

Eötvös Loránd Tudományegyetem Természettudományi Kar
Biológia Doktori Iskola
Elméleti és Evolúcióbíológia Doktori Program



Palatinszky Márton

Mikrobaközösségek klónkönyvtár analízise során
tapasztalható módszertani hibák elemzése, a preferenciális
ígálás kísérletes vizsgálata

– doktori értekezés –

Témavezető:
Márialigeti Károly
(ELTE Mikrobiológiai Tanszék)

Doktori iskola vezetője:
Dr. Erdei Anna
tanszékvezető egyetemi tanár
(ELTE Immunológiai Tanszék)

Doktori program vezetője:
Dr. Szathmáry Eörs
egyetemi tanár
(ELTE Növényrendszertani
és Ökológiai Tanszék)

Készült az ELTE TTK Mikrobiológiai Tanszékén, 2011-ben.

Tartalomjegyzék

Tartalomjegyzék	1
I. Bevezetés	4
II. Irodalmi áttekintés	6
II.1 Klónkönyvtár elemzés a mikrobaközösségek vizsgálatában	6
II.1.1 A nukleinsav alapú módszerek megjelenése	6
II.1.2 Korai klónkönyvtár alapú megközelítések	7
II.2 TA-klónozáson alapuló mikrobiális közösségvizsgálat	8
II.2.1 A mikrobiális ökológiában általánosan elterjedt TA-klónozó rendszerek	8
II.2.2 A TA-klónozás jelene és jövője	11
II.3 A közösségi nukleinsav kivonás hibalehetőségei	13
II.4 A multitemplát PCR közösségszerkezet torzító hatásai	15
II.4.1 Genomi tulajdonságok hatása a PCR amplifikációra	16
II.4.2 A PCR primerek szelektivitásából fakadó differenciális amplifikáció	18
II.4.2.1 Az oligonukleotid primerek	18
II.4.2.2 A primerek degeneráltsága	18
II.4.2.3 Primerek illeszkedési hibái és az annealációs hőmérséklet	20
II.5 A közösségi PCR termék TA-klónozása során fellépő hibalehetőségek	21
II.5.1 TA-klónozó vektorrendszerek méretpreferenciája	21
II.5.2 Inzertív szakaszok hatása a DNS – fehérje interakciókra	23
II.5.3 Az inzertált génszakasz hatása a transzformált sejtekre	24
II.5.4 Különböző klónozási stratégiák összehasonlítása szelektivitás szempontjából	24
II.6 Közösségelemzés során gyakran vizsgált gének és genomszakaszok	25
II.6.1 A riboszomális kis alegység RNS-ét kódoló gén	25
II.6.2 Riboszomális intergenikus spacer szakaszok	25
II.6.3 Funkciógének	27
III. Célkitűzések	29
IV. Anyag és módszer	31
IV.1 In silico hosszheterogenitás vizsgálatok	31
IV.1.1 A vizsgált adatbázisok	31
IV.1.2 A LengHet program, és az alkalmazott paraméterek	32
IV.2 Modellközösségek	33

IV.2.1	A modellrendszerekhez felhasznált törzsek, fenntartásuk, genomi DNS izolálás ..	34
IV.2.2	Az inzertek előállítás PCR segítségével (általános PCR protokoll).....	35
IV.2.3	A modellkísérletek inzertkeverékei, klónkönyvtárak áttekintése	37
IV.2.4	Az inzertek kópiaszámának meghatározása, az inzertkeverékek elkészítése	41
IV.2.5	Kapillaris elektroforézis és lézergesztesű fluoreszcencia intenzitásmérés.....	42
IV.2.6	A kvantifikálási és keverési eljárások validálása	43
IV.2.7	Klónkönyvtárak készítése.....	44
IV.2.7.1	Ligálás és transzformálás pGem-T vektorrendszerrel.....	44
IV.2.7.2	Ligálás és transzformálás TOPO-TA vektorrendszerrel.....	45
IV.2.8	A klónok azonosítása.....	46
IV.3	Környezeti minta.....	48
IV.3.1	Mintavétel, közösségi DNS kivonása.....	49
IV.3.2	Hosszheterogenitás-PCR (LH-PCR) vizsgálat.....	50
IV.3.3	Közösségi PCR termék előállítása, klónozás.....	51
IV.3.4	A klónkönyvtárak feldolgozása, klónok csoportosítása ARDRA-val.....	51
IV.3.5	Az ARDRA csoport reprezentáns klónok bázissorrend elemzése	52
IV.3.6	A reprezentáns klónok LH-PCR vizsgálata	54
IV.3.7	A klónkönyvtárak összehasonlító elemzése.....	54
IV.3.7.1	GC-tartalom, átlagos inzerthossz	54
IV.3.7.2	Közösség szerkezetek LH-PCR alapú összevetése	54
IV.3.7.3	A klónok filogenetikai megoszlásának elemzése.....	55
IV.3.7.4	Hasonlósági indexek, statisztikai elemzés	55
IV.3.7.5	Különbség szignifikancia teszt (Libshuff elemzés).....	55
V.	Eredmények és értékelésük.....	58
V.1	In silico hosszheterogenitás vizsgálatok.....	58
V.1.1	A teljes 16S rRNS gén hosszheterogenitása	58
V.1.2	A 16S rDNS első harmadának hosszheterogenitása.....	61
V.2	A modellrendszerek validálása	63
V.3	A két vektorrendszer vizsgálata modellközösségeken	64
V.3.1	A pGem-T vektorrendszer inzerthossz preferenciái.....	64
V.3.2	A pGem-T vektorrendszer preferenciális ligációjának függése a ligálási paraméterektől.....	67
V.3.3	A TOPO-TA és a pGem-T vektorrendszer preferenciális ligációjának összehasonlítása	68

V.4 A két vektorrendszer vizsgálata környezeti mintán	70
V.4.1 A pGem-T és TOPO-TA klónkönyvtárak jellemzése	70
V.4.2 A pGem-T és TOPO-TA vektorrendszerekkel nyert közösségszerkezeti képek összehasonlítása az LH-PCR eredményével.....	72
V.4.3 A pGem-T és TOPO-TA klónkönyvtárak filogenetikai összetételének összevetése	75
V.4.4 A pGem-T és TOPO-TA klónkönyvtárak filotípus tartalmának összehasonlító elemzése (Libshuff elemzés)	76
V.5 Ajánlások a pontosabb közösségszerkezet leírásokhoz.....	78
VI. Összegzés, kitekintés.....	79
VII. Idézett irodalom	81
VIII. Internetes hivatkozások jegyzéke.....	92
IX. Rövidítésjegyzék.....	93
X. Összefoglaló.....	94
XI. Abstract.....	95
XII. Köszönetnyilvánítás	96

I. Bevezetés

Az elmúlt két évtizedben a nukleinsav alapú közösségvizsgáló módszereknek köszönhetően gyökeresen megváltozott a prokarióták diverzitásáról alkotott kép. A klasszikus, tenyésztésen alapuló technikák szelektivitásukból kifolyólag a mikroorganizmusoknak csak igen szűk körét tették vizsgálhatóvá – legtöbb esetben kevesebb mint egy százalékát a közösségalkotó fajoknak (Torsvik *et al.*, 1990). A mikrobiális ökológia eszköztárában a kilencvenes évek elején jelent meg a polimeráz láncreakcióval (PCR-rel) felszaporított génszakaszok klónkönyvtárak elemzése, mely eljárás azután igen gyorsan és széles körben elterjedt, s mind a mai napig a közösségszerkezet leírás egyik alapvető eszköze maradt (Mering *et al.*, 2007).

Mint minden módszer, természetesen a közösségi nukleinsav izoláláson majd PCR amplifikáción alapuló klónkönyvtár elemzés sem mentes a hibáktól, a valós közösségszerkezeti képet torzító szelektivitástól. A technika komplexitása és több lépcsős természete megköveteli, hogy az egyes lépések hibalehetőségeit egyenként is mélységben vizsgáljuk. Ez elég korán megtörtént a nukleinsav kivonási eljárásokkal kapcsolatban – itt például a közösségalkotó fajok eltérő sejtfálszerkezetéből fakadó különböző hatékonyságú feltáródás torzítását kellett feltérképezni (Frostegard *et al.*, 1999). A technika meghatározó lépése, a PCR amplifikáció hibalehetőségeivel igen sok kutatócsoport, többek között az ELTE Mikrobiológiai Tanszéke is behatóan foglalkozott (Sipos *et al.*, 2007). Szakdolgozóként magam is részese lehettem a PCR primerek illeszkedési hibáiból fakadó preferenciális amplifikáció elemző munkának (Palatinszky, 2004). A téma szakirodalmában immáron igen gazdag, s a publikált eredmények fényében kijelenthető: a PCR hibáinak felderítésére komoly erőfeszítéseket tett a tudományos közösség az elmúlt két évtizedben.

A közösségi PCR termék klónozásának hibaelemző szakirodalmában viszont igen szegényes, s hasonló hiányosságokkal bír mint a polimeráz láncreakció kinetikájának korai, általános leírásai: mindkét módszert alapvetően homogén, egyféle DNS templátot tartalmazó rendszerek ligálása illetve amplifikációja esetén vizsgálták behatóbban. A környezeti mikrobiológiai alkalmazások során azonban legtöbbször igen heterogén DNS keverék szolgál a vizsgálatok kiindulásaként. A klónkönyvtár készítéséhez alkalmazott közösségi PCR termék amplitonjai mind szekvenciájukban, mind pedig

hosszukban is változatosak – mint később látni fogjuk – még az evolúciósan igen konzervált génszakaszok esetében is. Így fennáll a lehetősége annak, hogy a közösségszerkezeti arányok a ligálás esetleges szelektivitása miatt a klónozás során torzuljanak. Ezen túlmenően, a közösségszerkezeti vizsgálatok során leggyakrabban alkalmazott két TA-klónozó rendszer alapvetően különbözik a ligálást végző enzim tekintetében, felvetve annak a lehetőségét, hogy a kétféle vektorrendszerrel nyert eredmények esetleg eltérő jellegű torzításoknak vannak kitéve.

Úgy véljük, a TA-klónozás torzításainak feltárása elengedhetetlen az e módszerrel végzett közösségszerkezeti vizsgálatok hitelesebb interpretálásához, valamint a korábban publikált eredmények kritikus szemléletéhez.

II. Irodalmi áttekintés

II.1 Klónkönyvtár elemzés a mikrobaközösségek vizsgálatában

II.1.1 A nukleinsav alapú módszerek megjelenése

A mikrobiális ökológia klasszikus – törzsek izolálásán, fenotipikai, élettani, biokémiai tesztek elvégzésén alapuló – módszerei több elvi és gyakorlati problémával terheltek. A laborkörülmények között előállítható tenyésztési feltételek mellett önállóan általában csak a közösségalkotó szervezetek töredéke tud szaporodni. Nehézségként vizsgálni olyan fajokat, amelyek csak szoros szimbiózisban képesek fennmaradni, vagy speciális, ismeretlen növekedési faktorokat igényelnek. A tenyésztéses diverzitásvizsgálatok így azokat a fajokat azonosítják a közösség domináns tagjaiként, amelyek a tenyésztési körülményekhez legjobban képesek alkalmazkodni. Tekintve, hogy az egyes fajok (törzsek) a természetben igen változatos mikrohabitatokban, és szaporodásukhoz nélkülözhetetlen interakciókban élhetnek, a teljes mikrobiális diverzitást még csak megközelítőleg sem lehet ilyen módon leképezni.

Már a mikroszkópos sejtszámlálások is sokkal nagyobb faj- és egyedszámot sejtettek, mint ami az adott mintákból kitenyészthető volt – ezt a jelenséget „a nagy telepszámlálási anomália”-ként aposztrofálták („The great plate count anomaly”, Staley és Konopka, 1985) – ám a feltételezések alátámasztására a tenyésztéstől független módszerek megjelenéséig várni kellett.

A környezeti baktériumközösségek vizsgálatában az egyik legkorábbi nukleinsav alapú megközelítés a teljes közösségi genomiális DNS denaturációs és rehibridizációs tulajdonságainak vizsgálata volt, mellyel a diverzitás mértékét lehetett megbecsülni (Torsvik *et al.*, 1990), azonban az egyes közösségalkotók azonosítására nem volt alkalmas.

Carl Woese úttörő munkássága felfedte, hogy a Földön élő szervezetek filogenetikai rendszerbe sorolhatók riboszomális RNS-szekvenciák összehasonlító elemzése alapján (Woese, 1987). A korábbi filogenetikai vizsgálatok a citokró-m-c fehérje aminosavsorrend elemzésén alapultak. Az rRNS molekulát kódoló gének filogenetikai célú vizsgálatát a génszakasz több hasznos tulajdonsága is előnyössé tette. A gén minden sejt

életformában (univerzálisan) előfordul (a *Bacteria*, az *Archaea* és az *Eukarya* csoportokban is). Mivel a sejt működésében funkciója esszenciális, relatíve kicsi a mutációs rátája, evolúciósan igen konzervatív. Viszont a génen belül több, eltérő fokú variabilitást mutató régió is van, melyek elemzése révén különböző szintű illetve mélységű filogenetikai felbontás válik lehetővé.

Pace és munkatársai rámutattak arra, hogy az így szerzett filogenetikai ismeretanyag felhasználható olyan oligonukleotidok tervezéséhez, amelyek specifikusan hibridizálnak bizonyos bakteriális taxonok tagjainak örökítőanyagához (Pace *et al.*, 1986), s ezáltal megnyílt az út a nukleinsav alapú identifikálás irányába. A mikroorganizmusok 16S rDNS szekvenciáira specifikus fluoreszcensen jelölt oligonukleotidok alkalmazása a környezeti minták mikroszkópos vizsgálatában (Fluoreszcens In Situ Hibridizáció, FISH) igen hamar megjelent (Amann *et al.*, 1990). Azonban a közösségalkotók széles körű azonosításához és leírásához egy újabb technológia bevezetésére volt szükség.

A legújabbkori biológiai tudományra nézve talán legnagyobb hatású technológiai áttörést a Kary B. Mullis által 1985-ben feltalált polimeráz láncreakció (PCR) hozta. A felfedezés jelentőségét jelzi, hogy Mullis ezért 1993-ban a Nobel-díjban is részesült. A módszer mikrobiális ökológiai alkalmazására sem kellett sokáig várni. Elsőként Giovannoni és munkatársai tengerben élő planktonikus baktériumközösség szerkezetét tárták fel a 16S rDNS gén amplifikációja és klónkönyvtár elemzése segítségével (Giovannoni *et al.*, 1990). Ezután a riboszomális- vagy egyéb funkciógének amplifikációján és klónkönyvtár analízisén alapuló módszerek már a kilencvenes években igen széles körben elterjedtek (Hugenholtz *et al.*, 1998), és mind a mai napig megtartották központi helyüket a mikrobiális közösségleírásban.

II.1.2 Korai klónkönyvtár alapú megközelítések

A kilencvenes évek elején a klónkönyvtár alapú módszereknek még igen sokféle változatát alkalmazták. Egyes megközelítések a PCR amplifikáció mellőzésével, a közösségi DNS fragmentálásával, és véletlenszerű direkt klónozásával dolgoztak. Az egyik legelső ilyen munka során klónozó vektorként lambda-fág eredetű konstrukciókat alkalmaztak, és a 10-20 kilobázis hosszúságú DNS fragmentumokat tartalmazó klónok között radioaktívan jelölt oligonukleotidok hibridizálásával kerestek 16S rDNS

géneket (Schmidt *et al.*, 1991). Találkozhatunk még nagyméretű közösségi DNS fragmentumok BAC-klónkönyvtárainak vizsgálatával is (Stein *et al.*, 1996), mely módszer – bár egy ideig háttérbe szorult – napjainkban az új generációs, nagy áteresztőképességű bázisrendrend elemzési technikák megjelenésével mintha reneszánszát élné (Béja, 2004).

A 16S rDNS szakasz PCR amplifikációval kombinált klónkönyvtár elemzésének korai példái még nem TA-klónozást alkalmaztak, hanem „tompá végű” (blunt-end) vagy „ragadós végű” (sticky-end) klónozást. Ez utóbbihoz az amplifikációhoz használt primerek 5' végeire szintetizáltak adapter szekvenciákat, és ezzel komplementer végű vektorokat használtak (Liesack és Stackebrandt, 1992). A TA-klónozás kifejlesztésével (Holton és Graham, 1991) azonban a hangsúly egyre inkább ezen új technikára helyeződött, a tompa végű klónozásnál nagyságrendekkel nagyobb hatékonysága és a ragadós vég klónozásnál egyszerűbb kivitelezhetősége miatt (Osborn *et al.*, 2005). A 2000-es években és napjainkban a molekuláris mikrobiológia szótárában a „klónozás” (cloning) vagy „klónozás és bázisrendrend elemzés” (cloning and sequencing) gyakran használt terminusok, és alapvetően a PCR amplifikált génszakaszok TA-klónozását, valamint az inzertek Sanger félé szekvenálását jelentik.

II.2 TA-klónozáson alapuló mikrobiális közösségvizsgálat

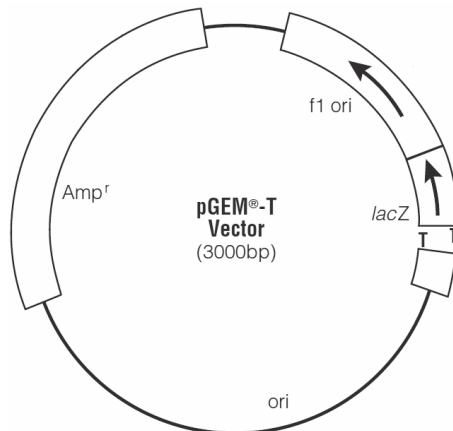
II.2.1 A mikrobiális ökológiában általánosan elterjedt TA-klónozó rendszerek

A közösségi PCR-termékek klónkönyvtáras elválasztására és elemzésére szinte kizárólag a pGem-T (Promega, WI, USA) illetve a TOPO-TA (Invitrogen, CA, USA) klónozó rendszereket alkalmazzák a mikrobiális ökológia terén (Osborn *et al.*, 2005).

A TA-klónozás során használt linearizált vektorok DNS szálai 3' végeiken egy-egy túlnyúló timin nukleotidot tartalmaznak. Bizonyos DNS polimerázok (pl. *Taq*, *Tfl*, *Tth*) terminális transzferáz aktivitással rendelkeznek, amelynek révén az általuk létrehozott DNS szál 3' végére egy túlnyúló adenint kötnek (Clark, 1988). A mindkét végükön túlnyúló 3' adeninekkel rendelkező PCR termékek megfelelő enzim segítségével képesek a túlnyúló 3' timinekkel rendelkező vektorokba bekötődni,

cirkuláris plazmidot létrehozva ezáltal. (Megjegyzendő, hogy az 5'-3' exonukleáz aktivitással rendelkező DNS polimerázok [pl. *Pfu*, *Pwo*] nem mutatnak terminális transzferáz aktivitást, így közvetlenül nem alkalmasak TA-klonozáshoz használható PCR termékek előállítására.) Az inzertálódás első lépéseként az inzert egyik végének túlnyúló adeninje kialakít egy kétszeres hidrogénhid kötést a vektor egyik végének túlnyúló timinjével. Ezt a TOPO-TA vektorrendszer esetében a cukor-foszfát lánc azonnali kovalens összekötése követi, amit a vektorhoz kovalensen kötött topozomeráz-I enzim katalizál. A pGem-T vektorrendszer esetében viszont az oldatban szabadon mozgó T4 DNS ligáz kötődése és aktivitása hozza létre a kovalens kapcsolatot az inzert és a vektor között.

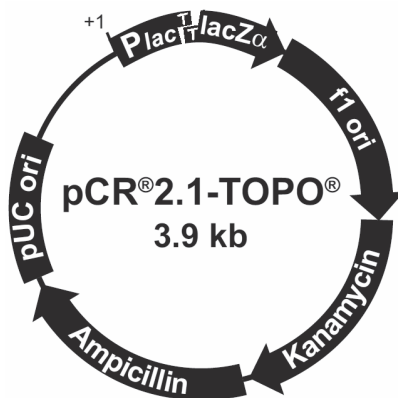
A pGEM-T vektor a pGEM-5Zf(+) plazmid vektor *EcoRV* enzimes hasításával és túlnyúló 3' timinek addíciójával jön létre. Hossza 3000 bázispár, tartalmaz ampicillin rezisztencia gént, valamint a β -galaktozidáz α -alegységét kódoló gént (ezen belül jön létre az *EcoRV* hasítás után az inzertálódási hely). A vektor sematikus felépítését az 1. ábra mutatja.



1. ábra. A pGem-T vektor felépítése. Ori: replikációs origó; *lacZ*: β -galaktozidáz α -alegység gén; *f1 ori*: F1-fág replikációs origó; *Amp^r*: ampicillin rezisztencia gén. (A kép forrása: pGEM-T and pGEM-T Easy Vector Systems Technical Manual, Promega).

A pGem-T klónozó rendszer az inzertek ligálására a T4 bakteriofág DNS ligáz enzimét használja, amely ATP jelenlétében képes foszfodiészter kötést létrehozni DNS szálak 5' (foszfát) és 3' (hidroxil) végei között.

A TOPO-TA vektorrendszer egy linearizált pCR2.1 fágemid vektort tartalmaz, amelynek végein a túlnyúló 3' timin nukleotidokon kovalensen kötött vírus (*Vaccinia*) eredetű topoizomeráz-I enzim van. Ez egy DNS giráz, amely specifikusan kötődik a kettősszálú DNS 5'-CCCTT bázissorrendű helyeihez, és a felismerőhely 3' végén a foszfodiészter kötést elhasítja, majd kovalens kötésben marad a DNS szállal a foszfát csoporton keresztül. Azonban ez a reakció reverzibilis: egy DNS szál 5' hidroxil csoportja az enzim és a DNS kovalens kötését megtámadva összekapcsolódhat a foszfát csoporttal, és a topoizomeráz ledisszociál (Shuman, 1994). A pCR-2.1-TOPO vektor hossza 3929 bázispár, tartalmaz ampicillin és kanamicin rezisztancia gént, valamint β -galaktozidáz α -alegységét kódoló gént (ezen belül van az inzertálódási hely). A vektor sematikus felépítését a 2. ábra mutatja.



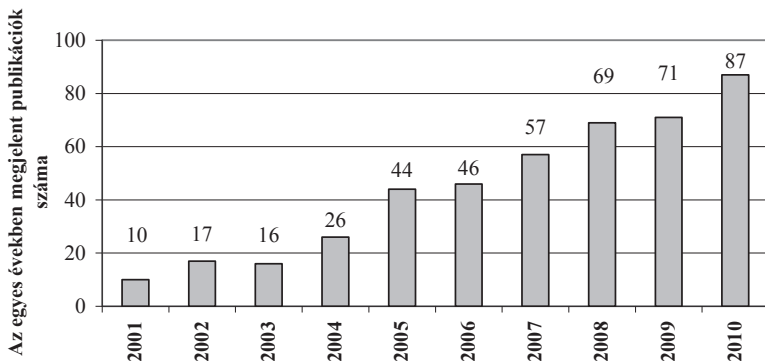
2. ábra. A pCR2.1-TOPO vektor felépítése. pUC ori: replikációs origó; *Plac*: *lacZ* α promóter; *lacZ* α : β -galaktozidáz α -alegység gén; *f1 ori*: F1-fág replikációs origó; (A kép eredetijének forrása: TOPO-TA Cloning User Manual, Invitrogen)

A transzformált kompetens sejtek szelekcióját mindkét vektorrendszer esetében ampicillin tartalmú táptalajon végzik. Az antibiotikum hatása miatt a nem transzformált sejtek képtelenek növekedni, viszont a transzformánsok szaporodása a vektoron levő ampicillin-rezisztancia gén segítségével biztosított. A sikeres ligálást mindkét vektorrendszer esetében úgynevezett „kék-fehér szelekció” útján ellenőrzik. A módszer

alapja, hogy sikeres ligálás esetén az inzert a LacZ α génbe épül be. Ezzel megszakítja a β -galaktozidáz α alegységét kódoló szekvenciát, és így nem expresszálódik működőképes β -galaktozidáz a sejtben. A szelektív táptalajban egy szabad állapotban kék színű molekulával kovalensen kötött szénhidrát van (5-bromo-4-kloro-3-indolil- β -D-galaktopiranozid, „X-Gal”), amelyet a β -galaktozidáz bontani képes. A hasítási termék (5-bromo-4-kloroindol) kék színű. A β -galaktozidáz expressziót a táptalajhoz adott izopropil- β -D-1-tiogalaktopiranoziddal (IPTG-vel) indukálják. Az inzertet tartalmazó vektorokkal bíró sejtek β -galaktozidáz enzim hiányában az X-Gal-t nem képesek hasítani, így telepeik fehér színt mutatnak, míg az inzertet nem tartalmazó klónokat a megjelenő hasítási termék kék színe jelzi.

II.2.2 A TA-klónozás jelene és jövője

A mikrobiális közösségek részletes leírásának legáltalánosabban használt módszere a kilencvenes évek végétől a PCR-rel amplifikált génszakaszok TA-klónkönyvtár elemzése volt (Mering *et al.*, 2007). Nem hagyhatjuk azonban figyelmen kívül, hogy az elmúlt években a mikrobiális ökológia területén egy újabb forradalom bontakozott ki: az új generációs, nagy áteresztőképességű szekvenálási technikák újabb távlatokat nyitottak a közösségek összetételének, genetikai diverzitásának, ahogyan a korszak hívószava összefoglalja, a metagenomjának megismerésében (Cardenas és Tiedje, 2008). Bár ezen új technikák sok szempontból felülmúlják a hagyományos, PCR alapú klónkönyvtár elemzés teljesítményét (Metzker, 2009), a módszer egyelőre nem mutatja az elavultság jeleit, használata, sőt használatának terjedése töretlen. Amennyiben az ISI Web of Science publikációs adatbázisában bakteriális közösségi klónkönyvtár analízist alkalmazó cikkekre keresünk (a keresőszó mezőben „clone AND library AND bacterial AND community”), azt tapasztaljuk, hogy az elmúlt tíz évben a vonatkozó publikációk évenkénti mennyisége folyamatosan növekedett (3. ábra).



3. ábra. Az elmúlt tíz évben az adott évfolyamokban megjelent, közösségi klónkönyvtár elemzést alkalmazó publikációk az ISI Web of Science adatbázisában. A keresőszó mezőben a „clone AND library AND bacterial AND community” kombinációt alkalmaztuk, tehát kizárólag a mind a négy kifejezést egyszerre tartalmazó absztraktokhoz tartozó cikkeket kerestük.

Mivel az egyes években megjelent publikációk számának alakulása még csak „plató fázist” sem mutat, úgy tűnik, hogy a klónkönyvtár analízis egyelőre kizorulófélben sincs, sőt továbbra is egyre szélesebb körben alkalmazzák. A jelenség háttérében több tényező is állhat. Az új generációs, nagy áteresztőképességű („High Throughput”, HT) bázissorrend elemző technikák egyik legnagyobb hátránya jelenleg, hogy bázissorrend leolvasási hosszai igen rövidek. Az Illumina SBS (Illumina, San Diego, CA, USA) technológia átlagosan körülbelül 75 bázispárt, míg a SOLiD (Applied Biosystems, Foster City, CA, USA) körülbelül 35 bázispárt olvas le egyetlen szekvenciatorédekről. Ennél lényegesen nagyobb átlagos leolvasási hosszai miatt a Roche emulziós PCR-rel kombinált „454” piroszekvenálás technikája terjed a leginkább a mikrobiális ökológiában (Hugenholtz és Tyson, 2008). Ennek legfejlettebb verziója, a GS FLX Titanium alkalmazásával körülbelül 400 bázispáros átlagos leolvasási hossz érhető el, viszont átlagosan 4 nukleotidnyi leolvasási hibával számolhatunk ezen a szakaszon (gyári adat, 454 Life Sciences, Roche, Branford, CT, USA). A jelen dolgozat írása idején legújabbnak számító technológia, az Ion Torrent (Life Technologies, Carlsbad, CA, USA) is csak 2012-re ígéri a 400 bázispár feletti leolvasási hossz elérését. A rövid leolvasási hosszokból fakadó hátrányt a HT bázissorrend elemző technológiák többszázszoros nagyságrendű leolvasással, erős redundanciával kompenzálják. Ez egy törzs teljes genomjának elemzése esetén működőképes

megközelítés, hiszen kellően nagy mennyiségű töredékből a teljes genom konszenzus szekvenciája egy idő után összeáll, az átfedő szakaszok nagy biztonsággal elhelyezhetőek, a leolvasási hibák felismerhetőek. Viszont egy baktériumközösség esetében még a gyakori, rokon szervezetekből származó, homológ gének között is sok apró eltérés van, így a jelenleg elképzelhető kapacitásoknál nagyságrendekkel nagyobb mennyiségű leolvasás kellene a teljes közösségi génállomány minden szakaszának többszörös leolvasású, redundáns elemzéséhez, a „konszenzus metagenom” összeállításához. A realitások talaján maradván, a leolvasási hibák által létrehozott bizonytalanság miatt még az evolúciósan konzervált gének esetében sem garantálható az egyes filotípusok biztonságos elkülönítése (Huse *et al.*, 2007). Ezen a gyengeségen bizonyos kutatások nem meglepő módon a hagyományos TA-klónozás és a piroszekvenálás alkalmazásának kombinációjával próbálnak felülkerekedni (Jones *et al.*, 2009).

A gyermekcipőben járó nagy áteresztőképességű bázissorrend elemző technikák természetesen további fejlődés után, és költségvonzataik nagymértékű csökkenése révén később a mikrobiális ökológia elterjedt és hatékony eszközeivé válhatnak. Nagyon valószínűnek tűnik azonban, hogy a TA-klónozáson alapuló eljárások nem fognak gyorsan visszaszorulni, és hosszú távon is maradnak olyan módszertani „niche”-ek, ahol alkalmazásuk továbbra is célszerű. Ilyen lehet a nagyszámú minta PCR alapú ujjenyomat módszerekkel történő feldolgozásának kiegészítése bázissorrend információkkal (Joo *et al.*, 2010), vagy még inkább az új primerek specificitási körének ellenőrzése (Gantner *et al.*, 2010). Ez utóbbi publikáció egy figyelemre méltó mondattal nyit: „Next generation sequencing technologies for in depth analyses of complex microbial communities rely on rational primer design based on up-to-date reference databases” – és kitűzött célját (újjonnan tervezett *Archea* specifikus 16S rDNS primerek validálását) klasszikus TA-kónozással, és Sanger-féle bázissorrend elemzéssel valósítja meg.

II.3 A közösségi nukleinsav kivonás hibalehetőségei

A klónkönyvtáras közösségleírások első lépése a DNS, illetve RNS kivonás, így ennek minősége alapvető jelentőséggel bír a vizsgálat sikerességének szempontjából. Fontos elvárás, hogy a DNS izolátumban az egyes közösségalkotók genomi DNS-einek

arányai minél pontosabban reprezentálják az eredeti közösségszerkezeti arányokat. Ezen kívül, a közösségi nukleinsav koncentrációja illetve tisztasága nagymértékben meghatározza a PCR (vagy reverz transzkripció, stb.) működését.

Szilárd fázist is tartalmazó mintatípusok (talaj, üledék, szövetek) esetében kétféle megközelítést alkalmaznak. Az úgynevezett indirekt eljárások során a sejteket előbb elválasztják a minta egyéb részeitől (például folyadékban rázatással), s csak ezután tárják fel a sejteket és izolálják a nukleinsav frakciót (Holben *et al.*, 1988). Ezzel szemben a direkt nukleinsav extrakció során a minta eredeti mátrixában történik meg a sejtek lizálása, majd a nukleinsavak elválasztása (Ogram *et al.*, 1987). Ez esetben a mintának igen fontos paramétere a DNS adszorpciós képessége, és adszorbeáló felületének nagysága. A feltárt sejtek örökítőanyaga a talajkolloidok felületére kikötődhet, s ez az extrahálható DNS mennyiségi arányait erősen befolyásolhatja. Másrészt, elhalt sejtek DNS-e igen hosszú időig képes épségben megmaradni agyagszemcsék felületén, és az extracelluláris DNS az élő sejtek örökítőanyagával együtt extrahálva hamis képet eredményezhet az aktív baktériumközösségről (Ogram *et al.*, 1994). Az indirekt eljárások előnye, hogy nagyobb tisztaságú, és kevésbé töredezett DNS nyerhető velük (Sorensen *et al.*, 2009), viszont hozamuk alulmarad a direkt módszerekével szemben. Ennél komolyabb hibájuk, hogy a direkt módszereknél erősebben torzítják a közösségszerkezeti képet, mivel az egyes fajok eltérő mértékben kötődhetnek a minta mátrixához, esetleg eleve zárt, nehezen hozzáférhető mikrokörnyezetekben élnek. Mindkét megközelítés esetében figyelembe kell venni, hogy az eltérő sejtfallszerkezetű közösségalkotók különböző hatékonysággal tárolhatnak fel az izolálás során, így az ellenállóbb sejtekkel bíró fajok alulreprezentáltak lehetnek a DNS izolátumban. Ennek elkerülésére a fizikai (ultrahang, sejtmalmos ráztatás), enzimatis (lizozim és proteináz-K kezelés) és kémiai (detergens, pl. SDS) kezelések kombinációját ajánlják (Frostegard *et al.*, 1999). A talajminták őrlése, vagy üvegyöngy hozzáadása mellett történő nagyfrekvenciájú ráztatása kellő mértékben feltárja a talaj aggregátumait, hozzáférhetővé téve ezáltal a mikrobákat a kémiai lízis számára. Az őrlés, vagy ráztatás a sejt mérettől függően a sejteltérést is elvégezheti, ám a túlzott mechanikus behatás a DNS fragmentálódásához vezethet. Magas huminsav tartalmú talajminták esetén előfordulhat, hogy a DNS izolálás és tisztítás után is a mintában maradó huminsavak a PCR-amplifikációt gátolják. Az ilyen mintákhoz az izolálás során polivinil-polipirrolidint (PVPP) szoktak adagolni, mivel a PVPP komplexálja, s ezáltal ártalmatlanítja a huminsavakat. Napjainkban több gyártó is kínál kifejezetten környezeti

mikrobiológiában előforduló mintatípusokra optimalizált nukleinsav izoláló kiteket, amelyek az említett eljárások mellett a DNS frakció szilikát mátrixos tisztítását is biztosítják – általában PCR amplifikációra alkalmas minőségű templátot nyújtva ezáltal.

A közösségi DNS kivonásánál nem lehet kizárni annak a lehetőségét, hogy a mintában levő kitarító képleteket, spórákat is feltárjuk. Ezért amennyiben kizárólag a közösség aktív tagjait szeretnénk leírni, célszerű RNS-alapon elvégezni a vizsgálatokat. A DNS-nél jóval rövidebb életidejű RNS molekulák izolálására is kidolgoztak kielégítő eredményt nyújtó eljárásokat (Hurt *et al.*, 2001).

Bár csak közvetve kapcsolódik a fejezet témájához, de meg kell említenünk a molekuláris közösségvizsgálatok során alkalmazott reagensekből származó DNS kontamináció veszélyét. Tanner és munkatársai (1998) kimutatták, hogy DNS minta hozzáadása nélkül elvégezve egy 16S rDNS alapú klónkönyvtár elemzést, több különböző baktériumtörzs riboszomális RNS génjeit is sikerült detektálniuk.

II.4 A multitemplát PCR közösségszerkezet torzító hatásai

Minden PCR alapú közösségvizsgálati módszer központi, leginkább meghatározó lépése a vizsgálandó génszakasz amplifikációja. A PCR reakciókinetikáját homogén templát amplifikálása esetén pontosan leírták (Kubista *et al.*, 2006), azonban a közösségi DNS egy diverz templátkeverék, melynek amplifikálása során egy sokkal komplexebb rendszerrel állunk szemben. Ideális esetben a PCR-termékben a különböző közösségalkotókból származó ampliconok mennyiségi arányai hűen tükrözik a kiindulási templátarányokat, ez azonban több tényező miatt sem valósul meg (von Wintzingerode *et al.*, 1997). Von Wintzingerode a főbb okok között említi a primerkötőhelyek hozzáférhetőségének és a primer-templát hibridek kötődési energiájának eltéréseit a különböző közösségalkotók esetében, valamint a lánchosszabbítás hatékonyságának függését a templátok bázisrendjétől.

A multitemplát PCR-t befolyásoló tényezőket a molekuláris környezeti mikrobiológiában elnyert központi szerepe miatt általában a 16S rRNS gén amplifikációja esetén vizsgálták, így alábbi áttekintésünk nagy része is ezen gén vizsgálatához kapcsolódik. Azonban a leírtak legtöbb esetben ugyanúgy vonatkoztathatóak egyéb funkciógének vagy gének közötti nem kódoló régiók (intergenikus spacer szakaszok) PCR amplifikációjának esetére is.

A preferenciális amplifikáció jelenségét Suzuki és Giovannoni (1996) kéttagú modellközösségekben, mesterségesen beállított templátarányokkal vizsgálta. Tapasztalataik szerint a PCR termékben az amplikonok arányai sok esetben az eredetileg beállított arányoktól függetlenül az 1:1 arány felé tolódtak el. A jelenség összefüggést mutatott a kiindulási templát összmennyiségével, illetve a PCR időtartamával, s háttérben feltételezhetően a PCR termékek reannulációjából fakadó gátlás áll (Mathieu-Daude *et al.*, 1996). Suzuki és munkatársai eredményeik alapján kinetikai modellt készítettek, amely jól prediktálta a PCR templát-reannulációból fakadó torzítását. A PCR során keletkező amplikonok a ciklusszám emelkedésével egyre nagyobb koncentrációban lesznek jelen a reakcióterfogatban. Az egymással szekvenciájukban teljesen komplementer amplikonok egymással nagyobb hatékonysággal annealálódnak, mint az amplikonoknál jóval rövidebb primerekkel. Így a primerek kötődése a nagy koncentrációban levő templátokhoz egyre kisebb hatékonysággal fog történni a PCR emelkedő ciklusszámával, míg a kezdetben kisebb koncentrációban jelenlevő templátok amplifikációja kevésbé lassul. Ez összességében a kezdeti változatos mennyiségi arányokat az 1:1 arány felé tolja el. A probléma megoldásaként a PCR ciklusszámának csökkentését, vagy a templát kiindulási koncentrációjának csökkentését javasolták. Más szerzők viszont diverz közösségi DNS templátok esetében ezt a torzítást még nagy ciklusszámok esetén sem tapasztalták klónkönyvtárak (Acinas *et al.*, 2005), illetve szemikvantitatív ujjlenyomat vizsgálatok során (Lueders *et al.*, 2003).

II.4.1 Genomi tulajdonságok hatása a PCR amplifikációra

A prokarióták evolúciójuk évmilliárdjai során nem csak morfológiájukban és fiziológiájukban, hanem genomjaik szerkezeti tulajdonságaiban is rendkívüli sokféleségre tettek szert. A genomméret, a genom kópiaszáma, bázisösszetétele (mol%-os G+C tartalma), a riboszomális RNS-eket kódoló operonok száma és eloszlása is igen nagymérvű heterogenitást mutat körükben, s ezen tulajdonságok szoros összefüggésben vannak az adott faj környezetének tulajdonságaival, életmenetével, szaporodási stratégiájával. A genomméret körülbelül 600-tól 13000 kilobázisig terjed. A sejtparazita, vagy szimbióta fajok általában kisebb méretű genomot tartanak fenn. A törzsek genommérete nem állandó, tápanyagforrások szűkében csökkenhet, vagy

mobilis genetikai elemek integrálódása miatt növekedhet. A genom kópiaszáma is változatos lehet, s akár eltérő méretű „genofórok” is lehetnek egyazon sejtben: a *Brucella melitensis* egy 2100 és egy 1150 kilobázis méretű „kromozómával” bír. Az ismert fajokban a 16S rRNS-t kódoló gének száma 1 (pl. *Rickettsia prowazekii*) és 15 (pl. *Clostridium paradoxum*) között változik, operonokba szerveződésük és eloszlásuk a genomon belül szintén igen változatos (Farrelly *et al.*, 1995). A nagyobb számú rRNS operon gyorsabb reagálást (rövidebb osztódási ciklust) tesz lehetővé a szaporodási feltételek kedvezőbbé válása esetén (Klappenbach *et al.*, 2000). A 16S rDNS alapú közösségvizsgálati módszerek alkalmazásánál ezek a tényezők nagymértékben nehezítik a közösségek összetételére vonatkozó mennyiségi következtetések levonását, hiszen a mintában az adott fajnak nem az egyedszáma, hanem az egyedszám és a 16S rDNS kópiaszámának szorzatára kapunk becslést. Az ismert 16S rDNS kópiaszámú fajok esetében korrigálni lehet a kapott értékeket, ám a – legtöbbször nagy számú – ismeretlen faj esetében erre nincs lehetőség (Farrelly *et al.*, 1995). Feltételezik azt is, hogy a genomban egymáshoz közel eső kópiák a PCR során összességében nagyobb határfokkal amplifikálódnak, mint a genomon belül szétszórt, bár ugyanolyan mennyiségű génszakaszok.

A közösségalkotó fajok eltérő genomi tulajdonságait tekintve a másik legfontosabb faktor az adott DNS molekula bázisösszetétele (mol%-os G+C-tartalma), amely a magas GC-tartalmú templátok nehezebb denaturálódása miatt bizonyítottan befolyásolja az amplifikáció hatékonyságát, a kis GC-tartalmú amplikonok túlzott reprezentációját eredményezve. Ez a torzítás csökkenthető, ha a PCR reakcióelegyhez például 5 térfogatszázalék acetamidot adnak (Reysenbach *et al.*, 1992). Az egyszálú DNS-templát által felvett másodlagos szerkezet, vagy a DNS-hez kötődő, a nukleinsav izolálás és tisztítás során el nem távolított fehérjék is nagymértékben csökkenthetik az extenzió sebességét. A primerkötő helyeket határoló régiók szerkezete szintén befolyásolhatja a primerkötés hatékonyságát, fajonként eltérő mértékben gátolva az amplifikációt a PCR kezdeti ciklusaiban. Ezt a torzítást nem sikerült sem a DNS denaturálódását befolyásoló anyagokkal, sem úgynevezett „touch down” PCR protokollal kiküszöbölni (Hansen *et al.*, 1998). (A „touch down” PCR elve az, hogy a kezdeti ciklusokban az annelációs hőmérséletet a primer számára javasoltnál magasabban tartják, majd fokozatosan csökkentik, így a kezdeti ciklusokban csak a primerekhez nagy pontossággal illeszkedő templátok amplifikálódnak. A későbbi ciklusokban az alacsonyabb annelációs hőmérséklettel az amplifikáció hatékonysága nő.

Ezáltal a PCR a hozam enyhe csökkenése mellett nagy biztonsággal csak a megcélzott génszakaszt szaporítja.)

II.4.2 A PCR primerek szelektivitásából fakadó differenciális amplifikáció

II.4.2.1 Az oligonukleotid primerek

A PCR egyik legalapvetőbb „paramétere” az alkalmazott primerpárok szekvenciája. Bár jelen áttekintésben csak az univerzális bakteriális 16S rDNS primerekkel foglalkozunk (minthogy teljes baktériumközösségek vizsgálatára ezek a legalkalmasabbak), meg kell jegyezni, hogy megfelelő szekvenciájú primerekkel egy szűkebb rokonsági kör is specifikusan felszaporítható, illetve bármely más (funkció)gén, lehetőséget adva például az egy közös enzimtípussal bíró csoportok, *guild*-ek (pl. ammónia oxidálók) vizsgálatára.

A jelenleg használt univerzális bakteriális primereket 16S rRNS gének összegyűjtött szekvenciaadatai alapján tervezték, arra törekedve, hogy a 16S rRNS-t kódoló szekvenciák legkonzervatívabb 15-20 nukleotidnyi részéhez minél több fajban, minél kevesebb bázispárosodási hibával (mismatch) legyenek képesek kötődni (Pace *et al.*, 1986). A primertervezést a riboszomális RNS-t kódoló gének szekvenciáit tartalmazó on-line adatbázisok segítik, mint például a Ribosomal Database Project II (Cole, 2007; Cole, 2009; HTTP-1 – lásd az internetes hivatkozások jegyzékét a 91. oldalon), vagy az ARB (Ludwig *et al.*, 2004). Olyan univerzális primert tervezni természetesen lehetetlen, amely minden ismert faj adott kötőhelyeihez tökéletesen illeszkedik, nem beszélve az ismeretlen bázissorrendű 16S rDNS szakasszal bíró szervezetekről.

A riboszomális gének szekvenciaadatait tartalmazó adatbázisok növekedésével párhuzamosan időről időre újabb primerpárokat terveznek, hogy az ismert mikrobiális diverzitás minél nagyobb részét detektálhatóvá tegyék, de az eddigi eredmények azt mutatják, hogy minden primer szelektív bizonyos rendszertani csoportokra.

II.4.2.2 A primerek degeneráltsága

Az univerzális primerek tervezése több szempontból is nehéz. Az ismert 16S rDNS szekvenciájú baktériumok szekvencaadataiból kiderül, hogy még ezen evolúciósan nagymértékben konzervált gén bázisainak is csak a 10%-a teljesen állandó, s ezek a pontok meglehetősen szétszórtan fordulnak elő a génen belül. A leghosszabb egybefüggő konzervált régió is csak 10 bázispár hosszú, s általában csak 3-4 bázispáros ilyen régiókat találunk. Tekintve, hogy a specificitás követelménye miatt a primerek legalább 15 (általában 15-25) bázis hosszúságúak kell legyenek, lehetetlen minden fajhoz tökéletesen illeszkedő univerzális bakteriális primerpárt tervezni. További paraméterek is nehezítik a kompromisszumot. A primerek ideális annelációs hőmérséklete – amelyet a hosszuk és a GC-arányuk határoz meg – nem különbözhet jelentősen. A két primer akárcsak részleges komplementaritását is el kell kerülni, mert ez nem kívánt műtermékek, úgynevezett primer-dimerek létrejöttét eredményezi a célszekvenciák amplifikációja rovására. Bár tiszta tenyészetből izolált DNS templát esetén már 70%-os primer illeszkedés mellett is sikerrel amplifikálhatunk, diverz környezeti mintákban a gyenge komplementaritás az adott templát alulreprézentiációját eredményezi a PCR-termékben (Baker *et al.*, 2003).

Egy primer sokkal nagyobb körben képes a különböző taxonok célszekvenciáihoz kötődni, ha bizonyos nukleotidjaira nézve degenerált. Ekkor a primer lényegében bizonyos pontokon különféle nukleotidokat tartalmazó oligonukleotidok keveréke. Polz és munkatársai a primerek degeneráltságának és preferenciális amplifikációjának összefüggését vizsgálták modellközösségeken. Tapasztalataik szerint a több G-C kapcsolatot tartalmazó primer–célszekvencia illeszkedés az adott templát preferenciális amplifikációját eredményezheti (Polz *et al.*, 1998). Lueders és munkatársai környezeti mintákon elemezték a degenerált primerek által okozott PCR torzítást, terminális restrikciós fragmenthossz polimorfizmus (T-RFLP) elemzés segítségével. Kimutatták, hogy nem degenerált primerrel a közösség összetétele nem módosult, ha különböző annelációs hőmérséklettel és ciklusszámmal végezték a PCR-t, míg degenerált primer esetében az annelációs hőmérséklet emelésével bizonyos csoportok részaránya nagymértékben lecsökkent a közösségen belül (Lueders *et al.*, 2003).

II.4.2.3 Primerek illeszkedési hibái és az annelációs hőmérséklet

Az amplifikáció során a legkisebb hatékonysággal azok a templátok amplifikálódnak, amelyek kötőhelyei és az alkalmazott primerek között az illeszkedés a legkevésbé teljes. Különösen fontos a primerek 3' végén található nukleotidok illeszkedése, hiszen az elongáció (az új DNS szál szintézise) ettől a nukleotidtól indul el. A primerek illeszkedésének hibájára („primer mismatch”) nézve talán az eddigi leginformatívabb eredményeket Hongoh és munkatársai írták le (Hongoh *et al.*, 2003). Kutatásuk során természetes bél-mikrobiótáját vizsgálták klónkönyvtár elemzés segítségével. Hat primerpár, különböző annelációs hőmérsékletek és PCR ciklusszámok alkalmazásával 14 klónkönyvtárat létesítettek. Azt tapasztalták, hogy míg a klónkönyvtárak nagy részében a legdominánsabb csoport a *Spirochaeta* volt (50% körüli részarány), a 63f-1389r primerpárral létesített klónkönyvtárak szinte egyetlen *Spirochaeta* klónt sem tartalmaztak. A *Spirochaeta* klónok bázissorrend elemzése során kiderült, hogy a 63f primer a csoportra nézve 6-8 illeszkedési hibával bír. A különböző annelációs hőmérsékletekkel létrehozott klónkönyvtárak összehasonlításából kiderült, hogy kisebb annelációs hőmérsékleten nagyobb diverzitást mutatnak a klónkönyvtárak. Ez azzal magyarázható, hogy a kisebb hőmérsékleten a primerek bekötődése a nem teljesen komplementer kötőhelyekre is lehetővé válik. A 63f primer módosításával (a 3' végén elhelyezkedő nukleotid citozinról timinre történő cseréjével) szintén teljesen eltérő közösség szerkezetet detektáltak, bizonyítva ezzel a 3' végén elhelyezkedő nukleotid illeszkedésének fontosságát.

Szintén a primer illeszkedési hiba által okozott torzítás és az annelációs hőmérséklet összefüggéseit vizsgálta Ishii és Fukui, analitikai módszerként a denaturáló grádiens gélelektroforézist (DGGE-t) alkalmazva (Ishii és Fukui, 2001). Templátként PCR-termékekből 1:1 arányban összekevert kéttagú modellközösséget alkottak, amelyeket tökéletesen illeszkedő primerpárokkal amplifikáltak. Hogy a primer illeszkedés hatását vizsgálni tudják, az egyik templát primerkötő helyén irányított mutagenézissel kicseréltek egy nukleotidot. Ebben az esetben valóban detektálható volt a preferenciális amplifikáció, ám az annelációs hőmérséklet 45°C-ra csökkentésével a torzítás megszűnt. Az analitikai módszerként alkalmazott DGGE gélkép elemzés nem alkalmas a PCR termékek relatív mennyiségi viszonyainak pontos meghatározására, így a torzítás mértékének megállapítására.

Ugyanebbe a kutatási vonulatba illeszkedett az ELTE Mikrobiológiai Tanszékén végzett kísérletsorozatunk, amelynek során kéttagú modellközösségekben vizsgáltuk a különböző mértékű primerilleszkedési hibák melletti preferenciális amplifikációt, azonban a korábbi megközelítéseknél pontosabb méréseket lehetővé tevő T-RFLP elemzés alkalmazásával (Palatinszky, 2004; Sipos *et al.*, 2007). Sikertült kimutatnunk, hogy a primer bázissorrendjével teljes egyezést mutató templátok jóval nagyobb hatékonysággal amplifikálódtak a PCR során, mint a három illeszkedési hibát tartalmazó templátok. A torzítás mértéke az annealációs hőmérséklettel csaknem exponenciális összefüggést mutatott, de megfelelően alacsony annealációs hőmérsékleten megközelítette a pontos illeszkedéssel kapott eredményeket.

A templát-termék aránymódosulások a PCR ciklusszámával csak kis mértékben függtek össze, viszont a három illeszkedési hiba hatása minden ciklusszám értéknél erőteljesen jelentkezett. A modellrendszerekkel szerzett tapasztalatokkal összhangban, diverz környezeti baktériumközösség esetén a különböző forward primerek és azonos reverz primerek alkalmazásával nyert közösségi T-RFLP ujjlenyomatok szignifikánsan eltértek. Az annealációs hőmérséklet változtatásával a közösség szerkezeti képekben is szignifikáns változások álltak be, kisebb hőmérsékleten nagyobb diverzitást mutatva. Lényeges következtetés, hogy multitemplát PCR szelektivitása miatt nem csak a közösség szerkezetben tapasztalható mennyiségi viszonyok változnak meg, hanem egyes közösségalkotó fajok vagy nagyobb rendszertani csoportok teljesen ki is maradhatnak a detektálható diverzitásból.

II.5 A közösségi PCR termék TA-klónozása során fellépő hibalehetőségek

II.5.1 TA-klónozó vektorrendszerek méretpreferenciája

A PCR termékek TA-klónozására általánosan alkalmazott pGem-T és TOPO-TA vektorrendszerek gyártói a klónozás hatékonyságát csak homogén inzertfélések esetén vizsgálták. A ligálás hatékonyságának meghatározásánál szempontjaik az üres és az inzertet tartalmazó plazmidok arányára, illetve a transzformánsok elérhető maximális számára korlátozódnak. A mikrobiális közösségvizsgálatok során előforduló,

bázissorrendjükben, illetve hosszaikban is heterogén inzertpopulációk ligálásának esetét – valószínűleg az alkalmazás számukra marginális volta miatt – nem vizsgálták.

A vektorok inzertméret preferenciáira csak közvetett információkat nyerhetünk a gyártók dokumentációi alapján. Mivel a vektorok képesek a bezáródásra (inzert nélkül is) ezért minimális ligálható inzertméret meghatározása értelmetlen. A TOPO-TA kit gyártója a pCR2.1-TOPO vektorba sikerrel ligálható inzertek hosszának felső korlátját 4000 bázispárban adja meg. A pGem-T vektorrendszer esetében ilyen információ sem adnak, de valószínűleg hasonló mérettartományra optimalizálták – hiszen mindkét vektorrendszert „PCR klónozásra” fejlesztették ki, így alapvetően általában maximálisan pár ezer bázis hosszúságú PCR amplikonok befogadására. A TOPO-TA rendszer esetében a gyártó honlapjáról letölthető kiegészítő dokumentáció (HTTP-2) megemlíti, hogy az 1500 bázispárnál hosszabb PCR termékeket érdemes megtisztítani a primer dimerektől, mivel a rövid PCR műtermékek nagyobb hatékonysággal ligálódnak a vektorba. A pGem-T gyártója is javasolja a primer dimerek eltávolítását, de ennél bővebb információt nem közöl. A Promega fejlesztői publikáltak egy „optimalizációs” kísérletet pGem-T vektorrendszer egy órás ligálási idejének 15 percre rövidítésére (sajnos szintén kizárólag a gyártó honlapján megjelentetett anyag, lásd HTTP-3). Ennek során a ligálási idő csökkentésének hatását vizsgálták az elérhető transzformánsok számát, illetve a kék/fehér telepek arányát illetően különböző méretű inzertek esetén. A kísérlet tervezése, értékelése és dokumentációja tudományos szempontból elég sok kívánnivalót hagy maga után. Talán annyit biztonsággal leszűrhetünk belőle tanulságként, hogy a ligálás ideje hatással van a sikeres inzertálódások mennyiségére: rövidebb idő alatt kevesebb inzertbeépülés történik, ami a diffúzió, mint tényező kiemelt szerepét sejteti.

Jóval érdekesebb az ugyancsak a Promega által (sajnos szintén kizárólag a honlapjukon) publikált vizsgálat, mely a pGem-T, a pGem-T Easy, és a TOPO-TA vektorrendszerek összehasonlítására tett kísérletet, különböző méretű inzertek alkalmazásával (HTTP-4). A kísérlet során 542, 750, 1400 illetve 3100 bázispár hosszú inzerteket ligáltak a különböző vektorokba, majd megvizsgálták az azonos térfogatnyi transzformált sejtszuspenzióból kinövő fehér, illetve kék telepek számát. A kísérleteket mindössze két párhuzamosban hajtották végre, s a ligálásokat bár ugyanabból a PCR termékből, de „különböző napokon” végezték (a leghosszabb inzert esetében például a második párhuzamosok minden vektorrendszer esetében minimum kétszer annyi sikeres transzformáns eredményeztek). De eltekintve a kísérlet gyengeségeitől (és a gyártó

elfoglaltságától) mindenképpen hasznos információt kaphatunk a vektorok inzertméret függő ligálási hatékonyságaira nézve. A pGem-T vektor a legnagyobb hatékonysággal az 542 bázispáros inzertet ligálta, a 750 illetve 1400 bázispáros inzerteket pedig körülbelül feleakkora hatékonysággal (a 3100 bázispáros inzert eredményei hatalmas szórásuk miatt nehezen kiértékelhetők). A TOPO-TA vektorrendszer a legnagyobb hatékonyságot szintén az 542 bázispáros inzert esetében mutatta, az 1400 és 3100 bázispáros inzertek esetében pedig drasztikusan lecsökkent a ligálási hatékonysága.

Mindezek az eredmények azonban csak fenntartásokkal, és közvetett információként kezelhetők a heterogén inzertkeverék esetében potenciálisan fellépő preferenciális ligálásra nézve.

II.5.2 Inzert szekvenciák hatása a DNS – fehérje interakciókra

A TA-klónozás sikerének elengedhetetlen feltétele, hogy az inzertek végeiken túlnyúló 3' adenin nukleotiddal rendelkezzenek. Ezt ideális körülmények között a *Taq* polimeráz terminális transzferáz aktivitása biztosítja. Brownstein, valamint Magnuson és munkatársai azonban kimutatták, hogy a *Taq* polimeráz terminális adenin addícióját befolyásolja a kiegészített lánc utolsó nukleotidja: láncvégi 3' adeninhez kisebb hatékonysággal köt újabb adenint, mint más nukleotidokhoz (Brownstein *et al.*, 1996; Magnuson *et al.*, 1996). Így azok a PCR amplikonok, amelyek amplifikálásánál az egyik, vagy mindkét primer 5' végén timin nukleotid szerepel, kisebb hatékonysággal fognak ligálódni. Ez a torzítási faktor természetesen nem feltétlenül jelentkezik egy közösségi PCR-termék esetében, hiszen egy meghatározott primerpárral amplifikálódik a teljes közösségi templát, így az amplikonok azonos végeiken azonos nukleotidokat fognak tartalmazni. Viszont amennyiben az alkalmazott primerpár akármelyik tagja 5' végén olyan degenerált pozícióval bír, ami timint is tartalmaz, a primerkötő helyeik 3' végi pozíciójában adenint tartalmazó közösségalkotók ki vannak téve a kisebb hatékonyságú terminális adenin addíció jelenségének, s így alulreprzentáltak lehetnek a klónkönyvtárban.

További tényezőt jelent, hogy az inzert kettősszálú DNS-ének bázissorrendje meghatározza a struktúra merevségét (Marko *et al.*, 2003), valamint „lélegzését” (időleges lokális széttekeredésének illetve renaturációjának gyakoriságát és

kiterjedtségét), ami befolyásolja a fehérje-DNS interakciókat (Altan-Bonnet *et al.*, 2003), s így akár a ligáz enzim kötődésének vagy működésének hatékonyságát is.

II.5.3 Az inzertált génszakasz hatása a transzformált sejtekre

A kereskedelmi forgalomban lévő TA-klónozó vektorok az inzertálódás helyének közelében általában *lac* promóter régiót tartalmaznak a kék-fehér szelekció lehetővé tételére (lásd II.2.1 fejezet). A promóter aktivitása miatt átíródó inzert akár RNS formájában, akár expresszáldva hatással lehet a transzformált sejtre. Az általánosan vizsgált riboszomális génszakaszok inzertjei például a sejt saját riboszomális RNS-eivel interferálhatnak (Taylor *et al.*, 2007).

II.5.4 Különböző klónozási stratégiák összehasonlítása szelektivitás szempontjából

Különböző TA-klónozó rendszerek heterogén PCR termékekkel végzett összehasonlító vizsgálatát közlő publikációról nem tudunk – jelenlegi szakirodalmi ismereteink alapján. Készült azonban igen alapos összehasonlítás a tompa végű klónozással és a TA-klónozással nyerhető közösségszerkezeti képekről gombaközösségek ITS-28S génszakasza esetén. Taylor és munkatársai 92 talajmintából izolált közösségi DNS-ből szaporították fel az ITS-2 és a vele szomszédos 28S rDNS régió egy részét gombákra specifikus primerpárral. A 92 közösségi PCR termékből egy keveréket hoztak létre, melyből két, 308 illetve 312 tagú klónkönyvtárat készítettek pCR4.0-TOPO (Invitrogen) TA-klónozó vektorral és perSMART-HC Kan (Lucigen, Middleton, WI, USA) tompa végű vektorral. Elemzésük alapján a két könyvtár által mutatott közösség filogenetikai összetételében és a közösségalkotók relatív mennyiségi viszonyaiban nem különbözött jelentősen, bár több olyan OTU-t is találtak ami az egyik vagy a másik könyvtárban erősen alulreprezentált volt (Taylor *et al.*, 2007).

II.6 Közösségelemzés során gyakran vizsgált gének és genomszakaszok

II.6.1 A riboszomális kis alegység RNS-ét kódoló gén

A közösségek leírása során legrégebb óta, és legáltalánosabban vizsgált génszakasz a riboszomális 16S rRNS-t kódoló gén, vagy másnéven 16S rDNS szakasz. A gén hossza átlagosan 1550 bázispár, nagyjából 200 bázispáros hosszheterogenitással a teljes szakaszon (García-Martínez *et al.*, 1999). A génen belül váltakozva találunk evolúciósan igen konzervatív és variábilisabb (V1-V9) szakaszokat (Baker *et al.*, 2003). Aránylag nagyobb hossz- és szekvenciaheterogenitása miatt gyakran vizsgálják önmagában a 16S rDNS első harmadát ujji lenyomat módszerekkel, például hosszheterogenitás PCR-rel (Suzuki *et al.*, 1998; Mills *et al.*, 2003), illetve denaturáló grádiens gélelektroforézissel (Muyzer *et al.*, 1998; Ellis *et al.*, 2003), és párhuzamosan alkalmazott klónkönyvtár elemzéssel. A különböző parciális 16S rDNS szakaszok filogenetikai felbontóképességének összehasonlító vizsgálatára jó példa Hutter és szerzőtársainak munkája, akik egy baktériumközösség 27f-519r illetve 515f-1525r primerpárokkal elkészített klónkönyvtárainak elemzését végezték el (Hutter *et al.*, 2003).

II.6.2 Riboszomális intergenikus spacer szakaszok

Bizonyos taxonok esetében a 16S rDNS filogenetikai felbontóképessége nem elégséges az egymással közeli rokonságban álló fajok elkülönítéséhez, mint például a *Bacillus* nemzetség esetében (Ranjard *et al.*, 1999), illetve vízi környezetekben élő cianobaktériumok – például a *Prochlorococcus* és *Synechococcus* genuszok – körében (Rocap *et al.*, 2002). Ezt a problémát bizonyos megközelítések azzal oldják fel, hogy nem az evolúciósan igen konzervatív 16S vagy 23S rRNS géneket vizsgálják, hanem a közöttük elhelyezkedő, mind hosszában, mind szekvenciájában igen variábilis „nem kódoló” régiót, másnéven 16S-23S *intergenikus spacer*-t (IS) (Barry *et al.*, 1991). A szakasz hossza meghaladhatja az 1500 bázispárt, bizonyos taxonokban viszont egészen rövid, sőt akár teljesen hiányozhat is (González *et al.*, 2003). Ez a régió is tartalmaz rövidebb konzervált szakaszokat, tRNS kódoló gének formájában, lehetővé téve ezzel belső primerek tervezését is. A régió PCR amplifikációját a spacer határoló 16S és 23S

gének végén illetve elején található igen konzervatív régiókhöz hibridizáló univerzális bakteriális primerekkel végzik (García-Martinez *et al.*, 1999). Az így létrehozott közösségi PCR termék relatíve egyszerűen, hossz polimorfizmus alapján elválasztható, hatékony ujjlenyomat módszert adva ezzel a mikrobiális ökológusok kezébe (Riboszomális Intergenikus Spacer Analízis, RISA). A módszer felbontása az IS régió rendkívül nagy hosszpolimorfizmusa miatt igen jónak mondható, ezt legjobban kapilláris-elektroforetikus elválasztással lehet kihasználni, gyors és nagyon részletes molekuláris ujjlenyomatot adva egy-egy közösségről (Fisher és Triplett, 1999).

A módszer elterjedését leginkább az IS szekvenciák identifikálását segítő adatbázisok igen szerény mérete és lassú bővülése akadályozta (Ranjard *et al.*, 1999). Ez a hiányosság kiküszöbölhető, ha a PCR során nem csak az IS régiót szaporítjuk fel, hanem a forward primert úgy választjuk, hogy a 16S rDNS szekvencia kb. 6-700 bázispárnyi szakaszával együtt amplifikálódjon az IS régió. Így amennyiben az amplikonokat a gélelektroforetikus elválasztás, vagy a TA-klonozás után bázissorrend elemzésnek vetjük alá, a 16S rDNS szekvenciákra vonatkozó adatbázisokból a közösségalkotókra nézve filogenetikai információt is nyerhetünk (Yu és Mohn, 2001).

A 16S-23S spacer régió klónkönyvtárak elemzésével (konkrétan a TOPO-TA vektorrendszerek alkalmazásával) gyakran találkozhatunk a tengeri cianobaktérium közösségek vizsgálata során (Ahlgren és Rocap, 2006; Chen *et al.*, 2006). E területen előfordul a parciális 16S rDNS és az IS régió együttes klónkönyvtárak vizsgálata is (Brown *et al.*, 2005; Haverkamp *et al.*, 2008).

Természetesen nem csak a prokarióta közösségek vizsgálatában alkalmazzák eme módszereket: a 2000-es években a környezeti mikológiában is elterjedt a használatuk, hiszen az eukarióta 18S rDNS régió szintén nem tartalmaz elég információt a nagy felbontású filogenetikai elemzéshez (Vandenkoornhuysen *et al.*, 2002). A gomba riboszomális intergenikus spacer (a mikológus terminológiában ITS-1 és ITS-2) régiók a 18S rRNS és a 28S rRNS gének között helyezkednek el, s közrefogják az 5S rRNS gént is. Az ITS-1 és ITS-2 régiók együttes hosszheterogenitására nézve Ranjard és munkatársai publikációjából kapunk információt, akik kapilláris elektroforetikus mérésekkel 300 és 1100 bázispár közötti hosszokat mértek (Ranjard *et al.*, 2001). A régió klónkönyvtárak vizsgálata is gyakori. Anderson és munkatársai gomba ITS-1 régió specifikus primerek validálásánál alkalmazták a pGem-T vektorrendszert (Anderson *et al.*, 2003), de találunk példát talaj gombaközösségek ITS régióinak TOPO-TA klónkönyvtárak vizsgálatára is (O'Brien *et*

al., 2005). A széleskörű közösségleírásokon kívül a technika taxonspecifikus primerpárok alkalmazásával alkalmas egy szűkebb filogenetikai csoport célzott, nagy filogenetikai felbontású vizsgálatára. Kitűnő hazai példaként említhetjük Kovács M. Gábor és kollégáinak munkáját, akik a virginiai holdruta (*Botrychium virginianum*) gyökerét kolonizáló, Glomeromycota törzsbe tartozó arbuskuláris mikorrhiza gombákat vizsgálták Glomeromycota-specifikus primerekkel, a 18S rDNA és ITS régiók körülbelül 900-1000 bázispár hosszú szakaszainak klónkönyvtár analízisével (Kovács *et al.*, 2007).

Tekintve, hogy mind a prokarióta mind az eukarióta riboszomális intergenikus spacer régiók hosszukban és szekvenciájukban is diverzek, a preferenciális ligálás jelensége hatással lehet a vonatkozó vizsgálatok eredményeire.

II.6.3 Funkciógének

Természetesen a mikrobiális ökológiában a filogenetikai diverzitás vizsgálata mellett sok esetben merülnek fel a közösség metabolikus potenciáljával kapcsolatos kérdések. A közösségi anyagcserét elemző kutatások gyakran csak egy bizonyos funkciót betöltő, meghatározott anyagcsere útvonalat használó csoportra, *guild*-re koncentrálnak. Ez esetben a vizsgálatokat az adott metabolikus funkcióban központi szerepet játszó enzimet kódoló funkciógénre, vagy az azt határoló nem kódoló régiókra célszerű összpontosítani. Ezáltal eleve szűrt információkat kaphatunk az adott guild tagjaira nézve, ami diverz közösségek esetén nagyban megkönnyíti a vizsgálatot. További előny, hogy funkciógének szekvenciahasonlóságainak elemzésével a riboszomális géneken alapuló rendszertantól független, adott esetben jobb felbontású filogenetikai információt nyerhetünk. A megközelítés előnyeiből kifolyólag a szakirodalomban számos példát találunk funkciógéneken alapuló klónkönyvtárs közösségvizsgálatokra – ezekből emelnék ki párat alább (a teljesség igénye nélkül).

A tengeri és édesvízi cianobaktérium közösségek vizsgálatában és többek között a *Prochlorococcus*, *Synechococcus*, és *Cyanobium* nemzetségek filogenetikai viszonyainak feltárásában központi szerepet játszik a fikocianin operon elemzése. Ez a *cpcBA*-IGS régió, azaz a fikocianin két alegységét kódoló gének (*cpcB* és *cpcA*) és a köztük lévő nem kódoló, és emiatt jelentős hosszpolimorfizmussal rendelkező szakasz vizsgálatát jelenti. A pikocianobaktériumoknál a *cpcB* és *cpcA* gének közötti nem

kódoló szakasz hossza taxonspecifikus érték, és a szakasz bázissorrend elemzése a konzerváltabb 16S rDNS alapú vizsgálatoknál jobb filogenetikai felbontást ad (Robertson *et al.*, 2001). A *cpbA*-IGS régió TA-klónkönyvtárak vizsgálatával találkozhatunk édesvízi cianobaktériumok által okozott vízvirágzások közösségeinek vizsgálatánál (TOPO-TA vektorrendszerrel) (Kim *et al.*, 2006), illetve a Balatoni fotoautotróf pikoplankton közösségek szezonális dinamikájának elemzésénél (Felföldi *et al.*, 2011). Ez utóbbi esetben a régió pGem-T vektorrendszerrel készített klónkönyvtárai mellett a közösségváltozás követésére hosszheterogenitás PCR-t is alkalmaztak. A Balti tenger pikocianobaktérium diverzitásának leírására Haverkamp és munkatársai a *cpbA*-IGS régió TA-klónkönyvtárak elemzése mellett a 16S rDNS gén egy részét és a 16S-23S spacer régiót együttesen tartalmazó TOPO-TA klónkönyvtár vizsgálatát alkalmazták (Haverkamp *et al.*, 2008).

A szén biogeokémiai ciklusában fontos szerepet betöltő metanotróf baktériumok célzott vizsgálatát a partikuláris metán-monooxigenáz enzim α alegységét kódoló *pmoA* gén elemzése is lehetővé teszi (McDonald *et al.*, 1997). A gén amplifikálására tervezett primerek közül leggyakrabban az A189 és A682 jelű oligonukleotid párt alkalmazzák (Bourne *et al.*, 2001), ami elméletileg egy 531 bázispár hosszú szakaszt szaporít fel (Pester *et al.*, 2004). Pester és munkatársai azonban a primerpár alkalmazásával létrehozott TOPO-TA könyvtárak vizsgálata során a génszakaszon körülbelül 100 bázispáros hosszheterogenitást tapasztaltak. Megjegyzendő, hogy Bourne és munkatársai szintén TOPO-TA klónkönyvtárakkal elemezték környezeti baktériumközösségeken a különböző *pmoA*-specifikus primerpárok szelektivitását, igen jelentős különbségeket tapasztalva a különböző primerpárokkal detektálható diverzitások között.

A nitrogén körforgásában kulcsfontosságú diazotróf baktériumok közösségvizsgálatában a nitrogenáz enzim egy részét kódoló *nifH* gén elemzése terjedt el (Lovell *et al.*, 2000). Sarita és munkatársai rizoszféra baktériumközösségek diazotróf tagjait vizsgálták a *nifH* gén elméletileg 466 bázispár hosszú szakaszát tartalmazó TOPO-TA klónkönyvtárakkal (Sarita *et al.*, 2008). Szintén Lovell és munkatársai a *nifH* gén vizsgálatára használt primerpárok összehasonlító elemzését is megkísérelték pGem-T klónkönyvtárak és Libshuff analízis segítségével, szignifikáns különbségeket tapasztalva a különböző primerpárokkal nyerhető közösségszerkezeti képek között (Lovell *et al.*, 2008). A jelen munka során is alkalmazott Libshuff elemzés részletes leírását a IV.3.7.5 fejezetben közöljük.

III. Célkitűzések

Az ELTE Mikrobiológiai Tanszékén korábban lezajlott, a multitemplát PCR torzításaival kapcsolatos módszertani kutatások (Palatinszky, 2004; Sipos *et al.*, 2007) folytatásaként munkánk során megkíséreltük kimutatni TA-klónozás közösségszerkezeti képet torzító hatását, majd részletesebben elemezni a torzítás függését az inzertek hosszheterogenitásától. A módszer elterjedtsége miatt bízhattunk abban, hogy eredményeink a mikrobiális ökológia művelőinek széleskörű érdeklődésére tarthatnak számot, s hogy munkánk alapján a módszert a korábinál nagyobb pontossággal, az eredmények kritikusabb interpretálásával lehet majd alkalmazni.

Miután a mikrobiális ökológiában és taxonómiában igen gyakran vizsgált 16S rRNS gén természetes hosszheterogenitásáról csak igen szórványos irodalmi adatok állnak rendelkezésre, megkíséreltük *in silico* módszerekkel felmérni ezen paraméter eloszlását a jelenleg ismert bakteriális világban.

A TA-klónozó rendszerek inzertméret preferenciája a módszerek dokumentációja alapján is feltételezhető, így alapvetően ezt a torzítási faktort tartottuk fontosnak megvizsgálni modellrendszerek segítségével a parciális 16S rDNS és a 16S-23S spacer régió hosszheterogenitási tartományaiban. Az ezen szakaszok vizsgálata során kapott eredmények természetesen iránymutatóak lehetnek más, hasonló hosszheterogenitással bíró szakaszok eseteire nézve is. A közösségvizsgálatok során a téma általunk eddig megismert szakirodalma alapján szinte kizárólag két TA-klónozó rendszert használnak, nevesen a pGem-T (Promega, WI, USA) illetve a TOPO-TA (Invitrogen, CA, USA) vektorokat. A különböző rendszerekkel eddig publikált eredmények összevethetőségének alapvető feltétele, hogy a két vektor esetleges eltérő torzításait megismerjük, így ezek komparatív elemzését is kiemelkedő fontosságúnak tartottuk.

A mikrobiális ökológus számára természetesen még inkább érdekes kérdés, hogy az általa alkalmazott módszer „éles helyzetben”, valódi közösségek vizsgálata során mennyire torzítja a valóságot. Ezért a két leggyakrabban alkalmazott vektorrendszer összehasonlító vizsgálatát elvégeztük egy természetes baktériumközösség leírása során is, referenciaként kiegészítve az összehasonlítást egy klónozástól független

közösségszerkezet elemző módszer eredményeivel. Munkánk célkitűzéseit részletesen az alábbi pontokban foglalhatjuk össze:

1. *In silico* vizsgálatok az általánosan vizsgált 16S rRNS génszakaszok természetes hosszheterogenitásának felmérésére.
2. Az inzerthosszfüggő preferenciális ligálás kimutatására alkalmas modellrendszer felállítása és validálása klónozástól független kvantitatív módszerrel, valamint azonos hosszúságú, de eltérő bázissorrendű inzertek ligálásával.
3. A pGem-T vektorrendszer inzerthossz preferenciájának vizsgálata parciális 16S rDNS-re és 16S-23S intergenikus spacerre jellemző hosszheterogenitás esetén, valamint a torzítás reprodukálhatóságának, és ligálási paramétereiktől való függésének kimutatása.
4. A pGem-T és TOPO-TA vektorrendszerek összehasonlító vizsgálata parciális 16S rDNS és 16S-23S intergenikus spacerre jellemző hosszheterogenitás esetén.
5. Egy természetes baktériumközösség pGem-T és TOPO-TA vektorrendszerekkel készített közepes (~300 klón) nagyságú parciális 16S rDNS klónkönyvtárainak összehasonlító elemzése (GC-arány preferencia, inzerthossz preferencia, filogenetikai szelektivitás), valamint összehasonlítása egy klónozástól független módszerrel (hosszheterogenitás-PCR-rel) nyert közösségszerkezeti képpel.

IV. Anyag és módszer

Tekintve, hogy e munka alapvetően módszertani jellegű, a dolgozatban az „Anyag és módszer” fejezet a szokásosnál nagyobb hangsúllyal szerepel, s nem pusztán a módszerek leírását tartalmazza, hanem a szükséges helyeken a technikák és konkrét paraméterek kiválasztásának indoklását is. Így olyan, szakirodalmi áttekintés jellegű részeket is tartalmaz, amelyeknek egy kevésbé módszer-centrikus disszertációban a klasszikus felépítésnek megfelelően nem az „Anyag és módszer” fejezetben lenne a helyük.

IV.1 *In silico* hosszheterogenitás vizsgálatok

IV.1.1 A vizsgált adatbázisok

Az *in silico* vizsgálatainkat két, 16S rRNS gének szekvenciáit tartalmazó nyilvános adatbázison végeztük el, amelyek egyike kizárólag típusörzsek adatait tartalmazza (így a legmegbízhatóbb szekvenciák forrásának tekinthető), másika pedig környezeti klónok bázissorrendjét is magába foglalja, így a diverzitás sokkal szélesebb körét reprezentálja. Az előbbi a Ribosomal Database Project II (Cole, 2007; Cole, 2009; HTTP-1) adathalmazából származó szekvenciatömeg volt, amelynek exportálásánál a következő szűrőfeltételeket alkalmaztuk (zárójelben az RDPII kezelőfelületén alkalmazott konkrét beállítások):

- csak típusörzsek szekvenciái (Strain: type)
- kizárólag izolátumok (Source: isolates) (ez persze a típusörzsek miatt evidens)
- 1200 bázispárnál hosszabb szekvenciák (Size: ≥ 1200)
- csak jó minőségű szekvenciák (Quality: good)

Az alkalmazott szűrésekkel 7448 szekvenciából álló FASTA adatbázist építettünk.

A másik adathalmazunk az ARB-SILVA típusörzsek, környezeti izolátumok és klónszekvenciák riboszomális kis alegységeket kódoló szakaszaiból épített, ellenőrzött minőségű, annotált, nem redundáns referencia adatbázisa volt (SSU Ref NR 104), amely 262092 szekvenciát tartalmaz (Pruesse, 2007; HTTP-5).

IV.1.2 A LengHet program, és az alkalmazott paraméterek

A jelenleg elérhető bázissorrend elemző és virtuális PCR szoftverek nem alkalmasak nagy mennyiségű szekvencia adott primerek közötti hosszheterogenitásának mérésére, legfeljebb csak egyenkénti, manuális vizsgálatokra. Sőt a legtöbb szoftver már a vizsgálandó adatbázisok méretének megfelelő szekvenciamentiség beolvasására sincs felkészítve. Ezért szükségesnek láttuk egy erre a célfeladatra megírt, automatizált adatbázis elemző szoftver kifejlesztését. A program megírását az általam összeállított rendszerparaméterek, bemeneti és kimeneti adattípus követelmények alapján Novák Ádám programtervező matematikus-biológus kollégám végezte el, megalkotva a LengHet v1.1. Java alkalmazást, mely az ELTE szerverein kapott végleges helyet, és szabadon letölthető (HTTP-6).

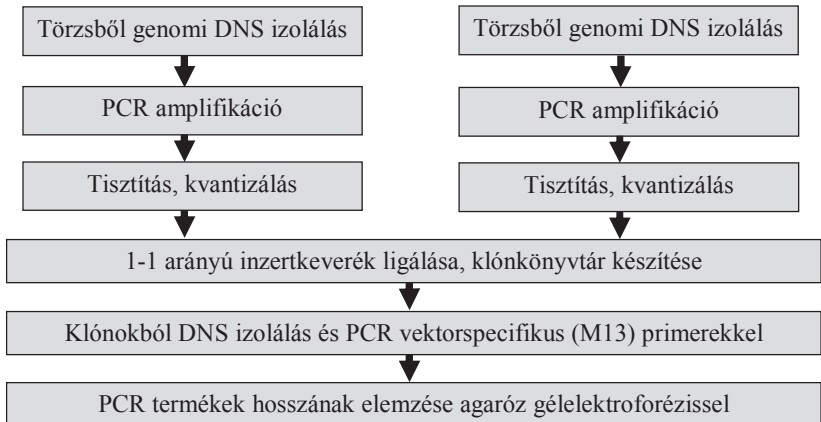
A szoftver a felhasználó által megadott forward illetve reverz primernek megfelelő kötőhelyeket keresi meg az adatbázisban szereplő szekvenciákon, majd az így kapható amplikon hosszát minden egyes szekvenciára meghatározza. A primerek illeszkedésének specificitását a felhasználó határozza meg, a megengedett illeszkedési hibák („mismatch”-ek) számának megadásával. Erre az úgynevezett „univerzális” primerek esetében is fennálló szelektivitás miatt van szükség, mely jelenséggel kutatócsoportunk korábban behatóan foglalkozott (Sipos *et al.*, 2007; Palatinszky 2004), s amely kutatás tapasztalatai alapján jelen esetben három bázispárnyi illeszkedési hiba megengedése mellett döntöttünk. Ennyi „mismatch” mellett egy szekvencia még fel tud szaporodni egy valós PCR során (akár közösségi vizsgálat esetén is, igaz kisebb mértékben, mint a tökéletesen illeszkedők), de ennél kevesebb illeszkedési hiba megengedésével az előzetes vizsgálataink alapján a szekvenciák túl nagy részét zárnánk ki az elemzésből.

A szoftver kimeneti oldalán egy az amplikonhosszok adatbázison belüli eloszlását tartalmazó táblázat, illetve grafikon jelenik meg. A program ezen kívül lehetőséget nyújt az egyes szekvenciákon történt primerillesztések ellenőrzésére is.

IV.2 Modellközösségek

A TA-klónozás mennyiségi torzításának vizsgálatát pontosan meghatározott összetételű PCR termék keverékekből álló egyszerű „modellközösségek” segítségével végeztük el, hogy a közösség szerkezeti vizsgálatok során általában fellépő – DNS-izolálással és PCR amplifikálással kapcsolatos – egyéb torzítások hatását kiküszöböljük. A kísérletek során tiszta tenyészetekből kivont genomi DNS-ből kiindulva meghatározott 16S rDNS szakaszokat amplifikáltunk PCR segítségével. Ezeket a termékeket megtisztítottuk, koncentrációjukat, és az amplikonok pontos molekulatömegének ismeretében mikroliterenkénti kópiaszámukat meghatároztuk. Ezután 1:1 kópiaszám arányú amplikon keverékeket készítettünk, melyekből klónkönyvtárakat hoztunk létre. A klónkönyvtárakban a kétféle inzert arányát felmérve meghatározhattuk a klónozó rendszer esetleges torzítását. Az inzertek tulajdonságait (hossz, GC-tartalom, végső nukleotidok) a felhasznált törzsek előzetesen meghatározott 16S rDNS bázissorrendje alapján, a kérdésfeltevéseinknek megfelelő szakaszok amplifikálásával tudtuk beállítani.

A modellrendszer alapvető sémáját a 4. ábra mutatja, a különböző paramétereket, és a speciális kísérlet típusokat (pl. validálás; különböző ligálási paraméterekkel, de azonos keverékből létrehozott könyvtárak) alább részletezzük. Az eredmények reprodukálhatóságának ellenőrzése végett minden kísérletet három, a DNS izolálástól kezdve teljesen független párhuzamosan végeztünk el.



4. ábra. A preferenciális ligálás elemzésére felállított modellrendszer általános vázlata.

IV.2.1 A modellrendszerekhez felhasznált törzsek, fenntartásuk, genomi DNS izolálás

A modellkísérletekben használt törzsek kiválasztásánál az alábbi kritériumrendszert tartottuk szem előtt: legyenek könnyen fenntartható, nem patogén, pontosan ismert 16S rDNS bázissorrendű szerkezetek, amelyek között több hasonló, illetve eltérő parciális 16S rDNS szakaszhosszokkal bíró törzs is van. Ezek figyelembevételével az alább felsorolt négy baktériumtörzs genomi DNS-éből állítottuk elő a modellközösségeket:

Aeromonas hydrophila ATCC 7966,
Bacillus cereus ATCC 14579,
Bacillus subtilis ATCC 6633,
Rhizobium gallicum RH1 környezeti izolátum.

Ez utóbbi törzs 16S rDNS génjének bázissorrendjét a FR728933 azonosítóval elhelyeztük a GenBank adatbázisban. A törzset egy, az ELTE Mikrobiológiai tanszékén

folyó párhuzamos kutatás izolátumai közül választottuk ki, 16S rDNS-ének igen rövid V1-V2-V3 szakaszhosszai alapján. A három típus törzs pontos 16S rDNS bázissorrend elemzését diplomamunkám elkészítése során már korábban elvégeztem (Palatinszky, 2004).

A törzseket húspepton agar (DSMZ-1) táptalajon tartottuk fenn, melynek összetétele:

Pepton	5,0 g
Húskivonat	3,0 g
Agar	15,0 g
Desztillált víz	1000 ml-re

A genomi DNS kivonását minden esetben frissen átoltott (24-48 óras) tenyészetekből végeztük, V-gene Bacterial Genomic DNA Mini-prep Kit (V-Gene Biotechnology Limited, Hangzhou, China) alkalmazásával, a gyártó utasításai szerint.

IV.2.2 Az inzertek előállítása PCR segítségével (általános PCR protokoll)

A PCR reakciókat GeneAmp[®] 2700 (Applied Biosystems, Foster City, CA, USA) készülékben futtattuk. A reakcióelegyek alapvető összetétele a következő volt (az esetleges eltéréseket az adott leírásoknál később jelezzük):

- 5 µl 10x *Taq* puffér [750 mM Tris-HCl, 200 mM (NH₄)₂SO₄, 0,1% Tween20; Fermentas, Vilnius, Litvánia]
- 4 µl MgCl₂ oldat (25 mM; Fermentas)
- 10 µl dNTP keverék (1 mM dATP, 1 mM dGTP, 1 mM dCTP, 1 mM dTTP; Fermentas)
- 0,5 µl forward primer (3,25x10⁻⁵ M-os törzsoldat)
- 0,5 µl reverz primer (3,25x10⁻⁵ M-os törzsoldat)
- 28 µl PCR tisztaságú DEPC kezelt víz
- 1 µl rekombináns *Taq* polimeráz (1 U/µl; Fermentas)
- 1 µl az izolált DNS-t tartalmazó oldatból (ez körülbelül 30 ng DNS-t jelent)

A vizsgálatok során alkalmazott primerek jelzéseit, szekvenciáit és eredeti hivatkozásait az 1. táblázat tartalmazza.

Primer jelzése	Oligonukleotid szekvencia	Eredeti közlés
63f	5'-CAGGCCTAACACATGCAAGTC-3'	Marchesi <i>et al.</i> , 1998
338r	5'-GCTGCCTCCCGTAGGAGT-3'	Amann <i>et al.</i> , 1990
519r	5'-ATTACCGCGGCTGCTGG-3'	Lane, 1991
1387r	5'-GGGCGGWTGTACAAGGC-3'	Marchesi <i>et al.</i> , 1998
1492r	5'-TACGGYTACCTTGTTACGACTT-3'	Lane, 1991

1. táblázat. Az inzertek előállításához felhasznált primerek bázissorrendje és hivatkozásai

A reagensek és a műanyagáru (PCR csövek és pipettahegyek) tisztaságának ellenőrzésére minden esetben negatív (templát DNS-t nem tartalmazó) kontrollt is használtunk. Az amplifikációt mindig 30 perces végső extenziós lépéssel zártuk, hogy a terminális adenin addíció a TA-klónozáshoz kellő mértékben végbemehessen. A PCR-ek általános hőprofilja a következő volt:

Kezdeti denaturáció	98°C	5 min
<i>Taq</i> pol. hozzáadása	94°C	a lépés elvégzéséig
denaturáció	94°C	30 sec
anneláció	52°C	30 sec
extenzió	72°C	1 min
		} 32 ciklusban
végső extenzió	72°C	30 min
hűtés	4°C	∞

A PCR termékeket agaróz gélelektroforézissel ellenőriztük (1%-os agaróz gélben, 100 V elektroforetikus feszültséggel, 20 perces futási idővel), etidiumbromiddal festve, áteső UV fényben detektálva. A PCR termékek méretének ellenőrzésére mérettől függően pUC Mix Marker-8 vagy Lambda DNA/*EcoRI*+*HindIII* Marker-3 (Fermentas) molekulásúly standardot futtattunk a mintákkal párhuzamosan.

A termékeket megtisztítottuk PCR-M Clean Up System (Viogene, Taipei, Taiwan) segítségével, majd 4°C-on tároltuk a további lépésekig. A klónkönyvtárak elkészítéséhez mindig friss PCR terméket használtunk, lehetőleg egy napon belül

elvégezve az amplifikációt, kvantizálást, keverést és a ligáló reakció összemérését, hogy elkerüljük az inzertek végein levő túlnyúló adeninek degradációját.

IV.2.3 A modellkísérletek inzertkeverékei, klónkönyvtárak áttekintése

A modellkísérletek „bemeneti oldalának” fontos paraméterei az inzertként szolgáló egyes amplikontípusok hossza, hosszaik különbsége, GC-tartalmuk, valamint végső nukleotidjaik. A megfelelő tulajdonságú inzerteket a rendelkezésre álló négy törzs különböző 16S rDNS szakaszainak előzetes *in silico* vizsgálatával és az alkalmas szakaszok amplifikálásával tudtuk előállítani. Az amplikonok hosszát Mega v4. (Tamura *et al.*, 2007) szoftverrel számoltuk ki, GC-tartalmuk meghatározásához BioXpress v1.0 programot használtunk. A modellrendszer megalkotásánál elsődlegesen a klónozás torzításaira ható változók minimalizálására törekedtünk, ezért olyan szakaszokat választottunk, melyek GC-tartalma közel azonos. Ezen túlmenően, amennyiben különböző primereket használtunk előállításukhoz, azokat úgy választottuk meg, hogy 5' nukleotidjaik is azonosak legyenek, elkerülendő a ligálás helyén levő nukleotidok különbözőségéből eredő esetleges torzítást. A modellrendszerekben alkalmazott inzertkeverékek amplikonjainak elnevezését, eredetét, az előállításukhoz felhasznált primerpárokat, illetve a szakaszok hosszát és GC-tartalmát a 2. táblázat részletezi.

Keverék	Inzertek elnevezése	Inzertek eredete (baktérium törzsek)	Forward primer	Reverz primer	Amplikon hossz	Inzert GC-tartalma
A1,A2,A3	AH1474bp	<i>Aeromonas hydrophila</i> ATCC 7966	63f	1492r	1474 bp	55.2 %
	BS1479bp	<i>Bacillus subtilis</i> ATCC 6633	63f	1492r	1479 bp	54.9 %
B1,B2,B3	AH1365bp	<i>Aeromonas hydrophila</i> ATCC 7966	63f	1387r	1365 bp	55.5 %
	BC322bp	<i>Bacillus cereus</i> ATCC 14579	63f	338r	322 bp	55.0 %
C1,C2,C3	RH433bp	<i>Rhizobium gallicum</i> RH1	63f	519r	433 bp	55.2 %
	BC501bp	<i>Bacillus cereus</i> ATCC 14579	63f	519r	501 bp	53.7 %

2. táblázat. A modellrendszerekben alkalmazott inzertkeverékek amplikonjainak elnevezése, eredete, az előállításukhoz felhasznált primerpárok, illetve a szakaszok hossza és GC-tartalma.

Célkitűzéseinknek megfelelően háromféle inzerthossz heterogenitás esetére állítottunk össze modellrendszeret (5. ábra). Bár a kvantizálási és keverési eljárásunk pontosságát és reprodukálhatóságát ellenőriztük és megbízhatónak találtuk, az egyes inzertheverékek pontos összetételét nem tudtuk más módszerrel ellenőrizni, ezért készítettünk minden keveréktípusból három párhuzamost. A kísérletek nagyrészt pGem-T vektorrendszerrel végeztük, két esetben pedig ugyanazon keverékből összehasonlító klónkönyvtárat készítettünk a TOPO-TA vektorrendszerrel is – itt az azonos kiindulási keverék miatt nem volt szükség több párhuzamos TOPO-TA könyvtár létesítésére.

Az első esetben („A” jelű párhuzamos keverékek) a teljes 16S rDNS felszaporítására általánosan használt eubakteriális primerek által közrefogott szakasz hosszának megfelelő inzerteket alkalmaztunk, melyek hosszban és GC-tartalomban szinte megegyeztek (0,5% alatti különbségek). Tekintve, hogy az inzertek bázissorrendjükben különböztek (77,5% szekvencia egyezés), ezen kísérlet során az inzerthossz preferenciától független, szekvenciafüggő preferenciális ligálás esetleges jelenlétére következtethettünk, illetve a pGem-T vektorrendszer esetében az elrendezést a különböző hosszúságú inzertekkel végzett kísérletek „negatív kontrolljaként” is vehettük.

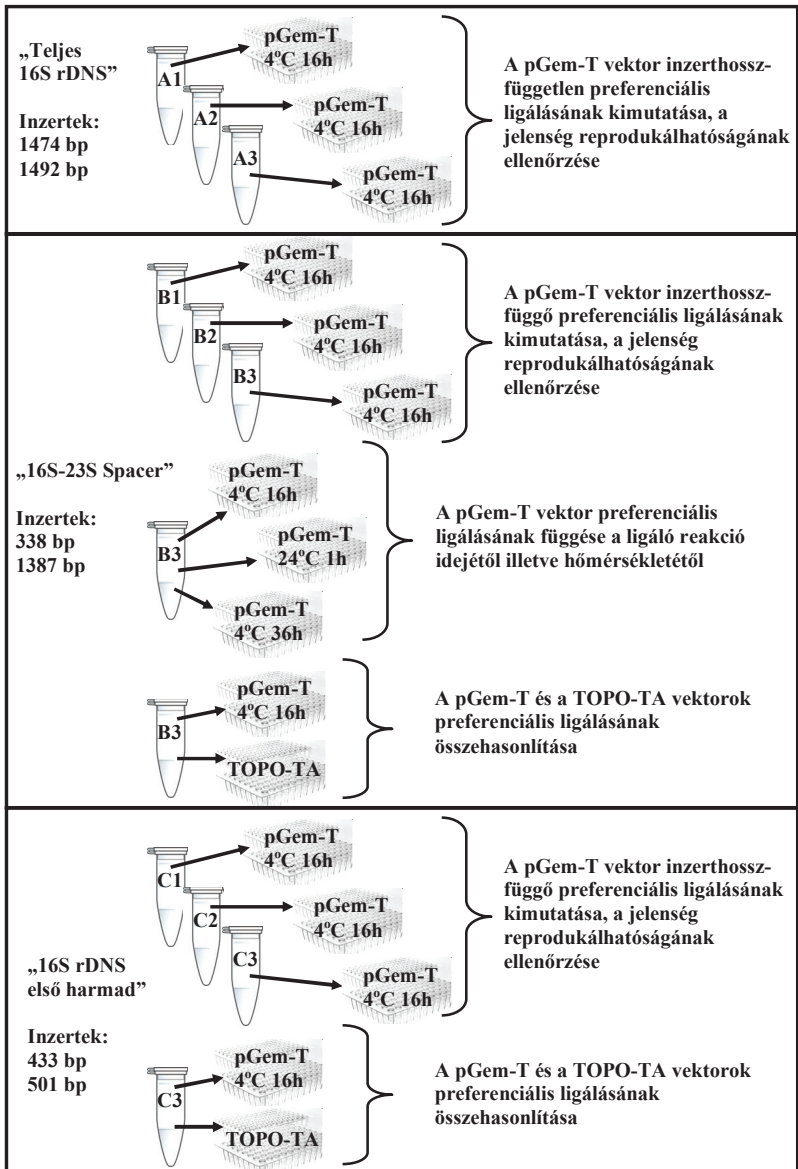
A „B” jelű keverékek a 16S-23S intergenikus spacer régió hosszheterogenitását imitálják, de modellként szolgálhatnak más, hasonló heterogenitású régiót tartalmazó klónkönyvtárak esetében is. A három párhuzamos keverékkel ez esetben is ellenőrizhettük a preferenciális ligálás jelenlétét, illetve ennek reprodukálhatóságát.

A pGem-T klónozó rendszer leírása a ligáló reakció idejére, illetve hőmérsékletére nézve kétféle protokollt is ajánl. Az alap protokoll 1 órányi 24°C-os inkubálást, míg a nagyszámú transzformánsra optimalizált protokoll 16 órányi, 4°C-on történő ligálást ír elő. Elméletileg a mikrobiális közösségelemzésekre értelemszerűen ez utóbbi, nagyszámú transzformáns adó protokollt célszerű alkalmazni, hiszen a vektorrendszer alapvető géntechnológiai alkalmazásai csak pár klón létrehozását igénylik, míg egy közösség leírása során több száz nagyságrendű transzformáns szoktak elemezni. Érdekesnek tartottuk összevetni a két protokoll esetleg eltérő torzítási hajlamát, valamint kideríteni azt, hogy az alap protokollal elérhető-e a szükséges többszáz nagyságrendű klónmennyiség. Ezen kívül egy harmadik

protokolli is alkalmaztunk, mely 4°C-on, de 36 óráig zajló ligálással dolgozik, így biztosan sok transzformáns eredményez, és megnövelt időtartamból kifolyólag az esetleg kisebb hatékonysággal ligálódó inzertek felvételére is több időt ad a vektorrendszernek. A „B3” jelű keverékből készített három, különböző ligálási paraméterekkel készített összehasonlító pGem-T könyvtár alkalmas adott a preferenciális ligálás (ligálási)idő-, illetve hőmérséklet-függésének kimutatására.

A „B3” jelű keverékből TOPO-TA vektorrendszerrel is készítettünk egy könyvtárat a két vektorrendszer torzítási hajlamának összevetése végett.

A „C” jelű keverékek a közösségszerkezeti vizsgálatok során gyakran alkalmazott parciális 16S rDNS (a variábilis V1-V2-V3 régiót tartalmazó első harmadot célzó) klónkönyvtárak természetes hosszheterogenitását reprezentálják, de modellként szolgálhatnak más, hasonló hosszheterogenitású génszakaszokra nézve is. A pGem-T rendszer preferenciális ligálásának statisztikailag is megbízható kimutatásán túl itt is elvégeztük a két vektorrendszer torzításának összehasonlító vizsgálatát (a „C3” keverék esetében).



5. ábra. A modellkísérletek során létrehozott keverékek, és a belőlük készített klónkönyvtárak, az egyes kísérleti elrendezések céljának rövid magyarázatával.

IV.2.4 Az inzertek kópiaszámának meghatározása, az inzertkeverékek elkészítése

A tisztított PCR termékek DNS koncentrációját 260 nm-es hullámhosszon mért abszorbanciájuk alapján határoztuk meg, Hellma Tray Cell (Hellma, Müllheim, Germany) típusú speciális mikroküvetta alkalmazásával, Lambda 35 UV-VIS spektrofotométerben (PerkinElmer Life and Analytical Sciences, Boston, MA), 320 nm-en mért alapvonal korrekció mellett. A minták DNS koncentrációját az $(A_{260\text{nm}} - A_{320\text{nm}}) * 500 = C_{\text{DNS}} \text{ (ng/}\mu\text{l)}$ képlet alapján számoltuk ki, ahol az 500-as szorzó a gyártó által megadott küvetta-specifikus faktor kettősszálú DNS koncentráció meghatározás esetére. Az abszorbancia alapú DNS koncentráció meghatározás legnagyobb pontossággal a 0,1 és 1,0 közötti $A_{260\text{nm}}$ tartományban érhető el (Sauer *et al.*, 1998), így szükség esetén a PCR termékeket kellő mértékben felhígítottuk. Minden mérést szobahőmérsékletű mintákkal, három ismétlésben végeztünk el, és a mért adatok átlagaival számoltunk. A számításokat Excel 2002 SP3 (Microsoft, Redmond, WA, USA) szoftverben készített, automatizált táblázatokkal végeztük. A DNS-ek tisztaságát minden esetben ellenőriztük az $(A_{260\text{nm}} - A_{320\text{nm}}) / (A_{280\text{nm}} - A_{320\text{nm}}) = T$ érték kiszámításával, ahol az elfogadható T érték az 1,7 és 1,9 közötti tartományba esik. Megjegyzendő, hogy tapasztalataim alapján ezzel a mérési technikával egy nagyságrenddel pontosabb, és általában kisebb szórású DNS koncentráció meghatározást lehet kivitelezni, mint a kifejezetten e célra gyártott NanoDrop 1000 (NanoDrop Technologies, Wilmington, DE, USA) készülékkel (ezirányú összehasonlító méréseim nem képezik szorosan a dolgozat tárgyát, így ehelyütt nem térek ki a részletekre).

A tisztított PCR termékek mikroliterenkénti amplikon kópiaszámait a DNS koncentrációk és az amplikonok bázisrendje alapján számított pontos molekulatömege alapján számítottuk ki. A törzsek 16S rDNS szekvenciáját a kutatás korábbi fázisában már meghatároztuk, így a primerek által meghatározott konkrét szakasz molekulatömegét a Mongo Oligo Mass Calculator v2.06 szoftverrel nagy pontossággal kiszámolhattuk (HTTP-7). Az egy mikroliterre eső kópiaszámot a koncentráció (ng/μl) és a moláris tömeg (g/mol) hányadosaként számoltuk. Az 1:1 kópiarányú keverékeket olyan térfogatokban mértük össze, hogy a lehető legnagyobb pontosságú tartományát használjuk ki a pipettának (5-10 μl). A művelethez Eppendorf Reference (0,5-10 μl) pipettát (Eppendorf AG, Hamburg, Germany) és speciális,

minimális folyadékviszataratású, ART-20E típusú (Thermo Scientific, Waltham, MA, USA) pipettahegyeket használtunk.

IV.2.5 Kapilláris elektroforézis és lézergyesztésű fluoreszcencia intenzitásmérés

A kvantifikálási és keverési módszerünk validálásához, valamint később a természetes közösségeken végzett vizsgálatok során is szükségünk volt egy a klónozástól független, de a közösségi PCR termékek mennyiségi viszonyainak pontos meghatározására alkalmas módszerre. A mikrobiális ökológiában elterjedt hosszheterogenitás-PCR (LH-PCR) illetve terminális restrikciós fragmenthossz polimorfizmus (T-RFLP) végső analitikai lépése, a kapilláris elektroforézissel kombinált lézergyesztésű fluoreszcencia intenzitásmérés (CE-LIF) szemikvantitatív jellege és nagyfokú reprodukálhatósága miatt optimális választásnak tűnt e célra. A T-RFLP módszer felbontóképességével és kvantitatív pontosságával behatóan foglalkoztunk korábbi kutatásaink során (Palatinszky, 2004; Sipos *et al.*, 2007), és az ekkor szerzett módszertani jártasság és pozitív tapasztalatok is megalapozták a módszer jelen kutatásban történő alkalmazását.

A CE-LIF eljárás során az előzetesen fluoreszcens primerrel amplifikált DNS szakaszokat hosszuk alapján elválasztjuk kapilláris elektroforézissel, és az elválasztás végén lézergyesztés mellett fluoreszcencia intenzitásuk alapján detektáljuk a különböző retenciós idővel beérkező amplikontömegek mennyiségét. A mért fluoreszcencia intenzitás egyenesen arányos a mintában az adott hosszúságú PCR termékek mennyiségével. A mintával együtt futtatott, más típusú fluoreszcens festékkel megjelölt molekulasúly standard lehetővé teszi az amplikonok méretének igen pontos meghatározását.

A kapilláris elektroforézis előtt a PCR termékek (illetve keverékek) 0,5 µl-ét 12 µl formamidot és 0,6 µl TAMRA-500 belső méret standardot (GeneScan 500 TAMRA Size Standard, Applied Biosystems) tartalmazó futási elegybe kevertük. Az így előkészített mintákat 98°C-on 5 percig denaturáltuk, majd 2 percre jégre helyeztük. Az elválasztást és detektálást ABI PRISM 310 Genetic Analyzer (Applied Biosystems) készülékben végeztük, 8 másodperc injektálás után, 30 percen keresztül POP-4™ polimerben (Performance Optimized Polymer 4, Applied Biosystems), 60°C-os

termosztálás mellett 15 kV-on futtatva. Minden minta esetben három ismételt futtatást végeztünk, illetve szükség esetén többszöri újrafuttatások során a mintákat a detektor optimális fluoreszcencia intenzitás tartományáig hígítottuk (minél nagyobb, de maximum 5000 fluoreszcencia egység nagyságú csúcsokra törekedve).

A kromatogramok elemzését GeneMapper v3.7 szoftverrel végeztük (Applied Biosystems). A kiértékeléshez a szoftver „Microsatellite” módszerét használtuk, a csúcsok detektálásánál a következő paraméterekkel dolgoztunk. Alapvonal: 1 fluoreszcencia egység; minimum csúcspéldányosság: 2 pt, az illesztett polinom fok: 5; ablakméret: 11 pt. A csúcspozíciók meghatározásához „Local Southern” algoritmust használtuk. Az 50 bp-nál rövidebb csúcsokat (a primereket és primerdimereket) kizártuk az elemzésből, az 500 bp-nál hosszabbakat viszont benne hagytuk. A belső méret standard segítségével a kromatogramokon szereplő csúcsokhoz tartozó amplitudonok hosszait ± 1 bp pontossággal határozhattuk meg. A kromatogramok csúcsainak görbe alatti területei az amplitudonok mennyiségeivel arányosak, így a közösségi PCR termékek illetve a modellközösség keverékek mennyiségi viszonyai pontosan meghatározhatóak voltak.

IV.2.6 A kvantifikálási és keverési eljárások validálása

A kvantifikálási és keverési módszerünk pontosságának és reprodukálhatóságának ellenőrzését „C” típusú inzertkeverékek segítségével végeztük (2. táblázat), melyek a klónozásra kerülő keverékektől annyiban különböztek, hogy a kiindulási amplitudonokat fluoreszcensen jelölt forward primerrel (TET63f) hoztuk létre. A „C” jelű keverék tagjai 433, illetve 501 bázispár hosszúak, így a többi inzertpárhoz képest a kapilláris elválasztás jobb felbontást adó mérettartományába esnek, valamint a technika szempontjából relative kicsi méretkülönbségük is nagy pontosságú kópiaszám arány meghatározást tesz lehetővé.

Az egyes fluoreszcens PCR termékek mennyiségi meghatározását, és az 1:1 kópiarányú keverékek összemérését a IV.2.4 fejezetben leírt módon végeztük. Öt független párhuzamos keveréket hoztunk létre, amelyek valós amplitudon arányait CE-LIF technikával mértük meg. Minden keverék esetében három ismételt futtatást végeztünk, és ezek átlagát vettük figyelembe a keverékre vonatkozó mérés

végeredményeként. Az átlagot és szórást a GeneMapper v3.7 programból exportált értékekkel az Excel 2002 SP3 (Microsoft) szoftverben számoltuk.

A kromatogramokon a keverék tagjait reprezentáló csúcsok az amplikonok ismert hossza alapján beazonosíthatóak voltak.

IV.2.7 Klónkönyvtárak készítése

A klónkönyvtárak létrehozásánál alapvetően a gyártók utasításainak pontos betartására törekedtünk. Ettől egyedül a pGem-T vektor esetében alkalmazott, korábban ismertetett, megnövelt ligálási idejű protokoll esetében tértünk el.

IV.2.7.1 Ligálás és transzformálás pGem-T vektorrendszerrel

A klónkönyvtárak elkészítésénél a „pGem-T Vector System” (katalógusszám: A3610, Promega, WI, USA) vektorrendszert, és a gyártó által mellékelte *Escherichia coli* JM109 kémiaiailag kompetens sejteket használtuk. A ligáló reakciók összemérésénél az inzertkeverékekből a használati útmutatóban optimálisként leírt mennyiségű inzert DNS-t adtunk az elegyhez. Ezek az egyes keveréktípusok esetében a következő mennyiségek voltak: „A”: 24,6 ng; „B”: 13,3 ng; „C”: 7,3 ng. A ligáló reakciók összetétele a következő volt:

2X Ligation Buffer	5 µl
pGEM-T Vektor (50 ng)	1 µl
T4 DNS Ligáz (3 Weiss unit/µl)	1 µl
Inzert DNS (koncentrációtól függően)	0,5-3 µl
PCR tisztaságú DEPC kezelt víz	0-2,5 µl (10 µl össztérfogatra kiegészítés)

A ligáló reakciókat a gyártó nagy mennyiségű transzformánsra ajánlott protokollja szerint, 4°C-on, 16 órán át inkubáltuk. Kivételt képezett ez alól a „B3” jelű keverékből készített két könyvtár, ahol egyik esetben a gyártó strandard ligálási protokollját (24°C-on, 1 órán át), illetve egy speciális, 4°C-on, 36 órán át történő ligálást alkalmaztunk.

A ligálás után a ligáló reakciókeverék 2 µl-ét egy jégen tartott 1,6 ml-es Eppendorf csőbe mértük át, majd hozzáértünk 50 µl, a -80°C-os tárolásból frissen felolvasztott kémiaiailag kompetens sejtuszpenziót. Óvatos keverés után 20 percen át jégen inkubáltuk az elegyet, majd a sejteket 45 másodpercen át pontosan 42°C-os vízfürdőben hősokkoltuk. A transzformálás után még 2 percre visszahelyeztük jégre a sejtuszpenziót, majd 950 µl szobahőmérsékletű SOC táplevest mértünk hozzá (a SOC médium pontos összetételét a vektorrendszer útmutatója tartalmazza). A transzformált sejteket 1,5 órán át 37°C-os rázótermosztátban inkubáltuk (150 fordulat/perc mellett). A sejtek kiszélesztéséhez előkészítettünk könyvtárként 6 darab 100 µg/ml ampicillint tartalmazó LB-táplemezt, melyeken előzetesen 20 µl 50 mg/ml X-gal-t, illetve 100 µl 100 mM IPTG-t egyenletesen eloszlattunk (a gyártó utasításai szerint). A szuszpenzióból 100-100 µl-t szélesztettünk, majd egy éjszakán át, 37°C-on történő inkubálás után a fehér színű (inzertet tartalmazó) telepekből körülbelül 320-at átpontoztunk ampicillin tartalmú LB agar táptalajra. Az átolított telepeket egy éjszakán át 37°C-on növesztettük DNS izolálás előtt.

IV.2.7.2 Ligálás és transzformálás TOPO-TA vektorrendszerrel

A klónkönyvtárak elkészítésénél „TOPO-TA Cloning Kit” (katalógusszám: K4500-01, Invitrogen, CA, USA) vektorrendszert, és a gyártó által mellékelte *Escherichia coli* One Shot TOP10F' kémiaiailag kompetens sejteket használtunk. A ligáló reakciók összemérésénél az inzertkeverékekből a használati útmutatóban optimálisként leírt mennyiségű inzert DNS-t adtunk az elegyhez. Ezek az egyes keveréktípusok esetében a következő mennyiségek voltak: „B”: 2,0 ng; „C”: 1,1 ng. A ligáló reakciók összetétele a következő volt:

Puffer (1,2 M NaCl; 0,06 M MgCl ₂)	1 µl
pCR [®] 2.1-TOPO Vektor (10 ng/µl)	1 µl
Inzert DNS (koncentrációtól függően)	0,5-4 µl
PCR tisztaságú DEPC kezelt víz	0-3,5 µl (6 µl össztérfogatra kiegészítés)

A ligáló reakció összetételénél és inkubálási paramétereinél (30 perc 22°C-on) a gyártó nagy mennyiségű transzformánsra ajánlott protokollja szerint jártunk el. A

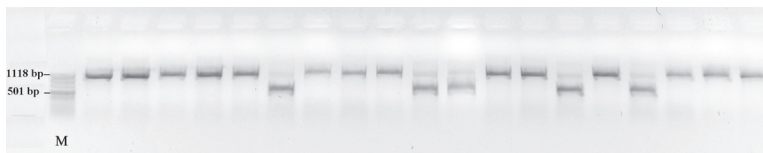
ligálás után az elegy 2 µl-ét hozzáadtuk 50 µl, a -80°C-os tárolásból frissen felolvasztott, kémiaiilag kompetens One Shot TOP10F' sejtuszpenzióhoz, a gyártó által a sejtek tárolására alkalmazott mikrocentrifuga csőben. Óvatos keverés után 10 percig jégen inkubáltuk, majd 30 másodpercig 42°C-os vízfürdőben hősokkoltuk a sejteket. A szuszpenziókat 2 percre visszahelyeztük jégre, majd 250 µl-nyi (a gyártó által biztosított) SOC médiumot adtunk hozzájuk. Ezután egy óráig rázattuk a sejteket 37°C-os termosztátban, 200 fordulat/perc sebességgel. Az inkubálás után a sejtuszpenziókat 50-50 µl-ét 6 darab 50 µg/ml ampicillint tartalmazó LB-táplemezre szélesztettük, amelyeken előzetesen 40 µl 40 mg/ml X-gal-t, illetve 40 µl 100 mM IPTG-t oszlattunk el egyenletesen (a gyártó utasításai szerint). Egy éjszakán át, 37°C-on történő inkubálás után a fehér színű (inzertet tartalmazó) telepekből körülbelül 320-at átpontoztunk ampicillin tartalmú LB agar táptalajra steril fogpiszkálók segítségével. Az áttoltott sejteket egy éjszakán át 37°C-on növesztettük DNS izolálás előtt.

IV.2.8 A klónok azonosítása

A táplemezeken kinőtt átpontozott telepeket steril fogpiszkálók segítségével átémeltük 96 lyukú PCR mikrotiter lemezek lyukjaiba (egy telep/lyuk), majd 50 µl HPLC tisztaságú vizet adtunk hozzájuk. A sejteket a lemezek lefedése után vortexeléssel szuszpendáltuk, majd 98°C-on 5 percig „forraltuk”. A sejttermelékét centrifugálással (4000 g, 5 perc) különítettük el a DNS-t tartalmazó felülúszótól. A centrifugáláshoz Jouan CR422 típusú centrifugát használtunk, mikrotiter lemez feltétes kilendülő rotorral.

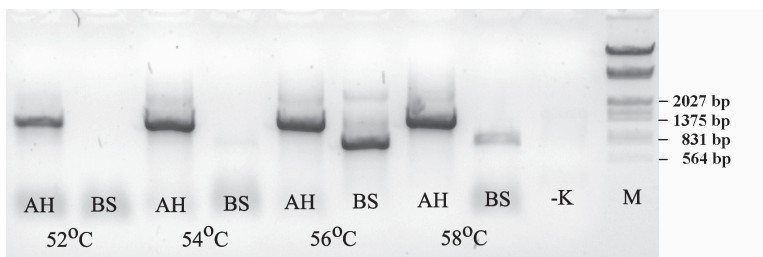
A „B” és „C” jelű keverékekből készített könyvtárakban a klónokban levő inzerteket hosszuk alapján azonosítottuk. Ehhez PCR-rel felszaporítottuk az inzerteket, vektorspecifikus, standard M13 primereket alkalmazva (M13f: 5'-GTAAAACGACGGCCAG-3'; M13r: 5'-CAGGAAACAGCTATGAC-3'), mely primerpár kötőhelye mindkét vektortípuson megtalálható az inzertálóhely közelében. A PCR összetétele és hőprofilja megegyezett a IV.2.2 fejezetben leírtakkal, azzal a különbséggel, hogy itt a *Taq* polimerázt a PCR premixbe már bemértük, és nem a kezdeti denaturáció után adtuk hozzá az elegyekhez (ez 96 lyukú PCR lemez esetén problematikus lett volna). A *Taq* polimeráz kímélése végett a kezdeti denaturációs lépést is 95°C, 3 percre változtattuk.

A PCR termékeket a IV.2.2 fejezetben leírt módon, agaróz gélelektroforézissel választottuk el, s a klónokban levő inzerteket hosszuk alapján azonosítottuk (6. ábra).



6. ábra. A klónok azonosítása agaróz gélelektroforézissel. A gyorsabb migrációjú PCR termékek (alsó sor) a rövidebb inzertet tartalmazó klónokra jellemzőek. A bal szélen pUC Mix Marker-8 molekulásúly standard futott („M”-el jelölve).

Az „A” jelű keverékekben az inzertek mérete kevesebb mint 0,5%-kal tért el, így a fentebb leírt egyszerű gélelektroforézissel történő klón-azonosítás nem volt lehetséges. Ezért az egyes klónok amplifikációjához az inzertek szekvenciája alapján olyan „differenciáló” reverz primereket terveztünk, amelyek azonos forward primerrel (63f) kombinálva a két inzertből különböző hosszúságú szakaszt szaporítanak fel (AHrev: 5'-GCAAGGCCCGGAAC-3' *Aeromonas hydrophila* specifikus, és BSrev: 5'-GTAAGGGGCGGAAAC-3' *Bacillus subtilis* specifikus primer). A primerkeverék működésének ellenőrzését és optimalizációját tiszta genomi DNS-eken, gradiens PCR készülék segítségével végeztük, különböző annealációs hőmérsékleteken (52°C, 54°C, 56°C, 58°C) tesztelve a primerkeverék működőképességét, specifitását (lásd 7. ábra).



7. ábra. A differenciáló primerkeverék működőképességének tesztelése és optimalizációja. „AH”: *Aeromonas hydrophila* templát DNS, „BS”: *Bacillus subtilis* templát DNS. A jobb szélen „-K”-val jelölve a negatív kontroll, „M”-mel jelölve a Lambda DNA/*EcoRI*+*HindIII* Marker-3 molekulásúly standard.

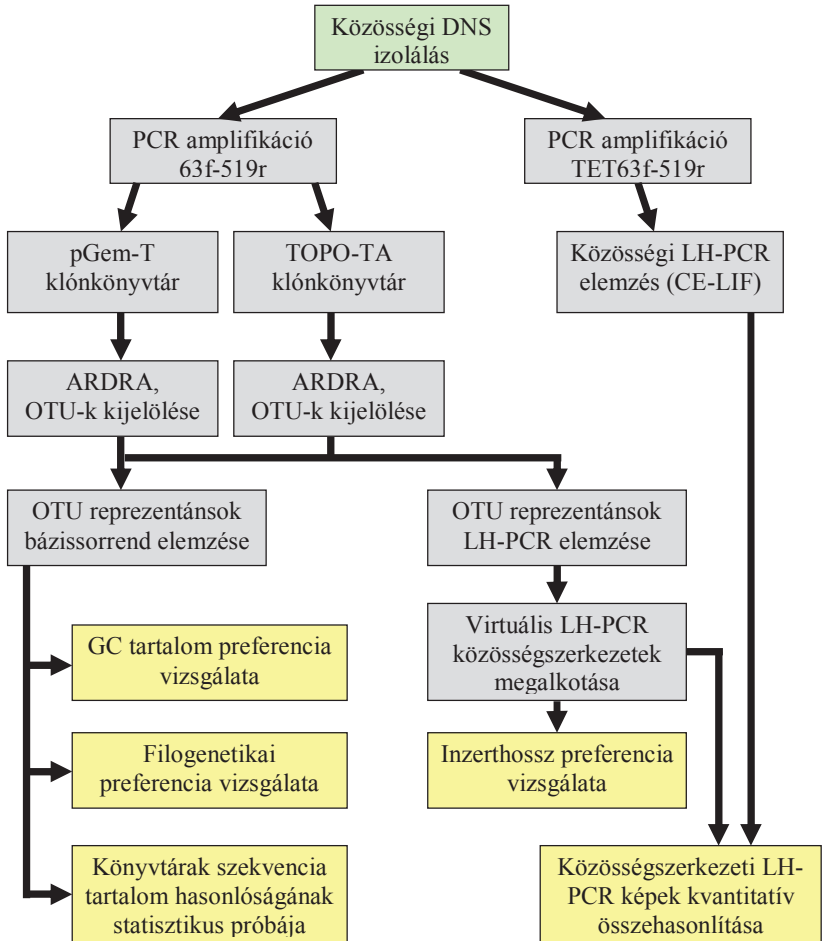
Az egyes klónokból a „differenciáló” primerkeverék használatával 56°C-os annealációs hőmérséklet alkalmazása mellett szaporítottuk fel az inzerteket. Így az

egyszerű gélelektroforézisen alapuló elválasztás és inzertazonosítás ez esetben is lehetővé vált.

IV.3 Környezeti minta

A különböző klónozó rendszerek életszerű komparatív vizsgálatát és klónozástól független módszerrel történő összehasonlítását egy természetes baktériumközösség leírása során végeztük el. Az ehhez leginkább megfelelő mintavételi helyet az ELTE Mikrobiológiai Tanszékén aktuálisan vizsgált élőhelyek közül választottuk ki, az addig nyert közösségszerkezeti ismeretek, részeredmények alapján. Célunk olyan közeg, minta, illetve baktériumközösség kiválasztása volt, ami környezeti mikrobiológusok által gyakran vizsgált mintatípus, és ahol relatíve nagy diverzitás tapasztalható, valamint a 16S rDNS első harmada nagy hosszheterogenitást mutat. Ezen kritériumoknak a Hévízi-tó üledékének az iszapfelszíntől számított 20 cm-es mélységében élő mikrobaközösség felelt meg leginkább (Krett és Palatinszky, 2009).

A környezeti mintán végzett komparatív vizsgálatok sematikus összefoglalását a 8. ábra mutatja.



8. ábra. A természetes baktériumközösségen végzett klónkönyvtár-, illetve LH-PCR-alapú összehasonlító vizsgálatok folyamatábrája.

IV.3.1 Mintavétel, közösségi DNS kivonása

A mintavétel 2007. október 29-én történt a Hévízi-tó üledékéből Hargrave-féle mintavevő készülékkel, a forráskráter keleti peremén, körülbelül 3 méteres mélységből.

A mintát steril eszközökkel a kiemelt iszaptömb közepéből, az iszapfelszín alatti 20 cm-es mélységű rétegből vettük.

Az üledékmintából a közösségi DNS-t UltraClean Soil DNA Kit (Mo Bio, Carlsbad, CA, USA) segítségével vontuk ki, a gyártó utasításainak megfelelően.

IV.3.2 Hosszheterogenitás-PCR (LH-PCR) vizsgálat

A közösségszerkezet parciális 16S rDNS hosszheterogenitás alapú kvantitatív vizsgálatához a közösségi DNS-ből a 16S rRNS gén első harmadát a IV.2.2 fejezetben leírt PCR paraméterekkel, TET63f - 519r primerpár alkalmazásával szaporítottuk fel.

A PCR terméket etanol precipitációval megtisztítottuk a kapillárisra történő felvitel előtt. Ennek során 600 µl-es Eppendorf-csőben 20 µl termékhez 80 µl precipitáló oldatot adtunk, melynek összetétele:

95% etanol	62,5 µl
3 M Na-oxalacetát	3 µl
DEPC kezelt víz	14,5 µl

Az így elkészített elegyet keverés után 15 percig állni hagyjuk szobahőmérsékleten. Ezután 4°C-on 20 percig 12000 g-vel centrifugáltuk, majd a felülúszót elöntöttük (a precipitált DNS ilyenkor a csövek falán marad). A csövekbe 250 µl 70%-os etanolt pipettáztunk, majd a mintákat 10 percen át 12000 g-vel centrifugáltuk. A felülúszót elöntöttük, majd a mintákat 30 percig vákuumcentrifugában szárítottuk, hogy eltávolítsuk az etanol legkisebb maradványát is. A kicsapódott DNS-t ezek után csövenként 30 µl DEPC kezelt vízben feloldottuk.

Az etanol precipitációval megtisztított közösségi PCR terméket a IV.2.5 fejezetben leírt módon kapilláris elektroforézissel elválasztottuk, és lézergeszteszt mellett fluoreszcencia intenzitás alapján detektáltuk. A GeneMapper v3.7 programból exportált adattáblázatokkal (csúcs pozíció, csúcs alatti terület) ezt követően az Excel 2002 SP3 (Microsoft) szoftverben dolgoztunk.

IV.3.3 Közösségi PCR termék előállítás, klónozás

A közösségi DNS-ből a parciális 16S rDNS klónkönyvtárak létrehozásához a 16S rRNS gén első harmadát a IV.2.2 fejezetben leírt PCR paraméterekkel, 63f-519r primerpár alkalmazásával szaporítottuk fel. A PCR terméket megtisztítottuk és kvantizáltuk a korábban leírt módon (IV.2.2 és IV.2.4 fejezetek). Ez esetben a kvantizálás csak a klónozó rendszerek optimális inzertmennyiségének beállításához kellett.

A PCR termékből a pGem-T és a TOPO-TA klónozó rendszerrel is létrehoztunk egy-egy 300 tagú klónkönyvtárat a IV.2.7 fejezetben leírtak szerint.

IV.3.4 A klónkönyvtárak feldolgozása, klónok csoportosítása ARDRA-val

A klónok csoportosítását (az azonos inzertet tartalmazó klónok megkeresését) amplifikált riboszomális DNS restrikciós elemzéssel (ARDRA) végeztük (Vaneechoutte *et al.*, 1993). Ez lényegében egy restrikciós fragmenthossz polimorfizmus (RFLP) analízis, amelyet alapvetően baktériumtörzsek 16S rRNS gén hasítási mintázat alapján történő csoportosítására fejlesztettek ki, de igen gyakran alkalmazzák klónok csoportosítására is. Tekintve, hogy jelen esetben is riboszomális génszakasszal dolgoztunk, a kétféle megnevezés – RFLP versus ARDRA – közül ez utóbbit használjuk.

A klónokból a DNS izolálást és a vektorspecifikus (M13) PCR-t a IV.2.8 fejezetben leírtak szerint végeztük. Ezek a PCR termékek az inzerteken kívül még tartalmaznak valamennyit a vektorokból is a két végükön, mivel a vektorspecifikus primerek kötőhelyei nem közvetlenül az inzertálódás helye mellett vannak. Ezen túl, miután az inzertek kétféle orientációban is be tudnak kötődni a vektorokba, az ARDRA során egy bizonyos inzert szekvenciára az M13-as PCR utáni közvetlen hasítás esetén kétféle hasítási mintázatot is kapnánk. Ezért a restrikciós hasítás előtt egy újabb PCR amplifikáció következett az eredeti primerpárral (63f-519r), melynek eredményeként kizárólag az eredeti inzerteket nyertük vissza termékként.

A PCR termékeket gyakran hasító restrikciós endonukleázok segítségével emésztettük. Első körben kétféle enzimmel (*Bsu*RI, *Msp*I) végeztük el a hasítást, majd a kiértékelés után a leggyakoribb ARDRA csoportra egy harmadik (*Hin*6I) enzimmel is

végeztünk egy ellenőrző-finomító mintázatelemzést. A felhasznált restriktációs endonukleázokat, a nekik megfelelő puffertípust, a hasítóhelyüket és a működés optimális hőmérsékletét a 3. táblázat részletezi.

Enzim (Fermentas)	Hasítóhely	Puffer (Fermentas)	Optimális hőmérséklet
<i>Bsu</i> RI	5' GG/CC 3' 3' CC/GG 5'	10x Puffer Red	37°C
<i>Hin</i> 6I	5' G/CGC 3' 3' CGC/G 5'	10x Puffer Tango	37°C
<i>Msp</i> I	5' C/CGG 3' 3' GGC/C 5'	10x Puffer Tango	37°C

3. táblázat: Az emésztéshez használt restriktációs endonukleázok és jellemzőik.

A restriktációs hasításokat 96 lyukú PCR lemezen végeztük, 20 µl-es végtérfogatban, az alábbi reakcióösszetétellel:

Enzim 0,3 µl
Puffer 2 µl
DEPC víz 7,7 µl
Templát 10 µl

Az inkubáció 37°C-os vízfürdőben, egy éjszakán át történt. A hasítási termékeket 2%-os agaróz gélben 80 percen át 80 V-on történő elektroforézissel választottuk el. A hasítási mintázatokat etídium-bromidos festés mellett áteső UV-fényben, digitális képalkotó eljárással dokumentáltuk. Minden sorban pUC Mix Marker-8 molekulaszú standardot is futtatunk a képek korrekt összevethetősége végett. A hasítási mintázatok alapján a klónokat ARDRA csoportokba soroltuk, majd ezekből a csoportok nagyságának függvényében egy vagy több reprezentáns klónt választottunk ki.

IV.3.5 Az ARDRA csoport reprezentáns klónok bázissorrend elemzése

A könyvtárak filogenetikai diverzitásának leírására az ARDRA csoport reprezentáns klónokat bázissorrend elemzésnek vetettük alá. Az ARDRA csoportosítás hitelességének ellenőrzése miatt minden csoportból a klónok legalább ötödén végeztünk

bázissorrend elemzést véletlenszerű kiválasztás alapján (tehát például a legnagyobb, 20-30 klónt számláló csoportokból 4 illetve 6 klónt választottunk ki).

A klónok M13-as PCR termékeire az eredeti (63f-519r) primerpárral újabb PCR-t készítettünk, majd megtisztítottuk a IV.2.2 fejezetben leírt módon. A tisztított PCR termékek bázissorrendjét BigDye Terminator v3.1 Cycle Sequencing (Applied Biosystems) alkalmazásával határoztuk meg, 519r primer felhasználásával. A reakciót a gyártó utasításait követve, de takarékosági okokból a megadott reakcióterfogat egynegyedében (10 µl) végeztük, a következő összetétellel:

Big Dye Terminator RR Mix	1 µl
5x Big Dye Puffer	1,5 µl
Primer ($3,25 \times 10^{-5} \text{M}$)	0,5 µl
DEPC kezelt víz	5 µl
Templát DNS	2 µl

A szekvenáló reakciót GeneAmp 2700 PCR készülékben futtattuk a következő hőprofilal:

denaturáció	96°C	10 sec	} 28 ciklusban
annezáció	50°C	5 sec	
extenzió	60°C	4 min	
hűtés	4°C	∞	

A bázissorrend elemzéshez szükséges kapilláris elektroforézist a Biomi Kft. (Gödöllő) végezte. A kromatogramokat Chromas Lite program (Technelysium Pty Ltd, Tewantin, Ausztrália) segítségével manuálisan javítottuk. Az esetleges kimérikus szekvenciákat a Mallard (Ashelford *et al.*, 2006) program segítségével azonosítottuk, s ezeket a további elemzésekbe nem vettük be.

A környezeti klónkönyvtárakból meghatározott összes szekvenciát publikáltuk az EMBL adatbázisban az FR728682-től FR728932-ig terjedő azonosító kódokkal.

IV.3.6 A reprezentáns klónok LH-PCR vizsgálata

Az OTU reprezentánsokat LH-PCR vizsgálatnak is alávetettük az IV.3.2 fejezetben leírtak szerint, mivel az egy irányú bázissorrend elemzés során kapott DNS szekvencia nem mindig tartalmazza a szekvenáló primerhez közeli szakaszt, így a teljes inzerthossz nem kiszámítható. Ellenben az LH-PCR-rel mért inzerthosszok nagy hitelességgel feleltethetők meg a közösségi LH-PCR kromatogramban detektált csúcsokkal.

IV.3.7 A klónkönyvtárak összehasonlító elemzése

IV.3.7.1 GC-tartalom, átlagos inzerthossz

A klónszekvenciák GC-arányait a BioXpress v1.0 (Progenus, Gembloux, Belgium) szoftverrel számoltuk ki. A klónkönyvtárakra vonatkozó GC-tartalom átlag és inzerthossz statisztikákat (átlag, rank-abundancia) Excel 2002 SP3 (Microsoft) szoftverben számoltuk.

IV.3.7.2 Közösségszerkezetek LH-PCR alapú összevetése

Az OTU reprezentánsok LH-PCR-rel mért inzerthosszainak és az OTU-kba tartozó klónok mennyiségeinek alapján mindkét klónkönyvtárra nézve alkottunk egy-egy virtuális LH-PCR közösségszerkezeti képet. Ezt a mintázatot összevetettük a (klónozástól függetlenül nyert) közösségi LH-PCR mintázattal. A pontos kvantitatív összehasonlítás érdekében a kapilláris LH-PCR kromatogram jól elkülöníthető, domináns csúcsai által lefedett hossztartományok alapján úgynevezett „csúcs csoportok”-at alkottunk az abba a hossztartományba eső klónokból. Az adott „csúcs csoport” közösségen belüli részaránya a kromatogramon a csúcs alatti terület nagyságával, míg a klónkönyvtárak esetében a „csúcs csoport”-ba tartozó klónok számával arányos.

IV.3.7.3 A klónok filogenetikai megoszlásának elemzése

A klónszekvenciáink illesztését és filogenetikai azonosítását a SINA Webaligner szoftverrel végeztük, mely az ARB on-line elérhető illesztőprogramja (ARB-Silva, Pruesse *et al.*, 2007). Az alkalmazás minden egyes klónszekvenciának megkeresi a legközelebbi rokon szekvenciáját az ARB adatbázisában, és a beküldött, illetve az ARB-ből származó szekvenciahalmazra illesztést végez. Az ARB-ből származó szekvenciák filogenetikai hovatartozása alapján megállapíthattuk klónjaink valószínű rendszertani helyét, és így a klónkönyvtáraink filogenetikai összetételét.

IV.3.7.4 Hasonlósági indexek, statisztikai elemzés

A klónkönyvtárak OTU összetételeinek hasonlósági indexeit, valamint az eredeti közösségek becsült fajszámait a SPADE szoftverrel számoltuk ki (Chao *et al.*, 2006). A hasonlósági indexek közül a mikrobiális közösségszerkezeti elemzésekre optimalizált Chao-Jaccard és Chao-Sorenson indexeket használtuk, a valódi fajszámbebecslésekhez pedig a „Chao1-bias corrected” algoritmust (Chao *et al.*, 2005). A módszereket Hughes és munkatársai remek áttekintő munkája nyomán választottuk ki (Hughes *et al.*, 2001).

IV.3.7.5 Különbség szignifikancia teszt (Libshuff elemzés)

A két klónkönyvtár összetételbeli különbségének szignifikanciaszintjét a kifejezetten 16S rDNS klónkönyvtárak szekvenciartalmának összehasonlítására kifejlesztett Libshuff nevű algoritmussal (Singleton *et al.*, 2001) becsültük meg, a WebLibshuff szoftver alkalmazásával (HTTP-8). A Libshuff elemzés azon nullhipotézis statisztikus próbája, hogy két 16S rDNS klónkönyvtár azonos közösséget mintáz – jelen esetben természetesen a közösség, sőt a klónozendó PCR termék is konkrétan ugyanaz, így amennyiben szignifikáns különbséget tapasztalunk, az a könyvtárak elkészítésénél alkalmazott eltérő klónozó rendszerek szelektivitásának jele lehet.

A Libshuff elemzés során a két könyvtár (jelöljük X-el illetve Y-al) által tartalmazott klónszekvencia halmazokat a valódi közösség egy-egy korlátos mintájaként fogjuk fel. Első lépésben az egyik klónkönyvtár általi lefedettséget (coverage: C_X) számoljuk ki az egyedi szekvenciák könyvtáron belüli részarányából:

$$C_X = 1 - (\text{egyedi szekvenciák száma} / \text{összes szekvencia})$$

Az „egyediség”-et DNS szekvenciák esetében meghatározhatjuk különböző mértékű szekvenciahasonlóságokkal, másként fogalmazva különböző mértékű megengedett evolúciós távolságokkal (D). Azaz, mondjuk a 100%-os egyezésnél ($D=0$) enyhébb, pl. 99%-os szekvenciahasonlóságot ($D=0,01$) is már azonos csoportként fogunk fel. Amennyiben a könyvtár lefedettségét (C_X) különböző megengedett evolúciós távolság (D) értékek függvényében ábrázoljuk, megkapjuk a könyvtár úgynevezett „homológ lefedettség görbét”.

Ezután az egyik könyvtár minden egyes szekvenciaféleségére megnézzük, hogy az a másik könyvtárban egyedinek számítana-e:

$$C_{XY} = 1 - (\text{egyedi szekvencia lenne az Y könyvtárban} / \text{az Y összes szekvenciája})$$

A C_{XY} -t ábrázolva különböző megengedett evolúciós távolság (D) értékek függvényében, megkapjuk a két könyvtár úgynevezett „heterológ lefedettség görbét”. A két görbét azonos grafikonon ábrázolva látható, hogy az Y könyvtár mennyire jól reprezentálja az X könyvtárat, ugyanis amennyiben a görbék teljesen azonos lefutásúak, a könyvtárak azonos mintának mondhatók. Ahhoz viszont, hogy meghatározhassuk a két könyvtár különbözőségének szignifikanciaszintjét, további elemzés szükséges. Ki kell számolnunk a C_X és C_{XY} értékek különbségeit minden D érték esetén, majd ebből a függvényből Cramér von-Mises statisztikával meg kell határozni egy konkrét – az adott könyvtár pára jellemző – ΔC értéket (ez az egyes D értékekhez tartozó C_X és C_{XY} értékek különbségeinek négyzetösszege).

Ezután a két könyvtárban levő szekvenciákat összekeverjük, majd véletlenszerűen szétosztjuk két, az eredetiekkel megegyező méretű mintára. A véletlenszerűen válogatott könyvtárpárra is meghatározzuk a ΔC értéket. Ha a két eredeti könyvtár erősen különbözött, akkor a kevert könyvtárak ΔC értékének jelentősen kisebbnek kell lennie. Amennyiben ezt a véletlenszerű keverés kísérletet sokszor elvégezzük, az eredeti minták ΔC értékének a véletlenszerű minták nagy részének ΔC -jénél nagyobbak kell lennie (ez a fajta elemzés a Monte-Carlo teszt egy alkalmazása). Hogy az eredeti könyvtárpár ΔC értékének és a kevert könyvtárak ΔC értékei különbségének szignifikanciáját meghatározhassuk, bevezetünk egy p értéket, ami

lényegében a ΔC értékek rangsorolásából határozható meg. Például ha 1000 ismétléssel elvégezzük a keveréses tesztet, és az eredeti ΔC értékünk az ötödik legmagasabb érték lesz az 1000 ΔC értékből, akkor $p=0,005$. Két klónkönyvtár szignifikánsan különböző összetételűnek mondható (95%-os megbízhatósággal), amennyiben ez a p érték 0,05 alatti (Singleton *et al.*, 2001).

A Libshuff analízis előkészítéseként a klónkönyvtáraink összes szekvenciáját a könyvtár összetételének megfelelő számban tartalmazó illesztést készítettünk a SINA Webaligner (HTTP-9) segítségével, majd fasta formátumban exportáltuk. Ezen illesztés szekvenciáinak távolsági mátrixát kiszámoltuk a DNADIST v3.5c programmal, Jukes-Cantor korrekció alkalmazásával (Felsenstein, 1989), majd ezt az adatfájlt a megfelelő formátumúra hozva adtuk meg a WebLibshuff bemeneti adathalmazaként. A Monte-Carlo teszt során 1000 véletlenszerű keverést alkalmaztunk.

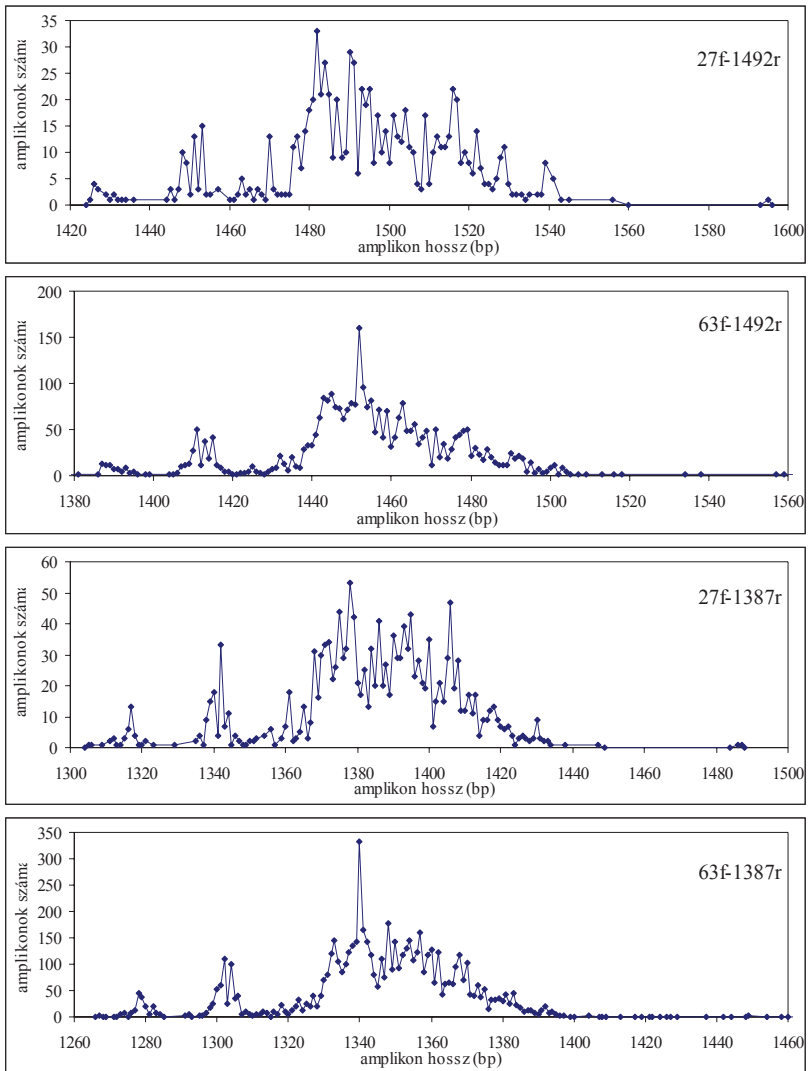
V. Eredmények és értékelésük

V.1 In silico hosszheterogenitás vizsgálatok

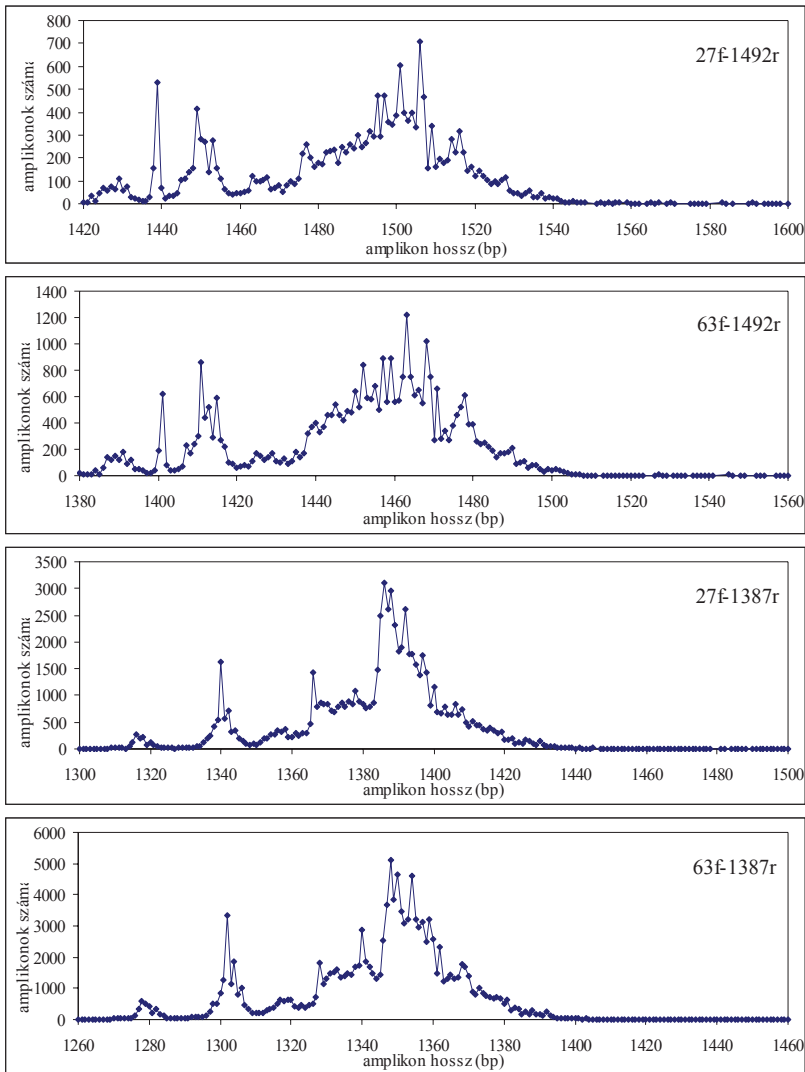
A LengHet szoftver segítségével felmértük a teljes 16S rRNS gén, illetve az első harmadát kitevő, a V1-V2-V3 variábilis régiókat magában foglaló szakasz természetes hosszheterogenitását két, ellenőrzött minőségű szekvenciákat tartalmazó adatbázis alapján. Ezek egyike a leírt típusörzsek 16S rDNS szekvenciáit tartalmazta (az RDPII részeként), a másik pedig az ARB adatbázisából származó, típusörzseket, izolátumokat és környezeti klónszekvenciákat tartalmazó, kiméramentes, másodlagos térszerkezetre nézve is ellenőrzött, annotált referencia adatbázis volt („SSU RefNR 104”).

V.1.1 A teljes 16S rRNS gén hosszheterogenitása

A 16S rDNS szakasz széléhez közeli 27f és 1492r primerkötőhelyeket az adatbázisokban levő szekvenciák jóval kisebb része tartalmazta, mint a szakasz kissé beljebb eső részén levő 63f és 1387r primerkötőhelyeket, ezért az illesztéseket mindegyik primerpárral, illetve ezek kombinációival is elvégeztük, annak érdekében, hogy minél teljesebb képet kaphassunk a hosszheterogenitásról. A többféle univerzális primerpár alkalmazásának további előnye, hogy a primerek szelektivitásából adódó esetleges torzítások is észlelhetőek. A kétféle adatbázisból a LengHet szoftverrel kapott eredményeinket a 9. és a 10. ábrák mutatják.



9. ábra. Az RDPII adatbázisban fellelhető 7448 típusörzs 16S rRNS génjének hosszheterogenitása különböző univerzális primerpárok kötőhelyei közötti szakaszon. Az alkalmazott primerpárokat a grafikonok jobb felső sarkában tüntettük fel.



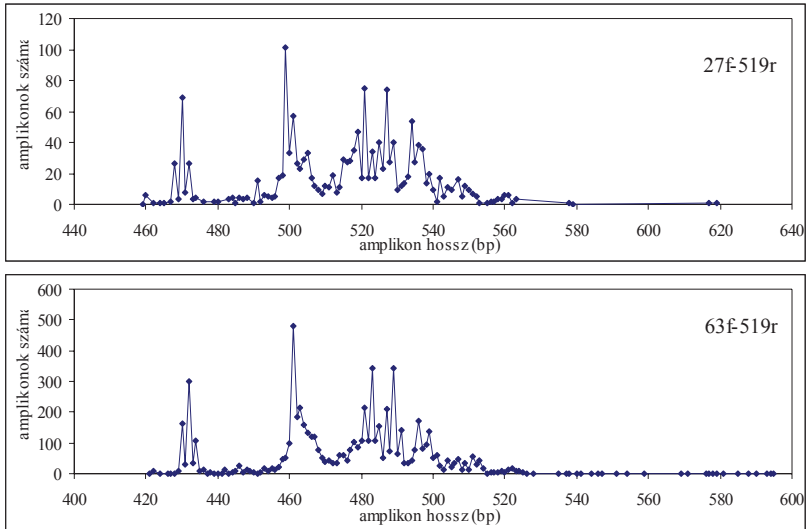
10. ábra. Az ARB referencia adatbázis 262092 válogatott, kiemelt minőségű 16S rDNS szekvenciájának hosszheterogenitása különböző univerzális primerpárok köthelyei közötti szakaszon. Az alkalmazott primerpárokat a grafikonok jobb felső sarkában tüntettük fel.

A teljes 16S rDNS szakaszon beljebb eső primerek alkalmazásával csaknem egy nagyságrenddel több szekvenciát sikerült bevonni a vizsgálatba. A különböző primerek alkalmazásával a hosszheterogenitás mintázatokban azonban nem jelentek meg új, szélsőséges amplikonhosszt adó csoportok, csak a különböző csúcsok arányai változtak, így a mi szempontunkból a primerek filogenetikai szelektivitása nem befolyásolta számottevő mértékben a vizsgálatot.

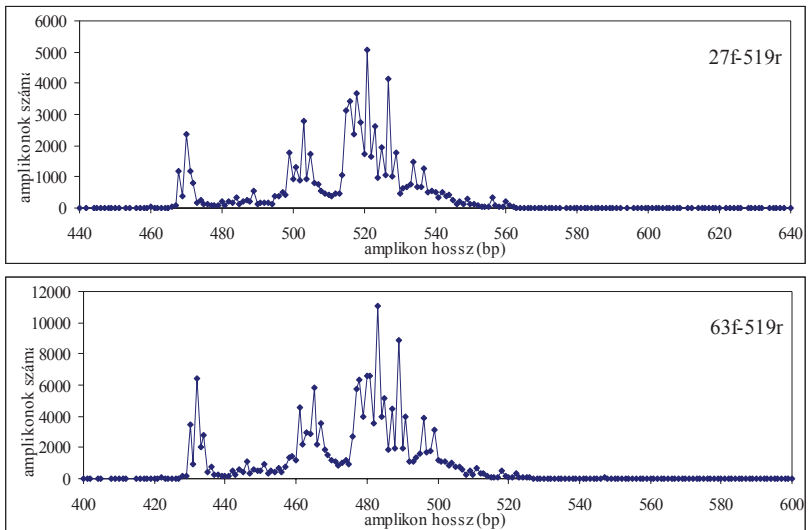
A teljes hosszon tapasztalt körülbelül 200 bázispáros hosszheterogenitás jól egybeesik a szakirodalom jóval korábbi, kis mintára alapozott becsléseivel (García-Martínez *et al.*, 1999). Mindenképpen említésre méltó, hogy a 10. ábra grafikonjainak hosszan elnyúló jobboldali lapos szakasza is 4-5 szekvenciát jelent minden egyes amplikonhossz esetében, sőt még az ábrázolt amplikonhossz tartományon kívül is igen sok találat volt. Konkrét példákat kiemelve az elviekben megbízható adatokkal rendelkező típusörzsek köréből: 27f és 1492r primerekkel illetve a *Thermoanaerobacter pseudethanolicus* ATCC 33223 típusörzs 1745 bázispár hosszú amplikont ad (szekvencia azonosító: CP000924), míg a *Nereida ignava* (T) 2SM4 típusörzs csak 1426 bázispár hosszút (szekvencia azonosító: AJ748748). Az összes szélsőséges esetre nézve persze nem lehet kizárni a hibás szekvenciák jelenlétét, illetve a nem specifikus primerilleszkedések miatti műtermékeket, de az adatbázisok szelektált jellege – típusörzsek hivatalos szekvenciái az RDPII esetében, illetve másodlagos szerkezet alapján is ellenőrzött szekvenciák az ARB adatbázis esetében – a találatok hiteles volta mellett szólnak.

V.1.2 A 16S rDNS első harmadának hosszheterogenitása

A „teljes” 16S rDNS felszaporítására alkalmas primereken kívül az igen gyakran elemzett első harmadát kitevő, a V1-V2-V3 variábilis régiókat magában foglaló szakasz amplifikációjánál használt 519r primert is bevettük a vizsgálatba, így a parciális 16S rDNS klónkönyvtárakra jellemző hosszheterogenitást is fel tudtuk térképezni. Az RDPII-ben szereplő típusörzsekre jellemző, illetve az ARB referencia adatbázisában tapasztalható hosszheterogenitás eloszlásokat a 11. és 12. ábrák mutatják.



11. ábra. Az RDPII adatbázisban fellelhető 7448 típusörzs parciális 16S rDNS hosszheterogenitása különböző univerzális primerpárok kötőhelyei közötti szakaszon. Az alkalmazott primerpárokat a grafikonok jobb felső sarkában tüntettük fel.



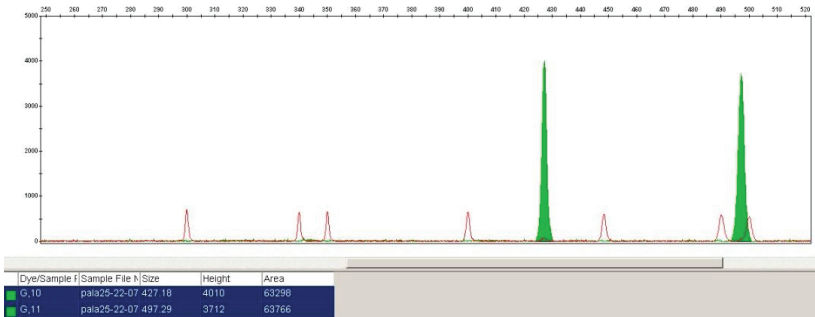
12. ábra. Az ARB adatbázis 262092 referencia minőségű szekvenciájának parciális 16S rDNS hosszheterogenitása különböző univerzális primerpárok kötőhelyei közötti szakaszon. Az alkalmazott primerpárokat a grafikonok jobb felső sarkában tüntettük fel.

A 16S rRNS gén első harmadának tapasztalt hosszheterogenitása a szekvenciák nagy részére nézve körülbelül megegyezik a korábban becült megközelítőleg 100 bázispáros értékkel (Suzuki *et al.*, 1998) – de jelentős számban vannak sokkal szélsőségesebb hosszokat mutató típusörzsek és klónszekvenciák is.

A teljes és parciális 16S rDNS szakasz hossza tekintetében tapasztalt és a szakirodalomban nem tárgyalt extremitások természetesen nem feltétlenül lepik meg a környezeti mikrobiológust, ám ismét felhívják a figyelmet az ismert, a detektálható, és a valós diverzitás között fennálló különbségekre. A PCR alapú közösségszerkezet vizsgálatok során az ennyire extrém rövid, vagy hosszú amplikonokat adó taxonok könnyebben vesznek el a módszerek szelektivitása miatt a közösségszerkezeti képből, akár a polimeráz láncreakció, akár az azt követő detektálási módszerek, vagy klónkönyvtár készítés során, hiszen ezen módszerek kifejlesztése, optimalizációja az átlagos amplikonhosszok figyelembe vételével történt. Összességében kijelenthető, hogy az *in silico* elemzéseink által kimutatott hosszheterogenitási eloszlások pontosították és kiegészítették a szakirodalom eredményeit a témában, és alátámasztották az inzerthossz függő preferenciális ligálás vizsgálatának létjogosultságát.

V.2 A modellrendszerek validálása

A DNS kvantizálási és keverési eljárásunk validálása alapján elmondhatjuk, hogy technikánk alkalmas az 1:1 kópiarányú inzertheverékek reprodukálható, nagy pontosságú előállítására. Az öt független keveréken alapuló mérésorozatot 1:0,988 átlagos kópiarányt és 1,1%-os szórást állapított meg. A 13. ábrán az egyik mérés GeneMapper v3.7 programból exportált kromatogramját láthatjuk.

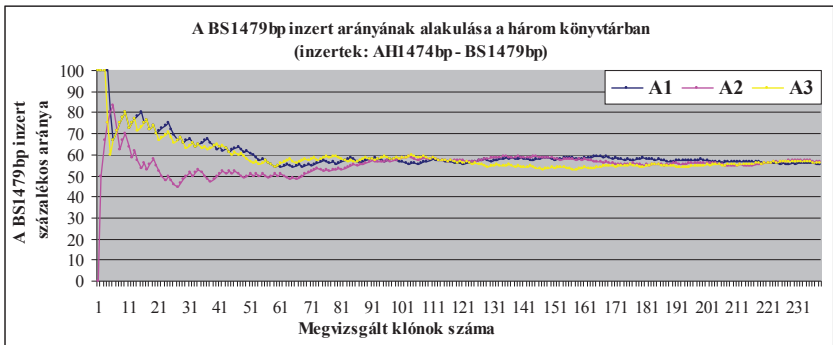


13. ábra. 1:1 kópiarányú ampikonkeverék kapilláris elektroforézissel kombinált lézergyújtású fluoreszcencia intenzitásméréssel készített kromatogramja. A függőleges tengely a fluoreszcencia intenzitást, a vízszintes tengely az elválasztott DNS-ek bázispárban mért hosszát jelzi. A zöld csúcsok a minta fluoreszcencia intenzitásai, a piros csúcsok a belső molekulásúly standard jelei. A kromatogram alatti táblázat a két ampikon által adott csúcs méretét (size, bp), magasságát (height), és görbe alatti területét (area) jelzi, – ez utóbbi alapján történt a mennyiségi összehasonlítás.

V.3 A két vektorrendszer vizsgálata modellközösségeken

V.3.1 A pGem-T vektorrendszer inzerthossz preferenciái

Az első kísérleti elrendezésben közel azonos hosszúságú és GC-tartalmú, de eltérő eredetű „teljes 16S rDNS” inzerteket tartalmazó pGem-T könyvtárakat készítettünk (2. táblázat, „A” jelű keverékek). A három párhuzamos könyvtárból 244, 240 illetve 238 klónt elemeztünk. Az inzertek arányának a vizsgált klónok mennyisége szerinti alakulását az egyes könyvtárakban a 14. ábra mutatja. Az eredetileg beállított 1:1 inzert arányhoz képest a három könyvtárban az 1479 bp hosszú inzert átlagos részaránya 56,02% lett, 0,6%-os szórással. Ez az igen kis szórás érték azt jelzi, hogy a klónozó rendszer kis mértékű szelektivitása is konzekvens, jól reprodukálható.



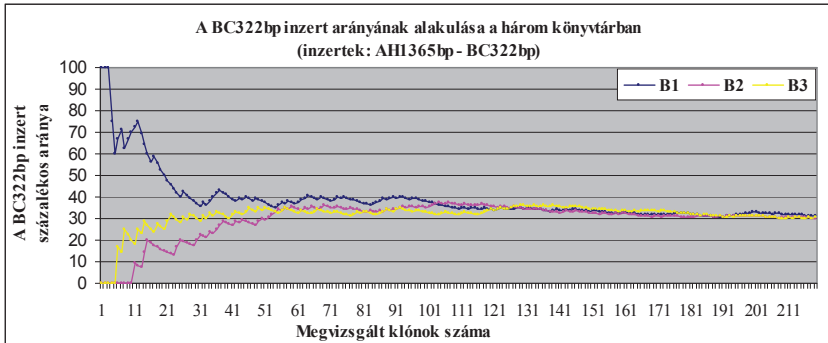
14. ábra. A BS1479bp inzert arányának alakulása a megvizsgált klónok számának függvényében az „A” jelű keverékekből készített három pGem-T könyvtárban (inzertek: AH1474bp - BS1479bp).

Tekintve hogy az inzertek hosszukban és G+C tartalmukban alig különböztek (5 bp hossz-, és 0,3% GC-arány különbség), a várt arányoktól való eltérés oka feltehetőleg az inzertek natív szerkezetének különbségében keresendő. Miután jelen esetben egy igen konzervatív génszakasszal dolgoztunk (16S rDNS), az inzertek szekvenciakülönbsége csekélyebb volt, mint a spacer régiók, vagy kevésbé konzervatív bázissorrendű funkciógének esetében. Feltételezhetjük tehát, hogy a szekvenciakülönbségből adódó aránytorzulás funkciógén vagy spacer klónkönyvtárak esetében még erőteljesebben befolyásolhatja az eredményt.

Az első modellközösségen alapuló kísérletsorozatban az eredmény szórásának megvizsgált klónmennyiség szerinti alakulása is fontos információ volt, hiszen ebből meg tudtuk határozni a későbbi klónkönyvtárak optimális méretét. A szórás tartósan 1% alá csökkenése könyvtáranként 177 klón megvizsgálása után következett be, így a további könyvtáraknál a minimálisan 180 klón analizisét tűztük ki célul a megbízható statisztikai elemzéshez. Mindemellett ennél még a következő könyvtárak esetén is jóval nagyobb klónmennyiségeket elemeztünk (lásd alább), hogy megállapításaink bizonyító ereje ne gyengüljön.

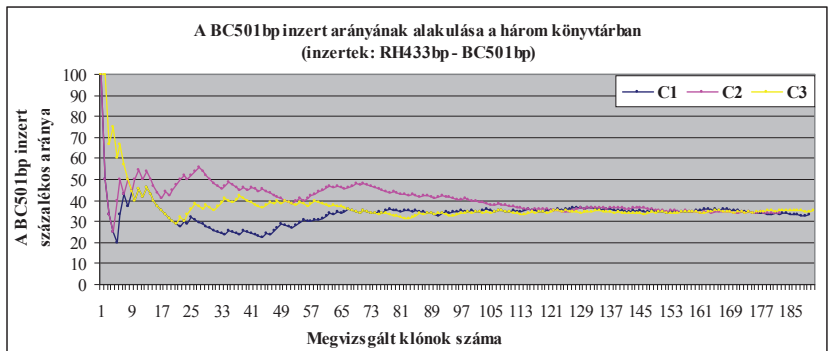
A 16S-23S intergenikus spacer klónkönyvtárakra általánosan jellemző hosszheterogenitást modellező „B”-jelű keverékekből készített klónkönyvtárakban az eredetileg beállított 1:1 inzert aránytól jóval nagyobb eltérést tapasztaltunk. 241, 224 és 220 klón megvizsgálása után a könyvtárakban a BC322bp inzert átlagos részaránya mindössze 30,79% lett, 0,64%-os szórással (15. ábra). Az eredmény felveti a torz

közösségszerkezeti képek lehetőségét nem csak a 16S-23S spacer klónkönyvtár alapú vizsgálatok (például Chen *et al*, 2006) esetében, hanem bármely nagy hosszheterogenitással bíró génszakaszokat tartalmazó pGem-T klónkönyvtárak esetén is (például kombinált parciális 16S és 16S-23S spacer könyvtárak, Brown *et al*, 2005).



15. ábra. A BC322bp inzert arányának alakulása a megvizsgált klónok számának függvényében a „B” jelű keverékekből készített három pGem-T könyvtárban.

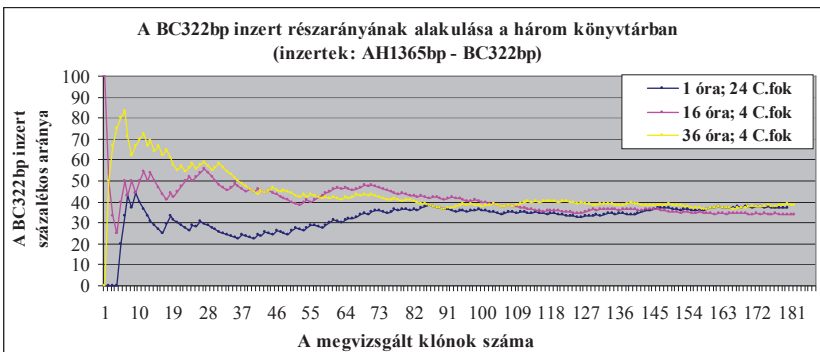
A 16S rRNS gén első harmadának (vagy hasonló méretű funkciógének) természetes hosszheterogenitását modellező „C” jelű keverékekből készített három pGem-T könyvtár szintén torzított módon adta vissza az eredetileg beállított inzert arányokat. A 191, 190 és 182 klón vizsgálata után a hosszabb (BC501bp) inzert átlagos részaránya 34,09% lett, 0,96%-os szórással (16. ábra).



16. ábra. A BC501bp inzert arányának alakulása a megvizsgált klónok számának függvényében a „C” jelű keverékekből készített három pGem-T könyvtárban.

V.3.2 A pGem-T vektorrendszer preferenciális ligációjának függése a ligálási paramétereiktől

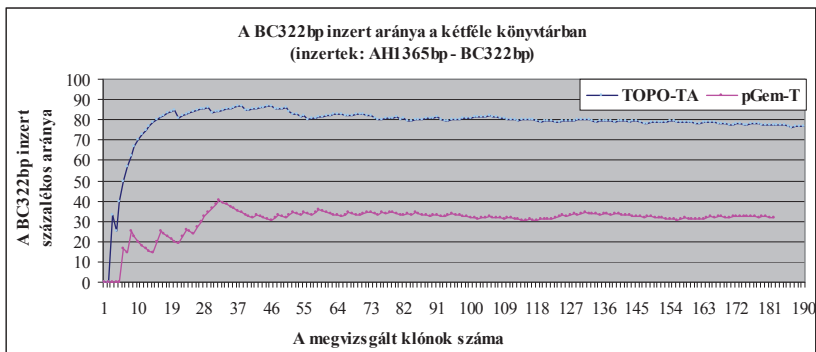
A „B3” jelű keverékből a pGem-T vektorrendszer gyártó által ajánlott kétféle ligálási protokolljával készített két klónkönyvtár eltérő inzert arányokat mutatott (18. ábra). A gyártó standard, géntechnológiai célú ligálási protokolljának (24°C, 1 óra) alkalmazásával készített 180 tagú könyvtárban a BC322bp inzert részaránya 37,22% lett. A nagyszámú transzformánsra optimalizált – s így mikrobiális közösségelemzésekre megfelelőbb – protokollal (4°C, 16 óra) készített 241 klón nagyságú könyvtár 31,06%-ban tartalmazta a BC322bp inzertet, jóval erősebb torzítást mutatva az elméletileg várható 50%-os arányhoz képest. Egy harmadik, nagyszámú transzformánsra optimalizált, de megnövelt idejű ligálással végzett alternatív protokollt is kipróbáltunk (4°C, 36 óra). Ez a 187 tagú könyvtár 37,43%-ban tartalmazta a BC322bp inzertet, közel ugyanolyan mértékű torzítást mutatva mint a standard protokoll.



18. ábra. A BC322bp inzert arányának alakulása a megvizsgált klónok számának függvényében a „B3” jelű keverékből létrehozott három, különböző ligálási paraméterekkel elkészített pGem-T könyvtárban.

V.3.3 A TOPO-TA és a pGem-T vektorrendszer preferenciális ligációjának összehasonlítása

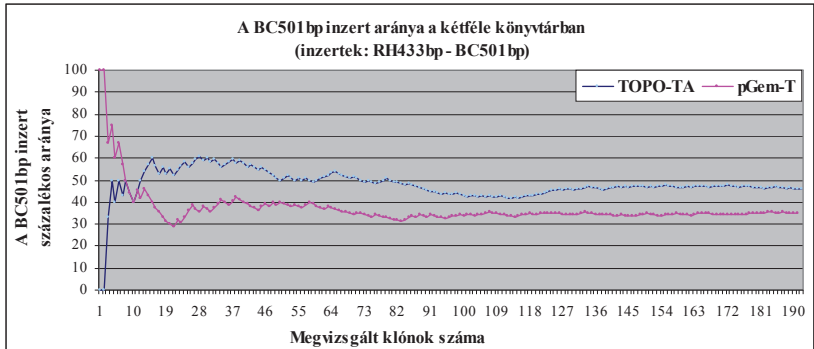
A két vektorrendszer direkt összehasonlítására megtervezett két kísérletsorozat közül az első a nagy hosszheterogenitású génszakaszok esetét reprezentáló „B3” keverékből (1365 bp és 322 bp hosszú inzertpárral) készített két, 190 (pGem-T) illetve 220 (TOPO-TA) tagú könyvtár elemzése volt. A két klónozó rendszer erős ám ellentétes előjelű torzítást mutatott: míg a pGem-T könyvtár 31,06%-ban tartalmazta a rövidebb inzertet, addig a TOPO-TA könyvtár 76,84%-ban (19. ábra).



19. ábra. A BC322bp inzert arányának alakulása a megvizsgált klónok számának függvényében a „B3” jelű keverékből létrehozott két, különböző vektorrendszerrel elkészített könyvtárban.

A tapasztalható inzertméret preferencia a TOPO-TA vektor esetében összeegyeztethető a Promega cég által végzett összehasonlító vizsgálatok következtetéseivel, miszerint a TOPO-TA vektorrendszer jóval nagyobb hatékonyságot mutat az 542 bázispáros inzertek ligálása során, mint az 1400 bázispár hosszúak esetében (HTTP-4).

Szintén eltérő mértékű torzítás volt tapasztalható a 16S rDNS első harmadának természetes hosszheterogenitását modellező „C3” keverékből készített kétféle könyvtárban (20. ábra). A 192 tagú TOPO-TA könyvtár 45,83%-ban tartalmazta az BC501bp inzertet, míg a 191 tagú pGem-T könyvtár csak 35,07%-ban.



20. ábra. Az BC501bp inzert arányának alakulása a megvizsgált klónok számának függvényében a „C3” jelű keverékből létrehozott két, különböző vektorrendszerrel elkészített könyvtárban.

A duplaszálú DNS szakaszok szekvenciájának és ligálódási hatékonyságainak összefüggéseire vonatkozó szakirodalom hiányában sajnos továbbra is csak feltételezésekre bocsátkozhatunk a jelenségek magyarázata során – bár tegyük hozzá, a torzítások biokémiai hátterének feltárása nem is szerepelt célkitűzéseink között. Mégis érdemesnek tartjuk végiggondolni a ligálás menetének mechanizmusát, és megfogalmazni az preferenciális ligálás mögött álló lehetséges okokat (vagy legalábbis egy részüket).

A két vektorrendszer eltérő mértékű (sőt akár irányú) szelektivitásának hátterében valószínűleg eltérő ligálási mechanizmusuk áll („szabadon lebegő” T4 ligáz a pGem-T, illetve vektorhoz kovalensen kötött topozimeráz a TOPO-TA rendszer esetében). Mindkét vektorrendszer esetében az inzert egyik végének túlnyúló adeninje először kétszeres hidrogénhid kötést alakít ki a vektor egyik végének túlnyúló timinjével. Viszont míg a TOPO-TA esetében a vektorhoz kötött topozimeráz-I azonnal kialakítja a kovalens cukor-foszfát kötést, a pGem-T vektor esetében még egy T4-ligáznak is meg kell találnia a helyszínt. Azonban a DNS szálak hőmozgása miatt a gyenge hidrogénhidas kölcsönhatás időről-időre megszakad (Altan-Bonnet *et al.*, 2003), így az inzert és a vektor eltávolodhat egymástól. A különböző hosszúságú inzert DNS-ek viselkedése ebből a szempontból eltérő: Sorlie és Pecora kimutatták, hogy 367 bázispár hosszú kettőszálú DNS-ek sokkal nagyobb diffúziós képességgel bírnak, mint a 2311 bázispár hosszúak (Sorlie és Pecora, 1990). A rövidebb inzertek jobb diffúziós képességük miatt nagyobb hajlamot mutathatnak a végleges disszociációra és

eltávolódásra a pGem-T vektortól. Más szempontból, gyorsabb diffúziójuk miatt (adott idő alatt hosszabb utat bejárva) nagyobb eséllyel találkoznak a vektorokkal a ligáló oldatban, s míg a TOPO-TA esetében azonnal kovalens kötésbe kerülnek, a pGem-T vektorról ledisszociálhatnak. Mindez magyarázatot adhat a 322 és 1365 bázispár hosszú inzerteket tartalmazó keverék esetén a pGem-T erősebb preferenciájára a hosszabb inzert felé, és a TOPO-TA preferenciájára a rövidebb inzert irányában.

Az pGem-T vektorrendszer preferenciális ligálásának időtől való függésével kapcsolatban a ligáló reakcióban szereplő molekulatípusok „populációdinamikáját” tekintve feltételezhető, hogy a hosszabb ligálási idő alatt a kisebb hatékonysággal ligálódó inzertek egy idő után relatív túlsúlyba kerülnek, mivel a preferált inzertek mennyisége lecsökken a még szabad inzert populációban. Ekkor a kisebb hatékonysággal ligálódó inzertek bekötésének megnő a relatív gyakorisága. Így a végső ligálási arány az idő növelésével bár továbbra is torz marad, de közelebb kerül az eredetileg beállított 1:1 arányhoz.

Természetesen mindez csak elmélet, melynek közvetlen kísérletes bizonyítására jelenlegi eszközeinkkel és kapacitásainkkal sajnos nem vagyunk képesek. Ezen kívül egyes kísérleti elrendezésekben tapasztalható torzítások nem magyarázhatók meg a fenti módon, tehát további tényezők, mint például az inzertek bázissorrendjéből adódó eltérő viselkedés is szerepet játszhatnak.

V.4 A két vektorrendszer vizsgálata környezeti mintán

V.4.1 A pGem-T és TOPO-TA klónkönyvtárak jellemzése

A könyvtárak feldolgozása, és az OTU-k bázissorrend elemzése után a kémikus szekvenciákat nem számítva a pGem-T klónkönyvtár 280 klónt tartalmazott, melyek 132 OTU-ba csoportosultak, míg a TOPO-TA könyvtár 269 klónt tartalmazott 133 OTU-ban. A kimerágyanus szekvenciák száma 3 illetve 2 volt a pGem-T illetve a TOPO-TA könyvtárban.

A két könyvtár szekvencia tartalmának átlagos GC aránya szinte teljesen megegyezett, a pGem-T esetében 56,32%, a TOPO-TA könyvtárban 56,26% volt. Ez az eredmény igazolta azon hipotézisünket, miszerint az inzertek preferenciális ligálása nem, vagy csak kis mértékben függ azok GC-tartalmától. Az inzertek LH-PCR-rel mért

hossza 430 illetve 549 bázispár között mozgott, domináns csoportokat alkotva még a 425 bp alatti, illetve az 515 bp feletti tartományokban is. Ezen nagy hosszheterogenitás ellenére a könyvtárakon belüli átlagos inzerthossz nem mutatott jelentős eltérést. A pGem-T könyvtár átlagos inzerthossza 478,39 bp volt, míg a TOPO-TA átlagosan 480,85 bp-t mutatott. Ez az eltérés bár csekély, de alátámasztja a modellrendszeres kísérletekben nyert eredményeket, mely szerint ebben a hosszheterogenitás tartományban a pGem-T vektorrendszer a rövidebb inzerteket erősebben preferálja mint a TOPO-TA.

A könyvtárak OTU abundancia alapú hasonlóság becslései az elméletileg várható értékhez (1) képest kis számokat mutattak. A Chao-Jaccard index 0,4423 míg Chao-Sorenson index 0,6134 értéket vett fel. Ez egyrészt lehet a klónozó rendszerek szelektivitásának következménye, másrészt a közösség alulmintázottságának jele is.

A könyvtárak OTU eloszlásai alapján Chao1-bc eljárással becsült eredeti teljes fajszám 406 volt a pGem-T könyvtár és 623 a TOPO-TA könyvtár alapján, ami reális érték egy ilyen közösség tekintetében (Torsvik *et al.*, 1996). A két becsült érték közötti jelentős eltérés oka abban keresendő, hogy az alkalmazott algoritmus alapvetően a könyvtárban egy illetve két példányban megtalált (kis részarányú) OTU-k arányait veszi figyelembe a fajszámbecslésnél (Chao *et al.*, 2006). A diverz mikrobiális közösségek összetételében általánosan tapasztalható igen nagy számú de kis részarányú filotípusok reprezentatív mintázásához viszont ezen közösség esetében is többtízezres nagyságrendű klónmennyiséggel operáló könyvtárelemzés lenne szükséges. Célunk azonban a (kísérletek megtervezése idején) általánosan használt méretű klónkönyvtárak esetében tapasztalható torzítások elemzése volt, vállalva azt a korlátot, hogy a 269-280 darabos klónmennyiség a kis részarányú filotípusokon belüli mennyiségi viszonyokat relatíve nagy pontatlansággal reprezentálja.

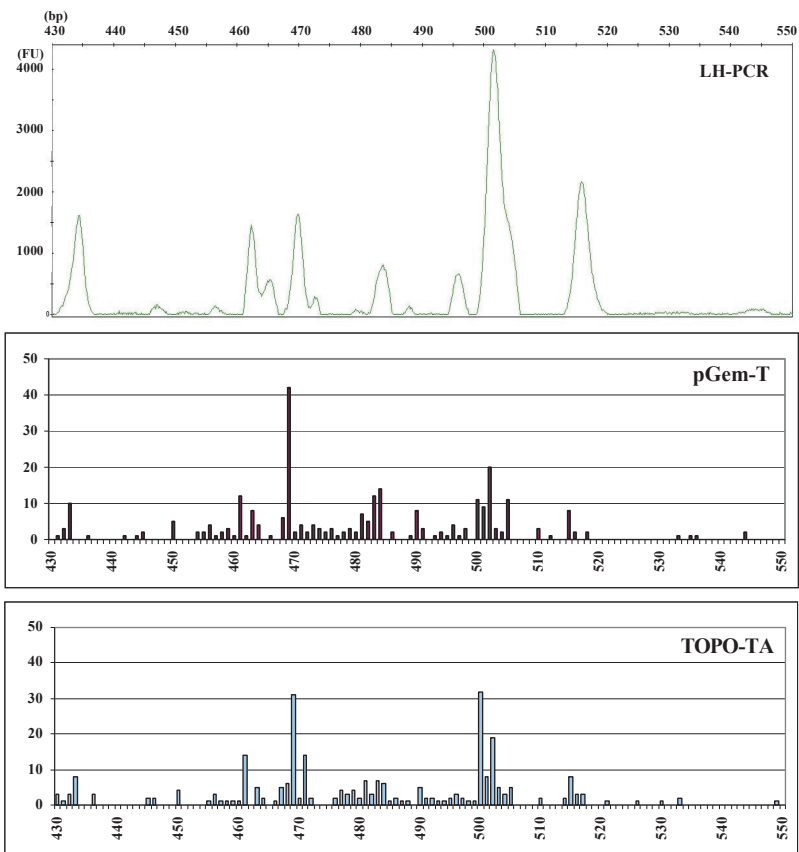
V.4.2 A pGem-T és TOPO-TA vektorrendszerekkel nyert közösségszerkezeti képek összehasonlítása az LH-PCR eredményével

A klónkönyvtárak alapján felállított közösségszerkezeti képeket a klónozástól független, LH-PCR-rel nyert képpel összehasonlítva szembevető különbségek észlelhetők (21. ábra, a következő oldalon). A szubjektív értékelésen túllépve, a közösségszerkezeteken belüli részarányok pontosabb összevethetősége érdekében létrehozott „csúcs csoportok” (lásd IV.3.7.2 fejezet) mennyiségi viszonyait a 22. ábra mutatja (a következő utáni oldalon).

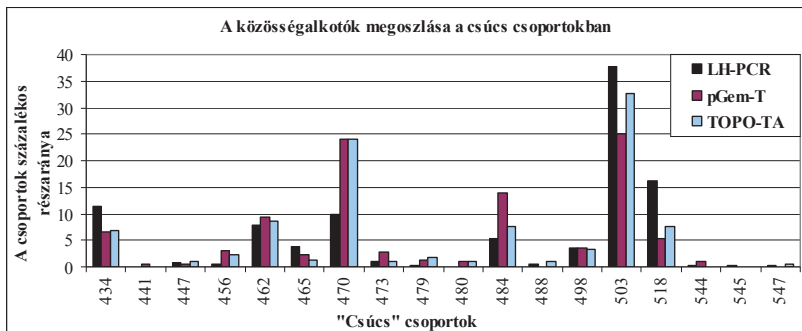
A csúcs csoporton belüli kisszámú klónból, illetve alacsony fluoreszcencia intenzitásból fakadóan rossz jel-zaj arányú, kis megbízhatóságú adatok mellőzése végett a továbbiakban csak a mindhárom közösségszerkezeti képben legalább 5%-os részarányal szereplő csúcs csoportokban tapasztalható torzításokat tárgyaljuk. Az LH-PCR-rel mérhető egyes csúcs csoport részarányokat mint referenciaértéket nézve kiszámolhatjuk a csúcs csoportoknak a klónkönyvtárakban tapasztalható alul- vagy felülreprezentáltságát (4. táblázat).

Csúcs csoport	434	462	470	484	503	518
Részarány az LH-PCR képhez képest a pGem-T könyvtárban	58,3%	120,3%	244,4%	254,7%	66,1%	33,0%
Részarány az LH-PCR képhez képest a TOPO-TA könyvtárban	59,1%	110,3%	243,1%	141,6%	86,1%	47,3%

4. táblázat. A mindhárom közösségszerkezeti képben legalább 5%-os részarányal szereplő csúcs csoportok relatív alul- illetve felülreprezentáltsága a klónkönyvtárakban az LH-PCRhez képest (az adott csúcs csoport LH-PCR-en belüli részarányát 100%-nak véve).



21. ábra. Az LH-PCR-rel nyert közösségszerkezeti kép, illetve a klónkönyvtárakból rekonstruált LH-PCR közösségszerkezeti képek összehasonlítása. A vízszintes tengely minden esetben a bázispárban mért amplikon hosszúságát, a függőleges tengely az LH-PCR elektroferogram esetében az abszolút fluoreszcencia intenzitást, a klónkönyvtárak esetében pedig az adott hosszúságú inzertek könyvtáron belüli, darabszámban mért részarányát mutatja.



22. ábra. Az LH-PCR kromatogram domináns csúcsainak és a hozzájuk rendelhető klónokból alkotott csúcs csoportok relatív mennyiségei. A függőleges tengely az LH-PCR elektroferogram esetében a kiválasztott csúcsok össz-fluoreszcencia intenzitásának százalékában, a klónkönyvtárak esetében pedig az adott csoportba tartozó klónok az összes csúcs csoportba sorolt klónmennyiségén belüli részarányának százalékában mutatja a mennyiségeket.

Mindkét könyvtárra jellemző volt, hogy az átlagos méretű inzerteket preferálták (mint a 470 és 484 bázispáros csúcs csoportok esetében), míg a szélsőségesen rövid (pl. 434 bp körüli) illetve hosszú (pl. 503 és 518 bp körüli) inzerteket pedig alulreprezentáltak az LH-PCR közösség szerkezeti arányaihoz képest. Bizonyos csúcs csoportok esetében az egyik vektorrendszer jóval erősebb torzítást mutatott. A pGem-T a 484 bp körüli inzerteket, míg a TOPO-TA könyvtár az 503-as csúcs csoportot preferálta kiemelkedően.

A 434-es csoport kivételével mindegyik kiemelt csúcs csoport esetében a tapasztalható torzítások mértékei összhangban vannak a modellközösségi vizsgálatok során megtapasztalt tendenciával a két vektorrendszer inzerthossz preferenciái tekintetében. A klónkönyvtárak közötti, illetve az LH-PCR közösség szerkezetéhez viszonyított torzítások sok esetben megközelítették a 2 – 2,5-szeres mértéket, megbízhatóan alátámasztva a preferenciális ligálás jelenlétét.

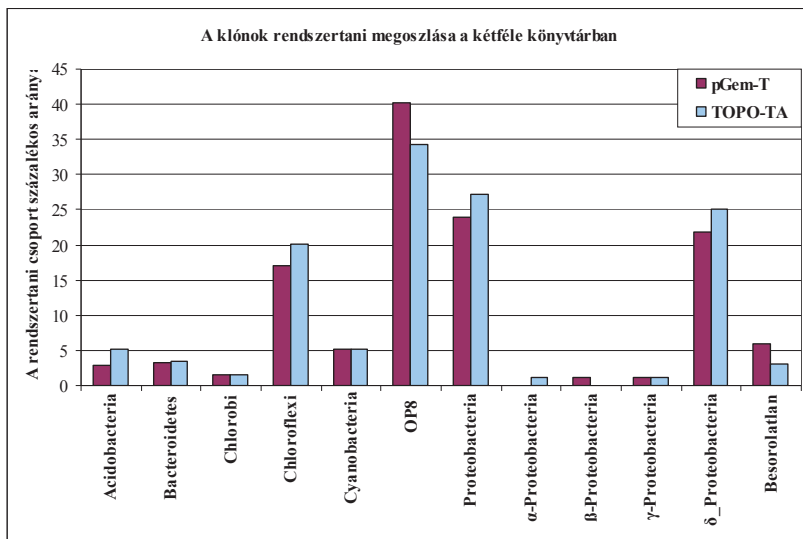
A TOPO-TA vektorral nyert közösség szerkezeti kép minden fő csúcs csoport esetében közelebb állt az LH-PCR-rel mért részarányokhoz. Ez az eredmény konzisztens a modellkísérletekből levont következtetésekkel, miszerint a TOPO-TA vektorrendszer a pGem-T-nél enyhébb torzítást mutat a szóban forgó inzerthossz tartományban.

Tekintve, hogy a 16S rDNS első harmadában található V1-V2-V3 variábilis régiók hossza szoros összefüggést mutat a filogenetikai rokonság mértékével (Suzuki *et al.*, 1998), a tapasztaltak alapján feltételezhetnénk, hogy a preferenciális ligálás mértéke taxonspecifikus érték. Viszont mivel az egyes csúcs csoportokat több, akár nagyon különböző rendszertani csoport is alkothatja, az adott csúcs csoportot alkotó különböző taxonokra ható eltérő mértékű – vagy akár irányú – preferenciális ligálás miatt az ílymódon detektálható torzítások filogenetikai összefüggéseit pusztán LH-PCR vizsgálatokkal nehézkes felderíteni.

V.4.3 A pGem-T és TOPO-TA klónkönyvtárak filogenetikai összetételének összevetése

A klónszekvenciák filogenetikai besorolásának ismeretében megvizsgáltuk a taxonómiai alapú megoszlásokat a klónkönyvtárakban. Mivel a könyvtáraink igen sok osztály vagy nemzetség szinten pontosan nem besorolható szekvenciát tartalmaztak, az elemzés során elsősorban phylum-szintű (illetve a Proteobacterián belül osztály szintű) torzításokat vizsgáltunk. A két könyvtár klónjainak rendszertani megoszlásait a 23. ábra mutatja (a következő oldalon).

Enyhe eltérést több rendszertani csoport esetében is tapasztaltunk, legerősebben a pGem-T vektorrendszer által preferált OP8 leszármazási ágba tartozó klónokkal kapcsolatban. Megvizsgálva az egyes rendszertani csoportokba tartozó klónok inzerthosszait, a modellrendszerekkel szerzett tapasztalatokkal egybevető eredményeket kapunk: a pGem-T által preferált, OP8 phylumba tartozó klónok inzerthosszai a 460-480 bázispáros hossztartományba esnek, míg a TOPO-TA vektorrendszer által preferált Chloroflexi és Proteobacteria klónok inzerthosszai az 500-as vagy előtti hossztartományba. Ez alapján valószínűsíthető, hogy a detektált filogenetikai preferencia is legalább részben az inzerthosszfüggő preferenciális ligálás számlájára írható.

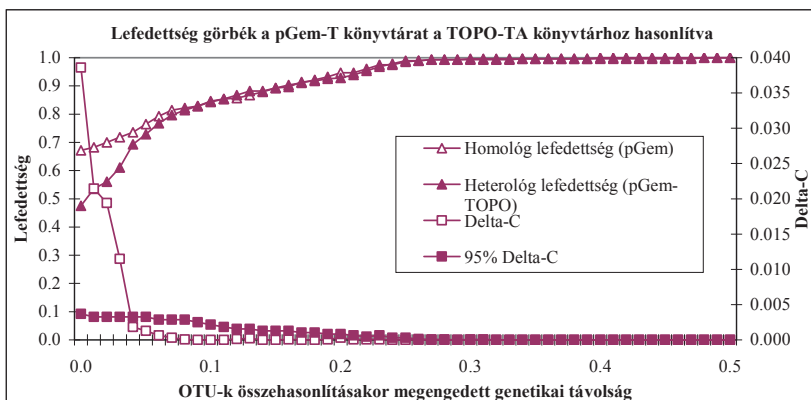
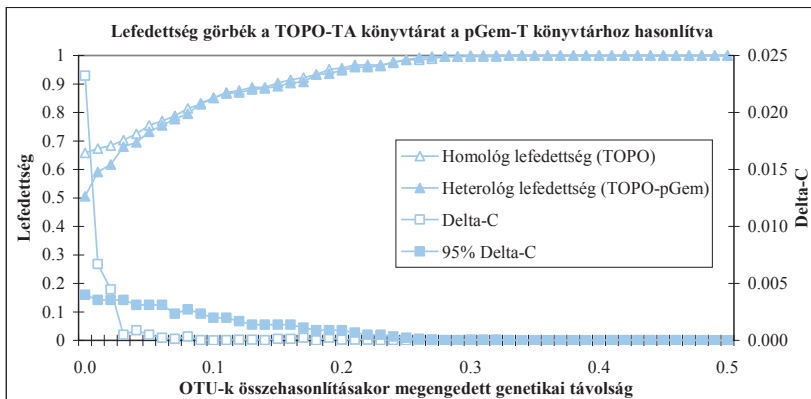


23. ábra. A természetes baktériumközösségből készített két klónkönyvtár klónjainak rendszertani megoszlása a könyvtárankénti teljes klónmennyiségek százalékában.

V.4.4. A pGem-T és TOPO-TA klónkönyvtárak filotípus tartalmának összehasonlító elemzése (Libshuff elemzés)

A közösségi 16S rDNS klónkönyvtárak összetételbeli különbségének szignifikanciaszint mérésére kidolgozott Libshuff elemzés megerősítette azon feltételezéseinket hogy a két könyvtár összetétele szignifikánsan eltér. A könyvtárak összehasonlításai során nyert lefedettség és különbség görbéket a 24. ábra mutatja.

Már a lefedettség görbék eltérő lefutásából is arra lehetett következtetni, hogy a könyvtárak kimutathatóan különböznek, aminek szignifikanciáját a Monte-Carlo teszt is megerősítette. Az ezerszeres ismétlésű teszt eredményeképp a két könyvtár összehasonlításai során nyert p érték 0,002 lett. Ez a kis érték arra enged következtetni hogy a két könyvtár közösségösszetétele 95%-os megbízhatóság mellett szignifikánsan különböző, (Singleton *et al.*, 2001).



24. ábra. A pGem-T és a TOPO-TA vektorrendszerrel készített könyvtárak Libshuff elemzése során nyert lefedettség és lefedettség különbség görbék.

Az alkalmazott módszert kifejezetten alulmintázott közösségek összehasonlítására találták ki, eredeti publikációja során 150 klónnál kisebb méretű könyvtárakkal mutatták ki megbízható működését. Ez alapján feltételezhetjük, hogy a kísérlet során létrehozott 269 illetve 280 tagú könyvtárak elegendő méretűek voltak a torzítások megbízható detektálásához.

V.5. Ajánlások a pontosabb közösség szerkezet leírásokhoz

A modellkísérletekkel, valamint a valós közösség vizsgálata során szerzett tapasztalatok alapján kijelenthető, hogy 4-500 bp hosszheterogenitású génszakaszok vizsgálatára enyhébb torzításából kifolyólag a TOPO-TA vektorrendszer alkalmasabb. A 16S rDNS első harmada mellett ilyen hossztartományban van például a *nifH* génszakasz vagy metanotróf közösségek esetében általában vizsgált *pmoA* génszakasz, mely a gyakran alkalmazott A189f és A682r primerpárok közötti szakaszon az elméleti 531 bp-os hossz mellett körülbelül 100 bázispáros varianciával bír (Pester *et al*,2004).

A nagymértékű, 300-1300 bázispáros hosszheterogenitást mutató régiók (pl. 16S-23S spacer) klónkönyvtárak elemzésére a pGem-T vektorrendszer jóval hitelesebb eszköznek minősült. A vektorhoz ajánlott ligálási protokollok közül pedig a standard (22°C, 1 óra) protokoll jóval kisebb torzítást mutatott, mint az egyébként közösségszerkezeti vizsgálatokra optimálisabbnak tűnő, nagyszámú transzformánst adó alternatív protokoll (4°C, 18 óra). Amennyiben a felhasználó mégis garantáltan nagy mennyiségű klónt kíván létrehozni, ajánlhatjuk az általunk „kigondolt” és ellenőrzött 4°C-os, 36 órás ligálást, ami még a standard protokollnál is enyhén kisebb torzítást mutatott. Meg kell jegyezzük azonban, hogy tapasztalataink szerint a klónkönyvtárak maximálisan elérhető méretét és a klónozás sikerességét sokkal nagyobb mértékben határozza meg a PCR termékek minősége (például a terminális adenin-addíció hatékonysága) valamint a kompetens sejtek állapota (inzertfelvétel képességének mértéke). Ezen kívül a könyvtárak mérete a kiszélesztett transzformánsok mennyiségével növelhető, sőt extrém nagymennyiségű klón igénye esetén egyetlen ligáló reakcióból nem csak a standard protokoll által meghatározott mennyiséget felhasználva transzformáláshoz, hanem a teljes mennyiséggel dolgozva (természetesen megfelelően felszorozott kompetenssejt mennyiséget alkalmazva) könnyen készíthető nagyságrendekkel nagyobb (akár tízezres méretű) klónkönyvtár. Erre mindkét vektorrendszer gyári kizserelése és protokolljai is lehetőséget nyújtanak.

VI. Összegzés, kitekintés

Eredményeink fényében kijelenthetjük, hogy a TA-klónozás nem mentes a szelektivitástól, s a különböző TA-klónozó rendszerek eltérő torzítást mutatnak. A preferenciális ligálás reprodukálható jelenség, és mértéke függ a ligálási körülményektől. Ez a torzítás úgy tűnik nincs összefüggésben az inzertek GC-tartalmával, okát az inzertek hosszának és bázissorrendjének különbségeiben kell keresnünk. A kétféle TA-klónozó vektor torzításának különbsége hasonló jellegű volt mint amire a pGem-T vektorrendszer gyártója által homogén inzerteken elvégzett összehasonlító vizsgálatból következtetni lehet.

Tapasztalataink összhangban vannak Rainey és munkatársainak feltételezésével, miszerint a különböző klónozó rendszerek eltérő közösségszerkezeti képet mutathatnak ugyanazon minta vizsgálata során (Rainey *et al.*, 1994). Az említett publikáció nem tartalmaz kísérletes bizonyítékot ezen állítás alátámasztására, így jelen munka a hipotézis sikeres bizonyításaként is szolgál.

Eredményeink összegezethetők Taylor és munkatársainak tapasztalataival (Taylor *et al.*, 2007), akik a TOPO-TA és a tompa végű klónozás összehasonlításánál azt találták, hogy bizonyos rendszertani csoportokat a módszerek egymáshoz képest alul-, illetve felülreprezentáltak. Igaz, végső következtetésük mégis az volt, hogy a két közösségszerkezeti kép nem tér el jelentősen. Az eltérő konklúziók oka a vizsgált genomszakaszok, közösségtípusok, illetve klónozó rendszerek különbségében keresendő. Az általuk elemzett gomba ITS-LSU szakasz hossza 1100 és 2000 bázispár között mozog, ami jelentősen eltér a modellrendszereinkben illetve közösségi vizsgálataink során célzott hossztartományoktól. Ezen kívül, az általuk alkalmazott klónozás előtti közösségi PCR termék 92 talajminta PCR termékeinek keveréke volt, így egy különösen nagy diverzitású (mesterséges) közösségnek fogható fel. Mivel a kis diverzitású közösségek TA-klónkönyvtárai elméletileg sokkal erősebb torzításokat mutathatnak, az ilyen nagy sokféleségű közösségek vizsgálatából levont következtetések alulbecsülhetik a módszer hibájának mértékét. Funkciógének vizsgálatakor például eleve egy szűkebb guild-et célzunk a közösségen belül, ami jóval kisebb diverzitást jelenthet az inzertek tekintetében. A Taylor és munkatársai által alkalmazott pCR4.0-TOPO TA-klónozó vektor is különbözött a jelen kutatás során – és

baktériumközösségek vizsgálatával foglalkozó publikációkban általánosan – alkalmazott pCR2.1-TOPO vektortól.

Összehasonlítva a TA-klónozás torzításainak mértékét a megelőző lépés (a multitemplát PCR) során jelentkező preferenciális amplifikáció lehetséges nagyságával, azt a „megnyugtató” következtetést vonhatjuk le, hogy a preferenciális ligálás úgy tűnik jóval kisebb mértékben torzítja tovább a közösségszerkezeti képet. Természetesen a kétféle torzítást nehéz kvantitatív módon összevetni, így csak a mindkét jelenséggel behatóan foglalkozó kutató szubjektív, összegző véleményét közölhetjük.

Végző konklúzióként e helyütt is ki kell emelnünk a mikrobiális közösségszerkezet leírások során a polifázikus megközelítés fontosságát. A klónkönyvtárakkal párhuzamosan célszerű más vizsgálóeszközöket is alkalmazni, s erre a mikrobiális ökológiában üdvözlendő módon terjedő szemikvantitatív ujjlenyomat módszerek (T-RFLP, LH-PCR) kitűnő lehetőséget adnak. Természetesen minden módszernek megvannak a maga hátrányai és torzításai, de az eltérő szelektivitású eljárások által szolgáltatott közösségszerkezeti képek együttes elemzése közelebb juttatja a kutatót a valóság megismeréséhez – különösen akkor, ha a módszerek potenciális hibaforrásait előzetesen kiismerte, s ezt az eredmények összevetésénél, interpretációjánál figyelembe tudja venni.

VII. Idézet irodalom

Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., Polz, M. F. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* **71**:8966–8969.

Ahlgren, A. N., Rocap, G. 2006. Culture isolation and culture-independent clone libraries reveal new marine *Synechococcus* ecotypes with distinctive light and N physiologies. *Appl. Environ. Microbiol.* **72**:7193–720.

Altan-Bonnet, G., Libchaber, A., Krichevsky O. 2003. Bubble dynamics in double-stranded DNA. *Phys. Rev. Lett.* **90**:138101–138104.

Amann, R. I., Krumholz, L., Stahl, D. A. 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J. Bacteriol.* **172**:762–770.

Anderson, I. C., Campbell, C.D., Prosser, J. I. 2003. Potential bias of fungal 18S rDNA and internal transcribed spacer polymerase chain reaction primers for estimating fungal biodiversity in soil. *Environ. Microbiol.* **5**:36–47.

Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., Weightman, A. J. 2006. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl. Environ. Microbiol.* **72**:5734–5741.

Baker, G. C., Smith, J. J., Cowan, D. A. 2003. Review and re-analysis of domain-specific 16S primers. *J. Microbiol. Meth.* **55**:541–555.

Barry, T., G. Colleran, M. Glennon, L. R. Dunican, and F. Gannon. 1991. The 16S/23S ribosomal spacer region as a target for DNA probes to identify eubacteria. *PCR Methods Appl.* **1**:51–56.

Béjà O., 2004. To BAC or not to BAC: marine ecogenomics. *Curr. Opin. Biotechnol.* **15**:187–190.

- Bourne, D. G., McDonald, I. R., Murrell, J. C.** 2001. Comparison of *pmoA* PCR Primer Sets as Tools for Investigating Methanotroph Diversity in Three Danish Soils. *Appl. Environ. Microbiol.* **67**:3802–3809.
- Brown, V. M., Schwalbach, M. S., Hewson, I., Fuhrman, J. A.** 2005. Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environ. Microbiol.* **7**:1466–1479.
- Brownstein, M. J., Carpten, J. D., Smith, J. R.** 1996. Modulation of non-templated nucleotide addition by *taq* DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **20**:1004–1006, 1008–1010.
- Cardenas, E., Tiedje, J. M.** 2008. New tools for discovering and characterizing microbial diversity. *Curr. Opin. Biotechnol.* **19**:544–549.
- Chao, A., Chazdon, R. L., Colwell, R. K., Shen, T. J.** 2006. Abundance-based similarity indices and their estimation when there are unseen species in samples. *Biometrics* **62**:361–371.
- Chao, A., Chazdon, R. L., Colwell, R. K., Shen, T. J.** 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol. Lett.* **8**:148–159.
- Chen, F., Wang, K., Kan, J., Suzuki, M. T., Wommack K. E.** 2006. Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA Internal Transcribed Spacer sequences. *Appl. Environ. Microbiol.* **72**:2239–2243.
- Clark, J. M.** 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucl. Acids Res.* **16**:9677–9686.
- Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Bandela, A. M., Cardenas, E., Garrity, G. M., Tiedje, J. M.** 2007. The ribosomal database project (RDP-II): introducing *myRDP* space and quality controlled public data. *Nucl. Acids Res.* **35**:169–172.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M., Tiedje, J. M. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucl. Acids Res.* **37**:141–145.

Ellis, J., Morgan P., Weightman, A., Fry, J. 2003. Cultivation-dependent and – independent approaches for determining bacterial diversity in heavy-metal-contaminated soil. *Appl. Environ. Microbiol.* **69**:3223–3230.

Farrelly, V., Rainey, F. A., Stackebrandt, E. 1995. Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl. Environ. Microbiol.* **61**:2798–2801.

Felföldi, T., Duleba, M., Somogyi, B., Vajna, B., Nikolausz, M., Présing, M., Márialigeti, K., Vörös, L. 2011. Diversity and seasonal dynamics of the photoautotrophic picoplankton in Lake Balaton (Hungary). *Aquat. Microbial Ecol.* **63**:273-287.

Felsenstein, J. 1989. PHYLIP – Phylogeny Inference Package (Version 3.2). *Cladistics* **5**:164–166.

Fisher, M., Triplett, E. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Appl. Environ. Microbiol.* **65**:4630–4636.

Frostegard, A., Courtois, S., Ramišse, V., Clerc, S., Bernillon, D., Le Gall, F., Jeanin, P., Nesme, X., Simonet, P. 1999. Quantification of bias related to the extraction of DNA directly from soils. *Appl. Environ. Microbiol.* **65**:5409–5420.

Gantner, S., Andersson, A.F., Alonso-Sáez, L., Bertilsson, S. 2010. Novel primers for 16S rRNA-based archaeal community analyses in environmental samples. *J. Microbiol. Meth.* **84**:12–18.

García-Martínez, J., Acinas, S. G., Antón, A. I., Rodríguez-Valera, F. 1999. Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *J. Microbiol. Meth.* **36**:55–64.

- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., Field, K. G.** 1990. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**:60–63.
- González, N., Romero, J., Espejo, R. T.** 2003. Comprehensive detection of bacterial populations by PCR amplification of the 16S-23S rRNA spacer region. *J Microbiol. Meth.* **55**:91-97.
- Hansen, M. C., Tolker-Nielsen, T., Givskov, M., Molin, S.** 1998. Biased 16S rDNA PCR amplification caused by interference from DNA flanking the template region. *FEMS Microbiol. Ecol.* **26**:141–149.
- Haverkamp, T, Acinas, S. G., Doeleman, M., Stomp, M., Huisman, J., Stal, L. J.** 2008. Diversity and phylogeny of Baltic Sea picocyanobacteria inferred from their ITS and phycobiliprotein operons. *Environ. Microbiol.* **10**:174–188.
- Holben, W. E., Jansson, J. K., Chelm, B. K., Tiedje, J. M.** 1988. DNA Probe Method for the Detection of Specific Microorganisms in the Soil Bacterial Community. *Appl. Environ. Microbiol.* **54**:703–711.
- Holton, T. A., Graham, M. W.** 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucl. Acids Res.* **19**:1156.
- Hongoh, Y., Yuzawa, H., Ohkuma, M., Kudo, T.** 2003. Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. *FEMS Microbiol. Lett.* **221**:299–304.
- Hugenholtz, P., Goebel, B. M., Pace, N. R.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**:4765–4774.
- Hugenholtz, P., Tyson, G. W.** 2008. Metagenomics. *Nature* **455**:481–483.
- Hughes, J. B., Hellmann, J. J., Ricketts, T. H., Bohannan, B. J.** 2001. Counting the uncountable: Statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67**:4399–4406.

- Hurt, R., Qiu, X., Wu, L., Roh, Y., Palumbo, A. V., Tiedje, J. M., Zhou, J.** 2001. Simultaneous recovery of RNA and DNA from soils and sediments. *Appl. Environ. Microbiol.* **67**:4495–4503.
- Huse, S. M., Huber, J. A., Morrison, H. G., Sogin, M. L., Welch, D. M.** 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**:R143.
- Hutter, G., Schlagenhauf, U., Valenza, G., Horn, M., Burgemeister, S., Claus, H., Vogel, U.** 2003. Molecular analysis of bacteria in periodontitis: evaluation of clone libraries, novel phylotypes and putative pathogens. *Microbiology* **149**:67–75.
- Ishii, K., Fukui, M.** 2001. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Appl. Environ. Microbiol.* **67**:3753–3755.
- Jones, R. T., Robeson, M. S., Lauber, C. L., Hamady, M., Knight, R., Fierer, N.** 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. *ISME J.* **3**:442–453.
- Joo, S., Lee, S. R., Park, S.** 2010. Monitoring of phytoplankton community structure using terminal restriction fragment length polymorphism (T-RFLP). *J. Microbiol. Meth.* **1**:61–68.
- Klappenbach, J. A., Dunbar, J. M., Schmidt, T. M.** 2000. rRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* **66**:1328–1333.
- Kim, S. G., Rhee, S. K., Ahn, C. Y., Ko, S. R., Choi, G. G., Bae, J. W.** 2006. Determination of cyanobacterial diversity during algal blooms in Daechung Reservoir, Korea, on the basis of *cpcBA* intergenic spacer region analysis. *Appl. Environ. Microbiol.* **72**:3252–3258.
- Kovács, G. M., Balázs, T., Péntzes, Z.** 2007. Molecular study of the arbuscular mycorrhizal fungi colonizing the sporophyte of the eusporangiate rattlesnake fern (*Botrychium virginianum*, *Ophioglossaceae*). *Mycorrhiza* **17**:597–605.
- Krett, G., Palatinszky, M.** 2009. A polyphasic study on the species diversity of the sediment microbiota of Lake Hévíz. *Acta Microbiol. et Immun. Hung.* **56**:339–355.

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., Sindelka, R., Sjöback, R., Sjögreen, B., Strömbom, L., Ståhlberg, A., Zoric, N. 2006. The real-time polymerase chain reaction. *Molec. Asp. Med.* **27**:95–125.

Lane, D. J. 1991. 16S/23S rRNA sequencing, *In: Nucleic Acid Technology in Bacterial Systematics.* (Eds: Stackebrandt, E., and Goodfellow, M.). John Wiley and Sons, New York, pp. 115–175.

Liesack, W., Stackebrandt, E. 1992. Occurrence of novel groups of the domain Bacteria as revealed by analysis of genetic material isolated from an Australian terrestrial environment. *J. Bacteriol.* **174**:5072–5078.

Lovell, C. R., Piceno, Y. M., Quattro, J. M. and Bagwell, C. E. 2000. Molecular analysis of diazotroph diversity in the rhizosphere of the smooth cordgrass, *Spartina alterniflora*. *Appl. Environ. Microbiol.* **66**:3814–3822.

Lovell, C. R., Decker, P. V., Bagwell, C. E., Thompson, S., Matsui, G. Y. 2008. Analysis of a diverse assemblage of diazotrophic bacteria from *Spartina alterniflora* using DGGE and clone library screening. *J. Microbiol. Meth.* **73**:160–171.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H. 2004. ARB: a software environment for sequence data. *Nucl. Acids Res.* **32**:1363–1371.

Lueders, T., Friedrich, M. W. 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and *mcrA* genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. *Appl. Environ. Microbiol.* **69**:320–326.

Magnuson, V. L., Ally, D. S., Nylund, S. J., Karanjawala, Z. E., Rayman, J. B., Knapp, J. I. 1996. Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase: implications for PCR-based genotyping and cloning. *Biotechniques* **21**: 700–709.

Marchesi, J. R., Sato, T., Weightman, A. J., Martin, T. A., Fry, J. C., Hiom, S. J., Dymock, D., Wade, W. G. 1998. Design and evaluation of useful bacterium-specific

PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl. Environ. Microbiol.* **64**:795–799.

Marko, J. F., Cocco, S. 2003. The micromechanics of DNA. *Physics World* **16**:37–41.

Mathieu-Daude, F., Welsh, J., Vogt, T., McClelland, M. 1996. DNA rehybridization during PCR: the 'Cot effect' and its consequences. *Nucl. Acids Res.* **24**:2080–2086.

McDonald I. R., Murrell, J. C. 1997. The particulate methane monooxygenase gene *pmoA* and its use as a functional gene probe for methanotrophs. *FEMS Microbiol. Lett.* **156**:205–210.

Mering, C., Hugenholtz, P., Raes, J., Tringe, S. G., Doerks, T., Jensen, L. J., Ward, N., Bork, P. 2007. Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science* **315**:1126–1130.

Metzker, M. L. 2009. Sequencing technologies – the next generation. *Nat. Rev. Gen.* **11**:31–46.

Mills, D. K., Fitzgerald, K., Litchfield, C. D., Gillevet, P. M. 2003. A comparison of DNA profiling techniques for monitoring nutrient impact on microbial community composition during bioremediation of petroleum-contaminated soils. *J Microbiol Methods.* **54**:57-74.

Muyzer, G., Smalla, K., 1998. Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Ant. Leeuwenh.* **73**:127–141.

O'Brien, H. E., Parrent, J. L., Jackson, J. A., Moncalvo, J. M., Vilgalys, R. 2005. Fungal community analysis by large-scale sequencing of environmental samples. *Appl. Environ. Microbiol.* **71**:5544–5550.

Ogram, A., Saylor, G. S., Barkay, T. 1987. The extraction and purification of microbial DNA from sediments. *J. Microbiol. Meth.* **7**:57–66.

Ogram, A. V., Mathot, M., Harsh, J., Boyle, J., Pettigrew, C. 1994. Effects of DNA polymer length on its adsorption to soils. *Appl. Environ. Microbiol.* **60**:393–396.

- Osborn, A. M., Smith, C. J.** 2005. *Molecular Microbial Ecology BIOS Advanced Methods* Taylor & Francis, Routledge, pp. 34–35.
- Pace, N. A., Stahl, D. A., Lane, D. J., Olsen, G.** 1986. The analysis of natural microbial communities by ribosomal RNA sequences. *Adv. Microbiol. Ecol.* **9**:1–55.
- Palatinszky, M.** 2004. Környezeti minták DNS alapú diverzitás vizsgálatának egy általános problémaköre: a multitemplát PCR torzításai. Diplomamunka, Eötvös Loránd Tudományegyetem, Természettudományi Kar
- Pester, M., Friedrich, M. W., Schink, B., Brune, A.** 2004. *pmoA*-based analysis of methanotrophs in a littoral lake sediment reveals a diverse and stable community in a dynamic environment. *Appl. Environ. Microbiol.* **70**:3138–3142.
- Polz, M. F., Cavanaugh, C. M.** 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl. Environ. Microbiol.* **64**:3724–3730.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B., Ludwig, W., Peplies, J., Glöckner, F. O.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl. Acids Res.* **35**:7188–7196.
- Rainey, F. A., Ward, N., Sly, L. L., Stackebrandt, E.** 1994. Dependence on the taxon composition of clone libraries for PCR amplified, naturally occurring 16S rDNA, on the primer pair and the cloning system used. *Experientia* **50**:796–797.
- Ranjard, L., Poly, F., Nazaret, S.** 1999. Monitoring complex bacterial communities using culture-independent molecular techniques: application to soil environment. *Res. Microbiol.* **151**:167–177.
- Ranjard, L., Poly, F., Lata, J. C., Mougel, C., Thioulouse, J., Nazaret, S.** 2001. Characterisation of bacterial and fungal soil communities by automated ribosomal intergenic spacer analysis fingerprints: Biological and methodological variability. *Appl. Environ. Microbiol.* **67**:4479–4487.
- Reysenbach, A., Giver, L., Wickham, G., Pace, N.** 1992. Differential amplification of rRNA genes by polymerase chain reaction. *Appl. Environ. Microbiol.* **58**:3417–3418.

- Robertson, B. R., Tezuka, N., Watanabe, M. M.** 2001. Phylogenetic analyses of *Synechococcus* strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content. *Int. J. Syst. Evol. Microbiol.* **51**:861–871.
- Rocap, G., D. L. Distel, J. B. Waterbury, and S. W. Chisholm.** 2002. Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl. Environ. Microbiol.* **68**:1180–1191.
- Sarita, S., Priefer, U. B., Prell, J., Sharma, P. K.** 2008. Diversity of *nifH* gene amplified from rhizosphere soil DNA. *Curr. Science* **94**:111–115.
- Sauer, P., Müller, M., Kang, J.** 1998. Quantitation of DNA. *Qiagen News* **2**:23–26.
- Schmidt, T. M., DeLong, E. F., Pace, N. R.** 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J. Bacteriol.* **173**:4371–8.
- Shuman, S.** 1994. Novel approach to molecular cloning and polynucleotide synthesis using Vaccinia DNA topoisomerase. *J. Biol. Chem.* **269**:32678–32684.
- Singleton, D. R., Furlong, M. A., Rathbun, S. L., Whitman, W. B.** 2001. Quantitative comparisons of 16S rDNA sequence libraries from environmental samples. *Appl. Environ. Microbiol.* **67**:4373–4376.
- Sipos, R., Székely, A. J., Palatinszky, M., Révész, S., Márialigeti, K., Nikolausz, M.** 2007. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol. Ecol.* **60**:341–350.
- Sorensen, J., Haubjerg Nicolaisen, M., Ron, E., Simonet, P.** 2009. Molecular tools in rhizosphere microbiology—from single-cell to whole-community analysis. *Plant and Soil* **321**:483–512.
- Sorlie, S. S., Pecora, R.** 1990. A Dynamic Light Scattering Study of Four DNA Restriction. Fragments. *Macromolecules* **23**:487–497.

- Staley, J. T., Konopka, A.** 1985. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu. Rev. Microbiol.* **39**:321–346.
- Stein, J. L., Marsh, T. L., Wu, K. Y., Shizuya, H., DeLong, E. F.** 1996. Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* **178**:591–599.
- Suzuki, M. T., Giovannoni, S. J.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol.* **62**:625–630.
- Suzuki, M., Rappe, M. S., Giovannoni, S. J.** 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. *Appl. Environ. Microbiol.* **64**:4522–4529.
- Tamura, K., Dudley, J., Nei, M., Kumar, S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
- Tanner, M. A., Goebel, B. M., Dojka, M. A., Pace, N. R.** 1998. Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl. Environ. Microbiol.* **64**:3110–3113.
- Taylor, D. L., Herriott, I. C., Long, J., O'Neill, K.** 2007. TOPO TA is A-OK: a test of phylogenetic bias in fungal environmental clone library construction. *Environ. Microbiol.* **9**:1329–1334.
- Torsvik, V., Goksoyr, J., Daae, F. L.** 1990. High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* **56**:782–787.
- Torsvik, V., Sørheim, R., Goksoyr, J.** 1996. Total bacterial diversity in soil and sediment communities – A review. *J. Ind. Microbiol. Biotechnol.* **17**:170–178.
- Vandenkoornhuysse, P., Baldauf, S. L., Leyval, C., Straczek, J., Young, J. P. W.** 2002. Extensive fungal diversity in plant roots. *Science* **295**:2051.

Vaneechoutte, M., De Beenhouwer, H., Claeys, G., Verschraegen, G., De Rouck, A., Paepe, N., Elaichouni, A., Portaels, F. 1993. Identification of *Mycobacterium* species with amplified rDNA restriction analysis. *J. Clin. Microbiol.* **31**:2061–2065.

von Wintzingerode, F., Gobel, U. B., Stackebrandt, E. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev.* **21**: 213-229.

Woese, C. R. 1987. Bacterial evolution. *Microbiol. Rev.* **51**:221–271.

Yu, Z., Mohn, E. 2001. Bacterial diversity and community structure in an aerated lagoon revealed by ribosomal intergenic spacer analyses and 16S ribosomal DNA sequencing. *Appl. Environ. Microbiol.* **67**:1565–1574.

VIII. Internetes hivatkozások jegyzéke

HTTP-1:

http://rdp.cme.msu.edu/hierarchy/hb_intro.jsp

HTTP-2

http://tools.invitrogen.com/content/sfs/productnotes/F_TOPO%20RD-MKT-TL-HL0506021.pdf

HTTP-3

<http://www.promega.com/resources/articles/pubhub/enotes/shorten-the-ligation-time-for-the-pgem-t-vector-systems/>

HTTP-4

<http://www.promega.com/resources/articles/pubhub/enotes/comparing-cloning-efficiency-of-pgemt-and-pgemt-easy-vectors-to-topo-ta-cloning-vectors/>

HTTP-5:

<http://www.arb-silva.de/projects/ssu-ref-nr/>

HTTP-6:

<http://phylogeny-cafe.elte.hu/LengHet/>

HTTP-7:

<http://library.med.utah.edu/masspec/mongo.htm>

HTTP-8:

<http://libshuff.mib.uga.edu/> vagy <http://whitman.myweb.uga.edu/libshuff.html>

HTTP-9:

<http://www.arb-silva.de/aligner/>

IX. Rövidítésjegyzék

ARDRA:	Amplified Ribosomal DNA Restriction Analysis (amplifikált riboszomális DNS restrikciós analízis)
ATP:	Adenozin-trifoszfát
CE-LIF:	Capillary Electrophoresis, Laser Induced Fluorescence (kapilláris elektroforézis és lézergyjlesztésű fluoreszcencia intenzitás mérés)
DEPC:	Dietil-pirokarbonát
DGGE:	Denaturing Gradient Gel Electrophoresis (denaturáló grádiens gélelektroforézis)
DNS:	Dezoxi-ribonukleinsav
dNTP:	Dezoxi-ribonukleotid-5'-trifoszfát
FISH:	Fluoreszcens In Situ Hibridizáció
HPLC:	High Performance Liquid Chromatography
HT:	High Throughput (nagy áteresztőképességű)
IGS:	Intergenikus spacer
IPTG:	Izopropil β -D-1-tiogalaktopiranozid
IS:	Intergenikus spacer (prokarióták riboszomális RNS operonja esetében)
ITS:	Intergenikus spacer (gombák riboszomális RNS operonja esetében)
LB:	Luria-Bertani táptalaj
LH-PCR:	Length Heterogeneity-PCR (hosszheterogenitás-PCR)
OTU:	Operational Taxonomical Unit (operatív rendszertani egység)
PCR:	Polymerase Chain Reaction (polimeráz lánreakció)
PVPP:	polivinil-polipirrolidon
rDNS:	Riboszomális RNS-t kódoló gén
RDPII :	Ribosomal Database Project II
RFLP:	Restriction Fragment Length Polimorphism (restrikciós fragmenthossz polimorfizmus)
RISA:	Riboszomális Intergenikus Spacer Analízis
rRNS:	Riboszomális RNS
SSU:	Small Subunit (kis alegység)
TAMRA:	N,N,N',N'-tetrametil-6-karboxi-rodamin
TET:	Tetraklór-fluoreszcein foszforamidit
T-RFLP:	Terminal Restriction Fragment Length Polimorphism (terminális restrikciós fragmenthossz polimorfizmus)
UV-VIS:	Ultra Violet – Visible light (ibolyántúli-, és látható fény)
X-Gal:	5-bromo-4-kloro-3-indolil- β -D-galaktopiranozid

X. Összefoglaló

Az elmúlt másfél évtizedben a nukleinsav alapú módszerek váltak a mikrobiális ökológiai kutatások alapvető eszközeivé. Ezen módszerek között is kiemelt szerepet tölt be a polimeráz láncreakcióval felszaporított génszakaszok TA-klónozáson alapuló klónkönyvtárak elemzése. Az e célra általánosan alkalmazott, különféle ligálási mechanizmusú TA-klónozó vektorrendszerek hatékonyságát csak homogén PCR termékek klónozása esetében vizsgálták, ami nem felel meg a környezeti mintákban található, hosszukban és szekvenciájukban is igen heterogén inzertkeverék tulajdonságainak. A mikrobiális közösségvizsgálatok során gyakran elemzett 16S rDNS génszakaszok hosszheterogenitásáról rendelkezésre álló szórványos szakirodalmi adatok kiegészítése végett *in silico* vizsgálatokkal megkíséreltük feltérképezni a régió természetes variabilitását.

Kutatásunk fő célja a két leggyakrabban használt vektorrendszer esetleges preferenciális ligálásának kimutatása, valamint a torzítás reprodukálhatóságának, és az inzertek hosszától, valamint a ligálási paramétereiktől való függésének elemzése.

Kéttagú modellközösségekkel nyert eredményeink fényében kijelenthetjük, hogy a TA-klónozás nem mentes a szelektivitástól, s a különböző TA-klónozó rendszerek eltérő mértékű (sőt adott esetben irányú) torzítást mutatnak. A preferenciális ligálás reprodukálható jelenség, és mértéke függ a ligálási körülményektől. Ez a torzítás úgy tűnik nincs összefüggésben az inzertek GC-tartalmával, okát az inzertek hosszának és bázissorrendjének különbségeiben kell keresnünk.

Egy természetes baktériumközösség parciális 16S rDNS alapú vizsgálata során is szignifikáns különbségeket tapasztaltunk a kétféle vektorrendszerrel, valamint egy, a klónozástól független módszerrel detektálható közösségszerkezetek között.

XI. Abstract

In the past decades of environmental microbiology, nucleic-acid based methodology gained a central role in the characterisation of microbial communities. Among these methods TA-cloning of PCR-amplified genes retained its renown as the “golden standard”. The two widely used TA-cloning vectors are differing in their ligation mechanism, but their performance was only tested with unimolecular inserts. No thorough comparative testing has been carried out with different TA-cloning systems on heterogeneous insert pools such as community PCR products. As the information in related literature on the length heterogeneity of 16S rDNA gene is scarce, we investigated the natural length variation of this region by *in silico* analysis.

The main aim of this study was to reveal and compare the potential selectivity of the two most widely used TA-cloning vectors, and to describe the reproducibility, insert length-, and ligation parameter-dependence of this bias.

Our results show that TA-cloning is prone to biases, and the different vector systems have characteristic and alternate selectivity. Preferential ligation is a reproducible phenomenon, and is dependent on the ligation circumstances such as time and temperature. The bias seems to be independent of the GC content of the inserts, but related to the insert’s differences in length and sequence.

Comparing the detectable community structure of a natural microbial assemblage, we found that the two TA-cloning systems give significantly different results, which are also differing from the community structure revealed by a cloning-independent method.

XII. Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani:

Témavezetőmnek, **Márialigeti Károlynak**, hogy munkámat értékes szakmai tanácsaival, elgondolkodtató kérdéseivel segítette teljesebbé tenni.

Nikolausz Marcellnek, korábbi témavezetőmnek s legfőbb konzulensemnek, hogy bevezetett a molekuláris mikrobiológiai módszerek rejtelseibe, s hogy rendkívül értékes szakmai útmutatásával és emberi támogatásával mindvégig segítette munkámat.

Társtémavezetőmnek, **Borsodi Andreának**, a doktorandusz éveim alatt mindvégig tanúsított szakmai és emberi támogatásáért, valamint a Hévízi-tó kutatása során történt gyümölcsöző együttműködésért.

Novák Ádám barátomnak, a kiváló LengHet program elkészítéséért, és a lelki támogatásért.

Szaktervezőmnek, **Sváb Domonkosnak**, a közös munka során tanúsított lankadatlan szorgalmáért.

Sipos Ritának, korábbi témavezetőmnek, a szakdolgozatom elkészítése során nyújtott segítségéért.

Az **ELTE Mikrobiológiai Tanszék** valamennyi jelenlegi és korábbi munkatársának, az elmúlt tíz évben kapott nélkülözhetetlen segítségükért, és a sok vidám pillanatért.

S legfőképp **családomnak**, a végtelen türelmükért, és támogatásukért.

Az **Oktatási és Kulturális Minisztériumnak**, hogy a **Deák Ferenc Ösztöndíjjal** (DFÖ 081/2010) támogatták munkám befejezését.