombe* esetében is.



**29. ábra.** A Rec12 hasítóhelyek előfordulási gyakorisága (Y-tengely) és az esszenciális gének lokális denzitásának (X-tengely) korrelációja a GO kategóriákban. **(A)** Korreláció a teljes genomra vetítve és **(B)** korreláció csak az aLCB-kben lokalizálódó génekre vonatkozóan. A fekete pöttyök a különböző GO kategóriákat reprezentálják, a piros vonal a regressziós egyenes. Mindkét esetben egy közepes erősségű, de szignifikáns negatív korreláció figyelhető meg.

Ezt a feltevést erősíti a GO kategóriánkénti értékek korrelációjának vizsgálata is, ami szerint egy nem túl erős, de szignifikáns negatív korreláció áll fenn az esszenciális gének kategóriánkénti száma és a Rec12 hasítóhelyek előfordulási gyakorisága között (29. ábra). Ez a megfigyelés független attól, hogy a teljes genomot (29. ábra A), vagy csak az aLCB-kben lokalizálódó génekre vonatkozóan vizsgáljuk a jelenséget (29. ábra B). A fenti adatok arra engednek következtetni, hogy bizonyos GO kategóriák génjeinek szignifikánsan gyakoribb előfordulása az aLCB-kben, az adott kategóriába tartozó esszenciális gének számának és lokalizációjának is lehet a következménye.

## 6. Eredmények megbeszélése

### 6.1. A *S. cryophilus* szuperkontigok összerendezése és evolúciós analízise

Bár napjainkra rutinszerűvé vált a szekvenálás és szinte tömegesen kerülnek publikálásra az új genomszekvenciák, egy genom precíz összeszerelése és annotálása továbbra is komoly kihívást jelent (Bradnam és mtsi., 2013; Liao és mtsi., 2019). Emiatt számos esetben a genomszekvenciák nem is kerülnek összeszerelésre, sok az úgynevezett „draft” genom állapotban marad. Pedig az átfogó evolúciós vonatkozású komparatív genomikai elemzésekhez okvetlenül szükségesek lennének a precízen összeszerelt genomszekvenciák.

Mivel a hasadó élesztők egy rendkívül érdekes (és értékes) csoportját alkotják a gombák királyságának és a nemzetség egyik tagjának, a *S. cryophilus*-nak a genomja korábban nem került kromoszóma-szintű összeszerelési állapotba, célul tűztük ki a szóban forgó genom összerendezését, az összerendezés validálását és komparatív genomikai összehasonlításokat közeli és távoli rokonfajokkal egyaránt.

#### 6.1.1. A *S. cryophilus* szuperkontigok összerendezése szinténikus kapcsolatok segítségével

Első lépésként megkíséreltük a *S. cryophilus* 9 nagyméretű szuperkontigját összeszerelni a rokon fajok genomszekvenciái segítségével. Mi magunk nem szekvenáltuk újra a genomot, az adatbázisba feltöltött szuperkontigok szekvenciáiról feltételeztük, hogy precízen összeállítottak (Rhind és mtsi., 2011). Mivel a korábbi munkák rámutattak, hogy a *S. cryophilus* legközelebbi rokonai valószínűleg a *S. octosporus* és *S. pombe*, mindemellett a géntartalom rendkívül konzervált a rokon fajok között, ezért e fajok genomszekvenciáit használtuk referenciaként (Helston és mtsi, 2010; Rhind és mtsi., 2011). A *S. japonicus* genomját nem vontuk be a vizsgálatokba, mivel ez a faj a nemzetség legkorábbi leágazása és egyben a legdivergensebb is, tehát

mind a génszekvenciáit, mind a genomzinténiát tekintve a *S. cryophilus* legtávolabbi rokonának tekinthető (Helston és mtsi, 2010; Rhind és mtsi., 2011).

A Mauve programmal végzett teljes genom illesztések rávilágítottak több különböző jelenségre is. Eredményeink megerősítették Helston és mtsi (2010) és Rhind és mtsi (2011) eredményeit, miszerint valóban a *S. octosporus* a *S. cryophilus* legközelebbi rokona, hiszen sokkal nagyobb kiterjedésű LCB-k voltak azonosíthatók a fajok között, mint az *S. cryophilus* – *S. pombe* fajpár esetében. Emellett, a szekvenciaillesztéseink azt is megmutatták, hogy a *S. cryophilus* genomszekvenciája nagyrészt a rokonfajok LCB-inek a mozaikjaiból tevődik össze. Ugyanakkor meggyőződhattünk arról is, hogy még ilyen nagyfokú genom konzerváltság mellett sem lehet pusztán a szinténikus kapcsolatokra hagyatkozva összeilleszteni egy genomot. Ugyanis a *S. cryophilus* szuperkontigjainak újrendezése merőben más sorrendet eredményezett a használt referenciafajok függvényében.

Ezért megpróbáltunk erősen konzervált „horgonypontokat” keresni a *S. cryophilus* és rokonfajai szekvenciáiban, amelyek támpontul szolgálhattak az összeillesztéshez. Rhind és mtsi (2011) tanulmányukban bemutatták, hogy a *S. octosporus* – *S. pombe* fajpár esetében a pericentromerikus géntartalom szinte megegyezik. Ennek az lehet az oka, hogy a centromert érintő vagy centromer közeli átrendeződések akár végzetesek is lehetnek a sejtek számára, ezért erős szelekciós hatás alatt állnak (Ellermeier és mtsi., 2010; Sasaki és mtsi., 2010; Fowler és mtsi., 2014). Bár a szubtelomerikus régió hajlamos az átrendeződésre, azonban a géntartalom meglehetősen konzervált ebben a régióban is a közeli rokon fajok esetében (Fabre és mtsi., 2005; Rhind és mtsi., 2011). A *S. pombe* és *S. octosporus* pericentromerikus és szubtelomerikus génjeinek segítségével sikerült beazonosítanunk a *S. cryophilus* összes pericentromerikus és néhány szubtelomerikus kontigvégeit. A pericentromerikusnak vélt kontigvégek esetében további bizonyítéknak számított, hogy tRNS kódoló gének klasztereit is megtaláltuk ezekben a régiókban, hasonlóan a közeli rokon fajokhoz (Kuhn és

mtsi., 1991; Rhind és mtsi., 2011; Tong és mtsi., 2019). Habár a szubtelomerikus régiók esetében nem sikerült az összeset beazonosítani, aminek háttérében vagy a feltételezhetően bekövetkezett átrendeződések magas számai vagy pedig egyszerűen csak a hiányos adatok állhattak, az Sc3, Sc4, Sc6 és Sc7 kontigokról megállapítottuk, hogy szubtelomerikus régiót tartalmaznak. Egy feltételezett szubtelomerikus vég esetében (Sc7) rDNS klaszterre utaló szekvenciákat is találtunk, amely további bizonyítékként szolgált az adott régió szubtelomerikus voltára (Pasero & Marilley, 1993).

Ez új ismeretek birtokában újrendeztük a *S. cryophilus* szuperkontigjait és a rokon fajokhoz történő hasonlítás során a korábbiakhoz képest egy sokkal rendezettebb képet kaphattunk. Eredményeink alapján a következő szuperkontig sorrendet javasoltuk: Sc3-Sc9-Sc1 (ChrI); Sc7-Sc5-Sc8-Sc6 (ChrII); Sc4-Sc2 (ChrIII). Mindazonáltal azon kontigok esetében, amelyek sem pericentromerikusnak, sem szubtelomerikusnak nem bizonyultak, összetartozásuk bizonyítására (vagy cáfolatára) további kísérleteket terveztünk.

#### 6.1.2. A *S. cryophilus* szuperkontigok összeszerelésének validálása molekuláris módszerekkel

A Sc9-Sc1 és Sc5-Sc8-Sc6 szuperkontigok összetartozásának bizonyítása érdekében PCR reakciókat terveztünk és hajtottunk végre. A sikeres reakciók alátámasztották a bioinformatikai eredményeinket az összetartozónak vélt szuperkontigok esetében.

Habár a bioinformatikai eredmények és PCR-es validáció alapján a javasolt szuperkontig sorrend megfelelőnek tűnt, további bizonyosság szerzése céljából PFGE elemzésnek is alávetettük a *S. cryophilus* kromoszómáit. Megerősítettük azt a felvetést, miszerint a *S. cryophilus* OY26 törzse is 3 jól elkülöníthető kromoszómával rendelkezik, mint az egyéb hasadó élesztő törzsek túlnyomó többsége (Rhind és mtsi., 2011). Ennek azért is volt jelentősége, mivel a *S. pombe* törzsek között előfordul 4 kromoszómával rendelkező változat is

(Brown és mtsi., 2014). Azt is megfigyelhettük, hogy a *S. cryophilus* kromoszóma méretei valamelyest eltérnek a *S. pombe* kromoszómáinak méreteitől. Habár a PFGE analízisek alapján nehéz pontosan megállapítani a kromoszómák méreteit, különösen, ha nincs megfelelő felbontású méretstandard, azonban valószínűsíthető volt, hogy a megfigyelt kromoszómaméretetek eltérnek a szekvencia adatoktól. Az eltérések valószínűleg a centromerek, telomerek és rDNS ismétlődések ismeretlen kiterjedéséből származtak.

Ezt a feltevést ellenőrizendő, megkíséreltük meghatározni a rDNS-ek (18S-5,8S-26S rDNS) lokalizációját a *S. cryophilus* kromoszómák végein. Kiindulásként a *S. cryophilus* – *S. octosporus* szinténikus kapcsolatait vettük alapul, majd a potenciális rDNS ismétlődések lokalizációját PCR-rel kívántuk megállapítani. Két feltételezett esetből (Sc3 és Sc7) egyet tudunk kísérletesen is bizonyítani (Sc7), de ettől függetlenül a közvetett bizonyítékok arra utaltak, hogy az általunk javasolt kromoszómafelépítés megfelelő lehet.

Egy nemrégiben megjelent tanulmány, amelyben harmadik generációs módszerrel újraszekvenálták a hasadó élesztő genomokat, megerősítette az eredményeinket (Tong és mtsi., 2019). Csupán egy eltérést mutattak ki, ugyanis egy centromerikus átkereszteződés az Sc7 és Sc4 kontigok cseréjét eredményezte. A centromer szekvenciák ismerete hiányában az általunk alkalmazott módszerek nem voltak képesek fényt deríteni e változásra. Mindazonáltal a későbbi eredményeket ez a változás nem befolyásolta.

### 6.1.3. A *S. cryophilus* genom összevetése közeli és távoli rokon fajokkal

A *S. cryophilus* szuperkontigjainak összerendezését és annak validálását követően kíváncsiak voltunk, hogy milyen típusú változások formálták e hasadó élesztő genomot.

Kimutattuk, hogy a *S. octosporus* genomhoz képest számos kromoszómákat érintő változás érte a *S. cryophilus* genomját, amelyek főleg

interkromoszómális transzlokációk voltak és nagyméretű régiókat érintettek. Tong és mtsi (2019) eredményei alapján még egy centromer régióban bekövetkező reciprok transzlokáció is keletkezhetett. Eredményeink arra az érdekességre is rávilágítottak, hogy az inverziók száma jelentősen megnövekedett a filogenetikai távolság növekedésével. Tehát a *S. cryophilus* – *S. pombe* összehasonlítása során már az inverziók száma emelkedett meg a transzlokációkkal szemben.

Rhind és mtsi (2011) komparatív genomikai tanulmányukban bemutatták, hogy a hasadó élesztő genomok géntartalma rendkívül konzervált például a hasonló divergencia idővel rendelkező *Saccharomyces* nemzetséghez képest. Kíváncsiak voltunk, hogy az átrendeződések tekintetében hogyan viszonyul a két csoport egymáshoz. Ezért megvizsgáltuk a közelebbi rokon *S. cerevisiae* – *S. uvarum* és távolabbi rokon *S. cerevisiae* – *N. castelli* fajpárokat is hasonló megközelítéssel, mint a hasadó élesztőket. Ha az összes megfigyelt átrendeződést figyelembe vettük, akkor számottevően több átrendeződés érthető a sarjadzó élesztő genomokat, mint a hasadó élesztőket. Viszont, ha kizártuk a szubtelomerikus régióban bekövetkező változásokat (mivel azok hajlamosak az átrendeződésekre, Fischer és mtsi, 2001), akkor egészen más következtetéseket vonhattunk le. A szigorúbb megközelítéssel feltárt átrendeződések száma a *S. cerevisiae* – *S. uvarum* esetében megegyezett a Fischer és mtsi (2000, 2001) által becsült értékekkel. Viszont ebben az esetben a hasadó élesztők kromoszómáin arányaiban sokkal több nagy méretű régiókat érintő átrendeződés következhetett be, mint a vizsgált sarjadzó élesztő genomokban. A hasadó élesztők konzervált géntartalma alapján arra számítottunk, hogy kevesebb átrendeződést figyelhetünk majd meg. Az átrendeződések típusait vizsgálva hasonló tendenciát figyelhetünk meg a sarjadzó élesztők esetében is, miszerint a filogenetikai távolsággal növekszik az inverziók száma.

E megfigyelések alapján felmerülhet a kérdés, hogy a transzlokációk tolerálhatóbbak lennének az inverzióknál kisebb evolúciós léptékben? Mivel

mind a transzlokációk, mind az inverziók egyaránt hatással lehetnek a teljes genom génextpressziós mintázatára, ezáltal fokozva vagy csökkentve az adott élőlény fitnessét, továbbá képesek reprodukív izoláció előidézésére, ez a feltevés nem valószínű (Lowry & Willis, 2010; Avelar és mtsi., 2013; Zanders és mtsi., 2014; Naumov és mtsi., 2015). Feltételezéseink szerint inkább az átrendeződéseket elősegítő vagy okozó hatások mechanizmusa kedvezhet a transzlokációknak. Ezzel egybehangzóan az is elképzelhető, hogy a nagyobb evolúciós távolságok esetében is túlsúlyban maradnak a transzlokációk, egyszerűen a bekövetkezett inverziók mossák el a transzlokációk nyomait (Seoighe és mtsi., 2000).

## **6.2. A stabil hasadó élesztőgomba genomok genetikai és evolúciós háttere**

Nyilvánvaló vált, hogy a nagy arányú géntartalom konzerváltság ellenére elég sok átrendeződés érhet az előzőekben vizsgált három hasadó élesztő faj genomját, továbbá a fajok meglepően gyors szekvencia evolúciós rátát mutatnak (Helston és mtsi., 2010; Rhind és mtsi., 2011). Ezért kíváncsiak voltunk, hogy mutat-e korrelációt a hasadó élesztők strukturális és szekvencia evolúciója, majd megvizsgáltuk, hogy milyen mértékű a genom konzerváltság a hasadó élesztőkben. Továbbá arra a kérdésre is kerestük a választ, hogy mi lehet a genetikai és evolúciós háttere a stabil genomjaiknak? A vizsgálatokhoz a *S. japonicus*-t használtuk referenciaként, hiszen ez a faj a klád legkorábbi leágazása.

### 6.2.1. Strukturális és szekvencia evolúció összehasonlítása a hasadó élesztőkben

Bebizonyítottuk, hogy a hasadó élesztők szekvencia- és strukturális evolúciója erős pozitív korrelációt mutat, hasonlóan például a *Lachancea*, *Verticillium* és számos Metazoa nemzetséghez, de eltérően az *Aspergillus* nemzetségtől (Burt és mtsi., 1999; Coghlan & Wolfe, 2002; Sharakhov és mtsi., 2002; Galagan és mtsi., 2005; Vakirlis és mtsi., 2016; Shi-Kunne és mtsi., 2018). Ugyanis az *A. fumigatus* és *A. oryzae* aminosav azonosságukat tekintve hasonló

értékeket mutatnak az *A. nidulans*-hoz viszonyítva, viszont az átrendeződések tekintetében az *A. oryzae* genomja több mint kétszer annyi változást szenvedhetett el az *A. fumigatus*-hoz képest (Galagan és mtsi, 2005).

Megállapítottuk azt is, hogy a strukturális evolúció szemszögéből a *S. japonicus*-hoz képest a másik három faj közel ugyanolyan mértékű divergenciával rendelkezik, ami arra enged következtetni, hogy nagyon egyenletes a változások dinamikája a genomjaikban. Egy más megközelítést alkalmazó tanulmány ugyanerre a következtetésre jutott egyéb Taphrinomycotina fajok bevonásával is (Rajeh és mtsi., 2018). Ez az eredmény mindenképp meglepő, ha figyelembe vesszük a nemzetségen belüli, egyébiránt jelentős filogenetikai távolságokat. Elképzelhetőnek tartjuk – mivel a Taphrinomycotina altörzsbe tartozó fajok többsége eléggé specializálódott életmódot követ – hogy az életmódjuk állhat a kimagaslóan erős szekvencia- és struktúra változások korrelációinak hátterében.

#### 6.2.2. Genom konzerváltság és ősi kolineáris blokkok azonosítása a hasadó élesztő genomokban

Eredményeink arra is rávilágítottak, hogy közel egyenlő mértékű a konzerváltság nukleotid szinten a fajok között a *S. japonicus* perspektívájából. Sőt, nemcsak a konzervált régiók összesített mérete, hanem a szegmensek egyéni méreteinek megoszlása is szignifikáns hasonlósággal bírt. Ez arra engedett következtetni, hogy hasonló régiók maradhattak változatlanok a hasadó élesztő genomokban. Ezt a feltevést erősítette meg a fajok azon genomi régióinak feltárása, amelyek minimum 5 (átlagosan 8) ugyanolyan sorrenddel és irányultsággal rendelkező ortológ génből álltak. Eredményeink alapján a vizsgált genomok géntartalmának 40-42%-a ilyen aLCB-kben helyezkedik el és valószínű, hogy ezek a blokkok a legközelebbi közös ős génsorrendjét tükrözik. Ha figyelembe vesszük, hogy például a fugu (gömbhal) és humán genom kisebb mértékű szekvencia divergenciát mutat, mint a vizsgált élesztők, ennek ellenére

maximum 2-3 génből álló szinténikus blokkok azonosíthatók a genomjaikban, az eredmények igen érdekesek (Sipiczki, 2000; Smith és mtsi., 2002; Rhind és mtsi., 2011). Mindemellett, Rajeh és mtsi (2018) adatainak alapos tanulmányozása alapján nem is igen találni egyéb olyan genust az Ascomycota törzsben, ahol ilyen mértékű szekvencia divergencia mellett ilyen mértékű génsorrend konzerváltság lenne megfigyelhető.

### 6.2.3. Az ősi kolineáris blokkok fennmaradása és a természetes szelekció kapcsolata

A kolineáris blokkok, szomszédos génpárok létezése, amelyek megfigyelhetők különböző genomokban, lehetnek a szelekció vagy a véletlen következményei (Hurst és mtsi., 2004). Az előbbi esetben előnyös, ha bizonyos gének szomszédosak maradnak a megfelelő reguláció vagy ko-expresszió miatt. A másik esetben a génpárok szétválhatnak, majd újra egymás mellé kerülnek a sok átrendeződésnek köszönhetően. Mivel a hasadó élesztők esetében a géntartalom 40-42%-a kolineáris blokkokban helyezkedik el, a véletlen hatása minimalizálható. Inkább az a kérdés, hogy a megfigyelt konzerváltság „csupán” az ősi génsorrend maradványa vagy netán a szelekció is közrejátszik a fenntartásukban?

A kérdés megválaszolására különböző „in silico” modelleket használtunk és mind a neutrális megközelítés, mint a filogenetikai módszer ugyanazt az eredményt sugallta. A szimulációk alapján az aLCB-k fennmaradása nem lehet csupán a véletlen műve, mivel egyetlen teszt-eredmény sem közelítette meg a valóságban megfigyelhető értékeket az aLCB-k átlagos géntartalmát tekintve. Feltevésünk szerint az aLCB-k fenntartó szelekció alatt állnak.

### 6.2.4. Az ősi kolineáris blokkok lehetséges eredete

Kíváncsiak voltunk, hogy az aLCB-k génsorrendje mennyire mutatkozik konzerváltnak, ha más fajokat is bevonunk a vizsgálatba? Ismert tény, hogy

kiterjedt kolineáris régiók nem figyelhetők meg különböző genuszok összehasonlítása során, sőt még csak két szomszédos ortológ gén sem található például a *S. pombe* és *S. cerevisiae* genomjában (Poyatos & Hurst, 2006). Azonban, ha az aLCB-k génsorrendje a legközelebbi közös őstől ered és szelekciós hatás alatt is áll, van némi esély arra, hogy hasonló génsorrendeket figyelhessünk meg egyéb Taphrinomycotina fajok esetében is. Habár kisebb méretű kolineáris régiókat megfigyelhettünk a vizsgált fajpárok között, releváns számú blokk nem volt azonosítható.

Filogenetikai szempontból távolabbi fajok bevonása is hasonló eredményt hozott, tehát a vizsgált aLCB-k génsorrendje csak a hasadó élesztőkre jellemző. Azonban mindenképp fontosnak tartjuk megjegyezni, hogy számottevően több szomszédos génpárt sikerült beazonosítani a fonalas fajok genomjaiban, mint a sarjadzó élesztő genomokban. Ez az eredmény egybevág azzal a feltételezéssel, miszerint a modern élesztők ősei fonalas gombák voltak és lehetséges, hogy a megfigyelt génsorrendbeli hasonlóságok a fonalas ősök génsorrendjét tükrözik (Berbee & Taylor, 1993; Sipiczki, 2000; Nagy és mtsi., 2014).

#### 6.2.5. Az ősi kolineáris blokkok konzerváltsága szekvencia szinten

Miután meggyőződünk arról, hogy a megfigyelt génsorrend leginkább a hasadó élesztőkre jellemző, kíváncsiak voltunk, hogy vajon az aLCB-kben lokalizálódó szekvenciák nagyobb konzerváltságot mutatnak-e a genom egyéb régióiban elhelyezkedő szekvenciákhoz képest. Nem törvényszerű, hogy gének, amelyek sorrendjüket tekintve változatlanok maradtak, azok a szekvenciáikat tekintve is kevésbé változzanak, de mivel az átrendeződési események maguk is mutagén hatásúak, így talán a változatlan régiók konzerváltabbak lehetnek szekvencia szinten is. Ezt vizsgálándó, összehasonlítottuk több mint 4200 protein szekvencia evolúciós rátáját az aLCB-ken belül, illetve azokon kívül is. Bár a protein szekvenciák nem tükrözik teljes mértékben a DNS szintjén

bekövetkező változásokat, a változások dinamikájáról szolgáltathatnak elegendő információt. Meglepetésünkre nem volt szignifikáns különbség a két csoport között. Ennek fényében megállapítható, hogy az aLCB-ken belül lokalizálódó gének szekvenciái hasonló mértékű változékonyságot mutatnak, mint a genom egyéb régióiban elhelyezkedők. Az intront vesztett vagy nyert gének vizsgálata is hasonló eredményre vezetett.

E megfigyelések mögött több dolog is állhat. Egyrészt az aLCB-ken kívüli genomi régiók sem egészen kaotikusak, mivel találhatóak kevesebb, mint 5 génből álló kolineáris blokkok is. Másrészt a robusztus átrendeződések nem csak lokálisan, hanem a teljes genomra képesek hatással lenni, ezáltal olyan régiókban is indukálódhatnak változások, amelyeket az átrendeződések nem is érintettek (Avelar és mtsi., 2013).

#### 6.2.6. A hasadó élesztőgombák stabil génsorrendjének lehetséges okai

Számos tényező játszhat közre a hasadó élesztők stabil génsorrendjének fenntartásában. Egy lehetséges ok az átrendeződések által potenciálisan okozott reprodukív izoláció. A nagy régiókat érintő átrendeződések, úgymint az interkromoszómális transzlokációk eltérő méretű kromoszómákhoz vezethetnek, amelyek negatív hatással lehetnek a szexuális ciklusra (pl. nem megfelelő kromatid párosodás a meiózisban) (Avelar és mtsi., 2013; Zanders és mtsi., 2014; Naumov és mtsi., 2015; Jeffares és mtsi., 2017). Függetlenül attól, hogy a képződő új struktúra-variánsok előnyökkel rendelkezhetnek bizonyos környezeti körülmények között, az életképes utód létrehozásának ellehetetlenülése nagy hátrányt jelenthet egy folyamatosan változó élettérben, különösen a haploid fajok esetében (Avelar és mtsi., 2013; Zanders és mtsi., 2014; Naumov és mtsi., 2015; Jeffares és mtsi., 2017). Továbbá, számos tanulmány bizonyítja, hogy a szexuális ciklus okozta genetikai változások jobban elősegítik az adaptációt, mint a spontán mutációk vagy átrendeződések (Takouridis és mtsi., 2015; McDonald és mtsi., 2016; Scheuerl & Stelzer, 2017;

Tusso és mtsi., 2019). Az átrendeződések által kiváltott reprodukív izoláció mellett az is az átrendeződések ellen hathat, hogy a *S. pombe* genomszekvenciájának közel 90%-a funkcióval rendelkezik (Gretch és mtsi., 2019). Így a funkcióval rendelkező régiókban bekövetkező változások ellen is hathat a szelekció, mivel a funkcióvesztés hosszú távon hátrányt jelenthet, bizonyos változások akár letálisak is lehetnek (Avelar és mtsi., 2013; Zanders és mtsi., 2014, Naseeb és mtsi., 2016). A hasadó élesztő és sarjadzó élesztő genomokban bekövetkezett átrendeződések becsült értékeinek összehasonlítása is azt mutatta, hogy függetlenül a hasonló divergencia időktől, a hasadó élesztőkben összességében kevesebb változás történt.

További ok lehet az is, hogy a szomszédos gének közös transzkripciós szabályozás alatt állnak, esetleg funkcionális csoportokat alkotnak. Ez elég gyakori jelenségnek számít a magasabb rendű eukarióta élőlények körében, de érdekes módon, az ilyen funkcionális csoportosulások hajlamosak az átrendeződésekre, megadván ezzel a lehetőséget új funkcionális csoportok kialakítására (Hurst és mtsi., 2004; Poyatos & Hurst, 2006; Liu & Han, 2009; Al-Shahrour és mtsi., 2010; Dávila López és mtsi., 2010; Noble & Andrianopoulos, 2013). Ebből következően, a funkcionális csoportosulás nem szolgálhat minden esetben magyarázattal a hosszú távú génsorrend megőrződésre. Mindemellett Tuller és mtsi (2009) nem tudtak kimutatni egyértelmű funkcionális csoportosulásokat a *S. pombe* genomjában. A mi eredményeinkből is ez a következtetés vonható le. Habár egyértelműen kimutattuk, hogy meghatározott funkcióval (GO kategóriák) rendelkező gének hajlamosabbak az aLCB-kben csoportosulni, a csoportosulás nem a funkcióhoz kötött, mivel egy aLCB-n belül ritkán voltak ugyanabba a GO kategóriába tartozó gének megtalálhatók és e gének ko-expressziója sem mutatott magasabb értékeket az egyéb régiókban lokalizálódó gének ko-expressziójához képest. Talán a kromoszómák 3D-s konformációi felfedhetnének funkcionális csoportosulásokat, azonban az ilyen típusú vizsgálatok túlmutatnak e doktori

tanulmány célkitűzésein (Tanizawa és mtsi., 2010; Gong és mtsi., 2015). Az adataink inkább azt sejtetik, hogy azon GO kategóriák génjei lokalizálódnak az aLCB-kben, amelyek között nagyobb arányú az esszenciális gének száma, továbbá kisebb arányú a közelükben elhelyezkedő Rec12 hasítóhelyek száma. Ezek az eredmények összhangban vannak Poyatos & Hurst (2006) eredményeivel, miszerint az esszenciális gének lokális denzitása, továbbá a rekombinációs forrópontok nagyban hozzájárulhatnak a génsorrend megőrzéséhez, illetve változásához a sarjadzó élesztő genomokban is.

## 7. Összefoglalás

Mára teljes mértékben elfogadottá vált a nézet, miszerint az eukarióta génsorrend sem véletlenszerű. Számos bizonyíték támasztja alá ezt a feltételezést, úgymint a hasonló funkcióval rendelkező vagy egy metabolikus útvonalhoz tartozó gének ko-lokalizációja, szomszédos gének ko-expressziója és ko-regulációja. További bizonyíték a nem véletlenszerű génsorrendre a lokális átrendeződések által okozott globális transzkripciós mintázat megváltozása. De ide sorolhatjuk a transzgen integrációk változatos kimeneteleit is.

A génsorrend konzerváltságát fenntartó tényezők vizsgálatára kiválóan alkalmasak a különböző élesztő modellszervezetek, mivel azok kisméretű, kompakt genommal rendelkeznek és a nagy egyedszámuk és gyors generációs idejük miatt a szelekció is megfelelően ki tudja fejteni a hatását az egyedfejlődésük során.

Jelen doktori tanulmányban a hasadó élesztő (*Schizosaccharomyces*) modellszervezetek genomszekvenciáit vettük górcső alá, vizsgálva a genomjaik által elszenvedett evolúciós léptékű változásokat és válaszokat keresve a rendkívül stabil génsorrendjük fenntartásával kapcsolatos kérdésekre. Munkánk során főleg számítógépes módszereket alkalmaztunk, azonban néhány kísérlettel is kiegészítettük a vizsgálatainkat.

Eredményeink összefoglalása következik:

- Sikerült kromoszóma szintű összeszerelési állapotba hozni a *S. cryophilus* kontigjait a rokon fajokkal való szinténikus kapcsolatok segítségével. Bioinformatikai eredményeink alapján a Sc3-Sc9-Sc1 (ChrI); Sc7-Sc5-Sc8-Sc6 (ChrII); Sc4-Sc2 (ChrIII) sorrendet javasoltuk.
- Munkánk során a pericentromerikus és szubtelomerikus régiók kolineáris génsorrendjeinek feltárása is támpontot adott az összeszereléshez.

- Molekuláris biológiai technikákkal validáltuk a számítógépes módszerekkel megállapított kontigsorrendeket: PCR módszerrel sikerült bizonyos kontigok szomszédosságát igazolni, míg az elvégzett PFGE analízis alapján kimutattuk, hogy a *S. cryophilus* is 3 kromoszómával rendelkezik. Ezek a kromoszómák méreteiket tekintve eltérnek a jól ismert *S. pombe* kromoszóma méreteitől. Továbbá sikeresen kimutattunk egy rDNS klasztert is az egyik kromoszóma karon.
- Számítógépes módszerekkel megbecsültük, hogy mennyi és milyen típusú átrendeződések formálhatták a hasadó élesztők közeli rokon fajainak genomjait. Vizsgálataink során a *S. octosporus*-hoz és *S. pombe*-hez hasonlítottuk a *S. cryophilus* genomját, amelynek eredményeként megállapítottuk, hogy a genomokat ért átrendeződések tekintetében is a *S. octosporus* a leghasonlóbb a *S. cryophilus*-hoz.
- Kimutattuk, hogy az átrendeződések közül az interkromoszómális transzlokációk voltak túlsúlyban, az inverziók száma csak a filogenetikai távolság növekedésével emelkedett.
- Összehasonlítottuk az átrendeződések típusa és mennyisége szempontjából a hasadó élesztő genomokat hasonló divergencia idővel rendelkező sarjadzó élesztő genomokkal. Bár összességében több átrendeződési esemény érhet a sarjadzó élesztőket, a hasadó élesztő genomok szenvedték el a nagyobb méretű régiókat érintő változásokat. Érdekes módon a sarjadzó élesztő genomok is hasonló tendenciát mutattak, miszerint eleinte a transzlokációk voltak túlsúlyban, az inverziók csak a filogenetikai távolság növekedésével kerültek túlsúlyba.
- További elemzéseinkbe bevontuk a hasadó élesztők legdivergensebb tagját, a *S. japonicus*-t is, és e fajt használva referenciaként kimutattuk, hogy a hasadó élesztő genomokban a strukturális- és szekvencia változások dinamikája rendkívül erős pozitív korrelációt mutat.

- Megállapítottuk, hogy a vizsgált fajok genomjaiban az átrendeződések nagyon egyenletesen következtek be, mivel a többi hasadó élesztő faj közel ugyanolyan mértékű kromoszóma-strukturális divergenciát mutatott a *S. japonicus*-hoz képest.
- Feltártuk a genomok legkonzerváltabb részeit és legalább 5 génből álló kolineáris régiókat azonosítottunk (aLCB-k). E blokkok génsorrendjét az élesztők valószínűleg a legközelebbi közös őstől örökölhették. Ilyen aLCB-kben lokalizálódik a genomok géntartalmának 40-42%-a.
- Számítógépes szimulációkat alkalmazva kimutattuk, hogy az aLCB-k génsorrendje nem csak az ősi génsorrend maradványa, hanem szelekciós hatás alatt is állhat.
- Megállapítottuk, hogy az aLCB-k génsorrendje bár hasadó élesztő specifikus, de tükrözheti egy fonalas ős génsorrendjét is, mivel a vizsgált aLCB-k génsorrendje jobban hasonlít a fonalas gombák génsorrendjére, mint a sarjadzó élesztő fajok génsorrendjére.
- Kimutattuk, hogy az aLCB-ken belül lokalizálódó gének szekvencia és strukturális (intron vesztés/nyerés) változásai nem történnek alacsonyabb intenzitással, mint a genomok egyéb régióiban elhelyezkedő gének esetében.
- Eredményeink azt sugallták, hogy bizonyos biológiai funkciókhoz (13 GO kategória) köthető gének szignifikánsan nagyobb arányban hajlamosak az aLCB-kben csoportosulni, mint mások. Mindazonáltal a GO kategóriák gényei is szétszórt elhelyezkedést mutatnak a kromoszómákon, tehát az egy biológiai funkcióhoz köthető gének nem szomszédosak az aLCB-kben.
- Az aLCB-kben elhelyezkedő gének ko-expressziós értékei nem magasabbak, mint a genom egyéb részeiben elhelyezkedő szomszédos gének ko-expressziós értékei, de szignifikánsan magasabbak, mint a teljes genomra vonatkoztatott átlagérték.

- Felfedtük, hogy az aLCB-kben csoportosuló GO kategóriák génjei között magasabb az esszenciális gének aránya, ugyanakkor kevesebb a gének környékén lokalizálódó Rec12 hasítóhelyek száma.

Eredményeink hozzájárulnak a hasadó élesztők, ezáltal az eukarióta génsorrend megőrzésében szerepet játszó jelenségek megismeréséhez, megértéséhez.

## 8. Summary

It is now fully accepted that the eukaryotic gene order is not random. Several pieces of evidence support this assumption, such as the co-localisation of genes with similar functions or belonging to the same metabolic pathway, co-expression and co-regulation of neighbouring genes. Further evidence for non-random gene order is the alteration of global transcriptional patterns caused by local rearrangements. But we can also include the diverse outcomes of transgene integration.

The use of various yeast model organisms is an excellent way to study the factors that maintain the conservation of the gene order, as they have a small, compact genome and, due to their large number of individuals and rapid generation time, selection can have a sufficient effect on their evolution.

In this present PhD study, we have investigated the genome sequences of the fission yeast (*Schizosaccharomyces*) model organisms, examining the evolutionary changes their genomes have undergone and seeking answers to questions about the maintenance of their highly stable gene order. Our work was mainly based on computational methods, but we also complemented our studies with certain experiments.

Our results are the following:

- The contigs of *S. cryophilus* have been successfully assembled at the chromosome level using syntenic relationships with related species. Based on our bioinformatic results, we proposed the order Sc3-Sc9-Sc1 (ChrI); Sc7-Sc5-Sc8-Sc6 (ChrII); Sc4-Sc2 (ChrIII).
- In the course of our work, we also explored the colinear gene sequences of pericentromeric and subtelomeric regions to support the assembly.
- Molecular biology techniques have been used to validate the contig order determined by computational methods: PCR has been used to confirm the neighbouring of certain contigs, while PFGE analysis has shown that

*S. cryophilus* also has 3 chromosomes. These chromosomes differ in the size from the well-known chromosome sizes of *S. pombe*. We also successfully detected an rDNA cluster on one chromosome arm.

- We used computational methods to estimate the number and type of rearrangements that could have shaped the genomes of closely related fission yeast species. Our studies compared the genomes of *S. cryophilus* with those of *S. octosporus* and *S. pombe*, and found that *S. octosporus* was also the most similar to *S. cryophilus* in terms of the rearrangements affecting the genomes.
- We showed that interchromosomal translocations were the most frequent among the rearrangements, with the number of inversions increasing only with increasing phylogenetic distance.
- We compared the type and number of rearrangements between fission yeast genomes and budding yeast genomes with similar divergence times. Although more rearrangement events may have occurred in the budding yeast genome overall, the changes affecting larger regions occurred in the fission yeast genomes. Interestingly, the budding yeast genomes also showed a similar trend, with translocations predominating initially, and inversions becoming predominant only as the phylogenetic distance increased.
- In our further analyses, we also included the most divergent member of the fission yeast *S. japonicus*, using as a reference species. We showed that the dynamics of structural and sequence changes in fission yeast genomes show a strong positive correlation.
- We found that the rearrangements in the genomes of the species occurred very uniformly, as the other fission yeast species showed almost the same degree of chromosomal structural divergence compared to *S. japonicus*.
- We have revealed the most conserved parts of the genomes and identified colinear regions of at least 5 genes (aLCBs). The gene order of these

blocks was probably inherited from their last common ancestor. The aLCBs contain 40-42% of the genome's whole gene content.

- Using computer simulations, we have shown that the gene order of aLCBs is not only a remnant of the ancestral gene order, but may also be under selection.
- We found that the gene order of aLCBs, though fission yeast specific, may reflect the gene order of a filamentous ancestor, as the gene order of the aLCBs studied more closely resembles the gene order of filamentous fungi than that of budding yeast species.
- We have shown that sequence and structural (intron loss/gain) changes in genes localized within aLCBs do not occur at lower intensities than for genes located in other regions of the genome.
- Our results suggested that genes associated with certain biological functions (13 GO categories) tend to cluster in aLCBs at significantly higher rates than others. However, the genes in these GO categories are dispersed across the chromosomes, implying that genes associated with the same biological function are not adjacent in the aLCBs.
- The co-expression values of genes located in aLCBs are not higher than the co-expression values of genes located in other parts of the genome, but they are significantly higher than the whole genome average.
- We discovered that genes from GO categories clustered in aLCBs have a higher proportion of essential genes but fewer Rec12 cleavage sites nearby.

Our results contribute to the knowledge and understanding of fission yeasts and thus to the phenomena involved in the conservation of the eukaryotic gene order.

## 9. Köszönetnyilvánítás

Köszönettel tartozom személyek sokaságának, de legfőképpen Istennek, aki végig vezetett ezen az úton.

Hála és köszönet illeti témavezetőmet, Gálné Dr. Miklós Idát kitartó támogatásáért, türelméért, nyitottságáért, aki tanácsaival mindvégig segítette a munkámat és tartotta bennem a lelket.

Köszönettel tartozom Dr. Sipiczki Mátyásnak, aki szakmai beszélgetéseink során észrevételeivel segítette a munkámat és biztatott a dolgozat megírására.

Köszönöm Dr. Papp László Attilának, kollégámnak, barátomnak a sok biztatást, szakmai beszélgetést és a PCR-es kísérletek, szkriptelések elvégzését. Szeretnék köszönetet mondani Dr. Batta Gyula Gábornak, első témavezetőmnek, aki megtanított az alapokra és megszerettette velem a molekuláris biológiát, genetikát. Köszönettel tartozom kollégámnak, Dr. Csoma Hajnalkának, aki sokat biztatott, lelkesített a doktori munka során. Köszönet illeti munkatársamat, Dr. Antunovics Zsuzsát a PFGE vizsgálatok elvégzéséért.

Szeretnék köszönetet mondani a Genetikai és Alkalmazott Mikrobiológiai Tanszék technikusainak: Lakatos Zoltánnénak, Oláh Anitának, Struba-Hegedűs Szilviának, Kiss Zoltánnénak és Bordánné Kovács Anitának, akik sok terhet levettek a vállamról a doktori munkán kívüli projekt-feladatok teljesítése során.

Végül, de nem utolsó sorban szeretnék köszönetet mondani a szüleimnek, feleségemnek türelmükért és támogatásukért.

A doktori munka megvalósulását elősegítették a TÁMOP-4.2.2/B-10/1-2010-0024; TÉT\_12\_JP-1-2014-0031 és EFOP-3.6.1-16-2016-00022 projekt-támogatások.

## 10. Irodalomjegyzék

- Akhtar W, de Jong J, Pindyurin AV, et al. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell*. 2013;154(4):914-927. doi:10.1016/j.cell.2013.07.018
- Alfa CE, Hyams JS. Distribution of tubulin and actin through the cell division cycle of the fission yeast *Schizosaccharomyces japonicus* var. *versatilis*: a comparison with *Schizosaccharomyces pombe*. *J Cell Sci*. 1990;96 ( Pt 1):71-77. doi:10.1242/jcs.96.1.71
- Allshire RC, Ekwall K. Epigenetic Regulation of Chromatin States in *Schizosaccharomyces pombe*. *Cold Spring Harb Perspect Biol*. 2015;7(7):a018770. Published 2015 Jul 1. doi:10.1101/cshperspect.a018770
- Alonso-Nuñez ML, An H, Martín-Cuadrado AB, et al. Ace2p controls the expression of genes required for cell separation in *Schizosaccharomyces pombe*. *Mol Biol Cell*. 2005;16(4):2003-2017. doi:10.1091/mbc.e04-06-0442
- Al-Shahrour F, Minguez P, Marqués-Bonet T, Gazave E, Navarro A, Dopazo J. Selection upon genome architecture: conservation of functional neighborhoods with changing genes. *PLoS Comput Biol*. 2010;6(10):e1000953. Published 2010 Oct 7. doi:10.1371/journal.pcbi.1000953
- Amsalem J, Cuomo CA, van Kan JA, et al. Genomic analysis of the necrotrophic fungal pathogens *Sclerotinia sclerotiorum* and *Botrytis cinerea*. *PLoS Genet*. 2011;7(8):e1002230. doi:10.1371/journal.pgen.1002230
- Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol*. 2006;55(4):539-552. doi:10.1080/10635150600755453
- Avelar AT, Perfeito L, Gordo I, Ferreira MG. Genome architecture is a selectable trait that can be maintained by antagonistic pleiotropy. *Nat Commun*. 2013;4:2235. doi:10.1038/ncomms3235
- Balazs A, Batta G, Miklos I, Acs-Szabo L, Vazquez de Aldana CR, Sipiczki M. Conserved regulators of the cell separation process in *Schizosaccharomyces*. *Fungal Genet Biol*. 2012;49(3):235-249. doi:10.1016/j.fgb.2012.01.003
- Batta G, Szilagyi Z, Laczik M, Sipiczki M. The involvement of the *Schizosaccharomyces pombe* sep9/spt8 gene in the regulation of septum

- cleavage. *FEMS Yeast Res.* 2009;9(5):757-767. doi:10.1111/j.1567-1364.2009.00522.x
- Beach D, Durkacz B, Nurse P. Functionally homologous cell cycle control genes in budding and fission yeast. *Nature.* 1982;300(5894):706-709. doi:10.1038/300706a0
- Beijerinck MW. *Schizosaccharomyces octosporus*, eine acht-sporige Alkoholhefe. *Zentralblatt Bakteriologie Parasitenkunde* 1894;16:49–58.
- Belokopytova P, Fishman V. Predicting Genome Architecture: Challenges and Solutions. *Front Genet.* 2021;11:617202. Published 2021 Jan 22. doi:10.3389/fgene.2020.617202
- Benito S. The impacts of *Schizosaccharomyces* on winemaking. *Appl Microbiol Biotechnol.* 2019;103(11):4291-4312. doi:10.1007/s00253-019-09827-7
- Berbee ML, Taylor JW. Dating the evolutionary radiations of the true fungi. *Botany.* 1993;71: 1114–1127.
- Berlaco M, Palumbo G, Piacentini L, Pimpinelli S, Fanti L. Position effect variegation and viability are both sensitive to dosage of constitutive heterochromatin in *Drosophila*. *G3 (Bethesda).* 2014;4(9):1709-1716. Published 2014 Jul 21. doi:10.1534/g3.114.013045
- Birnbaum K, Shasha DE, Wang JY, et al. A gene expression map of the *Arabidopsis* root. *Science.* 2003;302(5652):1956-1960. doi:10.1126/science.1090022
- Blumenthal T, Evans D, Link CD, et al. A global analysis of *Caenorhabditis elegans* operons. *Nature.* 2002;417(6891):851-854. doi:10.1038/nature00831
- Blumenthal T. Operons in eukaryotes. *Brief Funct Genomic Proteomic.* 2004;3(3):199-211. doi:10.1093/bfgp/3.3.199
- Boutanaev AM, Kalmykova AI, Shevelyov YY, Nurminsky DI. Large clusters of co-expressed genes in the *Drosophila* genome. *Nature.* 2002;420(6916):666-669. doi:10.1038/nature01216
- Bradnam KR, Fass JN, Alexandrov A, et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience.* 2013;2(1):10. Published 2013 Jul 22. doi:10.1186/2047-217X-2-10
- Brown WR, Liti G, Rosa C, et al. A Geographically Diverse Collection of *Schizosaccharomyces pombe* Isolates Shows Limited Phenotypic Variation

- but Extensive Karyotypic Diversity. *G3* (Bethesda). 2011;1(7):615-626. doi:10.1534/g3.111.001123
- Brown WR, Thomas G, Lee NC, et al. Kinetochores assembly and heterochromatin formation occur autonomously in *Schizosaccharomyces pombe*. *Proc Natl Acad Sci U S A*. 2014;111(5):1903-1908. doi:10.1073/pnas.1216934111
- Brysch-Herzberg M, Tobias A, Seidel M, et al. *Schizosaccharomyces osmophilus* sp. nov., an osmophilic fission yeast occurring in bee bread of different solitary bee species. *FEMS Yeast Res*. 2019;19(4):foz038. doi:10.1093/femsyr/foz038
- Bullerwell CE, Leigh J, Forget L, Lang BF. A comparison of three fission yeast mitochondrial genomes. *Nucleic Acids Res*. 2003;31(2):759-768. doi:10.1093/nar/gkg134
- Burt DW, Bruley C, Dunn IC, et al. The dynamics of chromosome evolution in birds and mammals. *Nature*. 1999;402(6760):411-413. doi:10.1038/46555
- Butler G, Rasmussen MD, Lin MF, et al. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature*. 2009;459(7247):657-662. doi:10.1038/nature08064
- Čadež N, Dlačhy D, Tome M, Péter G. *Novakomyces olei* sp. nov., the First Member of a Novel Taphrinomycotina Lineage. *Microorganisms*. 2021;9(2):301. Published 2021 Feb 2. doi:10.3390/microorganisms9020301
- Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000;17(4):540-552. doi:10.1093/oxfordjournals.molbev.a026334
- Caron H, van Schaik B, van der Mee M, et al. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science*. 2001;291(5507):1289-1292. doi:10.1126/science.1056794
- Chang SL, Lai HY, Tung SY, Leu JY. Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genet*. 2013;9(1):e1003232. doi:10.1371/journal.pgen.1003232
- Chen X, Zhang J. The Genomic Landscape of Position Effects on Protein Expression Level and Noise in Yeast. *Cell Syst*. 2016;2(5):347-354. doi:10.1016/j.cels.2016.03.009

- Cho RJ, Campbell MJ, Winzeler EA, et al. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell*. 1998;2(1):65-73. doi:10.1016/s1097-2765(00)80114-8
- Cissé OH, Almeida JM, Fonseca A, et al. Genome sequencing of the plant pathogen *Taphrina deformans*, the causal agent of peach leaf curl. *mBio*. 2013;4(3):e00055-e13. Published 2013 Apr 30. doi:10.1128/mBio.00055-13
- Coghlan A, Wolfe KH. Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Res*. 2002;12(6):857-867. doi:10.1101/gr.172702
- Cohen BA, Mitra RD, Hughes JD, Church GM. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet*. 2000;26(2):183-186. doi:10.1038/79896
- Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. 2010;464(7289):704-712. doi:10.1038/nature08516
- Cooper GM, Coe BP, Girirajan S, et al. A copy number variation morbidity map of developmental delay [published correction appears in *Nat Genet*. 2014 Sep;46(9):1040]. *Nat Genet*. 2011;43(9):838-846. Published 2011 Aug 14. doi:10.1038/ng.909
- Corradi N, Akiyoshi DE, Morrison HG, et al. Patterns of genome evolution among the microsporidian parasites *Encephalitozoon cuniculi*, *Antonospora locustae* and *Enterocytozoon bieneusi*. *PLoS One*. 2007;2(12):e1277. Published 2007 Dec 5. doi:10.1371/journal.pone.0001277
- Crécy-Lagard V, Hanson AD. Comparative Genomics☆, Reference Module in Biomedical Sciences, Elsevier, 2018, ISBN 9780128012383, <https://doi.org/10.1016/B978-0-12-801238-3.66095-6>.
- Dalquen DA, Anisimova M, Gonnet GH, Dessimoz C. ALF--a simulation framework for genome evolution. *Mol Biol Evol*. 2012;29(4):1115-1123. doi:10.1093/molbev/msr268
- Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. 2010;5(6):e11147. Published 2010 Jun 25. doi:10.1371/journal.pone.0011147
- Darwin C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. Murray. 1859.

- Dávila López M, Martínez Guerra JJ, Samuelsson T. Analysis of gene order conservation in eukaryotes identifies transcriptionally and functionally linked genes. *PLoS One*. 2010;5(5):e10654. Published 2010 May 14. doi:10.1371/journal.pone.0010654
- Deakin JE, Belov K. A comparative genomics approach to understanding transmissible cancer in Tasmanian devils. *Annu Rev Genomics Hum Genet*. 2012;13:207-222. doi:10.1146/annurev-genom-090711-163852
- Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science*. 2002;295(5558):1306-1311. doi:10.1126/science.1067799
- Dereeper A, Guignon V, Blanc G, et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res*. 2008;36(Web Server issue):W465-W469. doi:10.1093/nar/gkn180
- Dolatabadian A, Patel DA, Edwards D, Batley J. Copy number variation and disease resistance in plants. *Theor Appl Genet*. 2017;130(12):2479-2490. doi:10.1007/s00122-017-2993-2
- Domizio P, Liu Y, Bisson LF, Barile D. Cell wall polysaccharides released during the alcoholic fermentation by *Schizosaccharomyces pombe* and *S. japonicus*: quantification and characterization. *Food Microbiol*. 2017;61:136-149. doi:10.1016/j.fm.2016.08.010
- Dornburg A, Wang Z, Wang J, Mo ES, López-Giráldez F, Townsend JP. Comparative Genomics within and across Bilaterians Illuminates the Evolutionary History of ALK and LTK Proto-Oncogene Origination and Diversification. *Genome Biol Evol*. 2021;13(1):evaa228. doi:10.1093/gbe/evaa228
- Drillon G, Fischer G. Comparative study on synteny between yeasts and vertebrates. *C R Biol*. 2011;334(8-9):629-638. doi:10.1016/j.crvi.2011.05.011
- D'Souza CA, Kronstad JW, Taylor G, et al. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. *mBio*. 2011;2(1):e00342-10. Published 2011 Feb 8. doi:10.1128/mBio.00342-10
- Dujon B, Sherman D, Fischer G, et al. Genome evolution in yeasts. *Nature*. 2004;430(6995):35-44. doi:10.1038/nature02579
- Dujon B. Yeast evolutionary genomics. *Nat Rev Genet*. 2010;11(7):512-524. doi:10.1038/nrg2811

- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792-1797. Published 2004 Mar 19. doi:10.1093/nar/gkh340
- Elion EA, Brill JA, Fink GR. FUS3 represses CLN1 and CLN2 and in concert with KSS1 promotes signal transduction. *Proc Natl Acad Sci U S A.* 1991;88(21):9392-9396. doi:10.1073/pnas.88.21.9392
- Ellermeier C, Higuchi EC, Phadnis N, et al. RNAi and heterochromatin repress centromeric meiotic recombination. *Proc Natl Acad Sci U S A.* 2010;107(19):8701-8705. doi:10.1073/pnas.0914160107
- Enczi K, Yamaguchi M, Sipiczki M. Morphology transition genes in the dimorphic fission yeast *Schizosaccharomyces japonicus*. *Antonie Van Leeuwenhoek.* 2007;92(2):143-154. doi:10.1007/s10482-007-9142-x
- Fabre E, Muller H, Therizols P, Lafontaine I, Dujon B, Fairhead C. Comparative genomics in hemiascomycete yeasts: evolution of sex, silencing, and subtelomeres. *Mol Biol Evol.* 2005;22(4):856-873. doi:10.1093/molbev/msi070
- Fantes PA, Hoffman CS. A Brief History of *Schizosaccharomyces pombe* Research: A Perspective Over the Past 70 Years. *Genetics.* 2016;203(2):621-629. doi:10.1534/genetics.116.189407
- Farrer RA, Desjardins CA, Sakthikumar S, et al. Genome Evolution and Innovation across the Four Major Lineages of *Cryptococcus gattii*. *mBio.* 2015;6(5):e00868-15. Published 2015 Sep 1. doi:10.1128/mBio.00868-15
- Fawcett JA, Iida T, Takuno S, et al. Population genomics of the fission yeast *Schizosaccharomyces pombe*. *PLoS One.* 2014;9(8):e104241. Published 2014 Aug 11. doi:10.1371/journal.pone.0104241
- Fischer G, James SA, Roberts IN, Oliver SG, Louis EJ. Chromosomal evolution in *Saccharomyces*. *Nature.* 2000;405(6785):451-454. doi:10.1038/35013058
- Fischer G, Neuvéglise C, Durrens P, Gaillardin C, Dujon B. Evolution of gene order in the genomes of two related yeast species. *Genome Res.* 2001;11(12):2009-2019. doi:10.1101/gr.212701
- Fischer G, Rocha EP, Brunet F, Vergassola M, Dujon B. Highly variable rates of genome rearrangements between hemiascomycetous yeast lineages. *PLoS Genet.* 2006;2(3):e32. doi:10.1371/journal.pgen.0020032

- Florens L, Washburn MP, Raine JD, et al. A proteomic view of the *Plasmodium falciparum* life cycle. *Nature*. 2002;419(6906):520-526. doi:10.1038/nature01107
- Forsburg SL, Rhind N. Basic methods for fission yeast. *Yeast*. 2006;23(3):173-183. doi:10.1002/yea.1347
- Fowler KR, Sasaki M, Milman N, Keeney S, Smith GR. Evolutionarily diverse determinants of meiotic DNA break and recombination landscapes across the genome. *Genome Res*. 2014;24(10):1650-1664. doi:10.1101/gr.172122.114
- Franke M, Ibrahim DM, Andrey G, et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*. 2016;538(7624):265-269. doi:10.1038/nature19800
- Fudenberg G, Mirny LA. Higher-order chromatin structure: bridging physics and biology. *Curr Opin Genet Dev*. 2012;22(2):115-124. doi:10.1016/j.gde.2012.01.006
- Furuya K, Niki H. Hyphal differentiation induced via a DNA damage checkpoint-dependent pathway engaged in crosstalk with nutrient stress signaling in *Schizosaccharomyces japonicus*. *Curr Genet*. 2012;58(5-6):291-303. doi:10.1007/s00294-012-0384-4
- Gabriel A, Dapprich J, Kunkel M, Gresham D, Pratt SC, Dunham MJ. Global mapping of transposon location. *PLoS Genet*. 2006;2(12):e212. doi:10.1371/journal.pgen.0020212
- Galagan JE, Calvo SE, Borkovich KA, et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*. 2003;422(6934):859-868. doi:10.1038/nature01554
- Galagan JE, Calvo SE, Cuomo C, et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*. 2005;438(7071):1105-1115. doi:10.1038/nature04341
- Ghiurcuta CG, Moret BM. Evaluating synteny for improved comparative studies. *Bioinformatics*. 2014;30(12):i9-i18. doi:10.1093/bioinformatics/btu259
- Giannakou K, Cotterrell M, Delneri D. Genomic Adaptation of *Saccharomyces* Species to Industrial Environments. *Front Genet*. 2020;11:916. Published 2020 Aug 27. doi:10.3389/fgene.2020.00916

- Gilet J, Conte R, Torchet C, Benard L, Lafontaine I. Additional Layer of Regulation via Convergent Gene Orientation in Yeasts. *Mol Biol Evol.* 2020;37(2):365-378. doi:10.1093/molbev/msz221
- Goffeau A, Barrell BG, Bussey H, et al. Life with 6000 genes. *Science.* 1996;274(5287):546-567. doi:10.1126/science.274.5287.546
- Gómez-Gil E, Franco A, Madrid M, et al. Quorum sensing and stress-activated MAPK signaling repress yeast to hypha transition in the fission yeast *Schizosaccharomyces japonicus* [published correction appears in *PLoS Genet.* 2019 Jul 19;15(7):e1008282]. *PLoS Genet.* 2019;15(5):e1008192. Published 2019 May 31. doi:10.1371/journal.pgen.1008192
- Gong K, Tjong H, Zhou XJ, Alber F. Comparative 3D genome structure analysis of the fission and the budding yeast. *PLoS One.* 2015;10(3):e0119672. Published 2015 Mar 23. doi:10.1371/journal.pone.0119672
- Goodwin LO, Splinter E, Davis TL, et al. Large-scale discovery of mouse transgenic integration sites reveals frequent structural variation and insertional mutagenesis. *Genome Res.* 2019;29(3):494-505. doi:10.1101/gr.233866.117
- Gordon JL, Armisén D, Proux-Wéra E, ÓÉigeartaigh SS, Byrne KP, Wolfe KH. Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. *Proc Natl Acad Sci U S A.* 2011;108(50):20024-20029. doi:10.1073/pnas.1112808108
- Grallert A, Grallert B, Zilahi E, Szilagyí Z, Sipiczki M. Eleven novel sep genes of *Schizosaccharomyces pombe* required for efficient cell separation and sexual differentiation. *Yeast.* 1999;15(8):669-686. doi:10.1002/(SICI)1097-0061(19990615)15:8<669::AID-YEA411>3.0.CO;2-K
- Grech L, Jeffares DC, Sadée CY, et al. Fitness Landscape of the Fission Yeast Genome. *Mol Biol Evol.* 2019;36(8):1612-1623. doi:10.1093/molbev/msz113
- Gu Y, Yam C, Oliferenko S. Divergence of mitotic strategies in fission yeasts. *Nucleus.* 2012;3(3):220-225. doi:10.4161/nucl.19514
- Gu Y, Yam C, Oliferenko S. Rewiring of cellular division site selection in evolution of fission yeasts. *Curr Biol.* 2015;25(9):1187-1194. doi:10.1016/j.cub.2015.02.056
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies:

- assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307-321. doi:10.1093/sysbio/syq010
- Guo Y, Singh PK, Levin HL. A long terminal repeat retrotransposon of *Schizosaccharomyces japonicus* integrates upstream of RNA pol III transcribed genes. *Mob DNA.* 2015;6:19. Published 2015 Oct 9. doi:10.1186/s13100-015-0048-2
- Haerty W, Jagadeeshan S, Kulathinal RJ, et al. Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics.* 2007;177(3):1321-1335. doi:10.1534/genetics.107.078865
- Hammer Ø, Harper DAT, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 2001, 4, 9.
- Han MV, Zmasek CM. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics.* 2009;10:356. Published 2009 Oct 27. doi:10.1186/1471-2105-10-356
- Hane JK, Rouxel T, Howlett BJ, Kema GH, Goodwin SB, Oliver RP. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol.* 2011;12(5):R45. doi:10.1186/gb-2011-12-5-r45
- Harnqvist SE, Grace CA, Jeffares DC. Variables Influencing Differences in Sequence Conservation in the Fission Yeast *Schizosaccharomyces pombe*. *J Mol Evol.* 2021;89(9-10):601-610. doi:10.1007/s00239-021-10028-y
- Helston RM, Box JA, Tang W, Baumann P. *Schizosaccharomyces cryophilus* sp. nov., a new species of fission yeast. *FEMS Yeast Res.* 2010;10(6):779-786. doi:10.1111/j.1567-1364.2010.00657.x
- Hillier LW, Miller RD, Baird SE, et al. Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. *PLoS Biol.* 2007;5(7):e167. doi:10.1371/journal.pbio.0050167
- Hirano T, Funahashi S, Uemura T, Yanagida M. Isolation and characterization of *Schizosaccharomyces pombe* cutmutants that block nuclear division but not cytokinesis. *EMBO J.* 1986;5(11):2973-2979. doi:10.1002/j.1460-2075.1986.tb04594.x
- Ho SY, Chen AX, Lins LS, Duchêne DA, Lo N. The Genome as an Evolutionary Timepiece. *Genome Biol Evol.* 2016;8(9):3006-3010. Published 2016 Oct 5. doi:10.1093/gbe/evw220

- Hoffman CS, Wood V, Fantes PA. An Ancient Yeast for Young Geneticists: A Primer on the *Schizosaccharomyces pombe* Model System [published correction appears in Genetics. 2016 Mar;202(3):1241]. Genetics. 2015;201(2):403-423. doi:10.1534/genetics.115.181503
- Hurst LD, Williams EJ, Pál C. Natural selection promotes the conservation of linkage of co-expressed genes. Trends Genet. 2002;18(12):604-606. doi:10.1016/s0168-9525(02)02813-5
- Hurst LD. Evolutionary genomics and the reach of selection. J Biol. 2009;8(2):12. doi:10.1186/jbiol113
- Hurst LD, Pál C, Lercher MJ. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet. 2004;5(4):299-310. doi:10.1038/nrg1319
- International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution [published correction appears in Nature. 2005 Feb 17;433(7027):777]. Nature. 2004;432(7018):695-716. doi:10.1038/nature03154
- Jacobs JZ, Ciccaglione KM, Tournier V, Zaratiegui M. Implementation of the CRISPR-Cas9 system in fission yeast. Nat Commun. 2014;5:5344. Published 2014 Oct 29. doi:10.1038/ncomms6344
- Jeffares DC, Jolly C, Hoti M, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nat Commun. 2017;8:14061. Published 2017 Jan 24. doi:10.1038/ncomms14061
- Jeffares DC, Rallis C, Rieux A, et al. The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. Nat Genet. 2015;47(3):235-241. doi:10.1038/ng.3215
- Kalmykova AI, Nurminsky DI, Ryzhov DV, Shevelyov YY. Regulated chromatin domain comprising cluster of co-expressed genes in *Drosophila melanogaster*. Nucleic Acids Res. 2005;33(5):1435-1444. Published 2005 Mar 8. doi:10.1093/nar/gki281
- Kämper J, Kahmann R, Bölker M, et al. Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature. 2006;444(7115):97-101. doi:10.1038/nature05248
- Karanyicz E, Antunovics Z, Kallai Z, Sipiczki M. Non-introgressive genome chimerisation by malsegregation in autodiploidised allotetraploids during meiosis of *Saccharomyces kudriavzevii* x *Saccharomyces uvarum* hybrids.

- Appl Microbiol Biotechnol. 2017;101(11):4617-4633. doi:10.1007/s00253-017-8274-9
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-780. doi:10.1093/molbev/mst010
- Keeling PJ, Slamovits CH. Simplicity and complexity of microsporidian genomes. Eukaryot Cell. 2004;3(6):1363-1369. doi:10.1128/EC.3.6.1363-1369.2004
- Kimura, M., The Neutral Theory of Molecular Evolution, Cambridge: Cambridge Univ. Press, 1983.
- Kinnaer C, Dudin O, Martin SG. Yeast-to-hypha transition of *Schizosaccharomyces japonicus* in response to environmental stimuli. Mol Biol Cell. 2019;30(8):975-991. doi:10.1091/mbc.E18-12-0774
- Kitajima TS, Kawashima SA, Watanabe Y. The conserved kinetochore protein shugoshin protects centromeric cohesion during meiosis. Nature. 2004;427(6974):510-517. doi:10.1038/nature02312
- Klar AJ. *Schizosaccharomyces japonicus* yeast poised to become a favorite experimental organism for eukaryotic research. G3 (Bethesda). 2013;3(10):1869-1873. Published 2013 Oct 3. doi:10.1534/g3.113.007187
- Kleinjan DJ, van Heyningen V. Position effect in human genetic disease. Hum Mol Genet. 1998;7(10):1611-1618. doi:10.1093/hmg/7.10.1611
- Koch EN, Costanzo M, Bellay J, et al. Conserved rules govern genetic interaction degree across species. Genome Biol. 2012;13(7):R57. Published 2012 Jul 2. doi:10.1186/gb-2012-13-7-r57
- Kraft K, Magg A, Heinrich V, et al. Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat Cell Biol. 2019;21(3):305-310. doi:10.1038/s41556-019-0273-x
- Kruglyak S, Tang H. Regulation of adjacent yeast genes. Trends Genet. 2000;16(3):109-111. doi:10.1016/s0168-9525(99)01941-1
- Kudriawzew, W. Die Systematik der Hefen; Akademie: Berlin, Germany, 1960
- Kuhn RM, Clarke L, Carbon J. Clustered tRNA genes in *Schizosaccharomyces pombe* centromeric DNA sequence repeats. Proc Natl Acad Sci U S A. 1991;88(4):1306-1310. doi:10.1073/pnas.88.4.1306

- Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome [published correction appears in Nature 2001 Aug 2;412(6846):565] [published correction appears in Nature 2001 Jun 7;411(6838):720. Szustakowki, J [corrected to Szustakowski, JJ]. Nature. 2001;409(6822):860-921. doi:10.1038/35057062
- Lee JM, Sonnhammer EL. Genomic gene clustering analysis of pathways in eukaryotes [published correction appears in Genome Res. 2004 Dec;14(12):2510]. Genome Res. 2003;13(5):875-882. doi:10.1101/gr.737703
- Lee KM, Miklos I, Du H, et al. Impairment of the TFIIH-associated CDK-activating kinase selectively affects cell cycle-regulated gene expression in fission yeast. Mol Biol Cell. 2005;16(6):2734-2745. doi:10.1091/mbc.e04-11-0982
- Lee MG, Nurse P. Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. Nature. 1987;327(6117):31-35. doi:10.1038/327031a0
- Lefort V, Longueville JE, Gascuel O. SMS: Smart Model Selection in PhyML. Mol Biol Evol. 2017;34(9):2422-2424. doi:10.1093/molbev/msx149
- Legras JL, Galeote V, Bigey F, et al. Adaptation of *S. cerevisiae* to Fermented Food Environments Reveals Remarkable Genome Plasticity and the Footprints of Domestication. Mol Biol Evol. 2018;35(7):1712-1727. doi:10.1093/molbev/msy066
- Lercher MJ, Blumenthal T, Hurst LD. Coexpression of neighboring genes in *Caenorhabditis elegans* is mostly due to operons and duplicate genes. Genome Res. 2003;13(2):238-243. doi:10.1101/gr.553803
- Lercher MJ, Hurst LD. Co-expressed yeast genes cluster over a long range but are not regularly spaced. J Mol Biol. 2006;359(3):825-831. doi:10.1016/j.jmb.2006.03.051
- Lercher MJ, Urrutia AO, Hurst LD. Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet. 2002;31(2):180-183. doi:10.1038/ng887
- Li Y, Steenwyk JL, Chang Y, et al. A genome-scale phylogeny of the kingdom Fungi. Curr Biol. 2021;31(8):1653-1665.e5. doi:10.1016/j.cub.2021.01.074
- Liao X, Li M, Zou Y, et al. Current challenges and solutions of de novo assembly. Quant Biol. 2019;7:90–109. doi:10.1007/s40484-019-0166-9

- Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*. 2009;326(5950):289-293. doi:10.1126/science.1181369
- Linder T, Gustafsson CM. Molecular phylogenetics of ascomycotal adhesins--a novel family of putative cell-surface adhesive proteins in fission yeasts. *Fungal Genet Biol*. 2008;45(4):485-497. doi:10.1016/j.fgb.2007.08.002
- Lindner P. *Schizosaccharomyces pombe* n. sp., ein neuer Gahrungserreger. *Wochenschrift fur Brauerei* 1893;10:1298–300.
- Liti G, Carter DM, Moses AM, et al. Population genomics of domestic and wild yeasts. *Nature*. 2009;458(7236):337-341. doi:10.1038/nature07743
- Liu X, Han B. Evolutionary conservation of neighbouring gene pairs in plants. *Gene*. 2009;437(1-2):71-79. doi:10.1016/j.gene.2009.02.012
- Liu Y, Leigh JW, Brinkmann H, et al. Phylogenomic analyses support the monophyly of Taphrinomycotina, including *Schizosaccharomyces* fission yeasts. *Mol Biol Evol*. 2009;26(1):27-34. doi:10.1093/molbev/msn221
- Lock A, Rutherford K, Harris MA, et al. PomBase 2018: user-driven reimplementaion of the fission yeast database provides rapid and intuitive access to diverse, interconnected information. *Nucleic Acids Res*. 2019;47(D1):D821-D827. doi:10.1093/nar/gky961
- Lowry DB, Willis JH. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation [published correction appears in *PLoS Biol*. 2012 Jan;10(1):10.1371/annotation/caa1b7dd-9b6d-44db-b6ce-666954903625]. *PLoS Biol*. 2010;8(9):e1000500. Published 2010 Sep 28. doi:10.1371/journal.pbio.1000500
- Lundin LG. Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics*. 1993;16(1):1-19. doi:10.1006/geno.1993.1133
- Luo A, Ho SYW. The molecular clock and evolutionary timescales. *Biochem Soc Trans*. 2018;46(5):1183-1190. doi:10.1042/BST20180186
- Ma L, Chen Z, Huang da W, et al. Genome analysis of three *Pneumocystis* species reveals adaptation mechanisms to life exclusively in mammalian hosts. *Nat Commun*. 2016;7:10740. Published 2016 Feb 22. doi:10.1038/ncomms10740

- Ma LJ, Ibrahim AS, Skory C, et al. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet.* 2009;5(7):e1000549. doi:10.1371/journal.pgen.1000549
- Maguire SL, ÓhÉigeartaigh SS, Byrne KP, et al. Comparative genome analysis and gene finding in *Candida* species using CGOB. *Mol Biol Evol.* 2013;30(6):1281-1291. doi:10.1093/molbev/mst042
- Makarova M, Peter M, Balogh G, et al. Delineating the Rules for Structural Adaptation of Membrane-Associated Proteins to Evolutionary Changes in Membrane Lipidome. *Curr Biol.* 2020;30(3):367-380.e8. doi:10.1016/j.cub.2019.11.043
- Martín-Cuadrado AB, Dueñas E, Sipiczki M, Vázquez de Aldana CR, del Rey F. The endo-beta-1,3-glucanase eng1p is required for dissolution of the primary septum during cell separation in *Schizosaccharomyces pombe*. *J Cell Sci.* 2003;116(Pt 9):1689-1698. doi:10.1242/jcs.00377
- Mata J, Bahler J. Correlations between gene expression and gene conservation in fission yeast. *Genome Res.* 2003;13(12):2686-2690. doi:10.1101/gr.1420903
- Matsuda A, Asakawa H, Haraguchi T, Hiraoka Y. Spatial organization of the *Schizosaccharomyces pombe* genome within the nucleus. *Yeast.* 2017;34(2):55-66. doi:10.1002/yea.3217
- Mba IE, Nweze EI, Eze EA, Anyaegbunam ZKG. Genome plasticity in *Candida albicans*: A cutting-edge strategy for evolution, adaptation, and survival. *Infect Genet Evol.* 2022;99:105256. doi:10.1016/j.meegid.2022.105256
- McDonald MJ, Rice DP, Desai MM. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature.* 2016;531(7593):233-236. doi:10.1038/nature17143
- McGary KL, Slot JC, Rokas A. Physical linkage of metabolic genes in fungi is an adaptation against the accumulation of toxic intermediate compounds. *Proc Natl Acad Sci U S A.* 2013;110(28):11481-11486. doi:10.1073/pnas.1304461110
- Mégy K, Audic S, Claverie JM. Positional clustering of differentially expressed genes on human chromosomes 20, 21 and 22. *Genome Biol.* 2003;4(2):P1. doi:10.1186/gb-2003-4-2-p1

- Mezey JG, Nuzhdin SV, Ye F, Jones CD. Coordinated evolution of co-expressed gene clusters in the *Drosophila* transcriptome. *BMC Evol Biol.* 2008;8:2. Published 2008 Jan 7. doi:10.1186/1471-2148-8-2
- Michalak P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics.* 2008;91(3):243-248. doi:10.1016/j.ygeno.2007.11.002
- Michel AH, Hatakeyama R, Kimmig P, et al. Functional mapping of yeast genomes by saturated transposition. *Elife.* 2017;6:e23570. Published 2017 May 8. doi:10.7554/eLife.23570
- Miklos I, Szilagyi Z, Watt S, et al. Genomic expression patterns in cell separation mutants of *Schizosaccharomyces pombe* defective in the genes *sep10 (+)* and *sep15 (+)* coding for the Mediator subunits Med31 and Med8. *Mol Genet Genomics.* 2008;279(3):225-238. doi:10.1007/s00438-007-0296-z
- Millarini V, Ignesti S, Cappelli S, et al. Protection of Wine from Protein Haze Using *Schizosaccharomyces japonicus* Polysaccharides. *Foods.* 2020;9(10):1407. Published 2020 Oct 3. doi:10.3390/foods9101407
- Minnaar PP, Jolly NP, Paulsen V, Du Plessis HW, Van Der Rijst M. *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* yeasts in sequential fermentations: Effect on phenolic acids of fermented Kei-apple (*Dovyalis caffra* L.) juice. *Int J Food Microbiol.* 2017;257:232-237. doi:10.1016/j.ijfoodmicro.2017.07.004
- Mojardín L, Vázquez E, Antequera F. Specification of DNA replication origins and genomic base composition in fission yeasts. *J Mol Biol.* 2013;425(23):4706-4713. doi:10.1016/j.jmb.2013.09.023
- Mondo SJ, Dannebaum RO, Kuo RC, et al. Widespread adenine N6-methylation of active genes in fungi. *Nat Genet.* 2017;49(6):964-968. doi:10.1038/ng.3859
- Murray JM, Watson AT, Carr AM. Molecular Genetic Tools and Techniques in Fission Yeast. *Cold Spring Harb Protoc.* 2016;2016(5):10.1101/pdb.top087601. Published 2016 May 2. doi:10.1101/pdb.top087601
- Nagy LG, Ohm RA, Kovács GM, et al. Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. *Nat Commun.* 2014;5:4471. Published 2014 Jul 18. doi:10.1038/ncomms5471

- Naseeb S, Carter Z, Minnis D, Donaldson I, Zeef L, Delneri D. Widespread Impact of Chromosomal Inversions on Gene Expression Uncovers Robustness via Phenotypic Buffering. *Mol Biol Evol.* 2016;33(7):1679-1696. doi:10.1093/molbev/msw045
- Naumov GI, Kondratieva VI, Naumova ES. Hybrid Sterility of the Yeast *Schizosaccharomyces pombe*: Genetic Genus and Many Species in statu nascendi? *Mikrobiologija.* 2015;84(2):192-203.
- Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol.* 1970;48(3):443-453. doi:10.1016/0022-2836(70)90057-4
- Nei M, Suzuki Y, Nozawa M. The neutral theory of molecular evolution in the genomic era. *Annu Rev Genomics Hum Genet.* 2010;11:265-289. doi:10.1146/annurev-genom-082908-150129
- Ng MP, Vergara IA, Frech C, et al. OrthoClusterDB: an online platform for synteny blocks. *BMC Bioinformatics.* 2009;10:192. Published 2009 Jun 23. doi:10.1186/1471-2105-10-192
- Nguyen HV, Gaillardin C. Two Subgroups within the *Saccharomyces bayanus* Species Evidenced by PCR Amplification and Restriction Polymorphism of the Non-Transcribed Spacer 2 in the Ribosomal DNA Unit. *Systematic and Applied Microbiology.* 1997;20(2):286-294. [https://doi.org/10.1016/S0723-2020\(97\)80075-6](https://doi.org/10.1016/S0723-2020(97)80075-6)
- Nguyen TA, Cissé OH, Yun Wong J, et al. Innovation and constraint leading to complex multicellularity in the Ascomycota. *Nat Commun.* 2017;8:14444. Published 2017 Feb 8. doi:10.1038/ncomms14444
- Noble LM, Andrianopoulos A. Fungal genes in context: genome architecture reflects regulatory complexity and function. *Genome Biol Evol.* 2013;5(7):1336-1352. doi:10.1093/gbe/evt077
- Noé L, Kucherov G. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Res.* 2005;33(Web Server issue):W540-W543. doi:10.1093/nar/gki478
- Nozaki S, Furuya K, Niki H. The Ras1-Cdc42 pathway is involved in hyphal development of *Schizosaccharomyces japonicus*. *FEMS Yeast Res.* 2018;18(4):10.1093/femsyr/foy031. doi:10.1093/femsyr/foy031
- O'Donnell K. *Fusarium and Its Near Relatives*. In: Reynolds, D.R. and Taylor, J.W., Eds., *The Fungal Holomorph: Mitotic, Meiotic and Pleomorphic*

- Speciation in Fungal Systematic, CAB International, Wallingford, 1993: 225-233.
- Pál C, Hurst LD. Evidence for co-evolution of gene order and recombination rate. *Nat Genet.* 2003;33(3):392-395. doi:10.1038/ng1111
- Palazzo AF, Kejiou NS. Non-Darwinian Molecular Biology. *Front Genet.* 2022;13:831068. Published 2022 Feb 16. doi:10.3389/fgene.2022.831068
- Papp B, Pál C, Hurst LD. Dosage sensitivity and the evolution of gene families in yeast. *Nature.* 2003;424(6945):194-197. doi:10.1038/nature01771
- Papp LA, Ács-Szabó L, Póliska S, Miklós I. A modified culture medium and hyphae isolation method can increase quality of the RNA extracted from mycelia of a dimorphic fungal species. *Curr Genet.* 2021;67(5):823-830. doi:10.1007/s00294-021-01181-4
- Papp LA, Ács-Szabó L, Batta G, Miklós I. Molecular and comparative genomic analyses reveal evolutionarily conserved and unique features of the *Schizosaccharomyces japonicus* mycelial growth and the underlying genomic changes. *Curr Genet.* 2021;67(6):953-968. doi:10.1007/s00294-021-01206-y
- Papp L, Sipiczki M, Holb IJ, Miklós I. Optimal conditions for mycelial growth of *Schizosaccharomyces japonicus* cells in liquid medium: it enables the molecular investigation of dimorphism. *Yeast.* 2014;31(12):475-482. doi:10.1002/yea.3048
- Pasero P, Marilley M. Size variation of rDNA clusters in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. *Mol Gen Genet.* 1993;236(2-3):448-452. doi:10.1007/BF00277147
- Pevzner P, Tesler G. Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. *Genome Res.* 2003;13(1):37-45. doi:10.1101/gr.757503
- Pickeral OK, Li JZ, Barrow I, Boguski MS, Makałowski W, Zhang J. Classical oncogenes and tumor suppressor genes: a comparative genomics perspective. *Neoplasia.* 2000;2(3):280-286. doi:10.1038/sj.neo.7900090
- Poyatos JF, Hurst LD. Is optimal gene order impossible?. *Trends Genet.* 2006;22(8):420-423. doi:10.1016/j.tig.2006.06.003
- Poyatos JF, Hurst LD. The determinants of gene order conservation in yeasts. *Genome Biol.* 2007;8(11):R233. doi:10.1186/gb-2007-8-11-r233

- Princová J, Schätz M, Ťupa O, Převorovský M. Analysis of Lipid Droplet Content in Fission and Budding Yeasts using Automated Image Processing. *J Vis Exp.* 2019;(149):10.3791/59889. Published 2019 Jul 17. doi:10.3791/59889
- Rajeh A, Lv J, Lin Z. Heterogeneous rates of genome rearrangement contributed to the disparity of species richness in Ascomycota. *BMC Genomics.* 2018;19(1):282. Published 2018 Apr 24. doi:10.1186/s12864-018-4683-0
- Ralser M, Kuhl H, Ralser M, et al. The *Saccharomyces cerevisiae* W303-K6001 cross-platform genome sequence: insights into ancestry and physiology of a laboratory mutt. *Open Biol.* 2012;2(8):120093. doi:10.1098/rsob.120093
- Ranz JM, Maurin D, Chan YS, et al. Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biol.* 2007;5(6):e152. doi:10.1371/journal.pbio.0050152
- Revanna KV, Munro D, Gao A, Chiu CC, Pathak A, Dong Q. A web-based multi-genome synteny viewer for customized data. *BMC Bioinformatics.* 2012;13:190. Published 2012 Aug 2. doi:10.1186/1471-2105-13-190
- Rhind N, Chen Z, Yassour M, et al. Comparative functional genomics of the fission yeasts. *Science.* 2011;332(6032):930-936. doi:10.1126/science.1203357
- Ribár B, Bánrévi A, Sipiczki M. sep1+ encodes a transcription-factor homologue of the HNF-3/forkhead DNA-binding-domain family in *Schizosaccharomyces pombe*. *Gene.* 1997;202(1-2):1-5. doi:10.1016/s0378-1119(97)00390-9
- Rice AM, McLysaght A. Dosage-sensitive genes in evolution and disease. *BMC Biol.* 2017;15(1):78. Published 2017 Sep 1. doi:10.1186/s12915-017-0418-y
- Riley R, Haridas S, Wolfe KH, et al. Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci U S A.* 2016;113(35):9882-9887. doi:10.1073/pnas.1603941113
- Rodgers K, McVey M. Error-Prone Repair of DNA Double-Strand Breaks. *J Cell Physiol.* 2016;231(1):15-24. doi:10.1002/jcp.25053
- Rodríguez-Carballo E, Lopez-Delisle L, Zhan Y, et al. The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev.* 2017;31(22):2264-2281. doi:10.1101/gad.307769.117

- Rodríguez-López M, Cotobal C, Fernández-Sánchez O, et al. A CRISPR/Cas9-based method and primer design tool for seamless genome editing in fission yeast. *Wellcome Open Res.* 2017;1:19. Published 2017 May 5. doi:10.12688/wellcomeopenres.10038.3
- Rolland T, Dujon B. Yeasty clocks: dating genomic changes in yeasts [published correction appears in *C R Biol.* 2011 Dec;334(12):916]. *C R Biol.* 2011;334(8-9):620-628. doi:10.1016/j.crvi.2011.05.010
- Ruddle FH, Bentley KL, Murtha MT, Risch N. Gene loss and gain in the evolution of the vertebrates. *Dev Suppl.* 1994;155-161.
- Sacerdot C, Casaregola S, Lafontaine I, Tekaiia F, Dujon B, Ozier-Kalogeropoulos O. Promiscuous DNA in the nuclear genomes of hemiascomycetous yeasts. *FEMS Yeast Res.* 2008;8(6):846-857. doi:10.1111/j.1567-1364.2008.00409.x
- Salzberg, S. L. & Yorke, J. A. Beware of mis-assembled genomes. *Bioinformatics* 21, 4320–4321 (2005).
- Samejima I, Matsumoto T, Nakaseko Y, Beach D, Yanagida M. Identification of seven new cut genes involved in *Schizosaccharomyces pombe* mitosis. *J Cell Sci.* 1993;105 ( Pt 1):135-143. doi:10.1242/jcs.105.1.135
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977;74(12):5463-5467. doi:10.1073/pnas.74.12.5463
- Sasaki M, Lange J, Keeney S. Genome destabilization by homologous recombination in the germ line. *Nat Rev Mol Cell Biol.* 2010;11(3):182-195. doi:10.1038/nrm2849
- Scannell DR, Zill OA, Rokas A, et al. The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the *Saccharomyces sensu stricto* Genus. *G3 (Bethesda).* 2011;1(1):11-25. doi:10.1534/g3.111.000273
- Scheuerl T, Stelzer CP. Sex initiates adaptive evolution by recombination between beneficial loci. *PLoS One.* 2017;12(6):e0177895. Published 2017 Jun 2. doi:10.1371/journal.pone.0177895
- Sémon M, Duret L. Evolutionary origin and maintenance of coexpressed gene clusters in mammals. *Mol Biol Evol.* 2006;23(9):1715-1723. doi:10.1093/molbev/msl034

- Seoighe C, Federspiel N, Jones T, et al. Prevalence of small inversions in yeast gene order evolution. *Proc Natl Acad Sci U S A*. 2000;97(26):14433-14437. doi:10.1073/pnas.240462997
- Sharakhov IV, Serazin AC, Grushko OG, et al. Inversions and gene order shuffling in *Anopheles gambiae* and *A. funestus*. *Science*. 2002;298(5591):182-185. doi:10.1126/science.1076803
- Shen XX, Steenwyk JL, LaBella AL, et al. Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Sci Adv*. 2020;6(45):eabd0079. Published 2020 Nov 4. doi:10.1126/sciadv.abd0079
- Shi-Kunne X, Faino L, van den Berg GCM, Thomma BPHJ, Seidl MF. Evolution within the fungal genus *Verticillium* is characterized by chromosomal rearrangement and gene loss. *Environ Microbiol*. 2018;20(4):1362-1373. doi:10.1111/1462-2920.14037
- Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011;7:539. Published 2011 Oct 11. doi:10.1038/msb.2011.75
- Sims GE, Jun SR, Wu GA, Kim SH. Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions. *Proc Natl Acad Sci U S A*. 2009;106(8):2677-2682. doi:10.1073/pnas.0813249106
- Singer GA, Lloyd AT, Huminiecki LB, Wolfe KH. Clusters of co-expressed genes in mammalian genomes are conserved by natural selection [published correction appears in *Mol Biol Evol*. 2005 Apr;22(4):1160]. *Mol Biol Evol*. 2005;22(3):767-775. doi:10.1093/molbev/msi062
- Sipiczki M, Bozsik A. The use of morphomutants to investigate septum formation and cell separation in *Schizosaccharomyces pombe*. *Arch Microbiol*. 2000;174(6):386-392. doi:10.1007/s002030000214
- Sipiczki M, Grallert B, Miklos I. Mycelial and syncytial growth in *Schizosaccharomyces pombe* induced by novel septation mutations. *J Cell Sci*. 1993;104 ( Pt 2):485-493. doi:10.1242/jcs.104.2.485
- Sipiczki M, Takeo K, Grallert A. Growth polarity transitions in a dimorphic fission yeast. *Microbiology (Reading)*. 1998;144 ( Pt 12):3475-3485. doi:10.1099/00221287-144-12-3475

- Sipiczki M, Takeo K, Yamaguchi M, Yoshida S, Miklos I. Environmentally controlled dimorphic cycle in a fission yeast. *Microbiology (Reading)*. 1998;144 ( Pt 5):1319-1330. doi:10.1099/00221287-144-5-1319
- Sipiczki M. Splitting of the fission yeast septum. *FEMS Yeast Res*. 2007;7(6):761-770. doi:10.1111/j.1567-1364.2007.00266.x
- Sipiczki M. Where does fission yeast sit on the tree of life?. *Genome Biol*. 2000;1(2):REVIEWS1011. doi:10.1186/gb-2000-1-2-reviews1011
- Smith SF, Snell P, Gruetzner F, et al. Analyses of the extent of shared synteny and conserved gene orders between the genome of *Fugu rubripes* and human 20q. *Genome Res*. 2002;12(5):776-784. doi:10.1101/gr.221802
- Spellman PT, Rubin GM. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J Biol*. 2002;1(1):5. doi:10.1186/1475-4924-1-5
- Stechmann A. Genome evolution: the dynamics of static genomes. *Curr Biol*. 2004;14(12):R473-R474. doi:10.1016/j.cub.2004.06.012
- Sturtevant AH. The Effects of Unequal Crossing over at the Bar Locus in *Drosophila*. *Genetics*. 1925;10(2):117-147. doi:10.1093/genetics/10.2.117
- Su CH, Shih CH, Chang TH, Tsai HK. Genome-wide analysis of the cis-regulatory modules of divergent gene pairs in yeast. *Genomics*. 2010;96(6):352-361. doi:10.1016/j.ygeno.2010.08.008
- Sugino RP, Innan H. Natural selection on gene order in the genome reorganization process after whole-genome duplication of yeast. *Mol Biol Evol*. 2012;29(1):71-79. doi:10.1093/molbev/msr118
- Szilagyi Z, Batta G, Enczi K, Sipiczki M. Characterisation of two novel fork-head gene homologues of *Schizosaccharomyces pombe*: their involvement in cell cycle and sexual differentiation. *Gene*. 2005;348:101-109. doi:10.1016/j.gene.2004.12.043
- Szilagyi Z, Grallert A, Nemeth N, Sipiczki M. The *Schizosaccharomyces pombe* genes *sep10* and *sep11* encode putative general transcriptional regulators involved in multiple cellular processes. *Mol Genet Genomics*. 2002;268(4):553-562. doi:10.1007/s00438-002-0773-3
- Takouridis SJ, Tribe DE, Gras SL, Martin GJO. The selective breeding of the freshwater microalga *Chlamydomonas reinhardtii* for growth in salinity. *Bioresour Technol*. 2015;184:18-22. doi:10.1016/j.biortech.2014.10.120

- Tanaka K, Hirata A. Ascospore development in the fission yeasts *Schizosaccharomyces pombe* and *S. japonicus*. *J Cell Sci*. 1982;56:263-279. doi:10.1242/jcs.56.1.263
- Tanizawa H, Iwasaki O, Tanaka A, et al. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res*. 2010;38(22):8164-8177. doi:10.1093/nar/gkq955
- Tatusov RL, Mushegian AR, Bork P, et al. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr Biol*. 1996;6(3):279-291. doi:10.1016/s0960-9822(02)00478-5
- Teichmann SA, Veitia RA. Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics*. 2004;167(4):2121-2125. doi:10.1534/genetics.103.024505
- Tesler G. GRIMM: genome rearrangements web server. *Bioinformatics*. 2002;18(3):492-493. doi:10.1093/bioinformatics/18.3.492
- Todd RT, Wikoff TD, Forche A, Selmecki A. Genome plasticity in *Candida albicans* is driven by long repeat sequences. *Elife*. 2019;8:e45954. Published 2019 Jun 7. doi:10.7554/eLife.45954
- Tong P, Pidoux AL, Toda NRT, et al. Interspecies conservation of organisation and function between nonhomologous regional centromeres. *Nat Commun*. 2019;10(1):2343. Published 2019 May 28. doi:10.1038/s41467-019-09824-4
- Tsai HK, Su CP, Lu MY, Shih CH, Wang D. Co-expression of adjacent genes in yeast cannot be simply attributed to shared regulatory system. *BMC Genomics*. 2007;8:352. Published 2007 Oct 3. doi:10.1186/1471-2164-8-352
- Tsujino K, Okuzaki Y, Hibino N, et al. Identification of transgene integration site and anatomical properties of fluorescence intensity in a EGFP transgenic chicken line. *Dev Growth Differ*. 2019;61(7-8):393-401. doi:10.1111/dgd.12631
- Tuller T, Rubinstein U, Bar D, Gurevitch M, Ruppin E, Kupiec M. Higher-order genomic organization of cellular functions in yeast. *J Comput Biol*. 2009;16(2):303-316. doi:10.1089/cmb.2008.15TT
- Tusso S, Nieuwenhuis BPS, Sedlazeck FJ, Davey JW, Jeffares DC, Wolf JBW. Ancestral Admixture Is the Main Determinant of Global Biodiversity in

- Fission Yeast. *Mol Biol Evol.* 2019;36(9):1975-1989. doi:10.1093/molbev/msz126
- Vakirlis N, Sarilar V, Drillon G, et al. Reconstruction of ancestral chromosome architecture and gene repertoire reveals principles of genome evolution in a model yeast genus. *Genome Res.* 2016;26(7):918-932. doi:10.1101/gr.204420.116
- Veitia RA. Nonlinear effects in macromolecular assembly and dosage sensitivity. *J Theor Biol.* 2003;220(1):19-25. doi:10.1006/jtbi.2003.3105
- Veltri D, Wight MM, Crouch JA. SimpleSynteny: a web-based tool for visualization of microsynteny across multiple species. *Nucleic Acids Res.* 2016;44(W1):W41-W45. doi:10.1093/nar/gkw330
- Vyas A, Freitas AV, Ralston ZA, Tang Z. Fission Yeast *Schizosaccharomyces pombe*: A Unicellular "Micromammal" Model Organism [published correction appears in *Curr Protoc.* 2021 Jul;1(7):e225]. *Curr Protoc.* 2021;1(6):e151. doi:10.1002/cpz1.151
- Waalwijk, C., Vanheule, A., Audenaert, K. et al. *Fusarium* in the age of genomics. *Trop. plant pathol.* 42, 184–189 (2017). <https://doi.org/10.1007/s40858-017-0128-6>
- Wapinski I, Pfeffer A, Friedman N, Regev A. Automatic genome-wide reconstruction of phylogenetic gene trees. *Bioinformatics.* 2007;23(13):i549-i558. doi:10.1093/bioinformatics/btm193
- Watanabe Y, Nurse P. Cohesin Rec8 is required for reductional chromosome segregation at meiosis. *Nature.* 1999;400(6743):461-464. doi:10.1038/22774
- Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature.* 1953;171(4356):737-738. doi:10.1038/171737a0
- Weiler KS, Wakimoto BT. Heterochromatin and gene expression in *Drosophila*. *Annu Rev Genet.* 1995;29:577-605. doi:10.1146/annurev.ge.29.120195.003045
- Williams AH, Sharma M, Thatcher LF, et al. Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum* formae speciales facilitates identification of candidate effectors. *BMC Genomics.* 2016;17:191. Published 2016 Mar 5. doi:10.1186/s12864-016-2486-8

- Williams EJ, Bowles DJ. Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res.* 2004;14(6):1060-1067. doi:10.1101/gr.2131104
- Wolfe KH. Origin of the Yeast Whole-Genome Duplication. *PLoS Biol.* 2015;13(8):e1002221. Published 2015 Aug 7. doi:10.1371/journal.pbio.1002221
- Wong S, Wolfe KH. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. *Nat Genet.* 2005;37(7):777-782. doi:10.1038/ng1584
- Wood V, Gwilliam R, Rajandream MA, et al. The genome sequence of *Schizosaccharomyces pombe* [published correction appears in *Nature* 2003 Jan 2;421(6918):94. Cerrutti L [corrected to Cerutti L]]. *Nature.* 2002;415(6874):871-880. doi:10.1038/nature724
- Xu J, Yanagisawa Y, Tsankov AM, et al. Genome-wide identification and characterization of replication origins by deep sequencing. *Genome Biol.* 2012;13(4):R27. Published 2012 Apr 24. doi:10.1186/gb-2012-13-4-r27
- Yamada Y, Banno I. *Hasegawaea* gen. nov., an ascosporegenous yeast genus for the organisms whose asexual reproduction is by fission and whose ascospores have smooth surfaces without papillae and which are characterized by the absence of coenzyme Q mid by the presence of linoleic acid in cellular fatty acid composition. *J. Gen. Appl. Microbiol.* 1987, 33, 295–298.
- Yukawa M, Maki T. *Schizosaccharomyces japonicus* nov. spec. Fukuoka, Japan: La Bul Sci Falkultato Terkultura Kjusu Imp Univ, 1931;4:218–26.
- Zakharov IA. Location of genes in chromosomes: Random or not?. *Russ J Genet* 2010;46: 1165–1172. doi:10.1134/S1022795410100066
- Zanders SE, Eickbush MT, Yu JS, et al. Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. *Elife.* 2014;3:e02630. Published 2014 Jun 24. doi:10.7554/eLife.02630
- Zhang XR, He JB, Wang YZ, Du LL. A Cloning-Free Method for CRISPR/Cas9-Mediated Genome Editing in Fission Yeast. *G3 (Bethesda).* 2018;8(6):2067-2077. Published 2018 May 31. doi:10.1534/g3.118.200164
- Zhu T, Niu DK. Mechanisms of intron loss and gain in the fission yeast *Schizosaccharomyces*. *PLoS One.* 2013;8(4):e61683. Published 2013 Apr 17. doi:10.1371/journal.pone.0061683

Zilahi E, Miklós I, Sipiczki M. The *Schizosaccharomyces pombe* sep15+ gene encodes a protein homologous to the Med8 subunit of the *Saccharomyces cerevisiae* transcriptional mediator complex. *Curr Genet.* 2000;38(5):227-232. doi:10.1007/s002940000158

## 11. Függelék I.

Mivel jelen doktori disszertáció eredményeinek túlnyomó többsége számítógépes módszereken alapszik, a kiegészítő adatok (ábrák, táblázatok) terjedelmük miatt csupán online elérhető linkeken keresztül kerülnek bemutatásra. Ezek az adatok az értekezés alapját képező publikációk „Supplementary information” fájljaiként érhetők el.

Függelék 1. ábra: Supplementary Figure S1 c,d. A *S. cryophilus* szuperkontigjainak újra rendezése a Mauve program segítségével a szinténikus kapcsolatok alapján.

Függelék 2. ábra: Supplementary Figure S2. A szubtelomerikus régiók szinténikus kapcsolatai az *S. cryophilus* és *S. octosporus* esetében.

Függelék 3. ábra: Supplementary Figure S4 c-f. A sarjadzó élesztő genomok YASS és Mauve illesztései.

A Függelék 1., 2. és 3. ábrái a következő linken keresztül érhetők el:

[https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-018-32525-9/MediaObjects/41598\\_2018\\_32525\\_MOESM5\\_ESM.pdf](https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-018-32525-9/MediaObjects/41598_2018_32525_MOESM5_ESM.pdf)

Függelék 4. ábra: Figure S1 A-C. Páronkénti Mauve illesztések a *S. japonicus*-t használva referencia genomként.

Függelék 5. ábra: Figure S4 A,B. Az  $E \leq 1 \times 10^{-30}$  YASS illesztések statisztikai értékelése.

Függelék 6. ábra: Figure S3 A-D. A páronkénti YASS illesztések vizualizációja.

A Függelék 4., 5. és 6. ábrái a következő linken keresztül érhetők el:

<https://www.mdpi.com/2309-608X/7/10/864/s1?version=1634281401>

Függelék 1. táblázat: Supplementary Table S1. Közös pericentromerikus gének a *S. pombe*, *S. octosporus* és *S. cryophilus* genomjaiban.

Függelék 2. táblázat: Supplementary Table S2. Közös szubtelomerikus gének a *S. octosporus* és *S. cryophilus* genomjaiban.

A Függelék 1. és 2. táblázatai az alábbi linken keresztül érhetők el:

[https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-018-32525-9/MediaObjects/41598\\_2018\\_32525\\_MOESM4\\_ESM.xlsx](https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-018-32525-9/MediaObjects/41598_2018_32525_MOESM4_ESM.xlsx)

Függelék 3-8. táblázatok: Supplementary table S2-S7. Páronkénti YASS illesztések eredményei.

Függelék 9. táblázat: Supplementary table S9. A hasadó élesztők feltételezett ortológ génjei.

Függelék 10. táblázat: Supplementary table S10. A hasadó élesztő aLCB-k gyűjteménye.

Függelék 11. táblázat: Supplementary table S11. Feltételezett ortológok néhány random módon kiválasztott aLCB esetében.

Függelék 12. táblázat: Supplementary table S12. A filogenetikai fa építéséhez felhasznált 26 feltételezett ortológ protein listája.

Függelék 13. táblázat: Supplementary table S14. A különböző GO kategóriákba tartozó gének megoszlása az aLCB-ken belül és azokon kívül.

Függelék 14. táblázat: Supplementary table S15. A szignifikáns GO kategóriák meghatározása.

Függelék 15. táblázat: Supplementary table S16. A Rec12 hasítóhelyek lokalizációja a különböző GO kategóriákba tartozó gének mellett.

Függelék 16. táblázat: Supplementary table S17. Az esszenciális gének lokalizációja a különböző GO kategóriákba tartozó gének mellett.

A Függelék 3-18. táblázatai az alábbi linken keresztül érhetők el:

<https://www.mdpi.com/2309-608X/7/10/864/s1?version=1634281401>

## 12. Függelék II.

A disszertáció alapjául szolgáló, referált nemzetközi folyóiratban megjelent cikkek:

1. Ács-Szabó L, Papp LA, Antunovics Z, Sipiczki M, Miklós I. Assembly of *Schizosaccharomyces cryophilus* chromosomes and their comparative genomic analyses revealed principles of genome evolution of the haploid fission yeasts. *Scientific Reports* 8, 14629, 2018. **IF: 4,011 D1**
2. Acs-Szabo L, Papp LA, Sipiczki M, Miklos I. Genome comparisons of the fission yeasts reveal ancient collinear loci maintained by natural selection. *Journal of Fungi* 7(10), 864, 2021. **IF: 5,724 Q1**

Az értekezés témájában elhangzott konferencia előadások:

1. Ács-Szabó Lajos, Papp László Attila, Csoma Hajnalka, Sipiczki Mátyás, Miklós Ida. A génsorrend konzerváltságát meghatározó tényezők vizsgálata a hasadó élesztőkben. Genetikai Műhelyek Magyarországon XV. Jubileumi Minikonferencia, Szeptember 16, 2016, Szeged.
2. Ács-Szabó Lajos, Papp László Attila, Antunovics Zsuzsa, Karanyicz Edina, Hadházi Zsuzsanna, Sipiczki Mátyás & Miklós Ida. A *Schizosaccharomyces cryophilus* genomjának összeillesztése konzervatív tulajdonságok segítségével. Assembly of the genome of *Schizosaccharomyces cryophilus* using conservative genomic features. Magyar Mikrobiológiai Társaság 2016 évi Nagygyűlése, Keszthely, 2016.

Hazai és nemzetközi konferencián bemutatott poszterek az értekezés témakörében:

1. Lajos Ács-Szabó, László Attila Papp, Matthias Sipiczki, Ida Miklós. Conserved synteny and gene order determination in *Schizosaccharomyces*. Hungarian Molecular Life Sciences, Eger, 2015.
2. Lajos Ács-Szabó, Zsuzsanna Hadházi, Matthias Sipiczki, Ida Miklós. How could meiotic DNA breaks define the fate of gene order in related species? Magyar Mikrobiológiai Társaság 2015 évi Nagygyűlése, Budapest, 2015.

Egyéb publikációk:

Szakkikkek:

1. Balazs A, Batta G, Miklos I, Acs-Szabo L, Vazquez de Aldana CR, Sipiczki M. Conserved regulators of the cell separation process in *Schizosaccharomyces*. *Fungal Genet Biol* 49:235-249, 2012. **IF: 3,263 Q1**
2. Csoma H, Ács-Szabó L, Papp L, Sipiczki M. Application of different markers and data-analysis tools to examination of biodiversity can lead to different results: a case study with *Starmerella bacillaris* (synonym *Candida zemplinina*) strains. *FEMS Yeast Res.*18, 5, 2018. **IF: 2,458 Q1**
3. Papp LA, Ács-Szabó L, Póliska Sz, Miklós I. A modified culture medium and hyphae isolation method can increase quality of the RNA extracted from mycelia of a dimorphic fungal species. *Current Genetics* 67(5):823–830, 2021. **IF: 2,695 Q1**
4. Papp LA, Ács-Szabó L, Batta Gy, Miklós I. Molecular and comparative genomic analyses reveal evolutionarily conserved and unique features of the *Schizosaccharomyces japonicus* mycelial growth and the underlying genomic changes. *Current Genetics* 67(6):953–968, 2021. **IF: 2,695 Q1**
5. Plaszkó T, Szűcs Z, Cziáky Z, Ács-Szabó L, Csoma H, Géczi L, Vasas G, Gonda S. Correlations between the metabolome and the endophytic fungal community suggests importance of various metabolite classes in community assembly in horseradish (*Armoracia rusticana* L., Brassicaceae) roots. *Frontiers in Plant Science* 13, 921008, 2022. **IF: 6,627 D1**
6. Acs-Szabo L, Papp LA, Csoma H, Miklos I, Sipiczki M. New strain of *Cyphellophora olivacea* exhibits striking tolerance to sodium bicarbonate. *Diversity* 14, 1023, 2022. **IF: 3,029 Q1**

Konferencia előadások:

1. Ács-Szabó L., Miklós I., Sipiczki M. Isolation of *Schizosaccharomyces japonicus* mutants, which are defective in morphological transition. 30<sup>th</sup> ISSY Cell Surface & Organelles in Yeasts, Stará Lesná, Slovakia, 2013.

2. Sipiczki Matthias, Antunovics Zsuzsa, Csoma Hajnalka, Kállai Zoltán, Horváth Enikő, Szabó Adrienn, Ács-Szabó Lajos, Miklós Ida, Batta Gyula, Papp László Attila. Három évtized a borászati élesztők kutatásában. Magyar Mikrobiológiai Társaság 2018 évi Nagygyűlése, Eger, 2018.
3. Ács-Szabó Lajos, Papp László Attila, Csoma Hajnalka, Sipiczki Mátyás, Miklós Ida. Különböző gomba nemzetségek komparatív genomikai vizsgálata vonal-specifikus átrendeződési dinamikát sejtet. Biotechnológia a Debreceni Egyetemen Tudományos Szimpózium, Debrecen, 2019.
4. Papp László Attila, Ács-Szabó Lajos, Takács Szonja, Miklós Ida. Fonális növekedés genetikai vizsgálata *Schizosaccharomyces japonicus*-ban. Genetikai Műhelyek Magyarországon Minikonferencia, Szeged, 2020.

#### Konferencia poszterek:

1. Anita Balázs, Anett Madar, Anikó Mónus, Gyula Batta, Lajos Ács-Szabó, Ida Miklós, Carlos R. Vázquez de Aldana, Matthias Sipiczki. Investigation of the genes playing roles in the cell separation of *Schizosaccharomyces japonicus*. 38th Annual Conference on Yeasts, Szlovákia, 2010.
2. Lajos Ács-Szabó, Ida Miklós, Mátyás Sipiczki. Isolation of *Schizosaccharomyces japonicus* mutants, which are defective in morphological transition. 30<sup>th</sup> ISSY, Szlovákia, 2013.
3. Hajnalka Csoma, Lajos Ács-Szabó, Mátyás Sipiczki. THE MITOCHONDRIAL DNA RFLP ANALYSIS OF CANDIDA ZEMPLININA YEAST STRAINS. Magyar Mikrobiológiai Társaság 2015 évi Nagygyűlése, Budapest, 2015.
4. Lajos Ács-Szabó, Bianka Balogh, Matthias Sipiczki, Ida Miklós. Investigation of the function and structure of Ace2p - a conserved regulator of cell separation. Magyar Mikrobiológiai Társaság 2015 évi Nagygyűlése, Budapest, 2015.
5. Víg Ildikó, Ács-Szabó Lajos, Sipiczki Mátyás, Miklós Ida. *S. pombe* ace2 transzkripciósfaktort kódoló gén csonkolt mutánsainak elkészítése.

Genetikai Műhelyek Magyarországon XV. Jubileumi Minikonferencia, Szeged, Szeptember 16 2016.

6. László Attila Papp, Lajos Ács-Szabó, Ida Miklós. Non-homologous end joining is the dominant repair pathway in the dimorphic fission yeast *Schizosaccharomyces japonicus*. Magyar Mikrobiológiai Társaság 2016 évi Nagygyűlése, Keszthely, 2016.
7. Hajnalka Csoma, László Attila Papp, Lajos Ács-Szabó and Mátyás Sipiczki. EXAMINATION OF THE GENETIC DIVERSITY OF CANDIDA ZEMPLININA WINE YEAST. Magyar Mikrobiológiai Társaság 2016 évi Nagygyűlése, Keszthely, 2016.
8. Hajnalka Csoma, Lívia Dályai, Lajos Ács-Szabó, Mátyás Sipiczki and Ida Miklós. SEARCHING FOR INTRASPECIFIC POLYMORPHISM OF ZYGOSACCHAROMYCES LENTUS WINE SPOLAGE YEAST STRAINS. Magyar Mikrobiológiai Társaság 2016 évi Nagygyűlése, Keszthely, 2016.
9. Adrienn Szabó, Ildikó Vig, Lajos Ács-Szabó, Gyula Batta, Matthias Sipiczki, Ida Miklós. Identification of NES and NLS sequences in the *Schizosaccharomyces pombe*'s Ace2 protein. Hungarian Molecular Life Sciences, Eger, 2017.
10. Ildikó Vig, Adrienn Szabó, Lajos Ács-Szabó, László Papp, Matthias Sipiczki, Ida Miklós. Conserved region between amino acids 105 and 123 of the Ace2 transcription factor is important for proper function. Hungarian Molecular Life Sciences, Eger, 2017.
11. Ildikó Vig, Lajos Ács-Szabó, László Attila Papp, Matthias Sipiczki & Ida Miklós. Study of the human *BUD31* gene and its counterparts. Magyar Mikrobiológiai Társaság 2017 évi Nagygyűlése, Keszthely, 2017.
12. Lajos Ács-Szabó, László Attila Papp, Matthias Sipiczki & Ida Miklós. Comparison of sequential- and structural evolution reveals principles of genome evolution of the fission yeasts. Magyar Mikrobiológiai Társaság 2017 évi Nagygyűlése, Keszthely, 2017.
13. Hajnalka Csoma, Hajnalka Tóth, Lajos Ács-Szabó, Mátyás Sipiczki. GENETIC DIVERSITY OF THE *ZYGOSACCHAROMYCES LENTUS*

SPOILAGE YEASTS. Magyar Mikrobiológiai Társaság 2017 évi Nagygyűlése, Keszthely, 2017.

14. Lajos Ács-Szabó, László Attila Papp, Hajnalka Csoma, Matthias Sipiczki & Ida Miklós. Intergenic regions as determinants of gene order conservation in the fission yeasts. Magyar Mikrobiológiai Társaság 2018 évi Nagygyűlése, Eger, 2018.
15. Lajos Ács-Szabó, László Attila Papp, Zoltán Kállai, Ida Miklós & Matthias Sipiczki. Microbial communities in aquatic environments. Magyar Mikrobiológiai Társaság 2018 évi Nagygyűlése, Eger, 2018.
16. László Attila Papp, Lajos Ács-Szabó, Ida Miklós. Transcriptomic analysis of hyphal growth in *Schizosaccharomyces japonicus*. Magyar Mikrobiológiai Társaság 2018 évi Nagygyűlése, Eger, 2018.
17. László Attila Papp, Lajos Ács-Szabó, Szonja Takács, Ida Miklós. Deletion of the DNA ligase IV (lig4) results in stress sensitivity in *Schizosaccharomyces japonicus*. Magyar Mikrobiológiai Társaság 2018 évi Nagygyűlése, Eger, 2018.
18. Lajos Ács-Szabó, Ildikó Vig, László Attila Papp, Ida Miklós. Comparative bioinformatics analyses of the extremely conserved fungal spliceosome component Cwf14 (Bud31) and its other counterparts. Magyar Mikrobiológiai Társaság 2020 évi Nagygyűlése, Keszthely, 2020. IF:
19. Lajos Ács-Szabó, László Attila Papp, Hajnalka Csoma, Ida Miklós & Matthias Sipiczki. Stowaway or good mate? – Isolation of fungal strains from stock cultures of cyanobacteria. Magyar Mikrobiológiai Társaság 2020 évi Nagygyűlése, Keszthely, 2020. IF:
20. Szonja Takács, László Attila Papp, Lajos Ács-Szabó, Ida Miklós. Pleiotropic effects of Ligase 4 gene deletion in *Schizosaccharomyces japonicus* dimorphic fission yeast. Magyar Mikrobiológiai Társaság 2020 évi Nagygyűlése, Keszthely, 2020. IF:
21. László Attila Papp, Lajos Ács-Szabó, Szonja Takács, Ida Miklós. Replacement of agar by gelatine enhanced the quality of RNA isolated from *Schizosaccharomyces japonicus* hyphae. Magyar Mikrobiológiai Társaság 2020 évi Nagygyűlése, Keszthely, 2020. IF:

22. Livia Dályai, Madina Khamitova, Lajos Ács-Szabó, Ida Miklós, Hajnalka Csoma. Detection of pulcherrimin gene cluster in *Metschnikowia fructicola* type strain (CBS 8853). FIBOK, Debrecen, 2020.
23. Tamás Plaszkó, Zsolt Szűcs, Zoltán Cziáky, Lajos Ács-Szabó, Hajnalka Csoma, László Géczy, Gábor Vasas, Sándor Gonda. Metabolom–mikrobiom korrelációk vizsgálata különböző tomafajtákban. Fiatal Gyógynövénykutatók Fóruma, Budapest, 2022.
24. Lajos Ács-Szabó, László Attila Papp, Hajnalka Csoma, Ida Miklós, Matthias Sipiczki. A NEW STRAIN OF *CYPHELLOPHORA OLIVACEA* SHOWS STRIKING TOLERANCE FOR SODIUM BICARBONATE. Biotechnology at the University of Debrecen – 2022 International Symposium, Debrecen, 2022.
25. Lajos Ács-Szabó, Matthias Sipiczki. BIOPRESERVATION OF RAW MEAT – A MICROBIAL AND METABARCODING APPROACH. Biotechnology at the University of Debrecen – 2022 International Symposium, Debrecen, 2022.
26. Miklós Ida, Horváth Enikő, Csoma Hajnalka, Ács-Szabó Lajos, Papp László Attila. Mikrobiológia és genetikai kutatások az egészségesebb környezetért. International Symposium on Environment, Health and Safety, Debrecen, 2022.
27. Ács-Szabó Lajos, Papp László Attila, Mátyás Sipiczki, Emri Tamás, Domonkos Dávid, Miklós Ida. Mikrobiológiával az élhetőbb jövőért - hazai algatörzsek izolálása, szelektálása bioremediációs célokra. International Symposium on Environment, Health and Safety, Debrecen, 2022.
28. Ildikó Vig, Lajos Ács-Szabó, Zsigmond Benkő, László Attila Papp, Matthias Sipiczki, Ida Miklós. THE SCHIZOSACCHAROMYCES POMBE *cwf14* GENE IS EVOLUTIONARILY CONSERVED AND INVOLVED IN THE TOR PATHWAY. Biotechnology at the University of Debrecen – 2022 International Symposium, Debrecen, 2022.