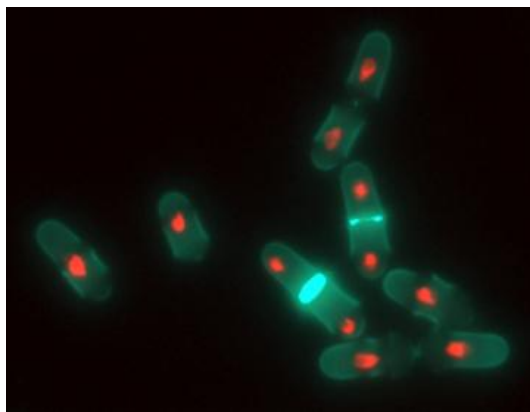


Budapesti Műszaki és Gazdaságtudományi Egyetem
Vegyésmérnöki és Biomérnöki Kar

A hasadó élesztő Sep15 sejtosztódási proteinjének filogenetikai analízise

TDK dolgozat



Készítette: **Mészáros Csilla**

IV. éves biomérnök BSc hallgató

Témavezetők: **Horváth Anna** doktoráns

BME-VBK Alkalmazott Biotechnológia és Élelmiszer-tudományi Tanszék

Dr. Sveiczter Ákos egyetemi docens

BME-VBK Alkalmazott Biotechnológia és Élelmiszer-tudományi Tanszék

Dr. Sipiczki Mátyás egyetemi tanár

DE-TTK Genetikai és Alkalmazott Mikrobiológiai Tanszék

Tartalomjegyzék

1. Bevezetés.....	3
2. Irodalmi áttekintés.....	5
2.1. A gombák életciklusa.....	5
2.1.1. A hasadó élesztőgomba sejtciklusa	8
2.1.1.1. A Sep15 protein szerepe a citokinézisben	9
2.1.2. Dimorfizmus a gombák világában.....	10
2.2. A Mediátor komplex	12
2.2.1. A Mediátor szerepe a transzkripcióban	13
2.2.2. A Sep15 protein szerepe a hasadó élesztőben	14
2.3. A filogenetika alapja: az evolúció.....	15
2.3.1. A gombák filogenetikája	16
2.4. Bioinformatikai módszerek.....	19
2.4.1. Szubsztitúciós mátrixok.....	19
2.4.2. Keresés adatbázisokban, a BLAST algoritmus	20
2.4.3. A doménszerkezet vizsgálata, a Pfam adatbázis	21
2.4.4. Többszörös illesztés.....	21
2.4.4.1. ClustalW	22
2.4.4.2. PRANK.....	22
2.4.4.3. MAFFT	23
2.4.5. Filogenetikai faszerkesztés	23
2.4.5.1. A neighbor joining módszer	24
2.4.5.2. A maximum parsimony módszer.....	24
2.4.5.3. Modellválasztás	25
2.4.5.4. A maximum likelihood módszer	26
2.4.4.5. Bootstrap analízis.....	26
3. Célkitűzés.....	28
4. Eszközök és módszerek.....	30
4.1. Adatgyűjtés	30
4.2. Ortológok keresése.....	30
4.3. Doménszerkezet azonosítása.....	33
4.4. Többszörös illesztés	34
4.5. Törzsfaszerkesztés.....	34

5. Eredmények.....	35
5.1. A Sep15 fehérje tulajdonságai	35
5.2. Ortológjelöltek keresése.....	35
5.2.1. Találatok a Taphrinomycotina altörzsben	35
5.2.2. Találatok a Saccharomycotina altörzsben	36
5.2.3. Találatok a Pezizomycotina altörzsben	37
5.2.4. Találatok a Basidiomycota törzsben.....	37
5.2.5. Találatok a Zygomycota törzsben	38
5.2.6. A külső csoport meghatározása	38
5.3. Az ortológjelöltek vizsgálata.....	38
5.3.1. Doménszerkezet azonosítása	38
5.3.2. Többszörös illesztés.....	39
5.4. Törzsfaszerkesztés.....	40
5.4.1. Neighbor joining módszer	41
5.4.2. Maximum likelihood módszer.....	45
5.4.3. Maximum parsimony módszer	48
6. Összefoglalás.....	51
7. Irodalomjegyzék.....	53
8. Alkalmazott adatbázisok és szoftverek	57
9. Függelék	58
10. Köszönetnyilvánítás	67

A borítókép forrása: European Molecular Biology Organization Workshop
(<http://cwp.embo.org/pc09-17/>)

1. Bevezetés

Minden élő organizmust sejtek építenek fel, amelyek önmagukban is élőknek tekinthetők, mivel képesek növekedni, szaporodni és anyagcserét folytatni. Az élet fenntartásának szempontjából alapvetően fontos a növekedés és az osztódás, amit a sejtek szigorúan szabályozott folyamatok sorozatával, a sejtciklussal hajtanak végre. A sejtek feltehetően egy közös őstől származnak, amelynek alapvető tulajdonságai konzerválódtak az evolúció során, így az egyszerűbb felépítésű élőlények működése és viselkedése nagyban hasonlít a magasabb rendű eukarióta sejtekéhez. Ez teszi lehetővé, hogy a kutatók a könnyebben szaporítható primitív élőlényeket alkalmazzák a kutatásaik alanyául, de az eredményekből az összetettebb szervezetek viselkedésére következtessenek (Alberts et al., 2009).

A 20. század második felében vált közkeletű mikroorganizmussá a hasadó élesztőként is ismert *Schizosaccharomyces pombe*. Kedvező fiziológiai tulajdonságai miatt gyakran alkalmazzák sejtbiológiai, genetikai és mikrobiális fiziológiai kutatásokban is. A sejtosztódás, a szeptáció, valamint a sejtszeparáció mechanizmusának megértésében nagy szerepet játszott olyan hőmérséklet-érzékeny mutánsainak izolálása, amelyek a sejtosztódásban vagy a szeptációban sérültek (Sipiczki, 2007). A hasadó élesztő jelentőségét növeli az is, hogy a teljes genomi szekvenciája elérhető 2002 óta (Wood et al., 2002), amely a genomikai kutatások alapját adhatja. Kutatásokban betöltött jelentős szerepét bizonyítja az is, hogy 2001-ben az orvosi Nobel-díjat az eukarióta sejtciklust tanulmányozó három kutatónak (Leland Hartwell, Tim Hunt, Paul Nurse) ítelték oda (Nasmyth, 2001). A kutatók egyike, Paul Nurse egész életében a hasadó élesztő sejtciklusát vizsgálta, és munkásságának legjelentősebb eredményeit is ezen a területen érte el. Eredményeik a sejtciklus szabályozásának megértése mellett a rák kezelésében teremtettek új lehetőségeket.

A gombák növekedése kétféleképpen mehet végbe: egysejtű élesztőszerű növekedéssel, amelynek során különálló sejtek képződnek; illetve fonalas növekedéssel, amely többsejtű struktúrát eredményez, mivel a sejtek nem válnak el. A külső körülmények változásához egyes gombák a két növekedési típus közötti váltással alkalmazkodnak, amit dimorfizmusnak nevezzük. Ez a jelenség elsősorban bizonyos humán patogén (pl. *Candida albicans*), illetve növény patogén (pl. *Taphrina*) fajok esetében jelentős, mivel a morfológiai átalakulás összefüggésben áll a fertőzőképesség kialakításával (Klein & Tebbets, 2007).

A sejtciklus tanulmányozásánál egyre inkább előtérbe kerülnek a bioinformatikai módszerek, amelyek olcsóbbak és kevésbé munka- és időigényesek, mint a klasszikus

módszerek. Ezeknek azonban nagy hátrányuk, hogy az eredményeik kevésbé megbízhatóak, ezért elsősorban a kísérletek tervezésében van jelentős szerepük.

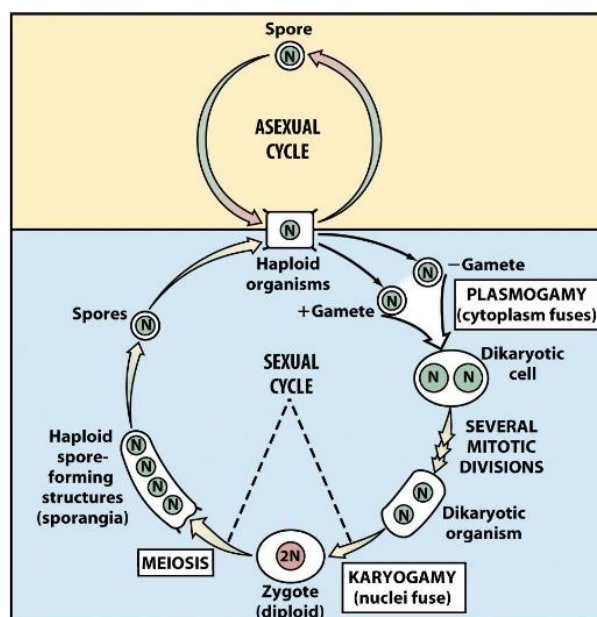
A bioinformatika igazi jelentőségét azonban az élőlények genomjának megszekvenálásából származó hatalmas adathalmaz feldolgozása és tárolása adja. A bioinformatikai eszköztár szolgál a filogenetikai kutatások alapjául, amelyek során a fajok, vagy fajokon belül egyes gének rokonságát próbálják felderíteni a genomok, illetve gének közötti hasonlóságok feltérképezésével. Az eltérő módon működő algoritmusok alkalmazásával egyre megbízhatóbb, a valóságot pontosabban leíró evolúciós leszármazásokat rekonstruálhatunk.

2. Irodalmi áttekintés

2.1. A gombák életciklusa

A gombák sejt- és telepmorfológiájukat tekintve nagyon változatos élőlénycsoportot alkotnak. Képviselőik lehetnek egysejtű szervezetek, valamint soksejtű fonalas sejteken keresztül a szövetses sejtszerveződést megközelítő álszövetses termőtesttel rendelkező nagygombák. Sejtmorfológiai szempontból a gombákat élesztő- és fonalas gombákra szokás osztani.

A gombák szaporodása végbemehet ivartalan (vegetatív) és ivaros (szexuális) szaporodással is. Rendszertani besorolásuk az életciklusuk alapján történik, amely perfekt gombák esetén az ivartalan és az ivaros szakaszokból, valamint azok váltakozásaiból áll. A vegetatív szaporodás tulajdonképpen a mitózis osztódás, amely élesztőgombák esetén sarjadzást vagy hasadást eredményez, fonalas gombák esetén pedig a hifák fejlődését, amely megnyilvánulhat a fonal hosszabbodásában vagy elágazásában. Az ivartalan szaporodás során is képződhetnek különböző szaporítóképletek, amelyek lehetnek konídiospórák vagy sporangiospórák. A szexuális ciklus három fő szakaszra bontható. A plazmogámia során két haploid sejt összeolvadása figyelhető meg, amelynek következtében egy sejten belül két haploid sejtmag (dikarion állapot) lesz jelen. A kariogámia során bekövetkezik a haploid sejtmagok fúziója, és létrejön a valódi diploid sejt. A meiózis fázisa alatt a diploid zigótából haploid kitartóképletek keletkeznek (1. ábra) (Kendrick, 2000).



1. ábra: A perfekt gombák szaporodásának egyik módszere (Black, 2012)

Az ivaros folyamat elvileg bármelyik lépés után megszakadhat, és a sejtek ivartalanul kezdenek szaporodni. Az egyes gombafajokra azonban jellemző, hogy melyik lépésben

szakad meg a folyamat, és ez alapján ötféle életciklust különböztetünk meg: haploid, diploid, dikarion, haploid-diploid és haploid-dikarion (Réczey et al., 2008).

Azonban ennek a rendszerezésnek két problémája is van. Az egyik, hogy egyes gombafajok életciklusában mindeddig nem sikerült ivaros szakaszt megfigyelni. Bizonyos feltételezések szerint az ilyen fajokban az ivaros folyamatokhoz szükséges gének inaktiválódtak az evolúciós fejlődés során, és ezáltal elvesztették ezt a képességüket. Ezeket a gombafajokat imperfektnek tekintik (Mehta et al., 2009). A rendszerezés másik problémája, hogy a vegetatív és a szexuális szakaszok közötti kapcsolat nem mindig egyértelmű. Ez redundanciát okozhat a rendszerben, mivel ugyanaz a faj két külön néven szerepelhet – egy anamorf (ivartalan) és egy teleomorf (ivaros) néven – mindaddig, amíg ki nem derül, hogy a két forma taxonómiaiilag azonos (Black, 2012).

A molekuláris tudományok fejlődése nyomán új nomenklátúra született, aminek tárgyalása túlmutat ennek a dolgozatnak a témáján. Azonban a későbbi keveredések elkerülése érdekében érdemes tisztázni, hogy az egykori tagozat helyett törzsről beszélünk, valamint esetenként altörzseket is megnevezünk. Az csak a magyar nyelv sajátossága, hogy a törzs két taxonómiai szintet is jelöl, egyrészt a királyság alatti, másrészt a faj alatti szintet.

A gombák (Fungi) alkotják az eukarióták egyik királyságát az állatok (Animalia) és a növények (Plantae) mellett. A rendszerezésnél figyelembe vett bélyegek közül a legfontosabb az életciklus típusa, de jelentősek a morfológiai, telepszerveződési, genetikai, biokémiai és citológiai tulajdonságok is. Ezek alapján a valódi gombákat öt törzsre szokás osztani: Chytridiomycota (rajzospórás gombák), Zygomycota (járomspórás gombák), Ascomycota (aszakuszos gombák), Basidiomycota (bazídiumos gombák) és Deuteromycota (imperfekt gombák). Az Ascomycota törzs három altörzsre bontható: Saccharomycotina (korábbi elnevezésén Hemiascomycetes, ami a perfekt sarjadzó élesztőket tartalmazza), Pezizomycotina (régebbi nevén Euascomycetes, ez nagyrészt fonalas gombákat foglal magában), valamint Taphrinomycotina (egykori nomenklátúra szerint Archiascomycetes) (Hibbett et al., 2007).

A legősibb gombacsoportot a Chytridiomycota törzs képviseli, amit másképpen vízi- vagy rajzospórás gombáknak is nevezünk. Ivaros szaporodásuk vízhez kötött, ami a nevüket adó rajzósejtek segítségével megy végbe. Az ezt követő ivartalan szaporodás során képződő morfológiailag változatos gombatelepek már szárazföldi életet élnek (Kendrick, 2000).

A Zygomycota tagozatba tartozó gombák már teljesen alkalmazkodtak a szárazföldhöz. Haploid életciklus jellemzi őket, az ivaros szaporodásukhoz két ellentétes párosodási típusú egyed találkozása szükséges. A haploid sejtmagokat tartalmazó

hifanyúlványokból (zigofórokból) a plazmogámia során kialakul a dikarion állapotú zigospóakezdemény. A folyamat során a zigospóra fala megvastagodik és rücskössé válik, miközben a kitartó képletben végbemegy az ivaros szaporodás másik két lépése is. Kedvező ökológiai körülmények között a zigospóra kicsírázik, belőle ún. meiosporangium fejlődik ki, abból pedig haploid meiospórák, ún. sporangiospórák képződnek, amelyek a vegetatív szaporítósejtekhez hasonlóak (Pesti et al., 2001).

Az aszkuszos gombákat egyesítő Ascomycota tagozat az ivaros szaporodás során létrejövő aszkusztól kapta a nevét, amely a kitartó képleteket (négy vagy nyolc aszkospóra) tartalmazza. A törzset három altörzssre szokás bontani, amelyek szaporodása eltérő. A valódi tömlős gombák (Pezizomycotina) osztálya haploid-dikarion életsiklust mutat. A haploid spóra kicsírázik, és pórusos szeptátságú haploid fonalak képződnek, amelyek mitózis révén osztódnak. Ha két ellentétes párosodási típusú szubsztrátmicélium párosodási szervecskéi konjugálódnak (plazmogámia), akkor létrejön a dikarion sejt, amiből mitózis ciklusok révén dikarion hifa képződik. A létrejött hifákból pórusos szeptátságú, dikarion sejteket tartalmazó légmicélium képződik, és ezután alakul ki a termőtest, ahol később a dikarion hifák csúcsi sejtjei tömlővé (aszkusz) alakulnak. Az aszkuszban végbemegy a kariogámia, aminek következtében az aszkusz diploiddá válik. Ezekből meiózissal és posztmeiotikus mitózissal spórák képződnek, így egy aszkuszban tipikusan nyolc aszkospóra jön létre endogén úton (Réczey et al., 2008).

A Basidiomycota törzsbe evolúciós szempontból a legfejlettebb gombák tartoznak. Az ellentétes szaporodási típusú hifák fúziójának eredményeként dikarion hifa képződik, amelyekben egy csatképzéses mechanizmus biztosítja, hogy sejtosztódáskor az utódok 1:1 arányban örököljék a szülői sejtmagokat. Megfelelő környezeti tényezők esetén a dikarion hifa jellegzetes termőtestet hoz létre, amelyben termőréteg alakul ki. A termőrétegben a fonalvégi sejtek bazídiummá alakulnak, amelyben lejátszódik a kariogámia és a meiózis, a két folyamat eredményeképpen pedig négy bazídiospóra képződik exogén úton (Kendrick, 2000).

Az ötödik, Deuteromycota törzsbe az imperfekt gombák tartoznak, amelyek nem tudnak ivaros szaporodni. Jelenleg vitatott, hogy ez a csoport különálló taxonnak tekinthető-e, ezért új fajokat már nem sorolnak ebbe a törzsbe (McNeill, J. et al, 2012). A molekuláris taxonómia fejlődésével egyre több anamorf fajnak találják meg az ivaros alakját, így az adott faj átkerülhet a tömlős vagy bazídiumos gombák közé (Black, 2012).

2.1.1. A hasadó élesztőgomba sejtciklusa

Hasadó élesztőknek az Ascomycota törzs, Taphrinomycotina altörzs, Schizosaccharomycetes osztály, *Schizosaccharomyces* nemzetségébe tartozó fajokat nevezzük (Hibbett et al., 2007). Négy fajt sorolunk ide: *S. pombe*, *S. japonicus*, *S. octosporus* és *S. cryophilus*. Elsőként közülük a *S. pombe*-t fedezték fel, ami a legismertebb és a legjelentősebb a négy közül (Sveiczner & Horváth, 2013).

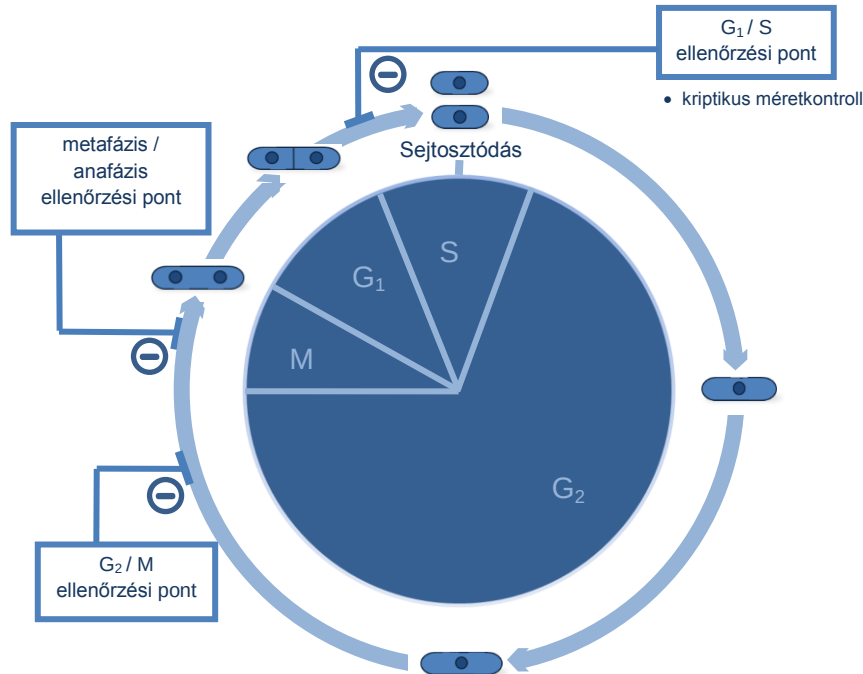
A *S. pombe*-t Paul Lindner izolálta 1893-ban egy kelet-afrikai köles sörből. A faj neve is afrikai eredetű, ugyanis a pombe az őslakosok által beszélt szuahéli nyelven sört jelent. Az 1940-es évek végén Urs Leupold használta először genetikai vizsgálatokra. Az 1950-es években Murdoch Mitchison felismerte a faj előnyös élettani tulajdonságait, és azóta a hasadó élesztőgomba modellorganizmusként szolgál a sejtbiológia, a mikrobiális fiziológia és a genetika területén (Egel, 2004).

Kedvező tulajdonságai nem csak az eukarióta sejtciklus vizsgálatára teszik alkalmassá, de a laboratóriumi munkákat is megkönnyítik. Ezek többek között a szimmetrikus osztódásban, a magasabb rendű eukariótákkal hasonlóbb fehérjékben, és a heterokromatin szerkezetben nyilvánulnak meg. Genetikai kutatásokban külön előnyt jelent a laboratóriumokban használt törzsek monolitikus származása, azaz mindegyik a Leupold által alkalmazott eredeti izolátum leszármazottja (Sunnerhagen, 2002).

A *S. pombe* ivaros szaporodása folyamán két ellentétes párosodási típusú sejt (h^- és h^+) konjugál egymással, és mivel a plazmogámia és a kariogámia szinte egyszerre megy végbe, diploid zigóta képződik. A diploid zigóta az esetek nagy részében rögtön aszkusszá alakul, amelyben meiózissal négy aszkospóra képződik. Kedvező ökológiai körülmények között az aszkospórák kicsírázásával vegetatív sejtek keletkeznek, amelyek ivartalanul osztódva tenyészeteket hoznak létre (Sipiczki, 1995). A hasadó élesztő rendelkezik ugyan diploid állapottal, ami szükséges a genetikai kutatásokhoz, de ez nem stabil. A laboratóriumokban ezért különböző eljárásokat kellett kifejleszteni, amelyekkel rögzíthető ebben az állapotban.

A hasadó élesztő sejtciklusa a magasabb rendű eukariótákéval majdnem azonos, eltérést az egyes szakaszok aránya mutat. A *S. pombe* sejtciklusa rövid G_1 , S és M fázisból áll, míg a G_2 fázis a teljes ciklusidő kb. 70%-át teszi ki, mivel alapvetően itt zajlik a sejtek növekedése (2. ábra). Az eukariótákban egy méretkontroll mechanizmus biztosítja, hogy a sejtek átlagos mérete az osztódások után, valamint az egész populáció tekintetében hosszú távon állandó maradjon. A méretkontroll teljes hiánya egyre kisebb sejteket eredményezne, ami valószínűleg a sejtek pusztulásával járna. A többi gombához hasonlóan a sejtosztódás

nem megy végbe közvetlenül a mitózis után, mert a sejtfal képzése (szeptáció) nagyjából 30-40 percet vesz igénybe. Mivel a G_1 fázis rövidebb, mint a szeptációhoz szükséges idő, ezért a két sejtmaggal rendelkező sejtek genetikai állományának replikációja megkezdődik, mielőtt befejeződik a citokinézis (Sveiczner & Horváth, 2013).



2. ábra: A hasadó élesztő sejtciklusa (Sveiczner & Horváth, 2013)

2.1.1.1. A Sep15 protein szerepe a citokinézisben

A citokinézis a sejtosztódás utolsó lépése, amely a különböző organizmusokban eltérő módon mehet végbe. Az állati sejtek a mitózis után szinte azonnal képesek osztódni a sejtmembrán befűződésével, ezzel szemben a növények és a gombák úgy osztódnak, hogy a mitózist követően új sejtfalat hoznak létre a két leánysejt között, ami időt vesz igénybe (Morgan, 2007).

Hasadó élesztőkben az új sejtfal (szeptum) három rétegből épül fel: az elsődleges szeptum fő alkotóelemei a lineáris 1,3- β -glükán és 1,3- α -glükán, míg a két oldalán elhelyezkedő másodlagos szeptumok 1,3- α -glükánból, 1,6- β elágazódásokat tartalmazó 1,3- β -glükánból, 1,6- β -glükánból és galaktomannánokból állnak. A sejtosztódás során az elsődleges szeptum és a szeptum-szegély (a szeptummal érintkező sejtfalrészlet) lebomlik, így jönnek létre a leánysejtek. Az utódsejtek sejtfalát a sejtvégen a másodlagos szeptumok képzik, amelyek összetétele nagyjából megegyezik a sejtfal többi részének összetételével (Balasubramanian & McCollum, 2004).

A *S. pombe* citokinézisében résztvevő gének aktiválódása egy kaszkád rendszeren keresztül megy végbe. Az elsőként azonosított sejtszeparációs gén a *sep1* volt, ami egy transzkripciós faktort kódol. A génben bekövetkező mutációk hatására sérül a sejtszeparáció, amelynek következtében elágazó fonalas (hifás) morfológia jön létre. A Sep1 fehérje aktiválja az *ace2* gén expresszióját, ami a folyamat következő lépésének cink-ujj transzkripciós faktorát kódolja. Hatására aktiválódnak a citokinézis lefolyásához szükséges gének, mint például az *agn1* (szeptum-szegély lebontása) és az *eng1* (elsődleges szeptum lebontása). Az *ace2Δ* mutánsok is hifát képeznek, ami a *sep1* mutánsoktól megkülönböztethetetlen. Ebből arra a következtetésre jutunk, hogy a Sep1 fehérje az *ace2* génen keresztül szabályozza a sejtszeparációt (Sipiczki, 2007).

A Sep15 fehérje a Mediátor komplex nyolcas számú alegységét alkotja, aminek felépítésére és működésére részletesebben a 2.2. fejezetben térek ki. A proteint szeptációs fehérjeként fedezték fel, így feltételezhető, hogy szerepet játszik a citokinézisben. Hőmérséklet-érzékeny mutánsait lassú növekedés, ionérzékenység, valamint részlegesen sérült sejtosztódás jellemzi. A mutáns tenyészet egyaránt tartalmaz élesztő sejteket és rövid hifákat, eszerint a protein a sejtszeparációban nem közvetlenül vesz részt (Zilahi et al., 2000). A citokinézist valószínűleg a kaszkádszabályozáson keresztül befolyásolja, ahol az Ace2 fehérjével áll fizikai kapcsolatban és kofaktorként van rá szükség a további gének aktiválásához. A fehérje az RNS polimeráz II Rpb4 alegységével is közvetlen kapcsolatban áll, amin keresztül a szabályozási információkat szállítja az Ace2 fehérje és az RNS polimeráz II között. Az Rpb4 csökkentett expressziója szintén a sejtek megnyúlásához és fonalas növekedéséhez vezet (Mehta et al., 2009).

2.1.2. Dimorfizmus a gombák világában

A dimorfizmus eredeti jelentése az ivari kétalakúság, azaz ugyanazon faj női és férfi formáinak eltérése, ami a testfelépítésében nyilvánul meg. Gombák esetében azonban a dimorfizmus alatt egyes fajok azon képességét értjük, hogy a külső körülményeknek megfelelően képesek az egysejtű élesztőszerű és a többsejtű fonalas növekedés között váltani. A jelenséget először Pasteur írta le, aki megfigyelte, hogy a *Mucor racemosus* fonalas alakja adott összetételű táptalajon szabályos artrospórákat képez, azaz különálló sejtekre esik szét a fonala, amelyek a körülményektől függően új fonalas telepet hoznak létre vagy pedig élesztőszerű növekedést mutatnak. A növekedési forma megváltozásában Pasteur szerint az oxigén parciális nyomásának változása játszott szerepet (Orlowski, 1991).

A *S. cerevisiae* képes élesztőszerű sejtek és pszeudohifa képzésére, azonban valódi hifát nem tud létrehozni. A pszeudohifa megnyúlt sejtek lánc, amely abban különbözik a valódi hifától, hogy a szeptumnál zsugorodás figyelhető meg, valamint a láncvégi sejtek mérete kisebb a többi sejthez képest. A *S. cerevisiae*-vel szemben a legtöbb di- és polimorf gomba valódi hifát képez, vagy legalábbis képes valódi hifa képzésére is a pszeudohifa mellett (Blacketer et al., 1993).

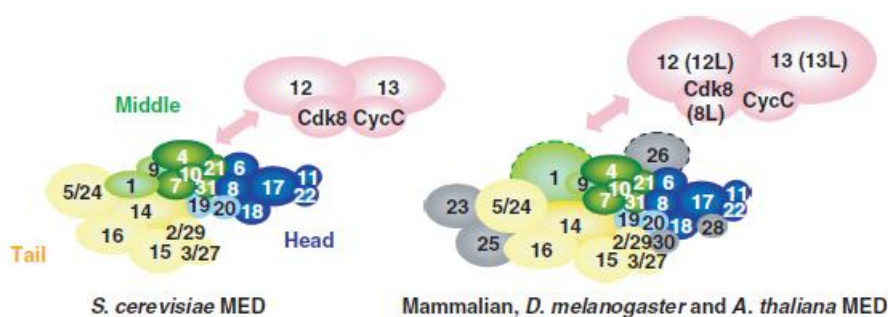
A *S. japonicus* külső ingerek hatására szintén képes váltani az élesztőszerű és a fonalas növekedés között. Folyékony közegben élesztőszerű sejteket hoz létre, míg bizonyos szilárd tápközegekben fonalas növekedést mutat. A nitrogénforrás megvonásával az élesztő sejtek valódi hifák és pszeudohifák képzésére készíthetők. Valódi hifái invazívak – ezzel biztosítja a megfelelő nitrogénellátottságot –, akárcsak egyes patogén fajok fonalas formái, azonban a *S. japonicus* nem patogén, ezért a vizsgálatok kedvelt modellorganizmusa lett (Enczi et al., 2007). A *Schizosaccharomyces* nemzetségbe tartozó másik három faj nem dimorf.

A dimorfizmus igazi jelentőségét azonban az adja, hogy egyes patogén fajok virulenciájához elengedhetetlen a váltás a két növekedési mód között. A humán patogének közé tartozó *Blastomyces dermatitidis*, *Coccidioides immitis*, *Histoplasma capsulatum* és *Paracoccidioides brasiliensis* radikális változáson megy keresztül, hogy életben maradjon a gazdatestben. Az átalakulást kiváltó inger az emlősök szervezetére jellemző hőmérséklet, ahol a nem patogén penész (fonalas) formáról a patogén élesztőre váltanak (Nemecek et al., 2006). Az átváltás a betegség kialakításához szükséges, azonban valószínűleg védelmet is nyújt a szervezet immunfehérjéivel szemben. A fenotípus váltás nem csak a sejt alakját módosítja, hanem megváltozik a sejtfal összetétele, az antigén molekulák jelenléte, valamint egyes virulenciáért felelős gének expressziója is (Klein & Tebbets, 2007). Feltehetően a *Candida albicans* és a növénypatogén *Taphrina* fajok patogenitása is összefüggésben áll a morfológiai átalakulással, csak fordítva. Esetükben a fonalas növekedés során képződött invazív hifák felelnek a virulenciáért.

A dimorfizmusban szerepet játszó gének vizsgálata egyre nagyobb jelentőséggel rendelkezik. A működés megértése a betegségek sikeresebb kezeléséhez vezethet új gombaellenes gyógyszerek megjelenésével (Nemecek et al., 2006).

2.2. A Mediátor komplex

A Mediátor egy több alegységből álló komplex, amely közvetítőként funkcionál a transzkripció egység és a szekvencia-specifikus transzkripció faktorok között. A komplexet a *S. cerevisiae*-ben izolálták először, amiben 25 alegységből épül fel. A Mediátor mind szabad formában, mind pedig az RNS polimeráz II-vel alkotott holoenzimként előfordul a sejtekben. A holoenzim vizsgálata során felfedezték, hogy a Mediátor három különböző, de nagyjából azonos méretű doménre oszlik. Ezek a globuláris polimeráz köré csavarodnak, és elhelyezkedésük alapján feji, középső, illetve farki doménnek nevezték el őket. Ez a három rész együttesen a Mediátor komplex magját adja. Azonban létezik egy negyedik domén is, ami a másik háromtól elkülönülve található, és a Cdk8 kináz modul nevet viseli (3. ábra) (Bourbon, 2008).



3. ábra: A Mediátor komplex szerkezete *S. cerevisiae*-ben és szövetes élőlényekben (Bourbon, 2008)

Az eukarióták evolúciója során a komplex szerkezete és funkciója konzerváltnak tekinthető, azonban felépítő alegységek száma és struktúrája fajonként változhat (Ansari & Morse, 2013). Ezt támasztja alá, hogy a sarjadzó élesztőt felépítő 25 alegységből 22 jellemző az eddig megszekvenált állati genomokra is, a maradék három eddigi ismereteink szerint csak a *Saccharomyces* nemzetség jellegzetessége. Azonban az emlősökben is előfordul nyolc olyan alegység, ami alacsonyabb rendű szervezetekben nem található meg. A *S. cerevisiae*-ben a komplex tíz alegységét (Med4, Med6, Med7, Med8, Med10, Med11, Med14, Med17, Med21 és Med22) a sejt életben maradásához esszenciálisnak találták. Az eddig vizsgált fajokban – legyen szó növényekről, gombákról, vagy állatokról – meg is találták az ezeket kódoló géneket. Az összehasonlító genomikai adatok alapján feltételezhető, hogy ez a tíz alegység alkotja a komplex konzervált szerkezeti vázát, és ehhez kapcsolódik az összes többi alegység (Bourbon, 2008).

2.2.1. A Mediátor szerepe a transzkripcióban

A transzkripció során a DNS-ben lévő információt RNS-re írja át a szervezet, hogy a későbbiekben fehérjét tudjon szintetizálni az elkészült transzkriptum alapján. A transzkripció elkezdéséhez az RNS polimeráznak fel kell ismernie a gén elejét, és ott erősen a DNS-hez kell kapcsolódnia. Ezeket a feltételeket biztosítja a promóter régió. Azonban az átíráshoz az eukarióta sejtekben a polimeráz önmagában nem elég, szükség van általános transzkripciós faktorokra is. Ezek biztosítják, hogy az RNS polimeráz megfelelő pozícióban legyen, valamint szétválasztják a DNS kettős hélix szerkezetét, és ezáltal szabaddá teszik a templátot. A komplementképzés mindaddig zajlik, amíg a polimeráz egy másik szignálhoz nem ér a DNS szálon, a terminátorhoz. Ebben a pontban a polimeráz leáll, és elengedi mind a kettő DNS szálát (Alberts et al., 2009).

A Mediátor nagy méretének köszönhetően egy összekötő felületként működik ebben a folyamatban: az RNS polimeráz II által katalizált transzkripció iniciálásához szükséges jeleket közvetíti az aktivátorok és az általános transzkripciós szerkezet között (Ansari & Morse, 2013). A DNS-nek azokat a szakaszait, ahová az aktivátorok kötődnek, enhancereknek nevezzük, mivel a jelenlétük jelentősen növeli a transzkripció sebességét. Az aktivátorok az enhancerektől több ezer nukleotid távolságra is képesek a transzkripció fokozására, valamint a géntől lefelé és felfelé elhelyezkedve is ki tudják fejteni hatásukat. Ennek háttérében az áll, hogy a DNS ilyen esetekben egy hurkot hoz létre, aminek a közepén a Mediátor helyezkedik el. A komplex kapcsolatot teremt az enhancer régióhoz rögzített aktivátor protein és a promóterhez kapcsolódott transzkripciós egység között. Ez a struktúra lehetővé teszi, hogy az aktivátorok közvetlenül befolyásolják a promóternél végbemenő folyamatokat akkor is, ha fizikailag távol vannak egymástól (Alberts et al., 2009). A kutatások alapján elsősorban azoknak a géneknek az aktivitása függ a farki modultól, amelyek rendelkeznek konszenzus TATA elemekkel a promóterükben. Ezzel szemben azok a gének, amelyek ilyen régiót nem tartalmaznak, főleg a feji és középső modulon keresztül kapcsolódnak a komplexhez. A kapcsolat mindkét esetben kialakulhat közvetlenül, valamint közvetetten, amikor az általános transzkripciós faktorok játsszák a központi szerepet (Ansari & Morse, 2013).

A Mediátor azonban nem csak a gének aktiválásában, hanem azok repressziójában is szerepet játszik a Cdk8 kináz modul egységen keresztül. Méreténél fogva elegendő nagyságú felülettel rendelkezik, hogy egyszerre több transzkripciós szabályozó elemmel is kölcsönhatásba kerüljön. Az aktivátorok és a represszorok mellett kofaktorokkal is interakcióba lép. Ezeket két csoportra lehet osztani: a kromatin-módosító faktorokra (hisztón-

és nukleoszóma-módosító enzimek csoportja), valamint általános kofaktorokra (stimulálják és szabályozzák a transzkripciós komplex összeállását). A transzkripció iniciálása mellett feltehetőleg szerepe van az elongációban és a terminációban is (Ansari & Morse, 2013).

2.2.2. A Sep15 fehérje szerepe a hasadó élesztőben

A *Schizosaccharomyces pombe*-ban a Sep15 protein volt az első azonosított Mediátor homológ, ami a komplex nyolcas számú alegységét alkotja, ezért gyakran Med8 fehérjeként is utalnak rá a szakirodalomban. Ez a *S. cerevisiae*-ből izolált alegységgel szerkezetileg nem teljesen egyezik, ugyanis nem tartalmazza a befejezetlen leucin-zipzárt, amelynek hiánya feltételezhetően funkcióbeli eltérésekben is megnyilvánul (Zilahi et al., 2000).

A rendelkezésünkre álló adatok szerint a *S. pombe* Mediátorának feji doménje lépcsős struktúrát mutat: a legszélén a Med27 helyezkedik el, ami a Med20-hoz kötődik, az a Med18-hoz kapcsolódik, ami pedig a Med8-cal áll kapcsolatban. Ezen kívül, amíg a *S. cerevisiae*-ben a farki modult hat alegység építi fel, addig az elektronmikroszkópos felvétel szerint az *S. pombe* esetében csak egy alegység, a Med15 van jelen (Linder et al., 2008).

A Sep15 proteint kódoló génben bekövetkező változások nem csak az előző fejezetben említett hifás növekedést mutató fenotípussal társulnak. Az egyetlen ismert mutáns esetében a második intron/exon határon bekövetkezett pontmutáció hatására sérül a splicing hatékonysága. Ennek következtében a transzkriptum jelentős része hasítatlan állapotban marad, azaz a génből nem vágódik ki az intron. Az intronban pedig egy stop kodon található, így a mutáns szekvencia alapján szintetizált fehérje akár az eredeti aminosav szekvencia 80%-át is elveszítheti. Vizsgálatok során azonban azt tapasztalták, hogy kétféle fehérje képződik, azaz a mutáns sejtek is elő tudnak állítani vad típusú Sep15 fehérjéket olyan mennyiségben, ami elég ahhoz, hogy a sejtek életben maradjanak (Miklós et al., 2007).

A Sep15 fehérje a Mediátor alegységeként különböző gének transzkripcióját befolyásolja. A *sep15* mutáns vizsgálatok során azt találták, hogy a szénhidrát metabolizmusban és a meiózisban résztvevő gének transzkripciós szintje megnövekedett, míg a riboszóma biogenezisben és a transzlációban résztvevő géneké lecsökkent. Az Ace2-dependens gének közül többnek is csökkent a transzkripciós szintje, mint például az *adg1*, az *adg2* és az *adg3* géneknek, amik kimondottan a sejtseparációban játszanak szerepet, vagy az *agn1*-nek és az *eng1*-nek, amelyek a szeptum lebontásáért felelnek (Miklós et al., 2007).

2.3. A filogenetika alapja: az evolúció

A biológiai kutatások egyik célja mindig is egy olyan rendszertan megalkotása volt, amely a természetes leszármazási kapcsolatrendszert tükrözi (Doolittle, 1999). Az első elfogadott rendszertan Carl Linné, 18. századi botanikus nevéhez fűződik, aki a hét szintből álló hierarchikus taxonómiai rendszerét a külső bélyegek alapján állította fel, azaz a morfológiai, strukturális, biokémiai hasonlóságokat vette alapul. Ez azonban a rendszerezés legnagyobb gyengeségét adja, mivel szubjektívvá teszi azt.

Charles Darwin híres könyve (A fajok eredete) ennek ellenére részben alátámasztotta Linné megállapításait, mivel a rendszerezés alapját adó bélyegek hasonlósága az evolúció miatt jöhetett létre. Darwin tehát megállapította, hogy minden faj egy közös őstől származik, csak azt nem tudta, hogyan. Elképzelése szerint az evolúció a természetes kiválasztódás révén megy végbe, amelyhez az alábbi négy feltételnek kell teljesülnie:

1. Minden generációban több utód születik, mint amennyit a környezet el tud tartani, ennek következtében az utódok egy része még azelőtt elpusztul, hogy saját utódokat nemzzen.
2. A populáció egyedei eltérnek a tulajdonságaikban.
3. Az eltérések egy része örökölhető különbségekből adódik.
4. Azok az egyedek, amelyek kedvező tulajdonságokkal rendelkeznek, nagyobb eséllyel maradnak életben és szaporodnak, mint azok, amelyeknek kevésbé kedvezőek a tulajdonságaik.

Ha mind a négy feltétel beteljesül, akkor a kedvezőbb tulajdonságú egyedek fognak elszaporodni a többiek rovására. Ez a természetes szelekció (Darwin, 2011).

Darwin megállapítása alapján a filogenetikai rendszerezés az egyetlen, amely a természetes kapcsolatrendszert tükrözi. Ennek alapját a rokonság, az evolúciós leszármazás képezi. A rendszertan kialakításához molekuláris filogenetikát alkalmaznak, amelynek két előnye is van. Egyrészt mikrobák esetében a származást főleg az örökítőanyag molekuláris szekvenciájára tud támaszkodni, mivel például a prokarióták nem is rendelkeznek komplex morfológiával és viselkedéssel, így hasonló bélyegek alapján nem – vagy csak nehezen – sorolhatók be. Másrészt a molekuláris szekvenciák határoznak meg minden organizmust, a gének kódolják az élőlények összes tulajdonságát (Doolittle, 1999).

A kezdetekben (1965-77) a mikrobiális filogenetika fehérje szekvenciákat használt markerként, leginkább ferredoxinokat és citokrómokat. Az 1970-es évek közepén Woese és társai egy adatbázison kezdtek dolgozni, amely a riboszóma nagy- és kis alegységeiben lévő

rRNS szekvenciákat tartalmazta. Az rRNS több szempontból is jobbnak bizonyult, mint a fehérje szekvencia: nagy mennyiségben fordul elő, vannak gyorsan és lassan változó régiói, ráadásul a struktúrája általánosan konzervált (Doolittle, 1999).

A törzsfák szerkesztése során leggyakrabban géneket (pl. a fent említett rRNS géneket) használnak. A közös őstől származó géneket homológoknak nevezzük. Ezen belül még két csoportot különböztetünk meg. Az ortológok különböző fajokban előfordul és közös őssel rendelkező gének, amelyek a divergens evolúció (fajképződés) során alakultak ki. Ezzel szemben a paralógok génduplikációval jöttek létre, így ugyanannak a fajnak a genomjában fordulnak elő. Míg az ortológok esetében a két gén azonos vagy hasonló biokémiai funkciót lát el, a paralógoknál az új gén eltérő funkcióval rendelkezhet (Zvelebil & Baum, 2008).

Az elmúlt években egyre inkább az egyszerre több gént, vagy akár a teljes genomot vizsgáló módszerek kerültek a középpontba, amelyek pontosabb eredményeket adhatnak a fajok filogenetikájáról. Ugyanakkor molekuláris biológiai kutatások szempontjából a gének (fehérjék) evolúciójának rekonstruálása is nagy jelentőséggel bírhat. A két módszer között azonban van egy jelentős különbség. A géneken alapuló törzsfa csak a gén evolúciós kapcsolatait ábrázolja, így nem egyezik meg a fajok valós leszármazásával. Ennek hátterében az áll, hogy a gének a fejlődés során eltűnhettek, vagy olyan mértékű átalakuláson mehetek keresztül, hogy már nem hasonlítanak kellő mértékben az eredeti szekvenciára. A másik jelentős problémát a laterális géntranszfer okozza, amely során a genetikai információ távoli rokonok között áramlik. Ezzel szemben a genomi szekvenciára épülő törzsfák a fajok valós leszármazását tükrözik, vizsgálatuk ezért kerülhetett előtérbe.

2.3.1. A gombák filogenetikája

A gombák több szempontból is fontos élőlények. Nem csak az iparban jelentős a szerepük, de gyakran alkalmazzák őket modellorganizmusként, valamint vizsgálatuk fontos a humán- és növénypatogén tulajdonságaik miatt is. Az elmúlt években egyre nagyobb hangsúlyt kapott a filogenetikai analízisük is.

A Taphrinomycotina altörzset eredetileg rRNS szekvencia alapján alkották meg, és olyan fajokat soroltak ide, amelyeknek a taxonómiai hovatartozása kérdéses volt. Ezáltal egy nagyon változatos altörzs jött létre, ami tartalmazza a *Schizosaccharomyces* nemzetséget, azaz a hasadó élesztőket; a növénypatogén *Taphrina* nemzetség tagjait; az anamorf *Saitoella* nemzetséget; a *Neolecta* nemzetséget, aminek tagjai egyedi típusú termőtestet képeznek; és a *Pneumocystis* nemzetségbe tartozó egysejtű tüdőpatogéneket. Az rRNS szekvenciát alkalmazó vizsgálatok azonban nem támasztották alá a csoport monofiletikus eredetét, vagyis

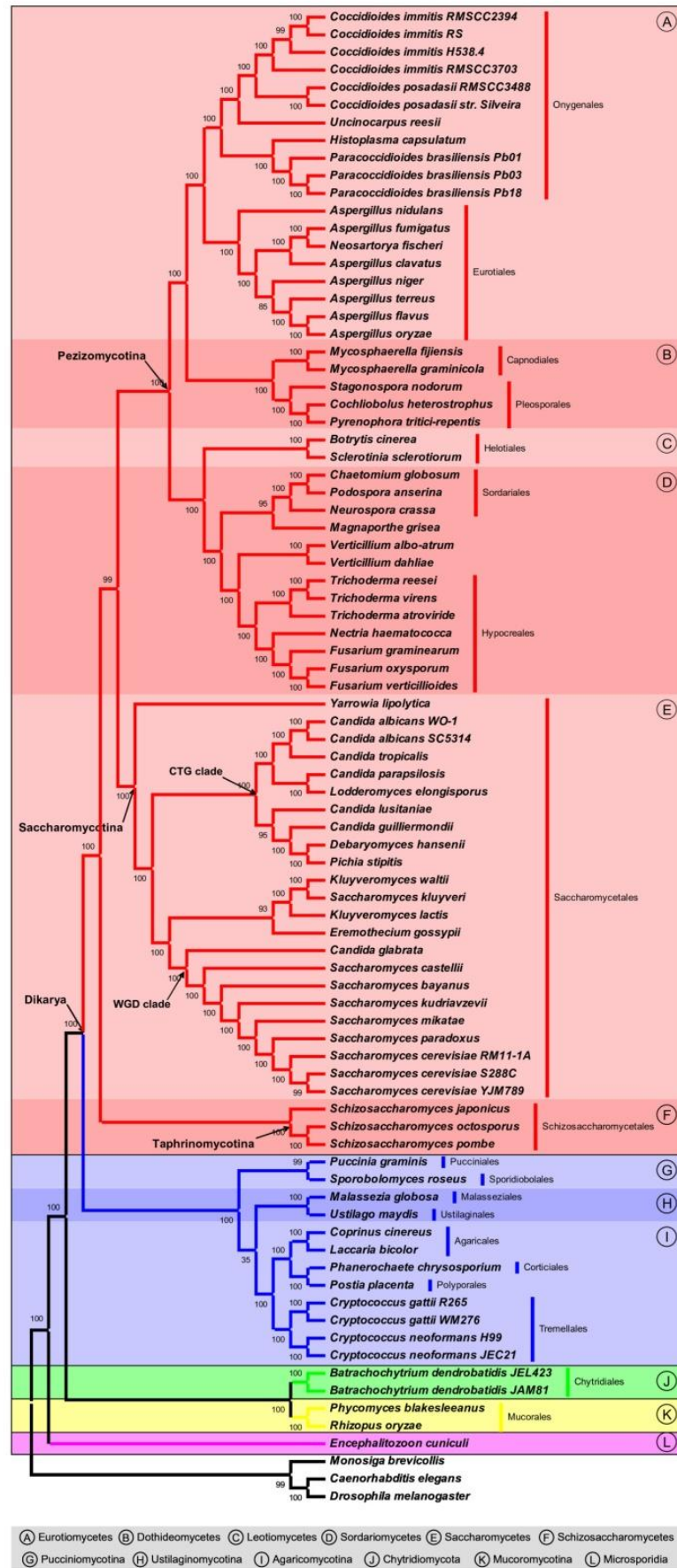
hogy a besorolt fajok egy közös őstől származnak. Liu és társai 113 sejtmagi fehérje szekvenciájából és 13 mitokondriális fehérje szekvenciájából készítettek törzsfákat, és arra a megállapításra jutottak, hogy a Taphrinomycotina altörzs a kezdeti elképzelésekkel ellentétben monofiletikus (Liu et al., 2009).

2009-ben Wang és társai egy páratlan, az egész királyság evolúcióját ábrázoló törzsfa szerkesztésére vállalkoztak (Wang et al., 2009). Az általuk vizsgált 82 faj a gombák királyságának minden fontosabb képviselőjét tartalmazta, és gének helyett teljes genomi szekvenciákkal dolgoztak. A törzsfák szerkesztéséhez egy szofisztikált, ún. CVTree (composition vector tree) módszert alkalmaztak. A filogenetikai fán (4. ábra) a törzsek és az altörzsek látványosan elkülönülnek, így a törzsek közötti evolúciós kapcsolatokról megbízható képet kapunk. Mivel a Chytridiomycota és a Zygomycota törzseket csak egy-két faj képviseli, nincs meg az a változatosság, ami a fejlődéstani következtetések levonásához szükséges.

A *Schizosaccharomyces* nemzetség elhelyezkedése a törzsfán máig vitatott. Az élesztőgombák ugyanis sok olyan tulajdonsággal rendelkeznek, amelyek elsősorban a magasabb rendű gombák jellegzetességei (pl. a gombasejtekre jellemző sejtfal, zárt mitózis), a hasadó élesztő fehérjéi viszont jobban hasonlítanak az emlősökben található homológokra, mint a *S. cerevisiae*-ben lévő proteinek. Ez annak tudható be, hogy a hasadó élesztők valószínűleg korán leváltak a fejlődés során, valamint a gombák gyorsabban változnak, mint a növények vagy az állatok. A különböző gombafajok is eltérő sebességgel változnak, ezzel lehet magyarázni, hogy az *S. cerevisiae* távolabb áll az állatoktól, mint a *S. pombe* (Sipiczki, 2000).

Feltételezések szerint az összes aszkuszos gomba egy fonalas őstől származik. A *S. japonicus* tulajdonsága, hogy az élesztőszerű és a fonalas növekedést váltakoztassa, arra utalhat, hogy a hasadó élesztők is fonalas őstől származnak. A *S. japonicus* valószínűleg hamarabb vált le, mint ahogy az egysejtű morfológia dominánssá vált volna (Sipiczki, 2004).

2012-ben a Debreceni Egyetem Genetikai és Alkalmazott Mikrobiológiai Tanszékén működő kutatócsoport a gombák szeptációjában résztvevő *ace2* és *sep1* gének filogenetikai fáját készítette el az aszkuszos gombákra szűkítve. Noha csak a gének evolúcióját ábrázolták, az eredményeik alapvetően megegyeznek a Wangék által bemutatott leszármazásnak (Balazs et al., 2012). A Budapesti Műszaki és Gazdaságtudományi Egyetemen működő Sejtciklus és Genomika Kutatócsoportban Vida Nóra az elsődleges szeptum lebontását végző Eng1 és Eng2 proteinek törzsfáját készítette el, és eredményei szintén a Wang és munkatársai által vázolt evolúciót tükrözték (Vida, 2012). A kutatócsoport a szeptum-szegély bontásában



4. ábra: A gombák filogenetikai fája teljes genomi szekvencia alapján (Wang et al., 2009)

résztevő Agn1 és a sporulációban szerepet játszó Agn2 fehérjéket is vizsgálta, amelyeknek filogenetikai fát Horváth Anna szerkesztette meg. Az ő eredményei több szempontból is eltértek a teljes genom szekvenciákból készült törzsfától. A *Saccharomycotina* altörzsben nem talált ortológokat, amit a sejtfalak különböző összetétele magyarázhat; valamint a *Schizosaccharomyces* nemzetség fehérjéi a bazídiumos gombákban talált ortológokhoz állnak a legközelebb a törzsfán. Ez ugyan ellentmond a rendszertani besorolásnak, azonban a sejtfalak összetétele alátámaszthatja az eredményt (Horváth, 2013).

Az általam vizsgált Sep15 fehérje evolúciós leszármazásáról nem találtam irodalmat. Azonban a konzervált struktúrája és funkciója miatt valószínűleg minden eukariótában esszenciális, ezért a fajok valós leszármazásához hasonló topológia valószínűsíthető.

2.4. Bioinformatikai módszerek

A filogenetikai elemzések a szekvenciák vizsgálatán alapulnak. Attól függően, hogy honnan származik a szekvencia, és milyen információt szeretnénk megtudni róla, eltérő szempontok alapján elemezhetjük. Különböző szekvenciákat illeszthetünk annak érdekében, hogy hasonlóságot találjunk közöttük, illetve a segítségükkel különböző adatbázisokban homológokat kereshetünk. De alkalmazhatóak a hasonlóság mértékének meghatározására és fehérjecsaládok felfedezésére is (Zvelebil & Baum, 2008).

2.4.1. Szubsztitúciós mátrixok

Két homológ szekvencia lehet annyira eltérő, hogy első ránézésre nem is tűnnek hasonlóknak, ezért a hasonlóság megállapításához először az optimális illesztést kell megtalálni. Azonban a lehetséges illesztések száma nagy, és ahhoz, hogy a legjobbat találjuk meg, szükség van egy pontozási rendszerre. Ezáltal a különböző illesztések egymáshoz mérhetőek, és a legmagasabb pontszámmal rendelkező lesz az optimális (Zvelebil & Baum, 2008).

A pontozási rendszer egyrészt minden reziduum (nukleotid vagy aminosav) párhoz egyedi értéket rendel, másrészt az inzerciókat és deléciókat büntetőpontokkal illeti. Mindkét reziduum vizsgálata során szubsztitúciós mátrixokat érdemes alkalmazni, amelyek fehérjék esetében minden aminosavpárhoz egy értéket rendelnek. A pontozási rendszerek általában egy 20x20-as mátrixot használnak alapul, amely a húsz aminosavat képviseli. A mátrix celláiban előforduló értékek arányosak annak a valószínűségével, hogy a mutáció eredményeként egy adott aminosav egy másik adott aminosavra cserélődik. Ha a két aminosav hasonló tulajdonságokkal rendelkezik (polaritás, méret, stb.), akkor azt a mátrix magasabb

pontszámmal illeti, mint az eltérőeket. A két leggyakrabban alkalmazott szubsztitúciós mátrix típus az extrapolációt alkalmazó PAM és a tapasztalati eredményekre épülő BLOSUM (Wheeler, 2003).

Az inzerciókat és deléciókat büntetőpontokkal kezelik, hiszen korlátlan mennyiségű szünet (gap) beillesztésével minden szekvenciaillesztésnél elérhető lenne a 100%-os azonosság. Filogenetikai szempontból egy aminosav törlődése vagy beépülése kevésbé jellemző, mint a hosszú szünetek jelenléte, ezért a szünetek megjelenéséért nagyobb büntetőpont jár, mint egy már meglévő szünet bővüléséért (Zvelebil & Baum, 2008).

2.4.2. Keresés adatbázisokban, a BLAST algoritmus

A szekvenciák keresése adatbázisokban többféle célt is szolgálhat. Egy adott DNS szekvenciával egy új genomban azonosíthatjuk a gént, fehérje szekvenciával pedig a lehetséges funkciót kereshetjük.

Az adatbázisok egyre növekvő mérete miatt a kereséshez használt algoritmusok mind valamilyen egyszerűsítést, heurisztikát alkalmaznak, hogy az illesztések megvalósíthatók legyenek. A legelterjedtebb algoritmus a BLAST (Basic Local Alignment Search Tool), amely egyszerűsítésként úgynevezett szavakra (words) bontja a szekvenciát. Ezt követően ezekhez keresi meg az összes olyan szót az adatbázisban, amely egy adott határérték (threshold value) feletti pontszámmal rendelkezik az alkalmazott szubsztitúciós mátrix alapján. Fehérjeszekvenciák esetében a szavak általában három aminosav hosszú részeket, mert ezt már viszonylag gyorsan tudja kezelni, de még nem jár jelentős adatvesztéssel (Altschul et al., 1990).

A találatoknál ellenőrizni kell, hogy valóban evolúciós kapcsolat van-e a két szekvencia között. Ezt különböző statisztikai vizsgálatokkal támasztja alá az algoritmus, amelynek során az optimális illesztést jellemző pontszámot (raw score) normalizálja. A normalizált pontszám a bit score, ami már független a keresési paraméterektől. Az adatbázis mérete és a bit score együtt határozza meg az ún. E-értéket (expect value), amely azoknak az illesztéseknek a száma, amelyek nagyobb vagy egyenlő bit score értéket adnak egy véletlenszerű keresés során. Kis E-értékek esetén ez közelítőleg azt írja le, hogy mekkora annak a valószínűsége, hogy a hasonlóság a véletlen műve, tehát minél kisebb egy találat E-értéke, annál valószínűbb, hogy a találat homológ (Zvelebil & Baum, 2008).

A program a BLAST keresés után az illesztett szekvenciák különböző paramétereit adja meg eredményként. Így a két szekvencia hosszát, azt hogy hány aminosavra illesztett, az azonos aminosavak számát és arányát (identities), a hasonló aminosavak számát és arányát

(positives), amibe beleszámítanak a teljesen egyezők is, a szünetek számát és arányát (gaps), az illesztést jellemző pontszámot (bit score), valamint az E-értéket (E-value).

2.4.3. Doménszerkezet vizsgálata, a Pfam adatbázis

A Pfam (Protein families) fehérjecsaldokat tartalmazó adatbázis, ami tulajdonképpen többszörösen illesztett konzervált régiók vagy domének szekvenciáit tartalmazza. A Pfam valójában két adatbázisból áll. A Pfam-A a kísérletekkel igazolt családokat foglalja magába, míg a Pfam-B a homológ szekvenciákból automatikusan generált fehérjecsaldokat tartalmazza (Finn et al., 2010).

Az adatbázisban minden domént jellemző többszörös illesztéshez egy rejtett Markov modell tartozik (HMM: Hidden Markov Modell). Egy HMM-t az állapotok száma, valamint a rájuk jellemző meghatározott számú állapotátmenetek jellemezik. A különböző állapotátmenetek (pl. aminosav cserék) valószínűségét megfigyelések alapján határozzák meg, és különböző valószínűséget társítanak az inzerciókhoz és deléciókhoz is. A fehérjeszekvenciák esetében ez azt jelenti, hogy minden pozícióban eltérő valószínűségértékek tartoznak az egyes aminosavakhoz. A beadott szekvenciák konzervált régióit a Pfam úgy azonosítja, hogy a rejtett Markov modelleket illeszti hozzájuk (Zvelebil & Baum, 2008).

Az algoritmus első lépésként megkeresi, hogy a szekvencia melyik részében lehet a domén (envelope), majd erre végezi el az illesztést, aminek eredményeként a valódi doménhosszt (alignment) kapjuk. A program kiadja a HMM hosszát (HMM length), valamint azt, hogy a beadott szekvenciát a HMM mekkora szakaszára illesztette (HMM). Ezen kívül rendelkezik egy pontozási rendszerrel, amely a bit score-t és az E-értéket adja meg. Ezek alapján a találatok egymással összehasonlíthatók.

2.4.4. Többszörös illesztés

Minden szekvencia vizsgálat pontos illesztési módszert igényel, az illesztések során elkövetett hibák ugyanis az evolúciós információk rossz értelmezéséhez vezethetnek. A filogenetikai fák szerkesztéséhez elengedhetetlen szempontokról, mint az egyes aminosavak pozíciója, a hasonlóság és az evolúciós kapcsolat, a többszörös illesztés szolgál megfelelő mennyiségű információval. Akárcsak a BLAST esetében, a többszörös illesztésre alkalmas algoritmusok is egyszerűsítéseket alkalmaznak, hogy kezelni tudják a nagy adatmennyiséget (Baxevanis & Ouellette, 2005). A különböző algoritmusok eltérő következtetésekre juthatnak azonos szekvenciák illesztése során. Ennek hátterében leggyakrabban az áll, hogy az illesztés

során keletkező szüneteket másképp kezelik (Löytynoja & Goldman, 2008), vagy eltérő heurisztikát alkalmaznak.

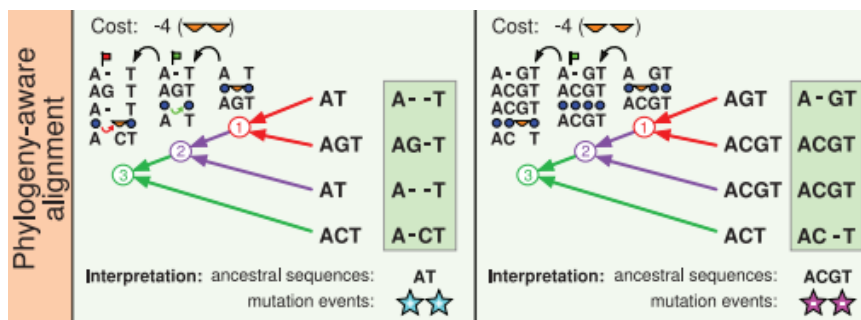
2.4.4.1. ClustalW

A leggyakrabban alkalmazott algoritmus a ClustalW, ami egy progresszív módszer. Az algoritmus először az összes lehetséges szekvenciapárt párosan illeszti, majd csoportanalízist végez, és egy távolságmátrixot számol. Ebből a 2.4.5. fejezetben tárgyalt neighbor joining módszerrel egy irányító fát (guide tree) hoz létre, ami alapján a többszörös illesztést hajtja végre. A módszer legnagyobb hátránya, hogy az első illesztéseknél elkövetett hibákat a későbbiekben nem lehet korrigálni. A ClustalW egy súlyozott algoritmus, azaz a közelebbi rokon szekvenciákat kisebb szorzóval számítja, mint a távolabbi rokonokat. A súlyozásokat az irányító fa ágaiból határozza meg. Ennek célja, hogy a közeli rokon szekvenciák ne szorítsák háttérbe a távolabbi rokonokét (Zvelebil & Baum, 2008).

A ClustalW, mint minden progresszív algoritmus, egy inzerció esetén az összes többi szekvenciába szünetet illeszt, míg deléció esetén csak az adott szekvenciában keletkezik egy gap. A pontozási rendszer egy új gap megjelenéséért több büntetőpontot ad, mint egy már létező gap bővüléséért. Ezt azért alkalmazza az algoritmus, hogy az egy lépésben bekövetkezett hosszabb inzerciókat ne büntesse túl, és ezáltal az evolúciót pontosabban rekonstruálja (Löytynoja & Goldman, 2008).

2.4.4.2. PRANK

A PRANK (Probabilistic Alignment Kit) a szünetek kezelésében teljesen más alapokon nyugszik, amit törzsfajlódás-tudatos (phylogeny-aware) algoritmusnak neveznek. Ez a módszer a progresszív algoritmusokkal ellentétben különbséget tesz az inzerció és deléció között. Az algoritmus megjelöli azokat a pozíciókat, ahol korábban egy gap keletkezett, majd a rokon szekvenciák evolúciós információi alapján állapítja meg, hogy a gap inzerció vagy deléció miatt jött-e létre (5. ábra).



5. ábra: Az inzerció (bal) és deléció (jobb) kezelése PRANK többszörös illesztési módszernél (Löytynoja & Goldman, 2008)

A pozíciók megjelölése egyfelől lehetővé teszi, hogy a későbbi illesztések során „újrahasználja” őket anélkül, hogy további büntetőpontokat adna értük. Másfelől a jelölés állandósítja az inzerciót, így ahhoz a pozícióhoz a későbbi szekvenciák nem illeszkedhetnek. Ennek egyszerű a magyarázata: az inzercióval bekerült karakterek nem leszármazottjai semmilyen korábbi karakternek, így hasonlóságot sem mutathatnak. Az algoritmus ezzel biztosítja, hogy az evolúciónak megfelelő leszármazást tükrözzön (Löytynoja & Goldman, 2008).

2.4.4.3. MAFFT

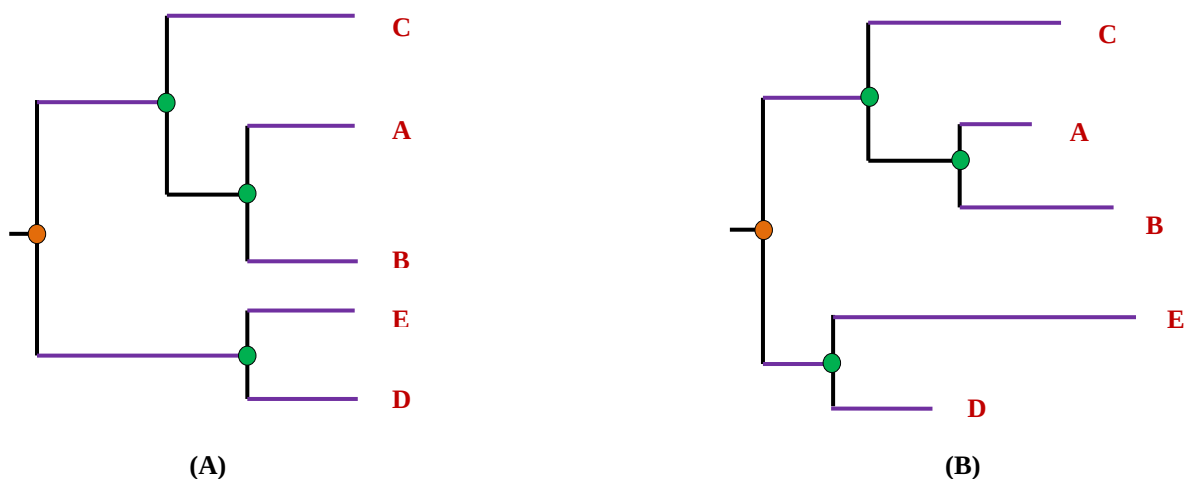
A MAFFT (Multiple Alignment East Fourier Transformation) szintén egy progresszív módszer, amit azért hoztak létre, hogy a többszörös illesztést gyorsabban végezze el, mint a többi algoritmus. Ehhez két új technikát alkalmaz. Egyrészt a homológ régiókat a gyors Fourier transzformációval (FFT: fast Fourier transformation) azonosítja. Mivel az aminosav-cserék gyakorisága az aminosavak fizikai-kémiai tulajdonságaitól függenek, ezért az FFT átalakítja az aminosav szekvenciát a legjelentősebb tulajdonságokra, azaz méret és polaritás értékekre. Ez egyben azt is jelenti, hogy a MAFFT kifejezetten szünetmentes régiókat keres. Másrészt egy egyszerűsített pontrendszert alkalmaz, ami jól alkalmazkodik a csökkentett futási időhöz, és emellett növeli az illesztések pontosságát (Katoh & Toh, 2008).

2.4.5. Filogenetikai faszerkesztés

A filogenetikai fa az evolúciós leszármazás kulcselemeit ábrázolja. A törzsfa tulajdonképpen egy diagram, amelynek a levelei vagy OTU-jai (operational taxonomic unit) általában a fajokat reprezentálják, de jelenthetnek kisebb-nagyobb taxonómiai egységeket és fehérjéket is. Az elágazások (nóduszok) a közös őseket jelölik, amiket szokás HTU-nak (hypothetical taxonomic unit) is nevezni. A leveleket és a csomópontokat összekötő ágak az evolúciós folyamatokat ábrázolják (6. ábra) (Zvelebil & Baum, 2008).

A filogenetikai analízisek során különbséget teszünk gyökereztetett és nem gyökereztetett fák között. Gyökér nélküli fa esetében ismeretlen az evolúció iránya, csak a rokonsági kapcsolatokat ábrázolja. Gyökereztetett fánál az egyik OTU-t megjelöljük gyökérként, vagyis az utolsó ismert közös ősként, így az evolúció irányát is rekonstruálhatjuk. A filogenetikai fákat ezen felül is többféleképpen ábrázolhatjuk. Ultrametrikus törzsfáknál a nóduszok az elválások sorrendjét reprezentálják, és az ághosszak nem korrelálnak az evolúciós távolsággal. Ezzel szemben additív törzsfa esetén az ághosszak az evolúciós távolságokat tükrözik (6. ábra) (Baxevanis & Ouelette, 2005).

Azonos adatsorokból kiindulva eltérő filogenetikai fákat készíthetünk attól függően, hogy milyen módszerrel próbáljuk rekonstruálni az evolúciós folyamatokat. A távolságalapú módszerek a szekvenciák közötti távolságot páronként hasonlítják össze, és abból hoznak létre egy távolságmátrixot, és ez alapján szerkesztik meg a törzsfát. A karakteralapú módszerek ezzel szemben nem távolságértékekkel operálnak, hanem közvetlenül a többszörös illesztések alapján szerkesztik meg a törzsfát (Zvelebil & Baum, 2008).



6. ábra: Az ultrametrikus (A) és additív (B) törzsfák

Az ábrán a narancssárga kör jelöli a legrégebbi közös őst, zöld körök pedig a csomópontokat (nóduszok) reprezentálják. A lila színnel kiemelt szakaszok az ágak, míg a piros betűk a törzsfá leveleit, azaz a taxonómiai egységeket szimbolizálják.

2.4.5.1. A neighbor joining módszer

A neighbor joining a távolságalapú módszerek közé tartozik. Az algoritmus alapját az a feltételezés adja, hogy minden taxonómiai egység egy csomópontból indul ki. Többszörös illesztés alapján páronként egy távolságértéket számol, és amelyik két szekvencia között a legkisebb a távolság, azt összeolvasztja egy új csomóponttá. Ezt követően ismét korrigált távolságértékeket számol, és a legközelebbi rokonokat megint egyesíti. Ez a két lépés egészen addig ismétlődik, amíg egy ággal összekötött két „levél” nem marad. A módszer azt a fát jeleníti meg eredményként, amely ágainak összege a legrövidebb. A módszer nagy előnye, hogy gyors és egyszerű. Hátránya viszont, hogy a szekvenciákban található információt távolságra konvertálja, ezért jelentős adatvesztéssel jár, valamint a karakteralapú módszerekkel ellentétben csak egy fát készít (Saitou & Nei, 1987).

2.4.5.2. A maximum parsimony módszer

A maximum parsimony a karakteralapú módszerek egyik típusa. Az algoritmus azt a fát keresi, amely egy közös őst feltételezve a lehető legkevesebb mutációval jöhetett létre. A módszer tehát nem egy modellt vesz alapul a szerkesztéshez, hanem egyszerűen megszámlálja

a végbement mutációkat. A legjobb fa megtalálásához szükség lenne az össze lehetséges fa megszerkesztésére, azonban sok szekvencia esetén ez nagyon sok időt venne igénybe, ezért az algoritmus különböző heurisztikákat alkalmaz. Ez viszont az eredményezheti, hogy a globális maximum helyett csak egy lokális maximumot talál meg a módszer (Langely & Fitch, 1974).

Az algoritmus előnyei közé tartozik, hogy az információkat nem cseréli le távolságértékekre, így nem jár jelentős adatvesztéssel, valamint több fa szerkesztéséből választja ki a legjobbat. Ugyancsak előnye, hogy nem csak a szekvenciákban tárolt információk alapján rekonstruál, hanem hozzá lehet adni új információkat (pl. külső jegyek) is, amiket szintén figyelembe vesz (Felsenstein, 1981). Azonban nagy hátránya, hogy a mutációk között nem tesz különbséget, azaz nem súlyoz a mutáció típusával, ami miatt aminosav-szekvenciák vizsgálatára kevésbé alkalmas. A módszer másik nagy hátránya, hogy nem tudja kezelni az evolúció felgyorsulását, így gyakran nem rokon szekvenciákat is közelít egymáshoz (Zvelebil & Baum, 2008).

2.4.5.3. Modellválasztás

A neighbor joining és a maximum parsimony módszerek több hátránnyal is rendelkeznek, így az elmúlt időszakban a modell alapú algoritmusok kerültek előtérbe, mint például a maximum likelihood és a bayesi módszer. Ezekhez meg kell keresni azt az evolúciós modellt, amely az adatokat a legjobban leírja. Az evolúciós modellek magukban foglalják az egyes nukleotidok vagy aminosavak közötti cserék valószínűségét, valamint a nukleotidok, ill. aminosavak gyakoriságát a szekvencián belül. Ezeknek léteznek kiegészítő paraméterei, amik a szekvencián belüli mutációs ráta variabilitását leíró gamma-eloszlást (G) és a nem-variábilis pozíciók arányát (I), valamint a különböző aminosavak gyakoriságának korrekcióját (F) veszik figyelembe (Zvelebil & Baum, 2008).

Általában igaz, hogy a több paraméteres modell jobban illeszkedik, pontosabban írja le az adatokat. Azonban előfordulhatnak olyan esetek, amikor túl kevés az adatunk, ezért a paraméterek értékét nem tudjuk meghatározni. A filogenetikai fák szerkesztése során egy ilyen túlillesztés torzíthatja a topológiát, ami téves következtetések levonásához vezethet. A túlparaméterezés elkerülése érdekében fejlesztették ki a modellszelekciós statisztikai kritériumokat, amik azt vizsgálják, hogy megéri-e hozzáadni egy új paramétert a jobb illeszkedés érdekében. Ilyen kritérium az AIC (Akaike Information Criterion) vagy a BIC (Bayesian Information Criterion).

A legjobb modell kiválasztásához minden evolúciós modellel készül egy törzsfa, amit egy maximális valószínűség értékkel (L) jellemez a program. Ezt az értéket általában $-\ln L$

formában adja meg. A két kritérium közötti különbség, hogy eltérően súlyoznak a paraméterszámmal: a BIC sokkal erőteljesebben veszi azt figyelembe, mint az AIC.

$$AIC = -2 \cdot \ln L_i + 2 \cdot p_i, \text{ ahol } i: \text{modell sorszáma}$$

L_i : az adott modellel készült fa valószínűsége
 p_i : az adott modell paraméterszáma

$$BIC = -2 \cdot \ln L_i + p_i \cdot \ln N, \text{ ahol } N: \text{az adott illesztésben a pozíciók száma}$$

A kritériumok a valószínűségi maximumot a $-\ln L_i$ minimumával keresik meg, így az a legjobb modell, aminek a legkisebb az AIC vagy a BIC értéke. Azonban minden esetben előre el kell dönteni, hogy melyik kritérium alapján akarjuk a legjobb modellt meghatározni, amelyet a különböző alkalmazások szabnak meg (Zvelebil & Baum, 2008).

2.4.5.4. A maximum likelihood módszer

A maximum likelihood egy modell alapú módszer, amely a kiválasztott evolúciós modell alapján szerkesztett legvalószínűbb törzsfát (amelynek a legnagyobb az L értéke) adja eredményül. Az L érték adja meg az adatsorunk (többszörös illesztés) valószínűségét az alkalmazott evolúciós modell és kapott topológia mellett. A módszer egy heurisztikát alkalmaz a legjobb fa megtalálásához, vagyis nem szerkeszti meg az összes lehetséges törzsfát. Így előfordulhat, hogy az eredményként kapott fa csak egy lokális maximumot ábrázol a globális maximum helyett (Felsenstein, 1981).

A módszer előnye, hogy a szekvenciákban tartalmazott információk jelentős részét felhasználja, valamint a legjobb fát statisztikai próbákkal választja ki. Hátránya viszont, hogy hatékonysága nagyon függ a kiválasztott modelltől, illetve időigényes (Zvelebil & Baum, 2008).

2.4.5.5. Bootstrap analízis

A neighbor joining, maximum parsimony és maximum likelihood módszerekkel szerkesztett törzsfák reprodukálhatóságát a bootstrap analízissel lehet vizsgálni. A módszer visszatevéses ismételt és random mintavételezéssel új adatsorokat, azaz többszörös illesztéseket hoz létre az eredetiből. Ezáltal előfordulhat, hogy egyes oszlopok teljesen kimaradnak, mások pedig többször szerepelnek. A módszert szemléltethetjük úgy, hogy egy pakli magyar kártya adatsorát a 32 lapja adja. Az új adatsor előállításához megkeverjük a paklit, majd egy lapot húzunk. A kártyát azonban nem tesszük félre, hanem csak felírjuk az értékét, és visszahelyezzük a többi közé. Ezt követően ismét megkeverjük a paklit, és egy újabb kártyát húzunk. Ennek is feljegyezzük az értékét, és visszatesszük a pakliba. Ezt

mindaddig folytatjuk, míg a feljegyzésünkben 32 érték nem szerepel. Mivel a kártyákat minden esetben visszaraktuk a pakliba, ezért előfordulhat, hogy az új adatsorban többször szerepel például a tök VIII-as, de egyszer sem került bele például a zöld felső.

Általában száz új adatsort állítunk elő, amelyek alapján törzsfákat készítünk. A száz filogenetikai fából egy konszenzus fát szerkesztünk, amelynek elágazásaihoz hozzárendeljük a bootstrap értéket. Ez az érték százas ismétlés esetén azt mutatja meg, hogy az adott elágazás hányszor fordult elő pontosan ebben az elrendezésben. Nagyobb számú ismétlés esetén a bootstrap százalékos értéknek tekintendő (Felsenstein, 1985). A szakirodalomban nincs egyetértés arra vonatkozólag, hogy a bootstrap érték pontosan mit mér. Eredetileg a reprodukálhatóság mérésére használták, manapság azonban inkább a fák megbízhatóságát jellemzik vele (Baxevanis & Ouelette, 2005).

3. Célkitűzés

TDK dolgozatom célja a *Schizosaccharomyces pombe* Sep15 (Med8) sejtszeparációs proteinjének filogenetikai analízise volt. A fehérje a transzkripcióhoz elengedhetetlen Mediátor komplex egyik alegysége, amely így a sejtszeparációban is fontos szerepet tölt be.

Munkám során célul tűztem ki, hogy körülbelül 70 gombafajban azonosítsam a Sep15 fehérje ortológjait fajok közötti BLAST keresés segítségével. Noha a gén minden eukariótában megtalálható, a keresést a gombák királyságára korlátoztam, mivel a kutatás elsősorban a szeptációs gének vizsgálatára koncentrált, ami hosszútávon a dimorfizmus megértését is segítheti. A fajok kiválasztásánál külön figyelmet fordítottam arra, hogy a fontosabb rendszertani egységeket megfelelően reprezentáljam, amelynek alapját Wang és munkatársai cikke (Wang et al., 2009) adta. A cikk megjelenése óta újabb fajokat szekvenáltak meg, így a listát ahhoz képest ki tudtam bővíteni. A kiválasztásnál fontos szempont volt, hogy az adott faj teljes genomi szekvenciája elérhető legyen az adatbázisokban, mivel részlegesen megszekvenált genom esetében nem lehetünk biztosak benne, hogy valóban az adott fehérje ortológját azonosítottuk. Ettől a kritériumtól azonban eltekintettem a munkám szempontjából fontos egy-két esetben. Ezen kívül csak azokat a fajok közötti találatokat fogadtam el ortológ jelöltnek, amelyek reciprok BLAST (visszakeresés a *S. pombe* genomjában a kapott szekvenciával) során a Sep15 fehérjét adták ki első találatként. Mivel a hasonlóságot jellemző érték függ az adatbázis méretétől, és a keresés során több adatbázist (NCBI, BROAD, JGI) is használtam, ezért standardizálni kellett az eredményeket. Ennek érdekében páros BLAST keresést alkalmaztam, amellyel páronként hasonlítottam össze a szekvenciákat.

Az ortológok pontos azonosítása érdekében megvizsgáltam, hogy a BLAST találatok doménszerkezete megegyezik-e a Sep15 fehérje doménszerkezetével. Csak azokat a fehérjeszekvenciákat fogadtam el valódi ortológoknak, amelyek egyben tartalmazzák a Med8 domént. Ehhez a Pfam-A internetes adatbázist használtam.

A filogenetikai analíziseknél nincs konszenzus a legjobb módszert illetően, ezért célul tűztem ki a különböző módszerek összehasonlítását. A valódi ortológok elfogadott találatok többszörös illesztését három különböző algoritmussal (ClustalX, PRANK és MAFFT) végeztem el. A többszörös illesztés adja a törzsfaszerkesztés alapját, ezért a több algoritmus alkalmazása pontosabb következtetések levonását eredményezi. Ezt követően elvégeztem a fehérjék törzsfájának szerkesztését. Akárcsak a többszörös illesztés során, itt is több módszert alkalmaztam a megbízhatóbb eredmények érdekében. A filogenetikai fákat a MEGA

programban neighbor joining, maximum parsimony és maximum likelihood módszerekkel is megszerkesztettem, és a fák megbízhatóságát bootstrap analízissel vizsgáltam. Annak érdekében, hogy megtaláljam az adataimat legjobban leíró többszörös illesztést és módszert, a törzsfaszerkesztéshez is különböző programokat alkalmaztam. Így a neighbor joining módszerrel ábrázolt fát PHYLIP szoftverrel, a maximum likelihood fát pedig PhyML programmal is elkészítettem.

4. Eszközök és módszerek

4.1. Adatgyűjtés

A munkámat adatgyűjtéssel kezdtem, annak érdekében, hogy a Sep15 fehérje tulajdonságait jobban megismerjem. Ahhoz, hogy minél részletesebb képet kapjak a proteinről, négy adatbázist használtam: a PomBase-t [1.], a Pfam-ot [2.], az UniProt [3.] és a BROAD Institute [4.] adatbázisát. (A szögletes zárójelek a munkám során alkalmazott adatbázisokat és szoftvereket jelölik, amelyeknek elérhetőségeit a 8. fejezetben tüntetem fel.)

Többek között utánajártam a vizsgált fehérje pontos méretének és funkciójának, az izoelektromos pontjának, a tömegének és a töltésének. De kíváncsi voltam arra is, hogy a sejtciklus alatt hogyan fejeződik ki, van-e poszttranszlációs módosulása, illetve milyen más fehérjékkel áll kölcsönhatásban.

4.2. Ortológok keresése

Munkám során nukleotid szekvenciák helyett fehérjeszekvenciákkal dolgoztam, mivel az általam végzett kutatásra sokkal alkalmasabbak. A nukleotid szekvenciákban minden mutáció nyomon követhető, de tartalmazzák a változások szempontjából jelentéktelen nem kódoló (intron) szakaszokat. Ezáltal közeli fajok evolúciós leszármazásának rekonstruálására megfelelnek, de a távolabbi fajok közötti rokoni kapcsolatok felderítésére a fehérjeszekvenciák alkalmasabbak. Emellett mivel húsz különböző elemből épülnek fel, az ortológok keresésére és a többszörös illesztésre jobban használhatók.

Az ortológok kereséséhez először a keresőszekvenciát, azaz a *S. pombe* Sep15 proteinjének aminosav-szekvenciáját kellett megkeresnem. Ezt a *S. pombe* PomBase adatbázisából töltöttem le. Az ortológok kereséséhez azonban három másik adatbázist használtam, a legelterjedtebben alkalmazott NCBI (National Center for Biotechnology Information) [5.], a JGI (Joint Genome Institute) [6.], valamint a BROAD Institute adatbázisát. Az utóbbi kettőre azért volt szükség, mert a Taphrinomycotina altörzsbe tartozó néhány faj teljes genomja nem elérhető a NCBI-ban. A JGI a *Saitoella complicata* szekvenciáját, a BROAD pedig a *Schizosaccharomyces* nemzetségbe tartozó négy fajt tartalmazza. Az egységesség érdekében a *S. octosporus*, a *S. cryophilus* és a *S. japonicus* szekvenciáját a BROAD adatbázisából töltöttem le.

Az ortológok kereséséhez a NCBI által biztosított BLASTp algoritmust alkalmaztam, amely lehetőséget ad különböző paraméterek egyéni beállítására. Manuálisan állítható a „max

target sequence”, ami a program által megjelenített találatok számát határozza meg; a „word size” állításával szabályozható, hogy mekkora egységekre darabolja a szekvenciát, ami alapján a hasonlóságokat keresi; az „expect threshold” pedig azt a felső értéket határozza meg, amelynél nagyobb E-értékű találatot nem jelenít meg. Megválasztható még ezen kívül az alkalmazott szubsztitúciós mátrix, a szünetekért járó büntetőpontok mértéke, valamint a maszkok és filterek alkalmazása is. Munkám során az alapbeállításokat alkalmaztam (1. táblázat).

1. táblázat: A BLAST beállításai

Max target sequences	100
Short queries	Automatically adjust parameters for short input sequences
Expect threshold	10
Word size	3
Max matches in a query range	0
Matrix	BLOSUM62
Gap costs	Existence:11, Extension:1
Filter	No
Mask	No

Az ortológok keresése során az *S. pombe* Sep15 fehérjeszekvenciájával fajok közötti BLASTp analíziseket végeztem. Eleinte a gombák királyságára, majd különböző gombatörzsekre, végül pedig az adott faj genomjára szűkítettem a teljes adatbázist. Ezt azért láttam szükségesnek, mert előfordult, hogy egy-egy keresés alkalmával nem jelent meg minden találat.

Mivel a keresés során több adatbázisból dolgoztam, és a hasonlóságot jellemző E-érték függ az adatbázis méretétől, ezért páros BLAST-ot végeztem el. Kereső szekvenciának minden esetben a *S. pombe* szekvenciáját használtam, a cél (target) szekvencia pedig az adott fajban talált aminosavszekvencia volt.

A keresés utolsó lépéseként reciprok (reverz) BLAST-ot hajtottam végre, amelynek során az adott faj szekvenciájával kerestem a *S. pombe* genomjában. Mint ahogy azt a célkitűzésben említettem, csak azt a találatot fogadtam el valós ortológnak, amely az *S. pombe* Sep15 proteinjét hozta vissza első találatként.

A keresés eredményeképpen vizsgált fajok listáját a 2. táblázat tartalmazza. A táblázatból kiolvasható, hogy az adott faj szekvenciáját melyik adatbázisból mentettem le, illetve jelöltem az irodalomban fontos perfekt gombák teleomorf és anamorf neveit is.

2. táblázat: A vizsgált fajok listája (taxonómiai egységenként, ABC sorrendben)

faj	adatbázis
Ascomycota – Taphrinomycotina (Archiascomycetes)	
<i>Pneumocystis jiroveci</i>	NCBI
<i>Saitoella complicata</i>	JGI
<i>Schizosaccharomyces pombe</i>	PomBase
<i>Schizosaccharomyces japonicus</i>	BROAD
<i>Schizosaccharomyces octosporus</i>	BROAD
<i>Schizosaccharomyces cryophilus</i>	BROAD
Ascomycota – Saccharomycotina (Hemiascomycetes)	
<i>Ashbya gossypii</i>	NCBI
<i>Candida albicans</i>	NCBI
<i>Candida dubliniensis</i>	NCBI
<i>Candida glabrata</i>	NCBI
<i>Candida orthopsilosis</i>	NCBI
<i>Candida tenuis</i>	NCBI
<i>Candida tropicalis</i>	NCBI
<i>Clavispora lusitaniae</i>	NCBI
<i>Debaryomyces hansenii</i> (<i>Candida famata</i>)	NCBI
<i>Eremothecium cymbalariae</i>	NCBI
<i>Kluyveromyces lactis</i>	NCBI
<i>Komagataella pastoris</i> (<i>Pichia pastoris</i>)	NCBI
<i>Millerozyma farinosa</i> (<i>Pichia farinosa</i>)	NCBI
<i>Saccharomyces cerevisiae</i>	NCBI
<i>Scheffersomyces stipitis</i>	NCBI
<i>Spathaspora passalidarum</i>	NCBI
<i>Yarrowia lipolytica</i>	NCBI
Ascomycota – Pezizomycotina (Euascomycetes)	
<i>Arthrobotrys oligospora</i>	NCBI
<i>Aspergillus flavus</i>	NCBI
<i>Aspergillus fumigatus</i>	NCBI
<i>Aspergillus kawachii</i>	NCBI
<i>Aspergillus nidulans</i>	NCBI
<i>Blastomyces dermatitidis</i> (<i>Ajellomyces dermatitidis</i>)	NCBI
<i>Botrytis cinerea</i> (<i>Botryotinia fuckeliana</i>)	NCBI
<i>Chaetomium thermophilum</i>	NCBI
<i>Coccidioides immitis</i>	NCBI
<i>Coccidioides posadasii</i>	NCBI
<i>Cochliobolus heterostrophus</i>	NCBI
<i>Exophiala dermatitidis</i> (<i>Wangiella dermatitidis</i>)	NCBI
<i>Fusarium pseudograminearum</i>	NCBI
<i>Glomerella graminicola</i> (<i>Colletotrichum graminicola</i>)	NCBI
<i>Histoplasma capsulatum</i> (<i>Ajellomyces capsulatus</i>)	NCBI
<i>Magnaporthe oryzae</i>	NCBI
<i>Myceliophthora thermophila</i>	NCBI
<i>Mycosphaerella fijiensis</i>	NCBI
<i>Nectria haematococca</i> (<i>Fusarium solani</i>)	NCBI
<i>Neosartorya fischeri</i>	NCBI
<i>Neurospora crassa</i>	NCBI
<i>Paracoccidioides brasiliensis</i>	NCBI
<i>Parastagonospora nodorum</i>	NCBI

A táblázat a 33. oldalon folytatódik

2. táblázat folytatása: A vizsgált fajok listája (taxonómiai egységenként, ABC sorrendben)

<i>Penicillium chrysogenum</i>	NCBI
<i>Pyrenophora tritici-repentis</i>	NCBI
<i>Thielavia terrestris</i>	NCBI
<i>Trichoderma reesei</i>	NCBI
<i>Tuber melanosporum</i>	NCBI
<i>Verticillium albo-atrum</i>	NCBI
<i>Verticillium dahliae</i>	NCBI
Basidiomycota	
<i>Coniophora puteana</i>	NCBI
<i>Coprinopsis cinerea</i>	NCBI
<i>Cryptococcus neoformans</i>	NCBI
<i>Dichomitus squalens</i>	NCBI
<i>Fibroporia radiculosa</i>	NCBI
<i>Fomitiporia mediterranea</i>	NCBI
<i>Laccaria bicolor</i>	NCBI
<i>Melampsora larici-populina</i>	NCBI
<i>Postia placenta</i>	NCBI
<i>Puccinia graminis</i>	NCBI
<i>Punctularia strigosozonata</i>	NCBI
<i>Schizophyllum commune</i>	NCBI
<i>Stereum hirsutum</i>	NCBI
<i>Trametes versicolor</i>	NCBI
<i>Ustilago maydis</i>	NCBI
<i>Wallemia sebi</i>	NCBI
Zygomycota	
<i>Mucor circinelloides</i>	NCBI
<i>Phycomyces blakesleeanus</i>	NCBI
<i>Rhizopus delemar (Rhizopus oryzae)</i>	NCBI
Chytridiomycota	
<i>Batrachochytrium dendrobatidis</i>	NCBI
Oomycota	
<i>Phytophthora sojae</i>	NCBI

4.3. A doménszerkezet azonosítása

A doménszerkezetet a Pfam internetes adatbázis segítségével vizsgáltam. Munkám során csak a Pfam-A adatbázist használtam, mert az tartalmazza a kísérletekkel igazolt fehérjecsaládokat. Eredeti elképzelés szerint csak azokat a találatokat fogadtam volna el ortológusnak, amelyek szignifikánsak, azonban munkám során ezen változtatni kellett. A Basidiomycota törzsbe tartozó fajok jelentős részében ugyanis az adatbázis nem találta szignifikánsnak a domént, így esetükben figyelmen kívül hagytam ezt a kritériumot. Az elfogadásnak azonban feltétele volt, hogy a fehérje megszakítás nélkül tartalmazza a Med8 domént.

4.4. Többszörös illesztés

A többszörös illesztést három különböző programmal végeztem el. Az algoritmusok másképp kezelik a szünetek büntetését, illetve eltérő heurisztikákat alkalmaznak, ezért alkalmasak az eredmények megerősítésére.

Munkám során a ClustalX [7.] programot használtam, ami a ClustalW algoritmust alkalmazza. A W az angol weighted szót szimbolizálja, mivel a program az illesztés különböző részeit eltérően súlyozza. A programban több paraméter is manuálisan állítható, amiket igyekeztem összhangba hozni a BLAST során alkalmazott értékekkel. Így az új szünetek megjelenéséért 11, míg a szünetek bővüléséért 1 büntetőpontot, az alkalmazott mátrixnak pedig a BLOSUM családot állítottam be. Aminek a BLAST-ban nem volt megfelelője, azt meghagytam alapbeállításként.

A ClustalX mellett az interneten is elérhető PRANK [8.] és MAFFT [9.] programokkal dolgoztam. Bár az internetes felületeken is lehet változtatni a paramétereken, én mindkét esetben az alapbeállításokat alkalmaztam.

4.5. Törzsfaszerkesztés

Munkám során elsősorban a MEGA [10.] programot használtam, amellyel a neighbor joining, maximum parsimony és maximum likelihood módszerekkel szerkesztett fákat készítettem el. A program lehetőséget ad a legjobb evolúciós modell kiválasztására és beállítására, valamint a bootstrap analízis alkalmazására is. A törzsfák szerkesztése során 100-as bootstrap értékkel dolgoztam. A neighbor joining fák esetén az alapértelmezett JTT (Jones-Taylor-Thornton) evolúciós modellt alkalmaztam, a maximum likelihood módszer esetén pedig mindig megkerestem a BIC kritérium alapján az adatokat legjobban leíró evolúciós modellt.

A végső fajlistából készült törzsfákat más programokkal is megszerkesztettem annak érdekében, hogy a MEGA programmal kapott eredményeket alátámasszam. A neighbor joining módszerrel készített fához a PHYLIP [11.], a maximum likelihood módszerrel készített fához pedig a PhyML programot [12.] alkalmaztam. A PHYLIP használata során is a JTT evolúciós modellt, míg a PhyML szoftver használata során a MEGA modellkeresése során kapott legjobb evolúciós modellt alkalmaztam. A MEGA programmal végrehajtott szerkesztésekhez hasonlóan ezekben az esetekben is 100-as bootstrap értékkel dolgoztam. A törzsfák megjelenítéséhez és szerkesztéséhez pedig a TreeView [13.] és a FigTree [14.] nevű programokat használtam.

5. Eredmények

5.1. A Sep15 fehérje tulajdonságai

A munkámat a vizsgált fehérje tulajdonságainak megismerésével kezdtem. A protein két néven ismert az irodalomban: Sep15-ként, mivel szeptációs fehérjeként fedezték fel a hasadó élesztőben; valamint Med8-ként, ami arra utal, hogy a Mediátor komplex nyolcas számú alegysége. Az eredmények tárgyalása során mind a kettő elnevezést használni fogom. A különböző adatbázisokból összegyűjtött fontosabb jellemzőket a 3. táblázat tartalmazza.

3. táblázat: A *S. pombe sep15* génjének és a Sep15 proteinjének főbb tulajdonságai

Gén lokalizációja	2. kromoszóma: 3219316-3222152 +
Gén hossza	2837 nukleotid
Fehérje hossza	200 aminosav
Fehérje tömege	23.3 kDa
Fehérje töltése	-7.5
Fehérje izoelektromos pontja	pH = 4.88

A fehérje kifejeződése nem mutat ciklikusságot a sejtciklus alatt, nem rendelkezik szignálpeptiddel, valamint poszttranszlációs módosulása sem ismert. A protein konzervált régiója két α -hélixet tartalmaz, amiken keresztül feltehetően a Mediátor komplex többi alegységéhez kapcsolódik. A fehérje N-terminálisa a Mediátor komplex feji doménjéhez kötődik, míg a C-terminálisa a Med18 alegységgel áll kapcsolatban.

5.2. Ortológjelöltek keresése

5.2.1. Találatok a Taphrinomycotina altörzsben

Kutatásom középpontjában a *Schizosaccharomyces pombe* áll, így az ortológjelöltek keresését a legközelebbi rokonaiban kezdtem. A többi hasadó élesztőben (*S. japonicus*, *S. octosporus*, *S. cryophilus*) megtaláltam a Sep15 fehérje ortológját, amelyek viszonylag nagymértékű hasonlóságot mutatnak. A találatokat jellemző E-értékek nagyon alacsonyak, ezért feltételezhető, hogy valódi ortológokról van szó (4. táblázat). A találatokkal reciprok BLAST-ot hajtottam végre, és minden esetben a Med8 proteint kaptam vissza első találatként.

A *Schizosaccharomyces* nemzetség vizsgálata után a Taphrinomycotina altörzs többi nemzetségében folytattam a keresést. A Sep15 fehérje ortológját a *Pneumocystis jiroveci* és a *Saitoella complicata* fajokban is megtaláltam. Ahogy a 4. táblázatban látható, mindkét ortológ nagyjából 50%-os egyezést mutat a *S. pombe* Sep15 proteinjével, azonban a *Schizosaccharomyces* fajoknál nagyságrendekkel nagyobb E-értékekkel rendelkeznek. A reciprok BLAST elvégzése során viszont itt is a Med8 fehérjét kaptam vissza első találatként.

Az altörzsbe tartozó *Taphrina* és *Neolecta* nemzetségeket dolgozatomban azért nem tárgyalom, mert jelenleg még nincs olyan fajuk, amelynek elérhető lenne a teljes genomi szekvenciája.

4. táblázat: A *S. pombe* Sep15 proteinjének ortológjai a Taphrinomycotina altörzsben

	<i>S. pombe</i>	<i>S. japonicus</i>	<i>S. cryophilus</i>	<i>S. octosporus</i>	<i>P. jiroveci</i>	<i>S. complicata</i>
Törzs	972h ⁻	yFS275	SCY4	SO6	SE6	NRRL Y-17804
Fehérje kódja	SPBC 21.04	SJAG_03622	SPOG_00748	SOCG_00147	CCJ29 994.1	lcl 36403
Fehérje hossza (as)	200	199	198	198	186	283
BLAST hasonlóság		140/193 (73%)	158/200 (79%)	153/199 (77%)	108/197 (54%)	54/97 (55%)
BLAST egyezés		110/193 (57%)	126/200 (63%)	124/199 (62%)	68/197 (35%)	33/97 (34%)
BLAST szünetek		3/193 (2%)	2/200 (1%)	2/199 (1%)	33/197 (16%)	1/97 (1%)
BLAST pontszám		222 bit (566)	256 bit (653)	236 bit (602)	111 bit (278)	60.1 bit (144)
E-érték		10 ⁻⁷⁷	8 · 10 ⁻⁹¹	4 · 10 ⁻⁸³	7 · 10 ⁻³⁵	10 ⁻¹⁵

A narancssárga háttérű fajokban talált ortológok különböző okok miatt nem szerepelnek a végső törzsfákon.

A Taphrinomycotina altörzs minden vizsgált fájában csak egy ortológot találtam, így feltételezhető, hogy génduplikáció során nem jöttek létre paralógok.

5.2.2. Találatok a Saccharomycotina altörzsben

Az ortológok keresését a Saccharomycotina altörzsben folytattam, ahová az élesztőszerű növekedést mutató aszkuszos gombák tartoznak. Munkám során 17 fajban találtam ortológot, amelyeknek részletes eredményeit a Függelék F1. táblázata tartalmazza. Egyes fajok esetében több törzsben is végeztem keresést. A *Saccharomyces cerevisiae* összes megszekvenált törzsében jelen van a fehérje, azonban csak három törzs szekvenciáját használtam fel a filogenetikai fák megszerkesztéséhez. A kiválasztott törzsek között szerepel a faj típustörzse (S288c), egy borászatban jelentős törzs (VL3), valamint egy anyagszer-mérnöki és rendszerbiológiai kutatásokban jelentős törzs (CEN.PK113-7D). A *Candida albicans* esetén is mindkét szekvenált törzsben (SC5314 és WO-1) megtaláltam az ortológot. Az SC5314 törzsben két kópiában is jelen van a protein, azonban ezek nem paralógok. A szekvenálás során egy diploid törzset használtak, ezért a törzs egész genomja duplán szerepel az adatbázisban.

A találatokra általánosságban igaz, hogy az E-értékük 10^{-20} körül ingadozik. A viszonylag magas érték oka az, hogy egy nagyon rövid - mindössze 200 aminosavból álló - fehérjéről van szó. A találatokkal reciprok BLAST-ot hajtottam végre annak érdekében, hogy megbizonyosodjak róla, valóban ortológ szekvenciákról van-e szó. Mivel minden alkalommal a Med8 proteint kaptam vissza első találatként, feltételezhető, hogy valóban evolúciós kapcsolat van a fehérjék között.

5.2.3. Találatok a *Pezizomycotina* altörzsben

A *Pezizomycotina* altörzset alkotó fonalas gombák között 27 fajban találtam meg a Sep15 fehérje ortológiát (Függelék, F2. táblázat). Azoknál a fajoknál (pl. *A. dermatitidis*, *A. fumigatus*, *P. brasiliensis*), amelyeknek több ismert törzsük van, a törzseken belül is elvégeztem az ortológok keresését. Az *A. capsulatus* esetén az öt ismert és megszekvenált törzséből csak hármat (NAM1, G186AR és H143) tüntettem fel a filogenetikai fákra. Egyrészt a G217B törzse nincs még annotálva, másrészt akárcsak a *S. cerevisiae* esetében, el akartam kerülni a fajok túltreprézntálását. A vizsgált fajok közül *P. tritici-repentis*, a *C. heterostrophus* és a *P. nodorum* esetében nem adódott ortológjelölt, aminek hátterében a nem megfelelő vagy részleges annotáció állhat. A genom annotálása alatt a gének és a kódoló régiók elhelyezkedésének, valamint a gének funkciójának azonosítását értjük. A fehérjék szekvenálása még nem számít rutin technikának, ezért a szekvenciákat csak a gének lefordításával kaphatjuk meg.

A találatok E-értékei ismét több nagyságrenddel nagyobbak, azonban a reciprok BLAST minden faj esetében a Med8 fehérjét hozta vissza első találatnak. Az *A. oligospora* viszonylag jó értékekkel rendelkezett, azonban a benne található ortológ hossza több mint háromszorosa a *S. pombe* Sep15 proteinjének, ami eléggé torzította a törzsfa topológiáját. Ezért a faj a végső fán nem szerepel.

5.2.4. Találatok a *Basidiomycota* törzsben

A bazídiumos gombák közötti keresés eredményeként 11 fajban találtam ortológot, amit a Függelék F3. táblázata foglal össze. Több fajnál (*Postia placenta*, *Cryptococcus neoformans*, *Melampsora larici-populina*, *Puccinia graminis* és *Ustilago maydis*) a keresés nem vezetett eredményre. Előfordulhat, hogy a megszekvenált genom nem megfelelően lett annotálva, vagy olyan mértékű változás következett be a fehérje szekvenciájában, hogy a BLAST már nem találja hasonlóknak a két szekvenciát.

A bazídiumos találatokra jellemző, hogy az E-értékük csak 10^{-10} körüli. Ahogy a többi esetben is, itt is mindig a Med8 proteint kaptam vissza első találatként a reciprok BLAST

során. A *C. cinerea* és a *D. squalens* szekvenciái túl hosszúnak bizonyultak, ezért a végső fán ezek a fajok sem szerepelnek.

5.2.5. Találatok a Zygomycota törzsben

A járomspórás gombák törzsében mindössze két fajban találtam ortológot, adataikat a Függelék F3. táblázata tartalmazza. A Zygomycota törzsbe tartozó fajok közül alig néhánynak érhető el a teljes genomi szekvenciája, illetve egyes megszekvenált fajok esetében (pl. *P. blakesleeanus*) nem teljes az annotáció. A találatok E-értéke alacsonyabb, mint a legtöbb bazídiumos gombában talált ortológé, valamint reciprok BLAST során mindkét esetben a Med8 fehérjét kaptam vissza első helyen. Ezek alapján valószínűsíthető, hogy a talált fehérjék a Sep15 fehérje ortológjai.

5.2.6. A külső csoport meghatározása

Külső csoportnak érdemes azt az ortológot választani, amelynek E-értéke nagyobb, mint a legtöbb belcsoportban lévő ortológé, de még elég kicsi ahhoz, hogy rokonnak tekinthessük. Először a petespórás (Oomycota) gombák között kerestem ortológokat, de az osztályt képviselő *P. sojae*-ban nem volt találatom. Ezután a Chytridiomycota törzsre szűkítettem a keresést, ahol találtam egy fajt (*B. dendrobatidis*), amiben jelen van a Sep15 protein ortológja. A találat E-értéke ugyan egy-két nagyságrenddel alacsonyabb néhány bazídiumos találatnál, azonban ez a törzsfák szerkesztése során nem torzította a topológiát. Reciprok BLAST során pedig a Med8 jött vissza első találatként, ami alátámasztja, hogy valódi ortológról van szó.

5.3. Az ortológjelöltek vizsgálata

5.3.1. Doménszerkezet azonosítása

Fontosnak tartottam, hogy a BLAST keresés során azonosított ortológjelölteknél ellenőrizsem, hogy tartalmazzák-e a *S. pombe*-ban jelen lévő Med8 domént (F4. táblázat). A *S. pombe*-ban található domén vége azért cikk-cakkos a 7. ábrán, mert a program nem tudta pontosan azonosítani, hogy hol ér véget a domén.



Significant Pfam-A Matches
[Show](#) or [hide](#) all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Med8	Mediator of RNA polymerase II transcript ...	Family	n/a	1	191	1	189	1	176	226	215.3	7.4e-64	n/a	Show

7. ábra: A *S. pombe* Sep15 fehérjeszekvenciájának azonosítása a Pfam-A adatbázisban [2.]

Egy találatot csak akkor tekintettem volna valódi ortológoknak, ha a domén szignifikánsan és megszakítás nélkül van jelen benne. Azonban a Pfam nem minősítette szignifikánsnak a domént a legtöbb bazídiumos gomba proteinjében. Ennek háttérében az állhat, hogy a Pfam adatbázisában a Med8 domént leíró rejtett Markov modellben a bazídiumos gombák alulreprezentáltak: a gombafajokból származó 117 szekvencia közül mindössze kettő tartozik a Basidiomycota törzsbe. A modell felépítésében az aszkuszos gombák mellett az állatok királysága, azon belül is a gerincesek fehérjéi játszanak fontos szerepet. Az említett két csoport szekvenciája jobban hasonlít egymáshoz, mint a bazídiumos gombákéhoz, így több Basidiomycota törzsbe tartozó faj bevonása valószínűleg torzítaná a kísérleti szempontból fontos fehérjék adatainak értékét. Ugyan a bazídiumos gombák fehérjeiben nem szignifikánsan, de jelen van a Med8 domén, azonban a domén hossza a legtöbb esetben száz aminosavnál is rövidebb (F4. táblázat). Mindezt figyelembe véve felülbíráltam a Pfam eredményét, és a végső törzsfák megszerkesztésekor a legtöbb bazídiumos gomba szekvenciáját felhasználtam.

A filogenetikai fák megszerkesztéséhez a Pfam által meghatározott doménszekvenciákat lementettem úgy, hogy a beillesztett szüneteket töröltem.

5.3.2. Többszörös illesztés

Az elfogadott ortológok többszörös illesztését három, a ClustalX, a PRANK és a MAFFT algoritmussal végeztem el. Az első két programot az összes törzsfa szerkesztése során alkalmaztam, a MAFFT algoritmust pedig a végső fák elkészítésénél használtam. A többszörös illesztésekhez az ortológok teljes fehérjeszekvenciái mellett a doménszekvenciákat is felhasználtam. A domének szekvenciái általában konzerváltabbak, aminek következtében az evolúciós változások lassabban mennek végbe bennük. Emellett általában kisebb eltérések mutatkoznak a szekvenciák hosszában, így kevesebb szünet beillesztésére van szükség. Ezeknek az előnyöknek köszönhetően pontosabb következtetéseket lehet levonni, ami megbízhatóbb eredményekhez vezet.

A WebLogo [15.] egy internetes felület, amely a többszörös illesztések szemléltetésére alkalmas. A 8. ábrán a négy hasadó élesztő, a *P. jiroveci*, a sarjadzó élesztő *S. cerevisiae*, valamint a fonalas *A. flavus* Sep15 aminosav-szekvenciáinak illesztésének részlete látható.

A logó a szekvencia minden pozícióját egyenként ábrázolja, amiről így könnyen leolvasható, hogy az adott pozíciókban milyen gyakorisággal fordulnak elő. Az aminosavakat jelölő betűk magassága az adott aminosav gyakoriságával, míg a vastagságuk a beillesztett szünetekkel arányos. A 8. ábrán a Sep15 protein egyik legkonzerváltabb régiója látható,

amely a 98. aminosavtól a 127. aminosavig tart. Egyes pozíciókban mind a hét faj azonos aminosavat tartalmaz, mint például az 103. pozícióban lévő prolin (P). Azonban még a legkonzerváltabb régiókban is előfordulnak olyan pozíciók, ahol majdnem mindegyik faj más aminosavat tartalmaz (pl. 101. pozíció).



8. ábra: WebLogo a Sep15 protein teljes aminosav-szekvenciájának egy konzervált régiójának ClustalX algoritmussal készült többszörös illesztéséről

5.4. Törzsfaszerkesztés

Annak érdekében, hogy a protein evolúciójáról a legpontosabb képet kapjuk, a törzsfákat kezdetben a többszörösen illesztett teljes fehérje- és doménszekvenciákkal is elkészítettem. Mint ahogy azt az 5.3.1. fejezetben említettem, a doménszekvencia sokkal konzerváltabb, így általában megbízhatóbb eredményeket ad. A törzsfák szerkesztése során azonban azt tapasztaltam, hogy a teljes fehérje szekvencia alkalmazásával jobb topológiát kapok. Ez nem csak a nagyobb bootstrap értékekben tükröződött, hanem az ágak elhelyezkedésében is, amelynek eredményeként többször előfordult, hogy a különböző törzsekbe tartozó gombafajok ortológjai keveredtek. Ennek az lehet az oka, hogy a Sep15 fehérje esetében a doménszekvencia hosszában sokkal nagyobb szórás mutatkozik, mint a fehérjék hosszában. Ezáltal pont a sok szünet beillesztését nem sikerül elkerülni, ráadásul a rövid doménszekvenciák nem is tartalmaznak elegendő információt a törzsfaszerkesztéshez.

A törzsfák szerkesztését ezután csak a teljes fehérjeszekvenciákkal hajtottam végre, azonban a többszörös illesztéseket a munkám során végig a ClustalX és PRANK algoritmusokkal végeztem el. A végső törzsfák elkészítése során pedig a MAFFT algoritmust is használtam. Alapvetően azt tapasztaltam, hogy a ClustalX programmal készült fák jobb bootstrap értékkel rendelkeztek, mint a másik két algoritmussal szerkesztett törzsfák; valamint az elágazások elhelyezkedése nagyjából megfelelt a Wang és munkatársai által leírt evolúciónak (4. ábra). A PRANK programmal készített illesztések leggyakoribb hibája a *Taphrinomycotina* altörzs evolúciójában jelentkezett, az algoritmus alapján a *S. pombe* Sep15 fehérjéje gyakran hamarabb vált el, mint a többi *Schizosaccharomyces* fajban talált ortológok. Ez ellentmond annak az elképzelésnek, miszerint a *S. japonicus* - szemben a többi hasadó élesztővel - azért képes váltani az élesztőszerű és a fonalas növekedési forma között, mivel

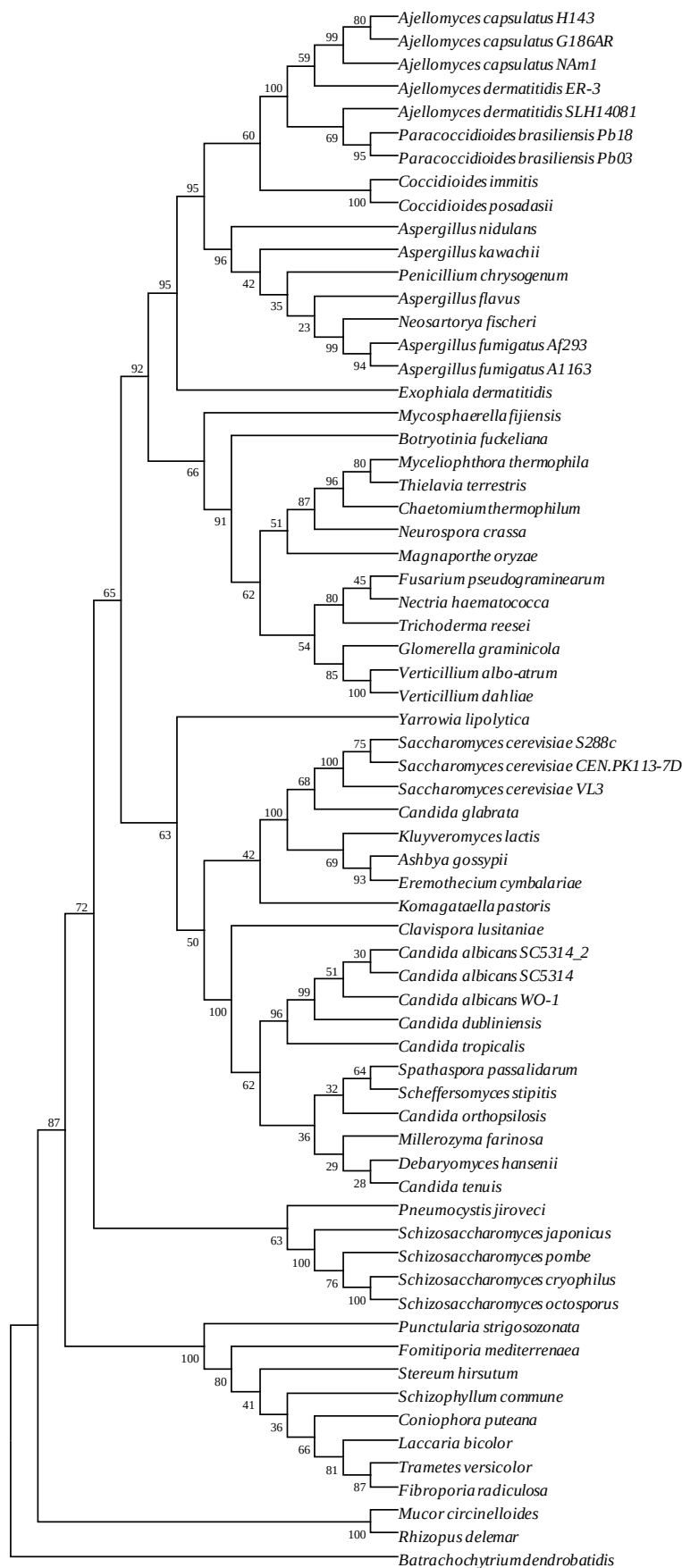
hamarabb vált el, mint ahogy a nemzetségre jellemző egysejtű morfológia dominánssá válhatott volna (Sipiczki, 2004). Egy olyan kutatásnál, amelynek középpontjában a Taphrinomycotina altörzs áll, ez nem elhanyagolható eltérés, azonban ezt leszámítva, a PRANK algoritmussal készült illesztések alátámasztják a ClustalX algoritmus többszörös illesztésén alapuló törzsfák eredményeit. A MAFFT programmal készített illesztések szinte azonos topológiájú törzsfákat adtak eredményül, mint a ClustalX algoritmus illesztéseinek alapulók, azonban kisebb bootstrap értékek szerepeltek az elágazásoknál.

A filogenetikai fák szerkesztése során több faj fehérjéje kikerült a végső listából, mivel torzították az elkészült fák topológiáját. Ilyen volt például a Taphrinomycotina altörzsbe tartozó *S. complicata* proteinje. A kutatás szempontjából fontos lett volna, hogy ez a fehérje szerepeljen a törzsfákon, hiszen a *Schizosaccharomyces* nemzetség egyik legközelebbi rokona. Azonban a protein általában rossz helyen helyezkedett el (pl. a járomspórás *R. delemar* proteinjével került egy ágra, vagy a fonalas aszkuszos gombákban talált ortológok közé ékelődött be), amivel torzította az eredményeket. Ezt okozhatta részben az, hogy a faj teljes genomi szekvenciája még nem elérhető, és az is elképzelhető, hogy a már meglévő részek nem megfelelően lettek annotálva, ezért nem valódi ortológot találtam. Egy másik problémás fehérje a *T. melanosporum*-ban talált ortológ volt, ami minden alkalommal más ágon helyezkedett el. Ez a protein rendelkezett az egyik legrövidebb doménszekvenciával, aminek az E-értéke az aszkuszos gombák találatai között a legalacsonyabb volt. Ez utalhat arra, hogy az ortológként azonosított fehérje nem hasonlít eléggé a *S. pombe* Sep15 proteinjére. A végső fákban a *W. sebi* proteinje sem szerepel. Bár a faj azon kevés bazídiumos gombák közé tartozik, amelyeknek fehérjéi szignifikáns Med8 doménnel rendelkeznek, jelenlétében a bazídiumos és a járomspórás fajok találatai gyakran keveredtek.

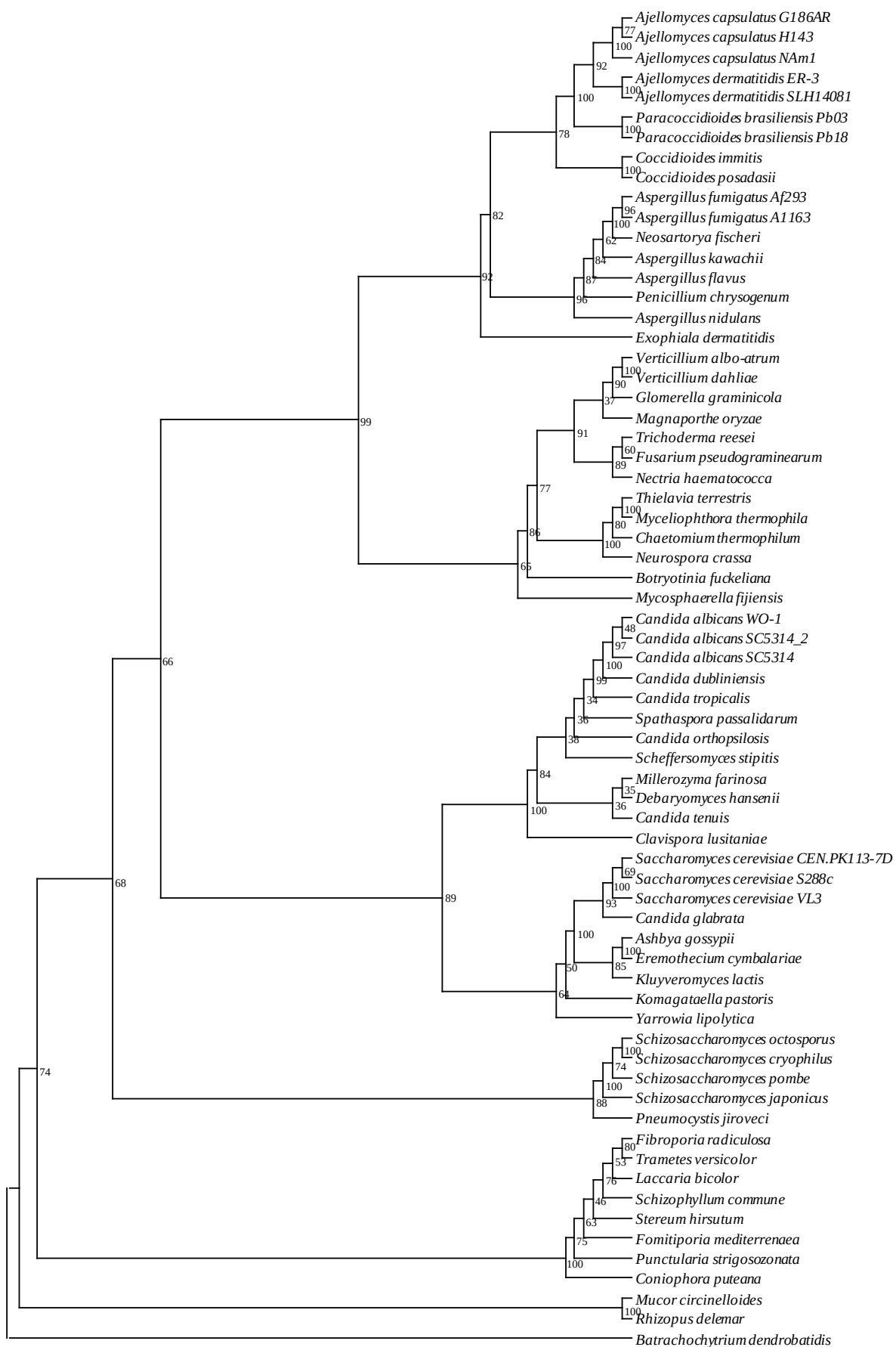
A törzsfákat mindig három különböző módszerrel (neighbor joining, maximum likelihood, maximum parsimony) készítettem el annak érdekében, hogy minél átfogóbb képet kapjak a vizsgált fehérje evolúciós lezármazásáról. A dolgozatomban a korábbi torz eredményeket (doménszekvencia alkalmazása, problémás fajok jelenléte), valamint a különböző illesztésekkel készült törzsfákat elegendő hely hiányában nem osztom meg, csak a végső eredményeket teszem közzé.

5.4.1. Neighbor joining módszer

A távolságalapú módszerek közé tartozó neighbor joining módszerrel készült törzsfát két különböző programmal is megszerkesztettem. A munkám során végig alkalmazott MEGA program (9. ábra) mellett a PHYLIP szoftverrel (10. ábra) is elkészítettem a filogenetikai fát.



9. ábra: A Sep15 fehérje filogenetikai fája (teljes fehérjeszekvencia, ClustalX illesztés, neighbor joining módszer, MEGA program)



10. ábra: A Sep15 protein filogenetikai fája (teljes fehérjeszekvencia, ClustalX illesztés, neighbor joining módszer, PHYLIP program)

Mindkét esetben látványosan elkülönülnek a találatok, mind a törzsek, mind pedig az altörzsek szintjén. A 9. ábráról leolvasható, hogy a közös őstől először a Zygomycota törzsben talált ortológok váltak el, utána a Basidiomycota törzs fehérjéi, végül pedig az Ascomycota törzs proteinjei. Az aszkuszos gombákban talált ortológok közül a hasadó élesztők és közeli rokonaik által alkotott Taphrinomycotina altörzs fehérjéi váltak el először. A törzsfá egyik érdekessége, hogy az *A. dermatitidis* két törzsében azonosított Sep15 ortológok nem egy ágon helyezkednek el. Ezt valószínűleg egy többszörös illesztéshiba (pl. egy rossz helyre illesztett szünet) okozta. A másik különlegessége, hogy a *P. chrysogenum*-ban talált homológ az *Aspergillus* nemzetség fehérjéivel egy ágon helyezkedik el. Egy 2011-es kutatás alapján a két nemzetség egy közös őssel rendelkezhetett, azonban a *Penicillium* nemzetségbe tartozó faj nem ékelődhetne az *Aspergillus* nemzetség tagjai közé (Houbraken & Samson, 2011). Elképzelhető, hogy a fajok annyira közeli rokonok, hogy nincs elég információ a vizsgált szekvenciákban az elválások pontos ábrázolásához, vagy itt is egy többszörös illesztéshiba állhat a háttérben.

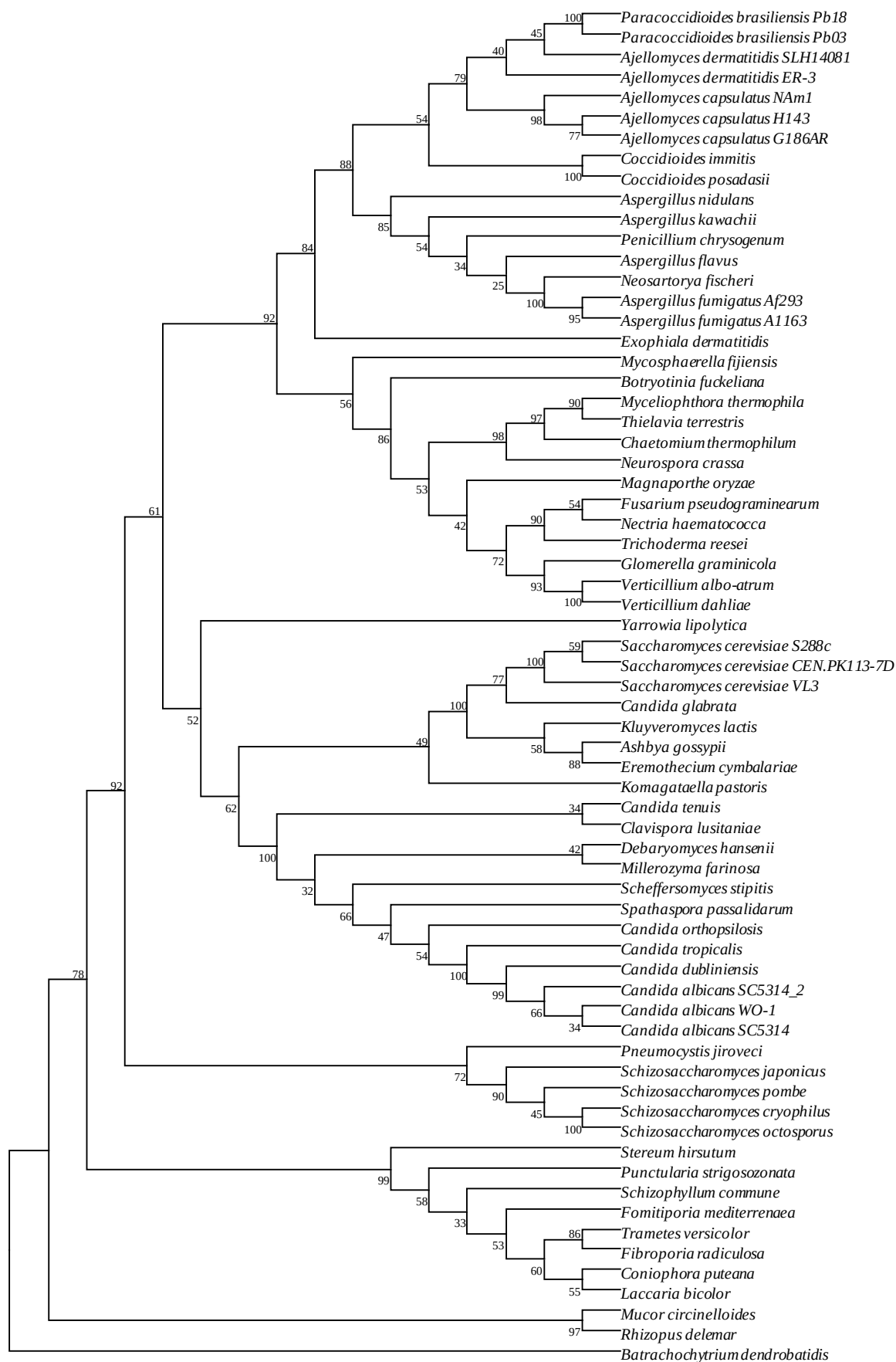
A PHYLIP szoftverrel szerkesztett fa (10. ábra) alapvetően azonos topológiával rendelkezik, mint a MEGA-val készített törzsfa, elsősorban a belső ágak elhelyezkedésében tapasztalható némi eltérés. A 10. ábrán például a *M. oryzae* fehérjéje a *Verticillium* nemzetségben talált ortológokkal helyezkedik el egy ágon, vagy a *C. puteana* proteinje vált el elsőként a bazídiumos gombák találatai között; ezzel szemben a 9. ábrán a *M. oryzae*-ban azonosított homológ a *N. crassa* proteinjével került egy ágra, a Basidiomycota törzs találatai közül pedig először a *P. strigosozonata* fehérjéje vált el. Míg a MEGA szoftvert alkalmazva a főelágazásoknál jobb bootstrap értékeket kaptam, addig a PHYLIP programmal készített törzsfa a belső elágazásoknál rendelkezik magasabb bootstrap értékekkel. A bootstrap értékeknél az 50-et tekintjük kritikusnak, vagyis az ennél kisebb értékkel rendelkező elágazásokat általában nem fogadjuk el. A két törzsfát összevetve az látható, hogy a MEGA szoftvert alkalmazva majdnem kétszer annyi alacsony értéket kaptam, mint a PHYLIP programmal. Összehasonlítva Wang és munkatársai által készített törzsfával (Wang et al., 2009) (4. ábra) elmondható, hogy a fő elágazások mindkét programot alkalmazva megegyeznek. A belső ágak elrendezése is nagyjából azonos, azonban ezek összehasonlítása már nehezebb, hiszen a vizsgált fajok eltérnek. A Basidiomycota törzsbe tartozó fajok proteinjeinek egymástól való elválásához egyáltalán nem szolgálhat referenciául a publikációjuk, mert csak egy olyan faj van (*L. bicolor*), ami mindkét kutatásban jelen van. A cikkben azok a fajok, amelyeknek fehérjéi a két programmal szerkesztve máshol

helyezkednek el (pl. *M. oryzae*, *Y. lipolytica*) a MEGA ábrázolásával azonosan találhatók meg.

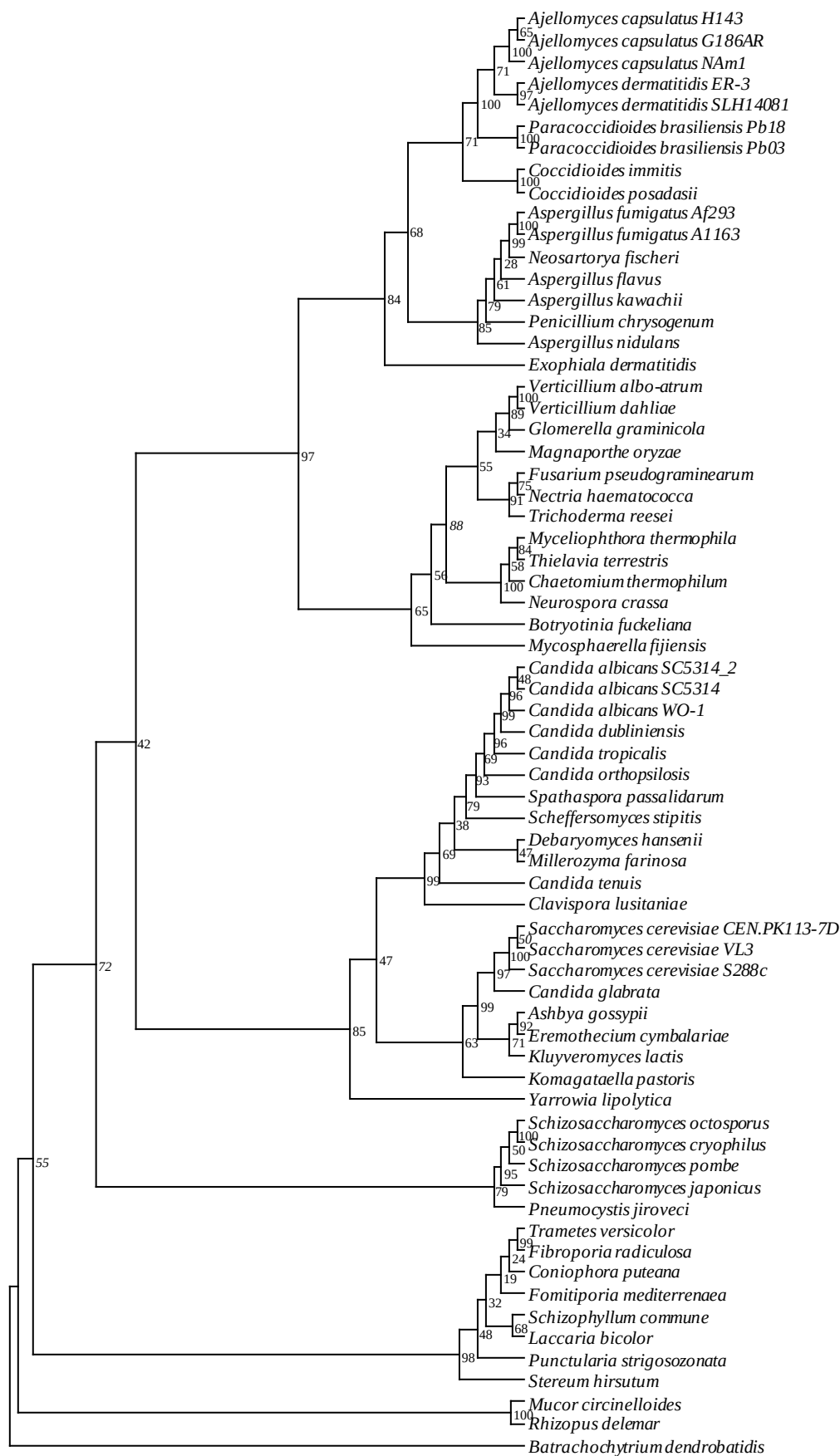
5.4.2. Maximum likelihood módszer

A modell alapú maximum likelihood módszerrel készült törzsfát is két különböző szoftverrel (MEGA, PHYLIP) szerkesztettem meg. Első lépésként a legjobb evolúciós modellt kellett meghatároznom, ami a MEGA programban a BIC kritérium alapján az rtREV+G lett. Az rtREV modellt elsősorban a reverz transzkriptázt (RT) tartalmazó szekvenciákra fejlesztették ki 2001-ben. A retrovírus elemek az általánostól eltérő evolúciós nyomásnak vannak kitéve, ezért alkalmasak egy specifikus modell létrehozására. A modellt 33 retrovírus *pol* gén szekvenciájával optimalizálták, amely a proteázt, a reverz transzkriptázt és az integrázt kódolja (Dimmic et al., 2001). A Sep15 fehérje azonban nem egy reverz transzkriptáz, így a BIC kritérium szerinti második (WAG+G) és harmadik (JTT+G) legjobb evolúciós modellel is megszerkesztettem a törzsfát. A +G paraméter minden esetben a már említett szekvencián belüli mutációs ráta variabilitását veszi figyelembe. Az elkészült filogenetikai fák között szignifikáns különbséget nem tapasztaltam, a feltételezett evolúciót az rtREV+G modell alapján készült fa követte a legmagasabb bootstrap értékekkel, ezért csak ezt mutatom be a dolgozatomban. Az rtREV modell feltételezhetően azért alkalmas az általam vizsgált fehérje leírására, mert a reverz transzkriptázhoz hasonlóan a Sep15 is állandó komponense a legtöbb eukarióta genomnak.

A négy törzs, valamint az Ascomycota törzs három altörzse mind a MEGA programmal (11. ábra), mind a PHYLIP szoftvert alkalmazva (12. ábra) látványosan elkülönül. A neighbor joining módszerhez hasonlóan külön ágat alkotnak a Taphrinomycotina altörzset képviselő fajok fehérjéi, amelyek az Ascomycota törzsbe tartozó gombák találatai közül elsőként válnak el. Összevetve a MEGA programmal készült, két különböző módszeren alapuló (neighbor joining és maximum likelihood) törzsfákat, a legnagyobb különbség a Saccharomycotina altörzs találatai között tapasztalható, ahol a belső elágazások topológiája eltér. Azonban az előző módszerhez hasonlóan itt sem kerültek egy ágra az *A. dermatitidis* törzseinek fehérjéi, valamint a *P. chrysogenum*-ban azonosított ortológ is ugyanott, az *Aspergillus* nemzetség találatai között helyezkedik el. Akárcsak a neighbor joining módszer esetében, itt sem tapasztaltam jelentős eltérést a két szoftver által készített fák topológiájában. A főelágazások teljesen megegyeznek, és a belső ágak esetén is csak kisebb különbségek adódtak (pl. a MEGA szoftverrel szerkesztett fán a *C. tenuis* fehérjéje és a *C. lusitaniae*



11. ábra: A Sep15 fehérje filogenetikai fája (teljes fehérjeszekvencia, ClustalX illesztés, maximum likelihood módszer, MEGA program)



12. ábra: A Sep15 protein filogenetikai fája (teljes fehérje szekvencia, ClustalX algoritmus, maximum likelihood módszer, PhyML program)

proteinje egy ágra került, vagy a PhyML programmal készített törzsfán a *M. oryzae* fehérjéje ismét a *Verticillium* nemzetségben azonosított ortológokkal szerepel együtt). Egy pontban, az *A. dermatitidis* törzseiben talált homológok ábrázolásában azonban a PhyML egyértelműen jobb eredményt ad, mint a MEGA.

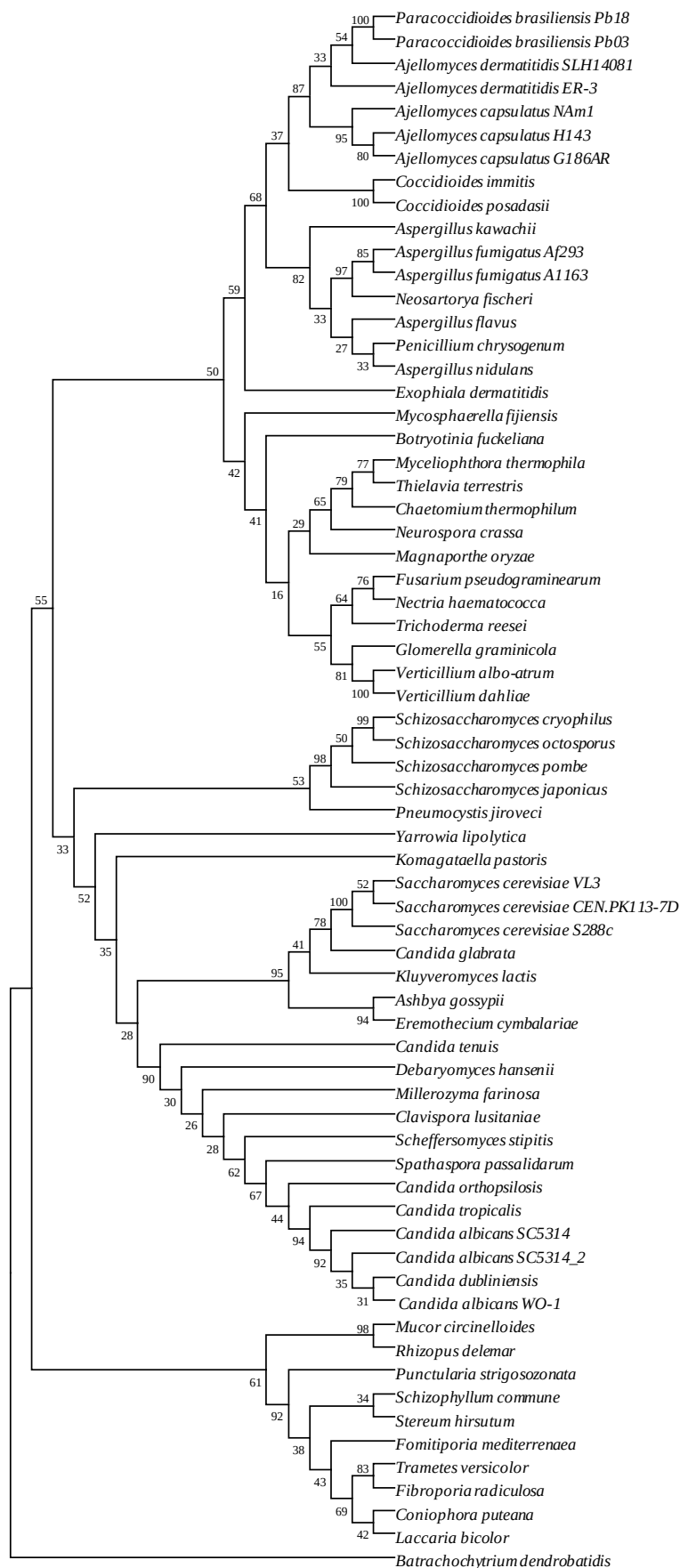
Az elágazásokban látható bootstrapek hasonlóan alakultak, mint a neighbor joining módszer esetén. A MEGA program a főelágazásokhoz, míg a PhyML általában a belső ágak elválásához rendelt nagyobb értékeket. Az előző módszerrel ellentétben közel ugyanannyi 50 alatti értékkel rendelkeznek, viszont a PhyML szoftverrel szerkesztett törzsfa esetében az egyik egy főelágazáshoz tartozik. A közel azonos topológia ellenére megjelenő különbségek a bootstrap értékekben valószínűleg a szoftverek által használt heurisztikákkal magyarázhatók.

Összevetve a Wang és munkatársai által készített törzsfával (Wang et al., 2009) (4. ábra) azt tapasztaltam, hogy a főelágazásokat tekintve mindkét program azonos eredményeket adott. A kutatás szempontjából jelentős Taphrinomycotina altörzsbe tartozó fajok fehérjei ugyanabban a sorrendben váltak el egymástól, mint a fajok: először a *P. jiroveci* proteinje, aztán a *S. japonicus* fehérjéje, majd a *S. pombe* Sep15 proteinje, végül pedig a *S. octosporus*-ban és *S. cryophilus*-ban azonosított ortológok.

5.4.3. Maximum parsimony módszer

Munkám során a legrosszabb eredményt mindig a karakter alapú maximum parsimony módszerrel kaptam (13. ábra). Mivel a maximum likelihood és a neighbor joining módszereknél nem tapasztaltam jelentős eltérést az eredményekben a különböző programok alkalmazása során, és ezzel a módszerrel nem volt igazán jó eredményem, amit érdemes lett volna egy másik szoftverrel is reprodukálni, ezt a törzsfát csak a MEGA szoftverrel szerkesztettem meg (13. ábra). Az ábrán jól látható, hogy sem a törzsek, sem az altörzsek nem különülnek el egymástól úgy, mint az előző két módszer esetében. A Taphrinomycotina altörzsben talált ortológok a Saccharomycotina altörzs fehérjéivel, a Zygomycota törzs találatai pedig a Basidiomycota törzs proteinjeivel került egy ágra. A módszerrel készített filogenetikai fa másik hátránya, hogy nagyon alacsony bootstrap értékekkel rendelkezik. Az összes érték több mint egyharmada a kritikus 50-nél kisebb, és nem csak a feltételezhetően rossz elágazásoknál. A legalacsonyabb érték (16) egy olyan elágazásnál látható, ami megegyezik a neighbor joining és maximum likelihood módszerekkel készített fák elágazásával.

A rossz eredmények hátterében az algoritmus működési elve állhat. Ahogy azt a 2.4. fejezetben említettem, a módszer azt az evolúciót keresi, ami a lehető legkevesebb mutációval



13. ábra: A Sep15 fehérje filogenetikai fája (teljes fehérjeszekvencia, ClustalX illesztés, maximum parsimony módszer, MEGA program)

ment végbe. Ehhez azonban nem egy modellt alkalmaz, hanem csak leszámolja a végbement mutációk számát anélkül, hogy súlyozna azok típusával. Ez a nukleotidok esetében, ahol csak négy építőelemet különböztetünk meg, nem jelent nagy problémát. A többszörösen illesztett aminosav-szekvenciák ábrázolását azonban erősen torzítja, mivel figyelmen kívül hagyja az aminosavak fizikokémiai tulajdonságait, a genetikai kód degeneráltságát, és azt, hogy az aminosav cserék háttérében gyakran több mutáció áll. A teljes genomszekvenciából szerkesztett törzsfával (Wang et al., 2009) (4. ábra) összehasonlítva elmondható, hogy a főelágazásokban nem mindenütt egyeznek meg, de a törzsfák egyes részei (pl. fonalas gombák) nagy hasonlóságot mutatnak.

6. Összefoglalás

Munkám során a *Schizosaccharomyces pombe* Sep15 sejtszeparációs proteinjének filogenetikai fáját szerkesztettem meg. A Sep15 fehérje a Mediátor komplex nyolcas számú alegységét alkotja, amely az RNS polimeráz II-dependens gének transzkripciójához elengedhetetlen. A protein többek között a sejtszeparációs gének transzkripciójáért felelős Ace2 fehérjével is kapcsolatban áll, így hiányában a sejtosztódás nem megy végbe megfelelően, ami hifás növekedést eredményez. A sejtosztódási fehérjék tanulmányozása a dimorfizmus jelenségének pontosabb megértéséhez vezethet.

A vizsgált 74 gombafajból fajok közötti BLASTp analízis segítségével 64 esetben sikerült a Sep15 protein ortológjait azonosítani. A fennmaradó eseteket alkotó fajok rendelkeznek ugyan megszekvenált genommal, azonban annak annotációja csak részleges. A 64 fehérjeszekvenciából további hatot nem ábrázoltam a végső törzsfákon, mivel különböző okok miatt torzították a topológiát.

Az ortológjelöltekkel reciprok BLASTp keresést is végeztem, amelynek során a találat szekvenciájával kerestem a *S. pombe* genomjában, hogy meggyőződjek arról, valóban a Sep15 fehérje ortológiát találtam meg. Minden esetben a Sep15 fehérjét kaptam vissza első találatként. Ezt követően páros BLAST-ot végeztem, hogy a különböző adatbázisokból származó adatok összehasonlíthatóak legyenek.

A potenciális ortológok doménszerkezetét is vizsgáltam a Pfam adatbázis segítségével. Eredeti elképzelés szerint csak azokat tekintettem volna valódi ortológoknak, amelyek szignifikánsan és megszakítás nélkül tartalmazzák a Med8 domént. A Pfam-A adatai alapján azonban a bazídiumos gombák jelentős részében a Med8 domén jelenléte nem volt szignifikáns. Ez feltételezhetően abból adódik, hogy a Basidiomycota törzs proteinei alulreprezentáltak, a rejtett Markov modellt elsősorban az Ascomycota törzsbe tartozó fajok és a gerinces állatok szekvenciái építik fel. Ezért a Basidiomycota törzs fehérjeinek esetében figyelmen kívül hagytam ezt a feltételt.

Az ortológok teljes fehérjeszekvenciái mellett a doménszekvenciákat is felhasználtam a törzsfák szerkesztéséhez. Munkám során azt tapasztaltam, hogy a teljes fehérjeszekvenciákkal készült fák jobb eredményeket adnak. Ezt az okozhatja, hogy a Sep15 fehérje doménszekvenciáinak hossza nagy szórást mutat, valamint a nagyon rövid domének nem tartalmaznak elegendő információt a törzsfaszerkesztéshez.

A valódi ortológ szekvenciák többszörös illesztését három különböző algoritmussal végeztem el, amelyek közül a ClustalX módszerrel kaptam a legjobb eredményeket. A

PRANK és MAFFT algoritmusokkal készült illesztések alátámasztják a ClustalX eredményeit, azonban az azok alapján szerkesztett fák gyengébb bootstrap értékekkel rendelkeznek és a belső elágazásoknál is tapasztalhatók kisebb eltérések. A törzsfákat minden esetben három módszerrel (neighbor joining, maximum likelihood, maximum parsimony) készítettem el. Az első két módszer esetében az alapvetően alkalmazott szoftver (MEGA) mellett más programokkal (PHYMLIP, PhyML) is megszerkesztettem a törzsfákat. A neighbor joining és maximum likelihood módszerekkel, az alkalmazott szoftvertől függetlenül, a Taphrinomycotina altörzs fehérjéi egy külön ágat alkotnak, az aszkuszos gombák találatai közül ezek váltak el először. A törzsfák a főelágazásokban minden esetben megegyeznek, kisebb eltérések csak a belső ágak topológiájában figyelhető meg. Ezek a filogenetikai fák nagyfokú hasonlóságot mutatnak a teljes genomi szekvenciák alapján készült törzsfával, amit Wang és munkatársai készítettek (Wang et al., 2009). A maximum parsimony módszerrel kaptam a legrosszabb eredményeket az egész munkám során, aminek hátterében az algoritmus működési elve állhat. A módszer alkalmazása során a törzsek és altörzsek nem váltak el egymástól, ezért az eredmény nem megbízható. Vizsgálataim alapján úgy gondolom, hogy a Sep15 fehérje evolúcióját a legjobban a ClustalX algoritmussal illesztett teljes fehérjeszekvenciák alapján, neighbor joining módszerrel és MEGA szoftverrel szerkesztett törzsfa írja le.

Ezek alapján feltételezhető, hogy a Sep15 protein evolúciója nagyon hasonlóan ment végbe, mint a fajok leszármazása. Valószínűsíthető, hogy a közös őstől először a Zygomycota törzs fehérjéi, majd pedig a Basidiomycota törzsben azonosított ortológok váltak el, amiket a Taphrinomycotina altörzs proteinjei követtek. A fonalas és élesztőszerű gombák fehérjeinek elválása egy feltételezhető közös őst követően mehetett végbe.

A közeljövőben tervezem a törzsfák elkészítését a modell alapú bayesi módszerrel is. Ezen kívül céljaim között szerepel, hogy a kutatásomat kiterjesszem az egész élővilágra, mivel a Sep15 protein feltehetően az összes eukarióta élőlényben esszenciális. Tervben van, hogy a filogenetikai analízisem eredményei alapján a Debreceni Egyetemen további komplementációs tesztek végeznének, ami megerősítené a bioinformatikai módszerekkel kapott eredményeimet. A laboratóriumi kísérletek során deléciós mutánsokat (amelyek hifás növekedésű fenotípussal rendelkeznek) hoznak létre, amit más fajokban talált ortológokkal komplementálnak. A kísérlettel azt vizsgálnák, hogy a gén különböző fajokban megtalálható homológjai azonos feladatot látnak-e el.

7. Irodalomjegyzék

Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2009). *Essential Cell Biology*. New York: Garland Publishing, Inc.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Ansari, S. A. & Morse, R. H. (2013). Mechanisms of Mediator complex in transcriptional activation. *Cell Mol Life Sci* **70**, 2743-2756.

Balasubramanian, M. K. & McCollum, D. (2004). Regulation of Cytokinesis. In *The Molecular Biology of Schizosaccharomyces pombe*, (ed. R. Egel), pp. 243-254. Berlin, Heidelberg: Springer-Verlag.

Balazs, A., Batta, G., Miklós, I., Acs-Szabo, L., Vazquez de Aldana, C. R. & Sipiczki, M. (2012). Conserved regulators of the cell separation process in *Schizosaccharomyces*. *Fungal Genet Biol* **49**, 235-249.

Baxevanis, A. D. & Ouelette, B. F. F. (2005). *Bioinformatics: a practical guide to the analysis of genes and proteins*. New Jersey: John Wiley & Sons, Inc.

Black, J. G. (2012). *Microbiology. Principles and Explorations*. New Jersey: John Wiley & Sons, Inc.

Blacketer, M., Koehler, C., Coats, S., Myers, A. & Madaule, P. (1993). Regulation of dimorphism in *Saccharomyces cerevisiae*: Involvement of the novel protein kinase homolog Elmlp and protein phosphatase 2A. *Mol Cell Biol* **13**, 5567-5587.

Bourbon, H.-M. (2008). Comparative genomics supports a deep evolutionary origin for the large, four-module transcriptional mediator complex. *Nucl Acid Res* **36**, 3993-4008.

Darwin, C. (2011). *On the Origin of Species*. Huntington: Empire Books.

Dimmic, M. W., Rest, J. S., Mindell, D. P. & Goldstein, R. A. (2001). rtREV: An amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *J Mol Evol* **55**, 65-73.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science* **284**, 2124-2128.

Egel, R. (2004). Fission yeast in general genetics. In *The Molecular Biology of Schizosaccharomyces pombe*, (ed. R. Egel), pp. 1-12. Berlin, Heidelberg: Springer-Verlag.

Enczi, K., Yamaguchi, M. & Sipiczki, M. (2007). Morphology transition genes in the dimorphic fission yeast *Schizosaccharomyces japonicus*. *Anton Leeuw* **92**, 143-154.

Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* **17**, 368-376.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791.

Finn, R. D., Mistry, J., Tate, J., Coghill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K. et al. (2010). The Pfam protein families database. *Nucl Acid Res* **38**, D211-D222.

Hibbett, D., Binder, M., Bischoff, J., Blackwell, M., Cannon, F., Eriksson, O., Huhndorf, S., James, T., Kirk, P., Lücking, R. et al. (2007). A higher-level phylogenetic classification of the *Funghi*. *Mycol Res* **3**, 509-547.

Horváth, A. (2013). A hasadó élesztőgombák Agn1 sejtszeparációs proteinjének filogenetikai törzsfája. BME-VBK Alkalmazott Biotechnológia és Élelmiszer-tudományi Tanszék. Budapest. (Diplomamunka)

Houbraken, J. & Samson, R. A. (2011). Phylogeny of *Penicillium* and the segregation of *Trichocomaceae* into three families. *Stud Mycol* **70**, 1-51.

Katoh, K. & Toh, H. (2008). Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform* **9**, 286-298.

Kendrick, B. (2000). The Fifth Kingdom. Newburyport: Focus Publishing.

Klein, B. & Tebbets, B. (2007). Dimorphism and virulence in fungi. *Curr Opin Microbiol* **10**, 314-319.

Langely, C. H. & Fitch, W. M. (1974). An examination of the constancy of the rate of molecular evolution. *J Mol Evol* **3**, 161-177.

Linder, T., Rasmussen, N. N., Samuelsen, C. O., Chatzidaki, E., Baraznenok, V., Beve, J., Henriksen, P., Gustafsson, C. M. & Holmberg, S. (2008). Two conserved modules of *Schizosaccharomyces pombe* mediator regulate distinct cellular pathways. *Nucl Acid Res* **36**, 2489-2504.

Liu, Y., Leigh, J. W., Brinkmann, H., Cushion, M. T., Rodriguez-Ezpeleta, N., Philippe, H. & Lang, B. F. (2009). Phylogenomic analyses support the monophyly of Taphrinomycotina, including *Schizosaccharomyces* fission yeasts. *Mol Biol Evol* **26**, 27-34.

Löytynoja, A. & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science* **320**, 1632-1635.

McNeill, J., Barrie, F. R., Buck, W. R., Demoulin, W., Greuter, W., Hawksworth, D. L., Herendeen, P. S., Knapp, S., Marhold, K., Prado, J., Prud'homme Van Reine, W. F., Smith, G. F., Wiersema, J. H. & Turland, N. J. (2012) International Code of Nomenclature for algae, fungi, and plants (Melbourne Code). Melbourne: Koeltz Scientific Books.

Mehta, S., Miklós, I., Sipiczki, M., Sengupta, S. & Sharma, N. (2009). The Med8 mediator subunit interacts with the Rpb4 subunit of RNA polymerase II and Ace2 transcriptional activator in *Schizosaccharomyces pombe*. *FEBS Letters* **583**, 3115-3120.

Miklós, I., Szilagyi, Z., Watt, S., Zilahi, E., Batta, G., Antunovics, Z., Enczi, K., Bahler, J. & Sipiczki, M. (2007). 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* **279**, 225-238.

Morgan, D. O. (2007). The Cell Cycle. London: New Science Press Ltd.

Nasmyth, K. (2001). A prize for proliferation. *Cell* **107**, 689-701.

Nemecek, J., Wütrich, M. & Klein, B. (2006). Global control of dimorphism and virulence in fungi. *Science* **312**, 583-588.

Orlowski, M. (1991). Mucor dimorphism. *Microbiol Rev* **55**, 234-258.

Pesti, M., Biró, S., Hornok, L., Kevei, F., Kucsera, J., Maráz, A., Szűcs, G. & Vágvölgyi, C. (2001). Általános mikrobiológia. Budapest-Pécs: Dialóg Campus Kiadó.

Réczey, J., Sipos, B. & Sveicz, Á. (2008). Részletes mikrobiológia. Budapest: Budapesti Műszaki és Gazdaságtudományi Egyetem (Oktatási segédlet: oktatas.ch.bme.hu/oktatas/konyvek/mezgaz/mikrobio/biomernok).

Saitou, N. & Nei, M. (1987). The Neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406-425.

Sipiczki, M. (1995). Phylogenesis of fission yeast. Contradictions surrounding the origin of a century old genus. *Anton Leeuw* **68**, 119-149.

Sipiczki, M. (2000). Where does fission yeast sit on the tree of life? *Genome Biol* **1**, 1011.1-1011.4.

Sipiczki, M. (2004). Fission yeast phylogenesis and evolution. In *The Molecular Biology of the Schizosaccharomyces pombe*, (ed. R. Egel), pp. 431-444. Berlin, Heidelberg: Springer-Verlag.

Sipiczki, M. (2007). Splitting of the fission yeast septum. *FEMS Yeast Res* **7**, 761-770.

Sunnerhagen, P. (2002). Prospects for functional genomics in *Schizosaccharomyces pombe*. *Curr Genet* **42**, 73-84.

Sveiczner, Á. & Horváth, A. (2013). Cell cycle, fission yeast. In *Encyclopedia of Systems Biology*, (eds W. Dubitzky, O. Wolkenhauer, K.-H. Cho & H. Yokota), pp. 349-353. Heidelberg - New York: Springer.

Vida, N. (2012). A hasadó élesztőgombák *eng1* és *eng2* génjeinek filogenetikai törzsfája. BME-VBK Alkalmazott Biotechnológia és Élelmiszer-tudományi Tanszék. Budapest. (Diplomamunka)

Wang, H., Xu, Z., Gao, L. & Hao, B. (2009). A fungal phylogeny based on 82 complete genomes using the composition vector method. *BMC Evol Biol* **9**, 195.

Wheeler, D. (2003). Selecting the right protein-scoring matrix. *Curr Protoc Bioinform*, 3.5.1-3.5.6.

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S. et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature* **415**, 871-880.

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

Zvelebil, M. & Baum, J. O. (2008). Understanding Bioinformatics. New York: Garland Science, Taylor & Francis Group, LLC.

8. Alkalmazott adatbázisok, szoftverek

- [1.] <http://www.pombase.org/>
- [2.] <http://pfam.sanger.ac.uk/>
- [3.] <http://www.uniprot.org/>
- [4.] <http://www.broadinstitute.org/>
- [5.] <http://www.ncbi.nlm.nih.gov/>
- [6.] <http://www.jgi.doe.gov/>
- [7.] <http://www.clustal.org/clustal2/>
- [8.] <http://www.ebi.ac.uk/goldman-srv/webPRANK/>
- [9.] <http://www.ebi.ac.uk/Tools/msa/mafft/>
- [10.] <http://www.megasoftware.net/>
- [11.] <http://evolution.genetics.washington.edu/phylip.html>
- [12.] <http://www.atgc-montpellier.fr/phyml/>
- [13.] <http://taxonomy.zoology.gla.ac.uk/rod/treeview.html>
- [14.] <http://tree.bio.ed.ac.uk/software/figtree/>
- [15.] <http://weblogo.threeplusone.com/create.cgi>

9. Függelék

F1. táblázat: A *S. pombe* Sep15 fehérjéinek ortológjai a *Saccharomycotina* altörzsben

	<i>S. pombe</i>	<i>C. glabrata</i>	<i>C. orthosilosis</i>	<i>C. albicans</i>			<i>C. tenuis</i>
Törzs	972h ⁻	CBS 138	Co 90-125	SC5314	SC5314 (2)	WO-1	ATCC 10573
Fehérje kódja	SPBC21.04	XP_449574.1	CCG22217.1	XP_712891.1	XP_712923.1	EEQ45871.1	EGV62121.1
Fehérje hossza (as)	200	233	227	239	239	239	201
BLAST hasonlóság		59/95 (62%)	83/155 (54%)	58/102 (57%)	58/102 (56%)	58/102 (56%)	84/156 (54%)
BLAST egyezés		38/95 (40%)	51/155 (33%)	38/102 (37%)	37/102 (36%)	38/102 (37%)	47/156 (30%)
BLAST szünetek		1/95 (1%)	11/155 (7%)	6/102 (6%)	6/102 (5%)	6/102 (5%)	17/156 (11%)
BLAST pontszám		75.5 bit (184)	70.5 bit (171)	68.9 bit (167)	67.4 bit (163)	68.9 bit (167)	65.5 bit (158)
E-érték		$3 \cdot 10^{-21}$	10^{-19}	$5 \cdot 10^{-19}$	$2 \cdot 10^{-18}$	$5 \cdot 10^{-19}$	$4 \cdot 10^{-18}$

F1. táblázat folytatása: A *S. pombe* Sep15 fehérjéinek ortológjai a *Saccharomycotina* altörzsben

	<i>S. pombe</i>	<i>C. dubliniensis</i>	<i>C. tropicalis</i>	<i>Y. lipolytica</i>	<i>K. pastoris</i>	<i>E. cymbalariae</i>
Törzs	972h ⁻	CD36	MYA-3404	CLIB122	GS115	DBVPG#7215
Fehérje kódja	SPBC21.04	XP_002418393.1	XP_002547295.1	XP_500410.1	XP_002491169.1	XP_003647717.1
Fehérje hossza (as)	200	243	224	194	193	222
BLAST hasonlóság		58/102 (57%)	59/106 (56%)	91/192 (47%)	68/122 (56%)	92/191 (48%)
BLAST egyezés		38/102 (37%)	37/106 (35%)	64/192 (33%)	48/122 (39%)	52/191 (27%)
BLAST szünetek		6/102 (6%)	6/106 (6%)	20/192 (10%)	6/122 (5%)	4/191 (2%)
BLAST pontszám		68.2 bit (165)	64.7 bit (156)	91.3 bit (225)	84.7 bit (208)	67.8 bit (164)
E-érték		$8 \cdot 10^{-19}$	$9 \cdot 10^{-18}$	$3 \cdot 10^{-27}$	$8 \cdot 10^{-25}$	10^{-18}

Az F1. táblázat az 59. oldalon folytatódik

F1. táblázat folytatása: A *S. pombe* Sep15 fehérjéinek ortológjai a Saccharomycotina altörzsben

	<i>S. pombe</i>	<i>S. stipitis</i>	<i>S. cerevisiae</i>			<i>M. farinosa</i>
Törzs	972h ⁻	CBS 6054	S288c	CEN.PK113-7D	VL3	CBS 7064
Fehérje kódja	SPBC21.04	XP_001384574.1	NP_009752.3	EIW12127.1	EGA87936.1	CCE84454.1
Fehérje hossza (as)	200	208	223	223	223	221
BLAST hasonlóság		87/154 (56%)	100/191 (52%)	100/191 (52%)	79/144 (55%)	61/102 (60%)
BLAST egyezés		54/154 (35%)	58/191 (30%)	58/191 (30%)	47/144 (33%)	40/102 (39%)
BLAST szünetek		14/154 (9%)	12/191 (6%)	12/191 (6%)	3/144 (2%)	5/102 (5%)
BLAST pontszám		80.5 bit (197)	77.0 bit (188)	77.0 bit (188)	74.7 bit (182)	75.9 bit (185)
E-érték		$4 \cdot 10^{-23}$	$7 \cdot 10^{-22}$	$7 \cdot 10^{-22}$	$4 \cdot 10^{-21}$	10^{-21}

F1. táblázat folytatása: A *S. pombe* Sep15 fehérjéinek ortológjai a Saccharomycotina altörzsben

	<i>S. pombe</i>	<i>D. hansenii</i>	<i>C. lusitaniae</i>	<i>A. gossypii</i>	<i>K. lactis</i>	<i>S. passalidarum</i>
Törzs	972h ⁻	CBS767	ATCC 42720	ATCC 10895	NRRL Y-1140	NRRL Y-27907
Fehérje kódja	SPBC21.04	XP_002770775.1	XP_002615074.1	NP_984609.1	XP_455017.1	EGW33037.1
Fehérje hossza (as)	200	225	211	223	236	237
BLAST hasonlóság		84/153 (55%)	105/204 (51%)	94/193 (49%)	56/97 (58%)	92/183 (50%)
BLAST egyezés		52/153 (34%)	53/204 (26%)	57/193 (30%)	36/97 (37%)	54/183 (30%)
BLAST szünetek		17/153 (11%)	22/204 (11%)	8/193 (4%)	1/97 (1%)	16/183 (9%)
BLAST pontszám		75.1 bit (183)	74.7 bit (182)	70.5 bit (171)	70.5 bit (171)	70.1 bit (170)
E-érték		$3 \cdot 10^{-21}$	$4 \cdot 10^{-21}$	10^{-19}	10^{-19}	$2 \cdot 10^{-19}$

F2. A *S. pombe* Sep15 fehérjéinek ortológjai a Pezizomycotina altörzsben

	<i>S. pombe</i>	<i>A. dermatitidis</i>		<i>T. melanosporum</i>	<i>A. flavus</i>	<i>N. fischeri</i>
Törzs	972h ⁻	ER-3	SLH14081	Mel28	NRRL3357	NRRL 181
Fehérje kódja	SPBC21.04	EEQ87974.1	XP_002627781.1	XP_002840444.1	XP_002376999.1	XP_001267185.1
Fehérje hossza (as)	200	255	255	282	244	252
BLAST hasonlóság		81/165 (49%)	81/165 (49%)	91/201 (45%)	79/153 (51%)	73/149 (49%)
BLAST egyezés		47/165 (28%)	47/165 (28%)	62/201 (31%)	43/153 (28%)	41/149 (28%)
BLAST szünetek		4/165 (2%)	4/165 (2%)	24/201 (12%)	4/153 (2%)	4/149 (3%)
BLAST pontszám		72.0 bit (175)	71.6 bit (174)	72.0 bit (175)	69.3 bit (168)	69.7 bit (169)
E-érték		5·10 ⁻²⁰	6·10 ⁻²⁰	7·10 ⁻²⁰	3·10 ⁻¹⁹	3·10 ⁻¹⁹

A narancssárga háttérű fajokban talált ortológok különböző okok miatt nem szerepelnek a végső törzsfákon.

F2. táblázat folytatása: A *S. pombe* Sep15 fehérjéinek ortológjai a Pezizomycotina altörzsben

	<i>S. pombe</i>	<i>A. fumigatus</i>		<i>P. chrysogenum</i>	<i>E. dermatitidis</i>	<i>A. oligospora</i>
Törzs	972h ⁻	Af293	A1163	Wisconsin 54-1255	NIH/UT8656	ATCC 24927
Fehérje kódja	SPBC21.04	XP_751989.1	EDP50215.1	XP_002559238.1	EHY52313.1	EGX45958.1
Fehérje hossza (as)	200	249	252	245	245	721
BLAST hasonlóság		72/149 (48%)	72/149 (48%)	78/167 (47%)	82/159 (52%)	56/97 (57%)
BLAST egyezés		42/149 (28%)	42/149 (28%)	48/167 (29%)	48/159 (30%)	36/97 (37%)
BLAST szünetek		4/149 (3%)	4/149 (2%)	9/167 (5%)	9/159 (6%)	0/97 (0%)
BLAST pontszám		68.9 bit (167)	69.3 bit (168)	68.9 bit (167)	67.0 bit (162)	65.5 bit (158)
E-érték		5·10 ⁻¹⁹	4·10 ⁻¹⁹	5·10 ⁻¹⁹	2·10 ⁻¹⁸	10 ⁻¹⁶

A narancssárga háttérű fajokban talált ortológok különböző okok miatt nem szerepelnek a végső törzsfákon.

Az F2. táblázat a 61. oldalon folytatódik

F2. táblázat folytatása: A *S. pombe* Sep15 fehérjéinek ortológjai a *Pezizomycotina* altörzsben

	<i>S. pombe</i>	<i>A. capsulatus</i>			<i>P. brasiliensis</i>		<i>M. fijiensis</i>
Törzs	972h ⁻	H143	NAm1	G186AR	Pb18	Pb03	CIRAD86
Fehérje kódja	SPBC21.04	EER45218.1	XP_001538616.1	EEH07092.1	EEH46986.1	EEH18086.1	EME89241.1
Fehérje hossza (as)	200	256	256	256	247	247	204
BLAST hasonlóság		61/111 (55%)	61/111 (54%)	61/111 (54%)	74/153 (48%)	74/153 (48%)	49/103 (47%)
BLAST egyezés		36/111 (32%)	36/111 (32%)	36/111 (32%)	43/153 (28%)	43/153 (28%)	33/103 (32%)
BLAST szünetek		1/111 (1%)	1/111 (0%)	1/111 (0%)	4/153 (3%)	4/153 (3%)	0/103 (0%)
BLAST pontszám		66.6 bit (161)	65.5 bit (158)	65.1 bit (157)	65.1 bit (157)	65.1 bit (157)	62.8 bit (151)
E-érték		3·10 ⁻¹⁸	10 ⁻¹⁷	10 ⁻¹⁷	10 ⁻¹⁷	10 ⁻¹⁷	4·10 ⁻¹⁷

F2. táblázat folytatása: A *S. pombe* Sep15 fehérjéinek ortológjai a *Pezizomycotina* altörzsben

	<i>S. pombe</i>	<i>V. dahliae</i>	<i>V. albo-atrum</i>	<i>A. nidulans</i>	<i>C. immitis</i>	<i>A. kawachii</i>	<i>C. thermophilum</i>
Törzs	972h ⁻	VdLs.17	VaMs.102	FGSC A4	RS	IFO 4308	DSM 1495
Fehérje kódja	SPBC21.04	EGY17269.1	XP_003005765.1	XP_659324.1	XP_001241689.1	GAA82781.1	EGS19264.1
Fehérje hossza (as)	200	268	229	260	262	268	290
BLAST hasonlóság		55/95 (58%)	55/96 (57%)	55/97 (57%)	51/96 (53%)	70/158 (44%)	54/97 (56%)
BLAST egyezés		33/95 (35%)	33/96 (34%)	33/97 (34%)	33/96 (34%)	42/158 (27%)	33/97 (34%)
BLAST szünetek		0/95 (0%)	0/96 (0%)	0/97 (0%)	0/96 (0%)	15/158 (9%)	0/97 (0%)
BLAST pontszám		62.4 bit (150)	61.6 bit (148)	60.5 bit (145)	60.5 bit (145)	58.2 bit (139)	58.5 bit (140)
E-érték		10 ⁻¹⁶	10 ⁻¹⁶	6·10 ⁻¹⁶	7·10 ⁻¹⁶	5·10 ⁻¹⁵	4·10 ⁻¹⁵

Az F2. táblázat a 62. oldalon folytatódik

F2. táblázat folytatása: A *S. pombe* Sep15 fehérjéinek ortológjai a Pezizomycotina altörzsben

	<i>S. pombe</i>	<i>G. graminicola</i>	<i>M. thermophila</i>	<i>M. oryzae</i>	<i>T. terrestris</i>	<i>C. posadasii</i>
Törzs	972h ⁻	M1.001	ATCC 42464	70-15	NRRL 8126	C735 delta SOWgp
Fehérje kódja	SPBC21.04	EFQ27846.1	XP_003658536.1	XP_003716133.1	XP_003656497.1	XP_003070639.1
Fehérje hossza (as)	200	265	310	315	273	265
BLAST hasonlóság		49/96 (51%)	50/97 (52%)	50/95 (53%)	50/97 (52%)	73/177 (41%)
BLAST egyezés		30/96 (31%)	33/97 (34%)	30/95 (32%)	31/97 (32%)	48/177 (27%)
BLAST szünetek		0/96 (0%)	0/97 (0%)	0/95 (0%)	0/97 (0%)	25/177 (14%)
BLAST pontszám		55.5 bit (132)	54.7 bit (130)	52.8 bit (125)	52.4 bit (124)	51.2 bit (121)
E-érték		$5 \cdot 10^{-14}$	10^{-13}	$5 \cdot 10^{-13}$	$6 \cdot 10^{-13}$	10^{-12}

F2. táblázat folytatása: A *S. pombe* Sep15 fehérjéinek ortológjai a Pezizomycotina altörzsben

	<i>S. pombe</i>	<i>N. haematococca</i>	<i>T. reesei</i>	<i>B. fuckeliana</i>	<i>F. pseudograminearum</i>	<i>N. crassa</i>
Törzs	972h ⁻	mpVI 77-13-4	QM6a	BcDW1	CS3096	OR74A
Fehérje kódja	SPBC21.04	XP_003049872.1	EGR48217.1	CCD56888.1	EKJ74825.1	XP_965268.1
Fehérje hossza (as)	200	248	258	228	261	278
BLAST hasonlóság		50/93 (53%)	69/134 (51%)	70/149 (47%)	68/142 (47%)	62/148 (42%)
BLAST egyezés		30/93 (32%)	38/134 (28%)	42/149 (28%)	37/142 (26%)	38/148 (26%)
BLAST szünetek		0/93 (0%)	4/134 (3%)	10/149 (7%)	4/142 (2%)	4/148 (3%)
BLAST pontszám		49.7 bit (117)	49.7 bit (117)	47.8 bit (112)	43.9 bit (102)	40.0 bit (92)
E-érték		$4 \cdot 10^{-12}$	$5 \cdot 10^{-12}$	$2 \cdot 10^{-11}$	$5 \cdot 10^{-10}$	10^{-8}

F3. táblázat: A *S. pombe* Sep15 fehérjéinek ortológjai a külső csoportként használt Chytridiomycota törzsben, valamint a Zygomycota és a Basidiomycota törzsekben

	<i>S. pombe</i>	<i>B. dendrobatidis</i>	<i>R. delemar</i>	<i>M. circinelloides</i>	<i>P. strigosozonata</i>	<i>S. commune</i>
Törzs	972h ⁻	JAM81	RA 99-880	B8987	HHB-11173 SS5	H4-8
Fehérje kódja	SPBC21.04	EGF83033.1	EIE88536.1	EPB85356.1	EIN05470.1	XP_003026806.1
Fehérje hossza (as)	200	215	245	255	298	317
BLAST hasonlóság		88/182 (48%)	57/107 (53%)	58/107 (54%)	66/132 (50%)	34/56 (61%)
BLAST egyezés		44/182 (24%)	28/107 (26%)	29/107 (27%)	39/132 (30%)	22/56 (39%)
BLAST szünetek		24/182 (13%)	5/107 (5%)	5/107 (4%)	20/132 (15%)	2/56 (4%)
BLAST pontszám		42.4 bit (98)	48.9 bit (115)	50.8 bit (120)	57.8 bit (138)	46.6 bit (109)
E-érték		10 ⁻⁹	9·10 ⁻¹²	2·10 ⁻¹²	10 ⁻¹⁴	8·10 ⁻¹¹

F3. táblázat folytatása: A *S. pombe* Sep15 fehérjéinek ortológjai a külső csoportként használt Chytridiomycota törzsben, valamint a Zygomycota és a Basidiomycota törzsekben

	<i>S. pombe</i>	<i>W. sebi</i>	<i>F. mediterranea</i>	<i>D. squalens</i>	<i>S. hirsutum</i>	<i>L. bicolor</i>
Törzs	972h ⁻	CBS 633.66	MF3/22	LYAD-421 SS1	FP-91666 SS1	S238N-H82
Fehérje kódja	SPBC21.04	EIM21155.1	EJD07728.1	EJF61606.1	EIM84004.1	XP_001879574.1
Fehérje hossza (as)	200	180	199	414	265	330
BLAST hasonlóság		93/196 (47%)	75/175 (43%)	81/186 (44%)	55/112 (49%)	54/119 (45%)
BLAST egyezés		49/196 (25%)	45/175 (26%)	50/186 (27%)	30/112 (27%)	31/119 (26%)
BLAST szünetek		31/196 (16%)	31/175 (18%)	33/186 (18%)	0/112 (0%)	16/119 (13%)
BLAST pontszám		43.1 bit (100)	42.0 bit (97)	42.4 bit (98)	41.2 bit (95)	39.3 bit (90)
E-érték		4·10 ⁻¹⁰	10 ⁻⁹	3·10 ⁻⁹	4·10 ⁻⁹	2·10 ⁻⁸

A narancssárga háttérű fajokban talált ortológok különböző okok miatt nem szerepelnek a végső törzsfákon.

Az F3. táblázat a 64. oldalon folytatódik

F3. táblázat folytatása: A *S. pombe* Sep15 fehérjének ortológjai a külső csoportként használt Chytridiomycota törzsben, valamint a Zygomycota és a Basidiomycota törzsekben

	<i>S. pombe</i>	<i>F. radiculosa</i>	<i>T. versicolor</i>	<i>C. puteana</i>	<i>C. cinerea</i>
Törzs	972h ⁻	TFFH 294	FP-101664 SS1	RWD-64-598 SS2	okayama7#130
Fehérje kódja	SPBC21.04	CCM05291.1	EIW56174.1	EIW75942.1	XP_001834921.2
Fehérje hossza (as)	200	283	288	184	401
BLAST hasonlóság		41/76 (54%)	37/61 (61%)	56/120 (47%)	39/75 (52%)
BLAST egyezés		25/76 (33%)	23/61 (38%)	32/120 (27%)	24/75 (32%)
BLAST szünetek		5/76 (7%)	3/61 (5%)	19/120 (16%)	11/75 (15%)
BLAST pontszám		38.5 bit (88)	36.2 bit (82)	35.8 bit (81)	35.0 bit (79)
E-érték		$4 \cdot 10^{-8}$	$2 \cdot 10^{-7}$	10^{-7}	$7 \cdot 10^{-7}$

A narancssárga háttérű fajokban talált ortológok különböző okok miatt nem szerepelnek a végső törzsfákon.

F4. táblázat: A Pfam-A programmal végzett doménazonosítás eredményei

Faj	Fehérje kódja	Illesztés		Domén hossza	E-érték
		kezdet	vég		
<i>P. jiroveci</i>	CCJ29994.1	7	149	142	$3,4 \cdot 10^{-24}$
<i>S. complicata</i>	lcl 36403	71	176	105	$8,2 \cdot 10^{-24}$
<i>S. pombe</i>	SPBC21.04	1	189	188	$6,8 \cdot 10^{-64}$
<i>S. japonicus</i>	SJAG_03622	10	184	174	$4,6 \cdot 10^{-43}$
<i>S. octosporus</i>	SOCG_00147	1	186	185	$2,5 \cdot 10^{-46}$
<i>S. cryophilus</i>	SPOG_00748	1	186	185	$2,1 \cdot 10^{-45}$
<i>A. gossypii</i>	NP_984609.1	19	195	176	$8,2 \cdot 10^{-39}$
<i>C. albicans SC5314</i>	XP_712891.1	25	237	212	$1,2 \cdot 10^{-80}$
<i>C. albicans SC5314_2</i>	XP_712923.1	25	237	212	$2,4 \cdot 10^{-79}$
<i>C. albicans WO-1</i>	EEQ45871.1	25	237	212	$1,3 \cdot 10^{-80}$
<i>C. dubliniensis</i>	XP_002418393.1	27	241	214	$2,5 \cdot 10^{-74}$
<i>C. glabrata</i>	XP_449574.1	23	164	141	$2,4 \cdot 10^{-45}$
<i>C. orthopsilosis</i>	CCG22217.1	22	225	203	$3,4 \cdot 10^{-50}$
<i>C. tenuis</i>	EGV62121.1	2	199	197	$2,2 \cdot 10^{-46}$
<i>C. tropicalis</i>	XP_002547295.1	23	222	199	$2 \cdot 10^{-62}$
<i>C. lusitaniae</i>	XP_002615074.1	10	209	199	$9 \cdot 10^{-40}$
<i>D. hansenii</i>	XP_002770775.1	17	222	205	$4,8 \cdot 10^{-50}$
<i>E. cymbalariae</i>	XP_003647717.1	25	176	151	$7,3 \cdot 10^{-40}$
<i>K. lactis</i>	XP_455017.1	46	201	155	$1,6 \cdot 10^{-40}$
<i>K. pastoris</i>	XP_002491169.1	10	150	140	$1,9 \cdot 10^{-35}$
<i>M. farinosa</i>	CCE84454.1	15	218	203	$2,2 \cdot 10^{-44}$
<i>S. cerevisiae S288c</i>	NP_009752.3	21	199	178	$2,5 \cdot 10^{-65}$
<i>S. cerevisiae CEN.PK113-7D</i>	EIW12127.1	21	199	178	$2,7 \cdot 10^{-65}$
<i>S. cerevisiae VL3</i>	EGA87936.1	21	200	179	$1,2 \cdot 10^{-64}$
<i>S. stipitis</i>	XP_001384574.1	8	208	200	$8,1 \cdot 10^{-57}$
<i>S. passalidarum</i>	EGW33037.1	29	235	206	$4,2 \cdot 10^{-57}$
<i>Y. lipolytica</i>	XP_500410.1	15	150	135	$1,6 \cdot 10^{-32}$
<i>A. capsulatus NAm1</i>	XP_001538616.1	1	252	251	$2,6 \cdot 10^{-45}$
<i>A. capsulatus H143</i>	EER45218.1	1	252	251	$5,2 \cdot 10^{-45}$
<i>A. capsulatus G186AR</i>	EEH07092.1	1	252	251	$8,1 \cdot 10^{-46}$
<i>A. dermatitidis ER-3</i>	EEQ87974.1	1	251	251	$2,6 \cdot 10^{-46}$
<i>A. dermatitidis SLH14081</i>	XP_002627781.1	1	251	251	$1,1 \cdot 10^{-46}$
<i>A. oligospora</i>	EGX45958.1	8	249	241	$3,3 \cdot 10^{-27}$
<i>A. flavus</i>	XP_002376999.1	1	242	241	$1,4 \cdot 10^{-62}$
<i>A. fumigatus Af293</i>	XP_751989.1	1	249	248	$2,5 \cdot 10^{-80}$
<i>A. fumigatus A1163</i>	EDP50215.1	1	252	251	$3,7 \cdot 10^{-77}$
<i>A. kawachii</i>	GAA82781.1	1	265	264	$2,9 \cdot 10^{-47}$
<i>A. nidulans</i>	XP_659324.1	4	117	113	$1,1 \cdot 10^{-40}$

Az F4. táblázat a 66. oldalon folytatódik

F4. táblázat folytatás: A Pfam-A programmal végzett doménazonosítás eredményei

Faj	Fehérje kódja	Illesztés		Domén hossza	E-érték
		kezdet	vég		
<i>B. fuckeliana</i>	CCD56888.1	5	205	200	$5 \cdot 10^{-24}$
<i>C. thermophilum</i>	EGS19264.1	11	218	207	$1,5 \cdot 10^{-25}$
<i>C. immitis</i>	XP_001241689.1	1	259	258	$5,7 \cdot 10^{-49}$
<i>C. posadasii</i>	XP_003070639.1	1	262	261	$7,7 \cdot 10^{-48}$
<i>E. dermatitidis</i>	EHY52313.1	1	244	243	$6 \cdot 10^{-35}$
<i>F. pseudograminearum</i>	EKJ74825.1	6	192	186	$1,1 \cdot 10^{-23}$
<i>G. graminicola</i>	EFQ27846.1	11	225	214	$2 \cdot 10^{-25}$
<i>M. oryzae</i>	XP_003716133.1	1	119	118	$6,9 \cdot 10^{-21}$
<i>M. thermophila</i>	XP_003658536.1	9	221	212	$1,2 \cdot 10^{-24}$
<i>M. fijiensis</i>	EME89241.1	9	168	159	$2,1 \cdot 10^{-27}$
<i>N. haematococca</i>	XP_003049872.1	11	189	178	$1,5 \cdot 10^{-24}$
<i>N. fischeri</i>	XP_001267185.1	1	252	251	$8,7 \cdot 10^{-76}$
<i>N. crassa</i>	XP_965268.1	16	152	136	$3,4 \cdot 10^{-22}$
<i>P. brasiliensis Pb03</i>	EEH18086.1	1	242	241	$2,7 \cdot 10^{-47}$
<i>P. brasiliensis Pb18</i>	EEH46986.1	1	242	241	$2,7 \cdot 10^{-47}$
<i>P. chrysogenum</i>	XP_002559238.1	1	244	243	$3,9 \cdot 10^{-53}$
<i>T. terrestris</i>	XP_003656497.1	11	214	203	$1,4 \cdot 10^{-22}$
<i>T. reesei</i>	EGR48217.1	9	189	180	$4,2 \cdot 10^{-25}$
<i>T. melanosporum</i>	XP_002840444.1	79	243	164	$2,7 \cdot 10^{-20}$
<i>V. albo-atrum</i>	XP_003005765.1	15	146	131	$5,3 \cdot 10^{-27}$
<i>V. dahliae</i>	EGY17269.1	15	219	204	$8,7 \cdot 10^{-28}$
<i>C. puteana</i>	EIW75942.1	11	70	59	$8,6 \cdot 10^{-5}$
<i>C. cinerea</i>	XP_001834921.2	31	88	57	$2 \cdot 10^{-4}$
<i>D. squalens</i>	EJF61606.1	165	222	57	$5,1 \cdot 10^{-3}$
<i>F. radiculosa</i>	CCM05291.1	20	124	104	$2,8 \cdot 10^{-2}$
<i>F. mediterranea</i>	EJD07728.1	13	196	183	$1,1 \cdot 10^{-3}$
<i>L. bicolor</i>	XP_001879574.1	23	81	58	$4,2 \cdot 10^{-4}$
<i>P. strigosozonata</i>	EIN05470.1	34	244	210	$7 \cdot 10^{-8}$
<i>S. commune</i>	XP_003026806.1	16	73	57	$7 \cdot 10^{-3}$
<i>S. hirsutum</i>	EIM84004.1	25	208	183	$1,5 \cdot 10^{-8}$
<i>T. versicolor</i>	EIW56174.1	32	214	182	$5,3 \cdot 10^{-4}$
<i>W. sebi</i>	EIM21155.1	11	167	156	$1,6 \cdot 10^{-7}$
<i>M. circinelloides</i>	EPB85356.1	9	181	172	$4,4 \cdot 10^{-11}$
<i>R. delemar</i>	EIE88536.1	11	166	155	10^{-10}
<i>B. dendrobatidis</i>	EGF83033.1	10	178	168	$5 \cdot 10^{-9}$

A piros színnel kiemelt fajok fehérjéiben a Med8 domén nem volt szignifikáns

10. Köszönetnyilvánítás

Elsősorban témavezetőimnek szeretnék köszönetet mondani, akik lehetőséget adtak munkám sikeres elvégzéséhez.

Hálával tartozom **Horváth Annának**, aki végtelen türelemmel, segítőkész támogatással, időt és energiát nem sajnálva foglalkozott velem, és elmagyarázta a bioinformatika alapjait.

Köszönetet szeretnék mondani **Dr. Sveiczer Ákosnak** a belém vetett bizalmáért, folyamatos támogatásáért és biztatásáért, valamint szakmai tanácsaiért.

Köszönettel tartozom **Dr. Sipiczki Mátyásnak** is, aki nem csak a szakmai gyakorlatom keretében, hanem külön konzultációs időpontok alkalmával bővítette szakmai ismereteimet.