

**Doktori (PhD) értekezés tézisei**

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

Ács-Szabó Lajos

Témavezető: Gálné Dr. Miklós Ida



**DEBRECENI EGYETEM**

Juhász-Nagy Pál Doktori Iskola

Debrecen, 2023



## 1. Bevezetés és célkitűzések

A feltevés, miszerint az élőlények változ(hat)nak az idők során, már az ókori gondolkodók körében is megfogalmazódott. Ma már tudjuk, hogy a változások három fő szintéren 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 kromoszómákkal 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 azonban 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 transzgen 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 (*S. pombe*, *S. octosporus* és *S. japonicus*). Egy további fajt megszekvenáltak ugyan, de nem szerelték teljesen össze a genomját (*Schizosaccharomyces cryophilus*), míg a legutóbb leírt faj (*S. osmophilus*) elérhető genom szekvenciával jelen dolgozat születésekor még nem rendelkezett.

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

Mivel a *S. cryophilus* megszekvenált genomjának összeszerelése csak részlegesen történt 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 genom konzerváltság mértékét. Lehetőségünk nyílt a fajok legközelebbi közös őstől származó genom-szegmensek felderítésére, majd a beazonosított szegmensek tulajdonságainak vizsgálatára. Végül megkíséreltük feltárni a konzervált régiók fennmaradásának néhány lehetséges okát, azok egymásra hatását.

## 2. Anyagok és módszerek

### 2.1. A kísérletek során felhasznált élesztő törzsek és tenyésztésük

A laboratóriumi kísérleteink során a *S. pombe* (L-972 h<sup>-</sup>), *S. cryophilus* (CBS11777) és *S. cerevisiae* (S288C) vad típusú törzseit használtuk fel, amelyeket komplex tápfolyadékokban (YEL, YPL) és táptalajokon (YEA, YPA) szaporítottunk, illetve tartottunk fent 30°C és 25°C-on.

### 2.2. Az alkalmazott molekuláris biológiai módszerek bemutatása

Specifikus DNS szakaszok irányított felszaporításához genomi DNS-t izoláltunk a *S. cryophilus* élesztő tenyészetből, majd különböző primerek segítségével amplifikáltunk DNS szekvenciákat DreamTaq polimeráz enzim (Thermo Scientific EP0711) felhasználásával. A sikeresen felszaporított DNS szakaszokat szekvenálásnak vetettük alá (GeneArt és Microsynth AG szolgáltatók). A *S. cryophilus* kromoszóma méreteinek meghatározása érdekében PFGE analízist végeztünk (Nguyen & Gaillardin, 1997; Naumov és mtsi., 2015).

### 2.3. Az alkalmazott bioinformatikai módszerek bemutatása

A bioinformatikai elemzésekhez felhasznált gomba genom szekvenciákat nemzetközi adatbázisokból töltöttük le: Broad Institute, NCBI, EnsemblFungi, JGI Genome Portal, *Saccharomyces* Database, *Saccharomyces sensu stricto* és Pombase.

A hasadó élesztőgombák genom szekvenciáit páronkénti és többszörös teljes genom illesztéseknek vetettük alá az online YASS (Noé

& Kucherov, 2005) és az asztali gépen futtatott Mauve (Darling és mtsi., 2010) programokkal, amelyek segítségével feltártuk a genomok konzervált régióit, továbbá ezek az adatok kerültek felhasználásra a GRIMM algoritmussal (Tesler, 2002) végzett kromoszóma átrendeződési vizsgálatokhoz is. Feltételezett ortológ szekvenciák azonosításához BLASTp kereséseket hajtottunk végre a korábban felsorolt adatbázisokban. Filogenetikai elemzésekhez az online Phylogeny.fr (Dereeper és mtsi., 2008), MAFFT+NJ (Katoh & Standley, 2013) és PhyML 3.0 (Guindon és mtsi., 2010) oldalakat/algoritmusokat használtuk. A szinténikus szakaszok vizualizációjához pedig a Genome Synteny Viewer GSV (Revanna és mtsi., 2011), SimpleSynteny (Veltri és mtsi., 2016) és OrthoClusterDB (Ng és mtsi., 2009) online platformok kerültek felhasználásra. Létrehoztunk egy adatbázist, amelyben a 4 hasadó élesztőgomba faj (*S. japonicus*, *S. pombe*, *S. octosporus* és *S. cryophilus*) feltételezett ortológ szekvenciái voltak megtalálhatók a genomi koordinátáikkal egyetemben. Ebben az adatbázisban olyan lokálisan kolineáris blokkokat kerestünk manuálisan, amelyek legalább 5 gént tartalmaztak és e gének sorrendje és irányultsága mind a 4 hasadó élesztő fajban megegyezett. E blokkokra ezt követően aLCB-ként hivatkozunk (aLCB – ancestral locally collinear blocks), mivel ezeket a legközelebbi közös őstől örökölhették a szóban forgó fajok.

Az aLCB-k fennmaradásának és a szelekció kapcsolatának vizsgálata érdekében „in silico” evolúciós modellezéseket végeztünk egy házon belül fejlesztett Python szkript és az Artificial Life Framework (ALF) (Dalquen és mtsi., 2012) segítségével. Emellett összehasonlítottuk az aLCB-kben és azokon kívül lokalizálódó gének evolúciós rátáit (Rhind

és mtsi., 2011) és génszerkezetét (intron vesztés/nyerés) (Zhu & Niu., 2013) is.

Az aLCB-k génsorrendjének fenntartására potenciálisan ható tényezők közül megvizsgáltuk a funkcionális csoportosulás (Pombase - biológiai funkciók GO kategóriái) (Lock és mtsi., 2018) és a ko-expresszió (Koch és mtsi., 2012) lehetőségét, továbbá a Rec12 hasító helyek (Fowler és mtsi., 2014) és esszenciális gének (Pombase) lokális denzitását.

Az adatok kiértékeléséhez számos normalizációs, randomizációs és statisztikai tesztet végeztünk el, amelyekhez a Microsoft Excel 2016 és Past4 (Hammer és mtsi., 2001) programokat használtuk.

### 3. Eredmények és megbeszélésük

#### 3.1. *A S. cryophilus szuperkontigok összerendezése és az összerendezés validálása*

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.

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.

Első lépésként megkíséreltük a *S. cryophilus* 9 nagyméretű szuperkontigját összeszerelni a rokon fajok genomszekvenciái segítségével. 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 Mauve programmal végzett teljes genom illesztések 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 lokálisan kolineáris blokkok (LCB-k) voltak azonosíthatók a fajok között, mint az *S. cryophilus* – *S. pombe* fajpár esetében. 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 referenciacajok függvényében.

Ezért megpróbáltunk erősen konzervált „horgonypontokat” keresni a *S. cryophilus* és rokonfajai szekvenciáiban, amelyek támpontul szolgálhattak az összeillesztéshez. A *S. pombe* és *S. octosporus* pericentromerikus és szubtelomerikus génjeinek segítségével sikerült beazonosítanunk a *S. cryophilus* összes pericentromerikus és négy 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, 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 kaptunk.

Eredményeink alapján a következő szuperkontig sorrendet javasoltuk: Sc3-Sc9-Sc1 (ChrI); Sc7-Sc5-Sc8-Sc6 (ChrII); Sc4-Sc2 (ChrIII).

A sem pericentromerikus, sem szubtelomerikus 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). Azt is megfigyelhettük, hogy a *S. cryophilus* kromoszóma méretei valamelyest eltérnek a *S. pombe* kromoszómáinak méreteitől, de szekvenciák alapján becsült méretekről is. 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 tudtunk 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.

### 3.2. 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. 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 volt jelentős 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. Ezért megvizsgáltuk a közelebbi rokon *S. cerevisiae* – *S. uvarum* és távolabbi rokon *S. cerevisiae* – *N. castellii* fajpárokat is hasonló megközelítéssel, mint a hasadó élesztőket.

Ha az összes megfigyelt átrendeződést figyelembe vettük, akkor számottevően több átrendeződés érhette 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 figyelhettünk 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 génexpressziós 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 filogenetikai távokon is túlsúlyban maradnak a transzlokációk,

egyszerűen a bekövetkezett inverziók mossák el a transzlokációk nyomait (Seoighe és mtsi., 2000).

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

Nyilvánvaló vált, hogy a nagy arányú géntartalom konzerváltság ellenére elég sok átrendeződés érhető 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? E vizsgálatokba már bevontuk a *S. japonicus*-t is referenciaként, hiszen ez a faj a nemzetség legkorábbi leágazása.

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).

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 spezializálódott életmódot követ – hogy az életmódjuk állhat a kimagaslóan erős szekvencia- és struktúra változások korrelációinak hátterében.

#### *3.4. 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 (aLCB-k). Eredményeink alapján a vizsgált genomok géntartalmának 40-42%-a ilyen aLCB-kben helyezkedik el és valószínű, hogy ezek a blokkok a legközelebbi közös 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 nemzetséget az Ascomycota

törzsben, ahol ilyen mértékű szekvencia divergencia mellett ilyen mértékű génsorrend konzerváltság lenne megfigyelhető.

### *3.5. 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 szétválnak, majd újra egymás mellé kerülnek a sok átrendeződésnek köszönhetően. Mivel a hasadó élesztők esetében a géntartalom 40-42%-a aLCB-kben 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 (ALF) 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.

### *3.6. Az ősi kolineáris blokkok lehetséges eredete*

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 (*Pneumocystis murina*, *Taphrina deformans*, *Protomyces lactucae-debilis*, *Saitoella complicata* és *Neolecta irregularis*) 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 (*Aspergillus nidulans*, *Botrytis cinerea*, *Neurospora crassa*), mint a sarjadzó élesztő genomokban (*Yarrowia lipolytica*, *Debaryomyces hansenii*, *Meyerozyma guilliermondii*). 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).

### 3.7. 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. Ezt vizsgálandó, összehasonlítottuk több mint 4200 protein szekvencia evolúciós rátáját az aLCB-ken belül, illetve azokon kívül is. 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 található 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).

### *3.8. 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 interkromozómális transzlokációk eltérő méretű kromozómákhoz 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 reproduktív izoláció mellett az is az átrendeződések ellen hathat, hogy a *S. pombe* genom szekvenciájának 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 becslött 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 transzkripció 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. Mindemellert 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.

## 4. 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 transzgén 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 élesztőgomba (*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.

Mivel az *S. cryophilus* genomját megszekvenálták ugyan, de összeállításra nem került (Rhind és mtsi., 2011), első lépésként e faj nagy méretű szuperkontigjait kívántuk összeszerelni. Ehhez közeli rokon fajok (*S. pombe* és *S. octosporus*) szinténikus kapcsolatait használtuk fel, majd

az összeszerelést különböző molekuláris biológiai módszerekkel (PCR és PFGE) validáltuk. A teljes genom szekvenciák összehasonlítása és a genomokat ért átrendeződések feltárása során megállapítottuk, hogy a *S. cryophilus* legközelebbi rokona a kromoszóma struktúra alapján is a *S. octosporus*. Kiderült, hogy a genomokat ért átrendeződések közül az interkromozómális transzlokációk voltak a leggyakoribbak, az inverziók száma csak a filogenetikai távolság növekedésével került túlsúlyba. Hasonló divergencia idejű Saccharomycotina fajok esetében is hasonló tendencia volt megfigyelhető, bár az esetükben nagyobb számú, de kisebb régiókat érintő változások következtek be.

A nemzetség legkorábbi leágazásának számító *S. japonicus* bevonásával megállapítottuk, hogy a hasadó élesztők esetében a szekvencia és strukturális változások erős pozitív korrelációt mutatnak, továbbá a másik 3 faj közel hasonló átrendeződési dinamikát mutat a referencia fajhoz képest. Emellett a teljes genom illesztések azt is előre vetítették, hogy számos esetben az eltérő fajok ugyanazon genomi régiói maradhattak változatlanok.

Ez a feltevés megerősítés nyert, amikor is a négy faj feltételezett közös ortológjainak beazonosítása során olyan lokálisan kolineáris blokkok (aLCB-k) kerültek feltárára, amelyek legalább 5 azonos sorrendű és irányultságú génből álltak. Ilyen aLCB-kben lokalizálódik a genomok géntartalmának 40-42%-a (2055 gén). Ezek az aLCB-k különböző számítógépes modellezések alapján nem lehetnek csupán a legközelebbi közös ős génsorrendjének maradványai, hanem valószínűleg a szelekció is hatással lehet a fenntartásukra. Viszont az aLCB-k génsorrendje a komparatív vizsgálatok alapján a hasadó

élesztőkre jellemző, mivel a vizsgálatba bevont egyéb gomba fajok genomjaiban nem lehetett hasonló aLCB-ket találni. Meglepő módon a *P. murina* és a fonalas gombák genomjaiban több szomszédos génpárt lehetett azonosítani, mint a sarjadzó élesztők vagy egyéb gombák genomjaiban.

Az aLCB-k és az azokon kívül eső régiók összehasonlítása során kiderült, hogy a szekvenciák változásainak (protein evolúciós ráta és intron vesztés/nyerés) dinamikája között nincs érdemi különbség, tehát hiába nagy a konzerváltság a génsorrendet tekintve, maguk a szekvenciák az aLCB-ken belül is ugyanolyan változékonyak, mint a genomok egyéb régiói.

Vizsgálataink során megállapítottuk, hogy bizonyos biológiai funkciókat ellátó gének (GO kategóriák) hajlamosabbak az aLCB-kben csoportosulni, mint mások. Bár a szóban forgó 13 GO kategória statisztikailag szignifikáns módon lokalizálódik az aLCB-kben, de elhelyezkedésük diszperz, tehát nem egy funkciót ellátó gének csoportosulnak egymás közelében. További érdekesség, hogy az aLCB-k génjei esetében a ko-expresszió értéke sem magasabb, mint az egyéb régiókban lokalizálódó gének értékei. Viszont megállapítottuk, hogy a szignifikáns GO kategóriák génjei között magasabb az esszenciális gének aránya és jóval kevesebb a környékükön található Rec12 hasítóhelyek száma. Ez arra enged következtetni, hogy az aLCB-k fennmaradásában nagy szerepet játszhat az esszenciális gének és Rec12 hasítóhelyek lokális denzitása. Meggyőződésünk, hogy eredményeink hozzájárulnak a hasadó élesztők, ezáltal az eukarióta génsorrend konzervációjában szerepet játszó jelenségek megismeréséhez, megértéséhez.

## 5. Tézisek

- 1) Sikeresen kromoszóma szintű összeszerelési állapotba hoztuk a *S. cryophilus* szuperkontigjait a rokon fajokkal való szinténikus kapcsolatok, továbbá a pericentromerikus és szubtelomerikus régiók kolineáris génsorrendjeinek segítségével. Bioinformatikai eredményeink alapján a Sc3-Sc9-Sc1 (ChrI); Sc7-Sc5-Sc8-Sc6 (ChrII); Sc4-Sc2 (ChrIII) sorrendet javasoltuk.
- 2) 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.
- 3) 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.
- 4) Kimutattuk, hogy az átrendeződések közül az interkromoszómális transzlokációk voltak túlsúlyban, az inverziók száma csak a filogenetikai távolság növekedésével emelkedett.

- 5) Összehasonlítottuk az átrendeződések típusa és mennyisége szempontjából a hasadó élesztő genomokat hasonló divergencia idővel rendelkező sarjadzó élesztő genomokkal. Bár összességében több átrendeződési esemény érhetette 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.
- 6) 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.
- 7) 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.
- 8) 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.

9) Számítógépes szimulációkat alkalmazva kimutattuk, hogy az aLCB-k génsorrendje nem csak az ősi génsorrend maradványa, hanem szelekciós hatás alatt is állhat.

10) Megállapítottuk, hogy az aLCB-k génsorrendje bár hasadó élesztő specifikus, de tükrözhetik egy fonalas ős génsorrendjét is, mivel a vizsgált aLCB-k génsorrendje jobban hasonlít a fonalas gombák génsorrendjére, mint a sarjadzó élesztő fajok génsorrendjére.

11) 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.

12) Eredményeink azt sugallták, hogy bizonyos biológiai funkciókhoz (13 GO kategória) köthető gének szignifikánsan nagyobb arányban hajlamosak az aLCB-kben csoportosulni, mint mások. Mindazonáltal e GO kategóriák génjei is szétszórt elhelyezkedést mutatnak a kromoszómákon, tehát az egy biológiai funkcióhoz köthető gének nem szomszédosak az aLCB-kben.

13) 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.

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

## 6. 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 prof. Dr. Sipiczki Mátyásnak, aki szakmai beszélgetéseink során észrevételeivel segítette a munkámat és biztatott a dolgozat megírására. Köszönöm Dr. Papp László Attilának, 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 Dr. Csoma Hajnalkának, aki sokat biztatott, lelkesített a doktori munka során. Köszönet illeti Dr. Antunovics Zsuzsát a PFGE vizsgálatok elvégzéséért. Szeretnék köszönetet mondani a Genetikai és Alkalmazott Mikrobiológiai Tanszék technikusainak: Lakatos Zoltánnának, Oláh Anitának, Struba-Hegedűs Szilviának, Kiss Zoltánnának 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 projekttámogatások.

**Short thesis for the degree of doctor of philosophy (PhD)**

**Comparative genomics of the fission yeast  
(*Schizosaccharomyces*)**

by Lajos Ács-Szabó

Supervisor: Dr. Ida Miklós



UNIVERSITY OF DEBRECEN

Juhász-Nagy Pál Doctoral School

Debrecen, 2023



## **1. Introduction and aims**

The assumption that organisms change over time was already held by ancient thinkers. Nowadays we know that change occurs at three main levels: phenotype, karyotype and DNA sequence. While adaptation is usually the main driver of phenotypic changes, many neutral changes, such as synonymous mutations, could occur at the DNA sequence level. The picture is perhaps most nuanced for karyotype, where we often observe different chromosome numbers and patterns within a species. Different chromosome numbers within a species are most common among plants, although in their case the different chromosome number is a consequence of polyploidy. However, both fungi and animals are known to be composed of geographically distinct populations with different numbers of chromosomes.

Nowadays, comparing whole genome sequences gives the opportunity to study the changes and their relationship together. This type of (comparative genomics) analysis has led, for example, to the discovery of whole genome duplications, the calibration of different molecular clocks or the refutation of random gene order. However, these studies do not only lead to theoretical conclusions, they also have many practical "benefits". For example, in the event of an epidemic, comparative genomics helps to predict certain changes that will facilitate control. In addition, comparing genomes of related species can also facilitate appropriate transgene integration, whether for research or economic purposes.

The aims of this present thesis are to describe some aspects of changes in the fungal genomes, focusing on the fission yeast (*Schizosaccharomyces*) genomes. Fission yeasts belong to a basal branch of the phylum Ascomycota and thus share many features with members of the Metazoa group, such as relatively large chromosome sizes, fission-type cell division, cell cycle control, metabolic pathways, RNA interference and many more. On the other hand, they are unique in that their genomes are highly conserved in terms of gene content, while their sequences change rapidly. The best-known member of the genus is the *Schizosaccharomyces pombe*, which has long been a popular model organism. To our current knowledge, the genus comprises five species, three of which have their genomes assembled to near chromosome level (*S. pombe*, *S. octosporus* and *S. japonicus*). One additional species has been sequenced but not fully assembled (*Schizosaccharomyces cryophilus*), while the most recently described species (*S. osmophilus*) did not have an available genome sequence at the time of writing.

In order to understand the genetic and evolutionary background of the stable genome organization in fission yeasts, we wanted to subject the four species of the genus with known genome sequences to different comparative studies.

As the assembly of the sequenced genome of *S. cryophilus* has only been partially assembled (Rhind et al., 2011), we aimed to assemble large supercontigs of the species and then validate the assembly using different molecular biology methods. With the assembled genome sequence, we were able to compare chromosome rearrangement dynamics in closely and distantly related species. Subsequently, we

wanted to investigate the relationship between the evolution of fission yeast sequence and chromosome structure, and the degree of genome conservation at the DNA level among the species. We were able to identify genomic segments from the closest common ancestor of the species and then investigate the properties of the identified segments. Finally, we attempted to explore some of the possible reasons for the persistence of conserved regions and their interactions.

## **2. Materials and methods**

### *2.1. Yeast strains used in the experiments and their cultivation*

In our laboratory experiments, wild-type strains of *S. pombe* (L-972 h-), *S. cryophilus* (CBS11777) and *S. cerevisiae* (S288C) were used. The strains were grown in liquid media (YEL, YPL) and maintained on solid media (YEA, YPA) at 30°C and 25°C.

### *2.2. Methods of molecular biology used in the study*

For the directed amplification of specific DNA sequences, genomic DNA was isolated from *S. cryophilus* yeast culture and amplified using different primers with the DreamTaq polymerase enzyme (Thermo Scientific EP0711). Successfully amplified DNA sequences were subjected to sequencing (GeneArt and Microsynth AG providers). PFGE analysis was performed to determine the chromosome sizes of *S. cryophilus* (Nguyen & Gaillardin, 1997; Naumov et al., 2015).

### *2.3. Presentation of bioinformatics methods used*

The fungal genome sequences used for bioinformatics analyses were downloaded from international databases: the Broad Institute, NCBI, EnsemblFungi, JGI Genome Portal, Saccharomyces Database, Saccharomyces sensu stricto and Pombase.

The fission yeast genome sequences were subjected to pairwise and multiple whole genome alignments using the online YASS (Noé & Kucherov, 2005) and desktop Mauve (Darling et al., 2010) programs to detect conserved regions of the genomes, and these data were also used

for chromosome rearrangement analyses using the GRIMM algorithm (Tesler, 2002). BLASTp searches were performed to identify putative orthologous sequences in the databases listed previously. For phylogenetic analyses, we used the online Phylogeny.fr (Dereeper et al., 2008), MAFFT+NJ (Katoch & Standley, 2013) and PhyML 3.0 (Guindon et al., 2010) sites/algorithms. For the visualization of syntenic regions, the online platforms Genome Synteny Viewer GSV (Revanna et al., 2011), SimpleSynteny (Veltri et al., 2016) and OrthoClusterDB (Ng et al., 2009) were used. A database was created containing putative orthologous sequences of the four fission yeast species (*S. japonicus*, *S. pombe*, *S. octosporus* and *S. cryophilus*) along with their genomic coordinates. In this database, we manually searched for locally colinear blocks that contained at least 5 genes and the sequence and orientation of these genes were the same in all four fission yeast species. These blocks are then referred to as aLCBs (ancestral locally collinear blocks), as they may have been inherited from the last common ancestor of the species.

To investigate the relationship between the persistence of aLCBs and selection, we performed *in silico* evolutionary modelling using an in-house built Python script and the Artificial Life Framework (ALF) (Dalquen et al., 2012). We also compared the evolutionary rates (Rhind et al., 2011) and gene structure (intron loss/gain) (Zhu & Niu., 2013) of genes localizing in and outside aLCBs.

Among the factors that potentially influence the maintenance of the gene order of aLCBs, we investigated the possibility of functional clustering (Pombase - GO categories of biological functions) (Lock et al.,

2018) and co-expression (Koch et al., 2012), as well as local density of Rec12 cleavage sites (Fowler et al., 2014) and essential genes (Pombase).

To evaluate the data, we performed several normalization, randomization and statistical tests using Microsoft Excel 2016 and Past4 (Hammer et al., 2001).

### 3. Results and discussion

#### 3.1. Assembly of the *S. cryophilus* supercontigs and validation of the assembly

Although nowadays sequencing has become routine and new genome sequences are published almost *en masse*, accurate assembly and annotation of a genome remains a major challenge (Bradnam et al., 2013; Liao et al., 2019). For this reason, in many cases genome sequences are not assembled completely, many remain in the so-called "draft" genome state.

Since fission yeasts constitute a highly interesting (and valuable) group of the fungal kingdom and the genome of one member of the genus, *S. cryophilus*, has not been previously assembled to a chromosome-level assembly state, we aimed to assemble this genome, validate the assembly, and perform comparative genomic comparisons with both closely and distantly related species.

As a first step, we attempted to assemble the 9 large supercontigs of *S. cryophilus* using genome sequences of related species. Previous work has shown that the closest relatives of *S. cryophilus* are most probably *S. octosporus* and *S. pombe*. Since the gene content is highly conservative between the aforementioned species, we used the genome sequences of these species as a reference (Helston et al., 2010; Rhind et al., 2011).

Whole-genome alignments using the Mauve program confirmed the results of Helston et al. (2010) and Rhind et al. (2011) that indeed *S. octosporus* is the closest relative of *S. cryophilus*, as much larger locally

collinear blocks (LCBs) were identified between the species than in the *S. cryophilus* - *S. pombe* species pair. At the same time, we were also convinced that even with such a high degree of genome conservation, it is not possible to assemble a genome based on syntenic relationships alone. In fact, rearrangement of the *S. cryophilus* supercontigs resulted in a completely different order depending on the reference species used.

Therefore, we tried to look for highly conservative "anchor points" in the sequences of *S. cryophilus* and its related species that could serve as a reference for alignment. Using the pericentromeric and subtelomeric genes of *S. pombe* and *S. octosporus*, we were able to identify all pericentromeric and four subtelomeric contig ends of *S. cryophilus*. For the contig ends that were thought to be pericentromeric, additional evidence was provided by the finding of clusters of tRNA-encoding genes in these regions, similar to those of closely related species (Kuhn et al., 1991; Rhind et al., 2011; Tong et al, 2019). Although not all subtelomeric regions could be identified, sequences of rDNA cluster were found for one putative subtelomeric end (Sc7), providing further evidence that the contig has a subtelomeric end (Pasero & Marilley, 1993).

With this new knowledge, we rearranged the supercontigs of *S. cryophilus* and compared them to the related species, resulting in a much more ordered genomic landscape than before. Based on our results, we proposed the following supercontig order: Sc3-Sc9-Sc1 (ChrI); Sc7-Sc5-Sc8-Sc6 (ChrII); Sc4-Sc2 (ChrIII).

PCR reactions were designed and carried out to demonstrate the neighbouring of neither pericentromeric nor subtelomeric Sc9-Sc1 and Sc5-Sc8-Sc6 supercontigs. The successful reactions supported our

bioinformatics results for the supercontigs that were assumed to be related.

Although the proposed supercontig sequence appeared to be appropriate based on bioinformatics results and PCR validation, we also subjected the *S. cryophilus* chromosomes to PFGE analysis to obtain further confirmation. We confirmed the hypothesis that *S. cryophilus* strain OY26 also has 3 distinct chromosomes, as do the vast majority of other fission yeast strains (Rhind et al., 2011). We also observed that the chromosome sizes of *S. cryophilus* differ somewhat from those of *S. pombe* chromosomes, but also from sizes estimated from sequences. The discrepancies were probably due to the unknown extent of centromeres, telomeres and rDNA repeats.

To test this hypothesis, we attempted to determine the localization of rDNAs (18S-5,8S-26S rDNA) at the ends of *S. cryophilus* chromosomes. As a starting point, we used the *S. cryophilus* - *S. octosporus* syntenic relationships and then sought to determine the localization of potential rDNA repeats by PCR. Of two putative cases (Sc3 and Sc7), we were able to experimentally prove one (Sc7), but the indirect evidence suggested that our proposed chromosome assembly might be appropriate.

A recent study using a third-generation method to resequence fission yeast genomes confirmed our results (Tong et al., 2019). Only one discrepancy was identified, as a centromeric crossover resulted in the replacement of the Sc7 and Sc4 contigs. In the absence of knowledge of centromeric sequences, our methods were unable to shed light on this change. However, subsequent results were not affected by this change.

### 3.2. Comparison of the *S. cryophilus* genome with closely and distantly related species

Following the assembly and validation of the *S. cryophilus* supercontigs, we were curious to know what types of changes shaped the genome of this fission yeast.

We have shown that, compared to the *S. octosporus* genome, the *S. cryophilus* genome underwent a number of chromosome rearrangement events: mainly interchromosomal translocations affecting large regions. Our results also revealed the interesting fact that the number of inversions increased significantly with increasing phylogenetic distance. Thus, in the comparison of *S. cryophilus* - *S. pombe*, the number of inversions was already significant compared to the number of translocations.

Rhind et al (2011) in their comparative genomics study showed that the fission yeast genomes are highly conserved compared to, for example, the genus *Saccharomyces* with similar divergence times. Therefore, we also examined the more closely related *S. cerevisiae* - *S. uvarum* and more distantly related *S. cerevisiae* - *N. castellii* species pairs using a similar approach as in the case of the fission yeasts. When all observed rearrangements were taken into account, significantly more rearrangements could have affected the budding yeast genomes than the fission yeasts. However, if we excluded changes in the subtelomeric region (as they are prone to rearrangements, Fischer et al., 2001), we could draw quite different conclusions. The number of rearrangements detected by a more rigorous approach for *S. cerevisiae* - *S. uvarum* was in agreement with the values estimated by Fischer and mtsi (2000, 2001).

However, in this case, proportionally more rearrangements involving large regions could occur in the fission yeast chromosomes than in the budding yeast genomes examined. Based on the conservative gene content of fission yeasts, we expected to observe fewer rearrangements. When looking at the types of rearrangements, we observed a similar trend in budding yeasts, with the number of inversions increasing with phylogenetic distance.

These observations might lead one to ask whether translocations would be more tolerable than inversions on a smaller evolutionary scale? Given that both translocations and inversions can affect the gene expression pattern of the whole genome, thereby increasing or decreasing the fitness of the organism, and are capable of inducing reproductive isolation, this hypothesis is unlikely (Lowry & Willis, 2010; Avelar et al, 2013; Zanders et al., 2014; Naumov et al., 2015). We hypothesize that rather, the underlying mechanisms that promote or cause translocations may favour translocations. Consistent with this, it is also possible that translocations remain predominant at longer phylogenetic distances, too, only the inversions that have occurred remove the traces of translocations (Seoighe et al., 2000).

### *3.3. Comparison of structural and sequence evolution in the fission yeasts*

It has become apparent that, despite the high degree of gene conservation, the genomes of the three fission yeast species studied previously may have undergone quite a lot of rearrangement, and the species show a surprisingly high rate of sequence evolution (Helston et

al., 2010; Rhind et al., 2011). We therefore wondered whether the structural and sequence evolution of fission yeasts showed any correlation? In these studies, we have already included *S. japonicus* as a reference, as this species is the earliest branch of the genus.

We have demonstrated that the sequence and structural evolution of fission yeasts show a strong positive correlation, similar to, for example, the genera *Lachancea*, *Verticillium* and several Metazoa, but different from the genus *Aspergillus* (Burt et al., 1999; Coghlan & Wolfe, 2002; Sharakhov et al., 2002; Galagan et al., 2005; Vakirlis et al., 2016; Shi-Kunne et al., 2018).

We also found that, from a structural evolution perspective, the other three species show nearly the same degree of divergence compared to *S. japonicus*, suggesting that they have almost uniform rearrangement rates in their genomes. A study using a different approach has reached the same conclusion using other Taphrinomycotina species (Rajeh et al., 2018). This result is certainly surprising when considering the otherwise substantial phylogenetic distances within the genus. It is conceivable, given that most of the species in the Taphrinomycotina subphylum follow a rather specialized lifestyle and their lifestyle could be behind the strong correlations of sequence and structural changes.

### *3.4. Genome conservation and identification of ancestral collinear blocks in the fission yeast genomes*

Our results also revealed that there is nearly equal conservation at the nucleotide level between species from the perspective of *S. japonicus*. Moreover, not only the overall size of conserved regions, but

also the distribution of individual segment sizes showed significant similarity. This suggested that similar regions may have remained unchanged in fission yeast genomes. This hypothesis was confirmed by the exploration of genomic regions of species that consisted of at least 5 (on average 8) orthologous genes with the same sequence and orientation (aLCBs). Our results suggest that 40-42% of the gene content of the genomes studied is located in such aLCBs and that these blocks may reflect the gene order of the closest common ancestor. Considering that, for example, the fugu (pufferfish) and human genomes show less sequence divergence than the fission yeasts studied, but nevertheless, in the former pair only small groups of between two and three genes remained adjacent, the results are very interesting (Sipiczki, 2000; Smith et al, 2002; Rhind et al., 2011). Besides, based on thorough examination of the dataset of Rajeh et al. (2018), there is almost no genus in the Ascomycota phylum with the same sequence divergence who exhibit as highly conserved a gene order as the fission yeasts do.

### *3.5. The relationship between the survival of ancestral collinear blocks and natural selection*

The existence of collinear blocks, pairs of adjacent genes that can be observed in different genomes, can be the result of selection or chance (Hurst et al., 2004). In the former case, it is advantageous if certain genes remain adjacent due to proper regulation or co-expression. In the other case, pairs of genes split and then become adjacent again due to rearrangement events. Since 40-42% of the gene content of the fission yeasts is located in the aLCBs, the effect of chance can be minimised.

Rather, the question is whether the observed conservation is "merely" a remnant of the ancestral gene order or whether selection is involved in their maintenance.

To answer this question, we used different *in silico* models and both the neutral approach and the phylogenetic method (ALF) suggested the same result. The simulations suggested that the persistence of aLCBs could not be a matter of chance, as none of the test results approximated the values observed in reality for the average gene content of aLCBs. We hypothesise that aLCBs are under maintenance selection.

### 3.6. Possible origin of the ancestral locally collinear blocks

If the gene order of aLCBs is derived from the last common ancestor and is under selection, there is some possibility that similar gene order could be observed in other Taphrinomycotina species (*Pneumocystis murina*, *Taphrina deformans*, *Protomyces lactucae-debilis*, *Saitoella complicata* and *Neoelecta irregularis*). Although small collinear regions were observed between the species pairs studied, no relevant number of blocks was identified.

The inclusion of phylogenetically more distant species yielded similar results, suggesting that the gene order of the aLCBs studied is specific to the fission yeasts. However, it is important to note that significantly more adjacent gene pairs were identified in the genomes of filamentous species (*Aspergillus nidulans*, *Botrytis cinerea*, *Neurospora crassa*) than in the genomes of budding yeasts (*Yarrowia lipolytica*, *Debaryomyces hansenii*, *Meyerozyma guilliermondii*). This result is in agreement with the hypothesis that the ancestors of modern yeasts were

filamentous fungi and it is possible that the observed gene order similarities reflect the gene order of the filamentous ancestors (Berbee & Taylor, 1993; Sipiczki, 2000; Nagy et al, 2014).

### *3.7. Conservation of ancestral collinear blocks at sequence level*

After confirming that the observed gene order is fission yeast specific, we wondered whether sequences localized in aLCBs show greater conservation compared to sequences located in other regions of the genome. To investigate this, we compared the evolutionary rates of more than 4200 protein sequences within and outside aLCBs. Strikingly, there was no significant difference between the two groups. In light of this, we conclude that the sequences of genes localised within aLCBs show a similar degree of variability as those located in other regions of the genome. The analysis of intron loss and gain also led to similar results.

There may be several reasons behind these observations. For one, genomic regions outside the aLCBs are not that chaotic, as collinear blocks of less than 5 genes can also be found. On the other hand, robust rearrangements can affect not only locally but also the entire genome, thereby inducing changes in regions that are not affected by rearrangements directly (Avelar et al., 2013).

### *3.8. Possible reasons behind the stable gene order of the fission yeasts*

Several factors may be involved in the maintenance of the stable gene order of the fission yeasts. One possible reason is the reproductive isolation potentially caused by rearrangements. Large-scale rearrangements such as interchromosomal translocations can lead to

chromosomes of different sizes, which can have negative effects on the sexual cycle (e.g., inappropriate chromatid pairing at meiosis) (Avelar et al., 2013; Zanders et al., 2014; Naumov et al., 2015; Jeffares et al., 2017). Regardless of the fact that the new structural variants that are formed may have advantages under certain environmental conditions, the inability to produce viable offspring can be a major disadvantage in a constantly changing habitat, especially for haplontic species (Avelar et al., 2013; Zanders et al., 2014; Naumov et al., 2015; Jeffares et al., 2017). Furthermore, several studies show that genetic changes induced by the sexual cycle are more conducive to adaptation than spontaneous mutations or rearrangements (Takouridis et al., 2015; McDonald et al., 2016; Scheuerl & Stelzer, 2017; Tusso et al., 2019). In addition to the reproductive isolation induced by rearrangements, the fact that nearly 90% of the genome sequence of *S. pombe* is functional may also act against rearrangements (Gretch et al., 2019). Thus, selection can also act against changes in regions with function, as loss of function can be a long-term disadvantage, with some changes even being lethal (Avelar et al., 2013; Zanders et al., 2014, Naseeb et al., 2016). Comparison of the estimated number of rearrangement events in the fission yeast and budding yeast genomes also showed that, overall fewer changes occurred in fission yeasts genomes regardless of the similar divergence times.

Another reason may be that neighbouring genes are under common transcriptional regulation, possibly forming functional clusters. This is a fairly common phenomenon in higher eukaryotic organisms, but interestingly, such functional clusters are prone to rearrangements, giving rise to the possibility of new functional groups (Hurst et al., 2004; Poyatos

& Hurst, 2006; Liu & Han, 2009; Al-Shahrour et al., 2010; Dávila López et al., 2010; Noble & Andrianopoulos, 2013). Consequently, functional clustering cannot always explain long-term gene order conservation. However, Tuller et al. (2009) were unable to detect clear functional clustering in the *S. pombe* genome. The same conclusion can be drawn from our results. Although we clearly showed that genes with specific functions (GO categories) tend to cluster in aLCBs, the clustering was not function-related, as genes belonging to the same GO category were rarely found within the same aLCB and the co-expression of these genes did not show higher values compared to the co-expression of genes localizing in other regions. Perhaps the 3D conformations of chromosomes could reveal functional clustering, but such studies are beyond the objectives of this PhD study (Tanizawa et al., 2010; Gong et al., 2015). Instead, our data suggest that GO categories with a higher density of essential genes and with a lower abundance of Rec12 cleavage sites nearby are inclined to cluster to the aLCBs. These results are in agreement with those of Poyatos & Hurst (2006), who found that the local density of essential genes and recombination hotspots may also contribute to the maintenance and variation of gene order in budding yeast genomes.

## 4. Summary

It is now fully accepted that the eukaryotic gene order is not random. A large body of evidence supports this assumption, such as 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 various yeast model organisms are well suited to investigate the factors that maintain the conservation of gene order, as they have small, compact genomes and, due to their large numbers and short generation times, selection can have a sufficient impact on their evolution.

In the present PhD study, we have investigated the genome sequences of the model genus fission yeasts (*Schizosaccharomyces*), examining the evolutionary scale changes their genomes have undergone and seeking answers to questions about the maintenance of their highly stable gene order. Our work was mainly carried out using computational methods, but we also complemented our studies with some experiments.

Since the genome of *S. cryophilus* has been sequenced but not assembled completely (Rhind et al., 2011), we first wanted to assemble the large supercontigs of this species. To this end, we used the syntenic relationships of closely related species (*S. pombe* and *S. octosporus*) and validated the assembly using different molecular biology methods (PCR and PFGE). By comparing whole genome sequences and revealing genome rearrangements, we found that *S. cryophilus* is most closely

related to *S. octosporus* based on chromosome structure, too. It was found that interchromosomal translocations were the most frequent genome rearrangements, with the number of inversions only becoming predominant with increasing phylogenetic distance. A similar trend was observed for Saccharomycotina species with similar divergence times, although in their case the changes were more numerous but affecting smaller regions.

By including *S. japonicus*, the earliest branch of the genus, we found that sequence and structural changes in the fission yeasts show a strong positive correlation, and the other three species show nearly similar rearrangement dynamics compared to the reference species. In addition, whole genome alignments also predicted that in many cases the same genomic regions of different species may have remained unchanged.

This hypothesis was confirmed when the identification of putative common orthologues of the four species revealed locally collinear blocks (aLCBs) consisting of at least 5 genes with identical order and orientation. In such aLCBs localise 40-42% of the genome's gene content (2055 genes). Our computational models suggest that these aLCBs may not only be remnants of the gene order of the last common ancestor, but are likely to be maintained by selection. However, the gene sequence of the aLCBs was found to be specific to fission yeasts based on comparative studies, as no similar aLCBs were found in the genomes of the other fungal species included in the study. Surprisingly, more adjacent gene pairs could be identified in the genomes of *P. murina* and filamentous fungi than in the genomes of budding yeasts or other fungi.

The comparison between aLCBs and regions outside the aLCBs revealed that there is no significant difference in the dynamics of sequence changes (protein evolution rate and intron loss/gain), so although there is a high degree of conservation in terms of gene order, the sequences themselves within aLCBs are as variable as in the other regions of the genome.

Our studies have shown that some genes with certain biological functions (GO categories) tend to cluster in aLCBs more than others. Although the 13 GO categories in question localise in aLCBs in a statistically significant manner, their location is dispersed, i.e. genes that have the same function do not cluster near each other. It is also interesting that the co-expression values for genes in aLCBs are not higher than those for genes localising in other regions. However, we found that genes in the significant GO categories have a higher proportion of essential genes and a much lower number of Rec12 cleavage sites in their vicinity. This suggests that the local density of essential genes and Rec12 cleavage sites may play a major role in the persistence of aLCBs. We believe that our results contribute to the understanding of the genome evolution of the fission yeast and thus to the phenomena involved in the conservation of the eukaryotic gene order.

## 5. Thesis

1) We have successfully assembled the *S. cryophilus* supercontigs at the chromosome level using syntenic relationships with related species and collinear gene order of pericentromeric and subtelomeric regions. Based on our bioinformatics results, we proposed the order Sc3-Sc9-Sc1 (ChrI); Sc7-Sc5-Sc8-Sc6 (ChrII); Sc4-Sc2 (ChrIII).

2) Molecular biology techniques have been used to validate the contig order determined by computer methods: PCR has been used to confirm the adjacency of certain contigs, while PFGE analysis has shown that *S. cryophilus* also has 3 chromosomes. These chromosomes differ in size from the well-known chromosome sizes of *S. pombe*. We also successfully detected an rDNA cluster on one chromosome arm.

3) We used computational methods to estimate the amount and type of rearrangements that may have shaped the genomes of closely related species of fission yeasts. 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.

4) We showed that interchromosomal translocations predominated, with the number of inversions increasing only with increasing phylogenetic distance.

5) We compared the type and amount of rearrangements between fission yeast genomes and budding yeast genomes with similar

divergence times. Although overall more rearrangement events may have occurred in budding yeasts, fission yeast genomes bore the changes affecting larger regions. Interestingly, the budding yeast genomes also showed a similar trend, with translocations initially predominating and inversions becoming predominant only as the phylogenetic distance increased.

6) In our further analyses, we included the most divergent member of fission yeasts, *S. japonicus*, and using this species as a reference, we showed that the dynamics of structural and sequence changes in fission yeast genomes show a very strong positive correlation.

7) We found that the rearrangements in the genomes of the species studied occurred very uniformly, with the other fission yeast species showing almost the same degree of chromosomal structural divergence compared to *S. japonicus*.

8) We have revealed the most conserved parts of the genomes and identified collinear regions of at least 5 genes (aLCBs). The gene order of these blocks was probably inherited from their closest common ancestor. In such aLCBs localise 40-42% of the genome's whole gene content.

9) 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.

- 10) We found that the gene order of aLCBs, although fission yeast specific, may also reflect the gene order of a filamentous ancestor, since the gene order of the aLCBs studied more closely resembles the gene order of filamentous fungi than that of budding yeast species.
- 11) 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.
- 12) 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 of these GO categories also show a scattered distribution on chromosomes, i.e. genes associated with a biological function are not adjacent in aLCBs.
- 13) The co-expression values of genes located in aLCBs are not higher than the co-expression values of neighbouring genes located in other parts of the genome, but are significantly higher than the whole genome average.
- 14) We discovered that genes of GO categories clustered in aLCBs have a higher proportion of essential genes, but fewer Rec12 cleavage sites localized around their genes.

## **6. Acknowledgements**

I am grateful to many people, but above all to God, who has guided me along this journey.

I am grateful and thankful to my supervisor, Dr. Ida Miklós for her persistent support, patience and openness, who helped me with her advice and kept my spirit alive. I would like to thank prof. Dr. Matthias Sipiczki, who helped me with his comments during our professional discussions and encouraged me to write my thesis. Many thanks to Dr. László Attila Papp, my colleague and friend, for the encouragement, professional discussions and for performing PCR experiments and scripts. I would also like to thank Dr. Gyula Gábor Batta, my first supervisor, who taught me the basics and made me like molecular biology and genetics. I would also like to thank my colleague Dr. Hajnalka Csoma, who encouraged and inspired me a lot during my doctoral work. I would like to thank my colleague Dr. Zsuzsa Antunovics for performing PFGE studies. I would also like to thank the technicians of the Department of Genetics and Applied Microbiology, Ilona Lakatos, Anita Oláh, Szilvia Struba-Hegedűs, Ilona Kiss and Anita Kovács-Bordán, who took a lot of burden off my shoulders during the completion of the project tasks outside the PhD thesis.

Finally yet importantly, I would like to thank my parents and my wife for their patience and support.

The realisation of the doctoral thesis was supported by the project grants TÁMOP-4.2.2/B-10/1-2010-0024; Tét\_12\_JP-1-2014-0031 and EFOP-3.6.1-16-2016-00022.

## 7. Hivatkozások/References

- 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
- 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
- Berbee ML, Taylor JW. Dating the evolutionary radiations of the true fungi. *Botany.* 1993;71: 1114–1127.
- 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
- 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
- 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

- 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
- 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
- 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
- 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
- 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
- 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
- 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
- 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
- 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
- Hammer Ø, Harper DAT, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 2001, 4, 9.
- 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

- 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
- 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
- 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
- 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
- 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
- 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
- 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

- Lock A, Rutherford K, Harris MA, et al. PomBase 2018: user-driven reimplementations 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
- 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
- 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? Mikrobiologiya. 2015;84(2):192-203.
- 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)
- 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
- 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
- Poyatos JF, Hurst LD. Is optimal gene order impossible?. Trends Genet. 2006;22(8):420-423. doi:10.1016/j.tig.2006.06.003

- 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
- 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
- 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
- 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
- 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
- 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
- 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
- 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
- Tesler G. GRIMM: genome rearrangements web server. *Bioinformatics.* 2002;18(3):492-493. doi:10.1093/bioinformatics/18.3.492
- 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

- 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
- 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
- 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
- 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



Nyilvántartási szám: DEENK/66/2023.PL  
Tárgy: PhD Publikációs Lista

Jelölt: Ács-Szabó Lajos  
Doktori Iskola: Juhász-Nagy Pál Doktori Iskola  
MTMT azonosító: 10038168

### A PhD értekezés alapjául szolgáló közlemények

#### Idegen nyelvű tudományos közlemények külföldi folyóiratban (2)

1. **Ács-Szabó, L.**, Papp, L. A., Sipiczki, M., Miklós, I.: Genome Comparisons of the Fission Yeasts Reveal Ancient Collinear Loci Maintained by Natural Selection.  
*J. Fungi*. 7 (10), 1-26, 2021. EISSN: 2309-608X.  
DOI: <http://dx.doi.org/10.3390/jof7100864>  
IF: 5.724
2. **Ács-Szabó, L.**, Papp, L. A., 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.  
*Sci Rep*. 8 (1), 1-12, 2018. EISSN: 2045-2322.  
DOI: <http://dx.doi.org/10.1038/s41598-018-32525-9>  
IF: 4.011

### További közlemények

#### Idegen nyelvű tudományos közlemények külföldi folyóiratban (6)

3. Plaszkó, T., Szűcs, Z., Cziáky, Z., **Ács-Szabó, L.**, Csoma, H., Gécz, L., Vasas, G., Gonda, S.: Correlations between the metabolome and the endophytic fungal metagenome suggests importance of various metabolite classes in community assembly in horseradish (*Armoracia rusticana* L., Brassicaceae) roots.  
*Front. Plant Sci*. 13, 1-34, 2022. EISSN: 1664-462X.  
DOI: <https://doi.org/10.3389/fpls.2022.921008>  
IF: 6.627 (2021)
4. **Ács-Szabó, L.**, Papp, L. A., Csoma, H., Miklós, I., Sipiczki, M.: New Strain of *Cyphellophora olivacea* Exhibits Striking Tolerance to Sodium Bicarbonate.  
*Diversity*. 14 (12), 1023-, 2022. EISSN: 1424-2818.  
DOI: <http://dx.doi.org/10.3390/d14121023>  
IF: 3.029 (2021)





5. Papp, L. A., **Ács-Szabó, L.**, Pólska, 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.* 67 (5), 823-830, 2021. ISSN: 0172-8083.  
DOI: <http://dx.doi.org/10.1007/s00294-021-01181-4>  
IF: 2.695
6. Papp, L. A., **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.* 67 (6), 953-968, 2021. ISSN: 0172-8083.  
DOI: <http://dx.doi.org/10.1007/s00294-021-01206-y>  
IF: 2.695
7. Csoma, H., **Ács-Szabó, L.**, Papp, L. A., Sipiczki, M.: Application of different markers and data-analysis tools to the examination of biodiversity can lead to different results: a case study with *Starmerella bacillaris* (synonym *Candida zemplinina*) strains.  
*FEMS Yeast Res.* 18 (5), 1-12, 2018. EISSN: 1567-1364.  
DOI: <http://dx.doi.org/10.1093/femsyr/foy021>  
IF: 2.458
8. Balázs, A., Batta, G., Miklós, I., **Ács-Szabó, L.**, Vazquez de Aldana, C. R., Sipiczki, M.: Conserved regulators of the cell separation process in *Schizosaccharomyces*.  
*Fungal Genet. Biol.* 49 (3), 235-49, 2012. ISSN: 1087-1845.  
DOI: <http://dx.doi.org/10.1016/j.fgb.2012.01.003>  
IF: 3.263

**A közlő folyóiratok összesített impakt faktora: 30,502**

**A közlő folyóiratok összesített impakt faktora (az értekezés alapjául szolgáló közleményekre): 9,735**

A DEENK a Jelölt által az iDEa Tudóstérbe feltöltött adatok bibliográfiai és tudományterületi ellenőrzését a tudományos adatbázisok és a Journal Citation Reports Impact Factor lista alapján elvégezte.

Debrecen, 2023.03.06.





Registry number: DEENK/66/2023.PL  
Subject: PhD Publication List

Candidate: Lajos Ács-Szabó

Doctoral School: Pál Juhász-Nagy Doctoral School of Biology and Environmental Sciences

MTMT ID: 10038168

### List of publications related to the dissertation

#### Foreign language scientific articles in international journals (2)

1. **Ács-Szabó, L.**, Papp, L. A., Sipiczki, M., Miklós, I.: Genome Comparisons of the Fission Yeasts Reveal Ancient Collinear Loci Maintained by Natural Selection.  
*J. Fungi*. 7 (10), 1-26, 2021. EISSN: 2309-608X.  
DOI: <http://dx.doi.org/10.3390/jof7100864>  
IF: 5.724
2. **Ács-Szabó, L.**, Papp, L. A., 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.  
*Sci Rep*. 8 (1), 1-12, 2018. EISSN: 2045-2322.  
DOI: <http://dx.doi.org/10.1038/s41598-018-32525-9>  
IF: 4.011

### List of other publications

#### Foreign language scientific articles in international journals (6)

3. Plaszkó, T., Szűcs, Z., Cziáky, Z., **Ács-Szabó, L.**, Csoma, H., Gécz, L., Vasas, G., Gonda, S.: Correlations between the metabolome and the endophytic fungal metagenome suggests importance of various metabolite classes in community assembly in horseradish (*Armoracia rusticana* L., Brassicaceae) roots.  
*Front. Plant Sci*. 13, 1-34, 2022. EISSN: 1664-462X.  
DOI: <https://doi.org/10.3389/fpls.2022.921008>  
IF: 6.627 (2021)
4. **Ács-Szabó, L.**, Papp, L. A., Csoma, H., Miklós, I., Sipiczki, M.: New Strain of *Cyphellophora olivacea* Exhibits Striking Tolerance to Sodium Bicarbonate.  
*Diversity*. 14 (12), 1023-, 2022. EISSN: 1424-2818.  
DOI: <http://dx.doi.org/10.3390/d14121023>  
IF: 3.029 (2021)





5. Papp, L. A., **Ács-Szabó, L.**, Pólska, 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.* 67 (5), 823-830, 2021. ISSN: 0172-8083.  
DOI: <http://dx.doi.org/10.1007/s00294-021-01181-4>  
IF: 2.695
6. Papp, L. A., **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.* 67 (6), 953-968, 2021. ISSN: 0172-8083.  
DOI: <http://dx.doi.org/10.1007/s00294-021-01206-y>  
IF: 2.695
7. Csoma, H., **Ács-Szabó, L.**, Papp, L. A., Sipiczki, M.: Application of different markers and data-analysis tools to the examination of biodiversity can lead to different results: a case study with *Starmerella bacillaris* (synonym *Candida zemplinina*) strains.  
*FEMS Yeast Res.* 18 (5), 1-12, 2018. EISSN: 1567-1364.  
DOI: <http://dx.doi.org/10.1093/femsyr/foy021>  
IF: 2.458
8. Balázs, A., Batta, G., Miklós, I., **Ács-Szabó, L.**, Vazquez de Aldana, C. R., Sipiczki, M.: Conserved regulators of the cell separation process in *Schizosaccharomyces*.  
*Fungal Genet. Biol.* 49 (3), 235-49, 2012. ISSN: 1087-1845.  
DOI: <http://dx.doi.org/10.1016/j.fgb.2012.01.003>  
IF: 3.263

**Total IF of journals (all publications): 30,502**

**Total IF of journals (publications related to the dissertation): 9,735**

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

06 March, 2023

