Az Escherichia coli K-12 inszerciós elemeinek hatása a genom stabilitására

Ph.D. értekezés

Umenhoffer Kinga

Témavezető: Dr. Pósfai György

Biológia Doktori Iskola

MTA Szegedi Biológiai Központ Biokémiai Intézet

SZTE TTIK

2010

Szeged

1. Bevezetés	3
2. Szakirodalmi áttekintés	4
2. 1. Az Escherichia coli baktérium	4
2. 2. A baktériumsejtek változékonysága	6
2. 3. Az inszerciós szekvenciák szerkezete és működése	7
2. 4. Az inszerciós szekvenciák szerepe a genom változékonyságában	10
2. 5. IS-ek okozta hátrányok biotechnológiai munkákban: klónozási műtermékek	13
2. 6. Közvetlen előzmények: IS-ek az E. coli K-12 genomban	14
3. Célkitűzések	16
4. Anyagok és módszerek:	17
4. 1. Felhasznált törzsek, táptalajok és plazmidok	17
4. 2. Kompetens sejtek és transzformáció	18
4. 3. Növekedési görbék automatizált meghatározása	18
4.4. Indukálható toxikus és kontroll plazmid konstrukciók növekedésgátló hatásán	ak
vizsgálata folyadékkultúrában	19
4. 5. Mutációs ráta vizsgálatok	19
4. 6. Mutációk típusainak vizsgálata	20
4. 7. Plazmid konstrukciók létrehozása	20
4. 8. Reverz-transzkripcióhoz kapcsolt, szemikvantitatív PCR (RT-PCR)	21
4. 9. Mikroszkópos sejtanalízis	21
4. 10. Az E. coli IS elemek előfordulásának bioinformatikai analízise a shotg	un
szekvenálási adatbázisokban	22
5. Eredmények	23
5. 1. A pCTXVP60 klónok heterogén fenotípusának és genotípusának vizsgálata	23
5. 2. A transzpozíciós gyakoriság összehasonlítása normál és fehérjetermelés okoz	zta
stressz-körülmények között	26
5. 2. 1. Toxikus fehérje termelése okozta stressz hatása a mutációs rátára	26
5. 2. 2. Nem-toxikus fehérjetermelés okozta stressz hatása a mutációs rátára	27
5. 3. A pCTXVP60 plazmid toxicitásának vizsgálata	29
5. 3. 1. A pCTXVP60 plazmid toxicitása folyadékkultúrában	29
5. 3. 2. Ritka kodon használat és toxikusság kapcsolatának vizsgálata	29
5. 3. 3. A ctxvp60 gén és az azon belüli, mesterséges ORF toxikusságának vizsgálata	30

5. 4. IS inszerció hatása a leucingazdag ORF238 átíródására 32
5. 5. pCTXVP60 plazmidot tartalmazó törzsek növekedési/szelekciós dinamikája
5. 6. Az E. coli genomi IS-ek és exogén szekvenciák interferenciájának bioinformatikai
tanulmányozása a shotgun szekvenálások adatbázisaiban
6. Az eredmények megvitatása 40
6. 1. Az IS-ek részesedése a mutációs folyamatokban 40
6. 2. A <i>ctxvp60</i> toxikus gén analízise
6.3. Az IS-ek közvetítette mutációk létrejöttének és érvényre jutásának dinamikája 42
6. 4. Az IS-ek okozta mutációk speciális szerepe 43
6. 5. Az E. coli IS elemek előfordulása a shotgun szekvenálások adatbázisaiban
6. 6. IS elem: a gazdasejt védekező rendszere?
6. 7. Gyakorlati következtetések 45
Köszönetnyilvánítás:
Szakirodalomi hivatkozások 47
Saját publikációk jegyzéke
Summary 59
Összefoglalás
Táblázatok és Függelékek
F1. AZ ORF238 fehérje DAS-TMfilter szerkezetpredikciója67
F2. AZ ORF238 fehérje HMMTOP szerkezetpredikciója
F3. pCTX plazmid térképe 69
F4. pCTXVP60 plazmid térképe 70
F5. pCTXVP60Frameshift plazmid térképe71
F6. pSG1144 plazmid térképe 72
F7. pSG1144-ORF238 plazmid térképe
F8. pSG1144-CTXVP60 plazmid térképe
F9. pSG1144-CTXVP60OPTt plazmid térképe75
F10. pSG1144-CTXVP60DEZOPT plazmid térképe76
F11. Felhasznált primerek szekvenciái 77

1. Bevezetés

PhD munkámat az MTA Biokémiai Intézet Genommérnöki Csoportjában végeztem. A csoport fő projektje az *E. coli* baktérium genomjának nagy léptékű, racionális alapokon nyugvó átalakítása, egyszerűsítése. A laboratóriumi körülmények között nélkülözhető, többnyire horizontális géntranszferrel létrejött genomi szigetekben található gének kiejtésével egy olyan törzssorozat készült, amely fizikailag is fokozatosan közelíti az összehasonlító genomikai elemzés alapján megállapított, a változatos *E. coli* törzsekben közös génekből álló teoretikus vázgenomot.

A munka célja kettős. Egyrészt az egyszerűsített genomú törzsek vizsgálata számos tudományos kérdésre adhat választ (Mi a genomi szigetek globális szerepe? Mennyire integráns részei a genomnak a mobilis genetikai elemek? Hogyan változik a járulékos genom elvesztésével a sejt alkalmazkodóképessége? stb.). Másrészt - feltevésünk szerint - a redukált genomú törzsek stabilabb, az erőforrásokat hatékonyabban használó gazdasejtként alkalmazhatók szintetikus biológiai / biotechnológiai célokra.

A gének, géncsoportok kiejtése során különös figyelmet kaptak a mobilis genetikai elemek: profágok, inszerciós szekvenciák (IS). Az IS-ek olyan, viszonylagos önállósággal bíró genetikai struktúrák, amelyek saját fennmaradásuk biztosítása mellett alkalmanként a gazdasejt alkalmazkodóképességéhez is hozzájárulhatnak. A vadtípusú MG1655 törzs nagyszámú IS-t (44) tartalmaz. Az ebből készült redukált genomú törzssorozat 42 deléciót hordozó tagja, az MDS42 elvesztette az összes IS-t. Az IS-mentes törzs egyedi kontroll-lehetőséget nyújt ahhoz, hogy az IS-ek és a gazdasejt kapcsolatát vizsgáljuk. Feladatom az volt, hogy megvizsgáljam különféle (normál és stressz-) helyzetekben az IS-ek szerepét a genomstabilitás alakulásában. Ehhez a munkához további segítséget jelentett az egyik együttműködő partnerünk, G. Keil által létrehozott plazmid, amely váratlan módon igen nagy gyakorisággal vált a gazdasejtben IS transzpozíciók célpontjává. A dolgozat jelentős részét annak vizsgálata teszi ki, hogy milyen mechanizmus okozza ennek a plazmidnak az instabilitását. Eredményeinket shotgun szekvenálási adatbázisok bioinformatikai analízisével kiegészítve megállapítottuk, hogy az IS-eknek az előzetesen feltételezettnél nagyobb mértékű, gyakorlati szempontokból is jelentős hatása van a genom stabilitására.

2. Szakirodalmi áttekintés

2. 1. Az Escherichia coli baktérium

Az Escherichia coli baktérium egy az Enterobacteriaceae családba tartozó Gram-negatív, fakultatív anaerob, nem spórázó baktérium. Sejtjei pálcika alakúak, átlagosan 2 µm hosszúak és 0.5 µm átmérőjűek. Normál körülmények között megtalálhatók az emberek és az állatok alsó béltraktusában, ahol a normál baktériumflóra részét képezik. Kiürülve a természetes környezetbe is képesek túlélni, ez széleskörű elterjedést biztosít a fajnak [1]. A legtöbb szerotípusuk ártalmatlan, de előfordulnak patogén változatok is. A kórkép patomechanizmusa és klinikai tünetei alapján hat olyan csoportot különíthetünk el, melyek kórokozó képességek (mint pl. adhéziós-, inváziós képesség és toxintermelés) tekintetében egymástól lényegesen különböznek: enterotoxikus (ETEC), enteropatogén (EPEC), enteroinvazív (EIEC), enterohaemorrhagiás (EHEC), enteroaggregatív (EAggEC), enteroaddherens (EAEC) E. coli. Az E. coli K-12-es törzsét 1922-ben izolálták először egy torokgyíkos beteg székletmintájából [2]. Azóta napjaink legfontosabb prokarióta modellorganizmusa lett. Könnyű tenyészthetősége, valamint alaposan ismert, manipulálható genomja miatt a leggyakrabban használt laboratóriumi törzs. Széleskörű molekuláris biológiai módszertan áll rendelkezésre a tanulmányozásához. Nem meglepő, hogy ezt a baktériumot kezdték meg először szekvenálni a Wisconsini Egyetemen (Wisconsin Madison, USA), Frederick R. Blattner és kollégái részvételével 1992-ben [3]. Számos tanulmány az Escherichia coli MG1655-ös izolátumának szekvenciáját választotta referenciának, mert izolálása óta minimális genetikai változtatást végeztek el rajta [4]. Jó tanulmányozhatósága, könnyű kezelhetősége miatt a modern szintetikus biológiai munkákban első számú gazdasejt [5][6]. Rendkívül fontos szerepe van olyan biotechnológiai alkalmazásokban, mint különböző metabolitok, rekombináns fehérjék előállítása (pl. aminosavak, gyógyszeralapanyagok, bioüzemanyagok, hormonok). A modern rekombinációs technikák segítségével génterápiához, DNS vakcinákhoz, interferencia RNSekhez, sőt akár a műanyag alapanyaggyártáshoz is felhasználható [7]

Az MG1655 genomja cirkuláris, duplaszálú DNS molekula, mely 4639221 bázispárból áll, ennek 87,8 %-a fehérjét kódoló gén, 0,8 %-a stabil RNS, 0,7%-a nem kódoló szekvenciaismétlődés és 11 %-a szabályozó és egyéb más funkciójú gén. A genom GC tartalma 50,8 %. Annotációja során a fehérjét kódoló mintegy 4288 bizonyított és feltételezett fehérjét kódoló gén egyharmadának eddig nem sikerült a funkcióját kideríteni. A legtöbb gén a központi anyagcsere-folyamatokban vesz részt, főként a növekedésért felelős folyamatokban. A gének 10 %-a kódol feltehetőleg transzport és kötőfehérjéket, melyek szerepe a tápanyagok felvétele, valamint az anyagcseretermékek kijuttatása a környezetbe. A gének 5,7%-a az energia-anyagcserében, 18,9%-a különböző bioszintetikus anyagcserefolyamatokban, 1,3 %-a a transzkripcióban, míg 4,2 %-a a transzlációban vesz részt.

Az Escherichia coli baktérium kiváló modellorganizmus a genom szerkezete és a baktérium életvitele közötti összefüggés tanulmányozása szempontjából, hiszen törzsei közt megtalálhatóak patogén, szabadon élő nem-kórokozó, valamint különböző fajokhoz alkalmazkodott, ún. kommenzalista változatok is. A genom érdekes, mozaikos szerkezetet mutat [8], [9]. A genusra jellemző, alapvető anyagcserefolyamatokért felelős erősen konzervált vázgének sorát ("core" genom) változatos géntartalmú, törzs-specifikus genomi szigetek (járulékos genom) szakítják meg [10], [11]. A szigetek tipikusan különböző virulencia faktorokat, mobilis genetikai elemeket, valamint esetenként olyan metabolikus géneket kódolnak, melyek segítik az adott törzs speciális életkörülményeihez való alkalmazkodást[12], [13]. Ezek a genomi szigetek a dinamikus törzsfejlődési folyamatok során keletkeztek horizontális génátvitel, genomredukció vagy újrarendeződési események során [14]. Az MG1655 törzsben mintegy 100, legalább génnyi méretű genomi sziget azonosítható. Ezek összességükben a genom mintegy 25%-át teszik ki. Az MG1655 törzs tízféle, összesen 44 db IS-et, nyolc kriptikus vagy hibás profágot tartalmaz genomjának különböző helyein; ezek az összes DNS tartalom 2% -át teszik ki [3].

Napjainkban már több mint 100 *E. coli* és vele nagyon közeli rokon faj genomjainak szekvenáló adatait találhatjuk meg a GenBank adatbázisában. Az adatok bioinformatikai elemzésével egyre közelebb jutunk a "core" (egyfajta közös, mindegyik izolátumban egyaránt azonosítható gének csoportja) valamint a "pán" genom (az adott fajon belül előforduló, de nem minden egyes törzsben megtalálható gének összessége [15]) pontos, kvantitatív meghatározásához. Ugyanakkor nem mindegy, mely törzseket választjuk összehasonlítási alapnak. A "core genom" és "pán genom" számításokat a fajon belül a hasonló életvitelt mutató törzsek csoportjain is elvégezhetjük. A K-12 törzs a kommenzalisták közé tartozik; ebben a csoportban a közös géneket tartalmazó "core genom" mintegy 3200 gént jelent (**1. ábra**). Kiterjesztve az összes 13296 géncsaládból 1472 alapvető géncsaládot alkotja a váz genomt [16].



1. ábra A "pán" és "core" genom négy *E. coli* alcsoport törzseiben. Piros: 6 kommenzalista, fekete: 13 emésztőrendszeri patogén, zöld: 7 tápcsatornán kívüli patogén, kék: 6 Shigella törzs [17].

2. 2. A baktériumsejtek változékonysága

Minden élő sejt változik, a genetikai anyagban kisebb-nagyobb változások mennek végbe. A változások mértéke nem véletlenszerű, hanem hosszú evolúciós folyamat által behatárolt. A mutációs ráta finoman hangolt: kellően alacsony, hogy a sejt működését ne veszélyeztesse, ugyanakkor kellően magas, hogy a környezet változásaihoz való alkalmazkodáshoz megfelelő genetikai variabilitást biztosítson. A változékonyság mértékéről, mechanizmusáról - bakteriális rendszereken végzett tanulmányokból - jó néhány általános megállapítást lehet tenni[18], [19], [20]:

• A mutációk számának van felső, elméleti limitje (6 mutáció/genom/replikáció mezofil organizmusokban), efölött a sejt degradálódik, működése lehetetlenné válik.

• A mutációs ráta a genommérettel fordítottan arányos.

• A gyakorlatban a mutációs gyakoriság a felső limitnél mintegy ezerszer kisebb a proofreading és egyéb javító mechanizmusok működésének köszönhetően. A magasabb mutációs ráta fenntartása "kockázatos" és energetikailag sem előnyös.

• Stressz hatására,, amikor nagyobb genetikai variabilitás válhat szükségessé a populáció egy részének túléléséhez, a mutációs ráta megemelkedik, elsősorban a javító mechanizmusok gátlása, ill. hibázó polimerázok indukciója révén. Ez a képesség – a generált adaptív mutációk "potyautasaként" - azáltal marad fenn, hogy a sejtek időről-időre kedvezőtlen körülményeknek vannak kitéve.

• A mutációs ráta emelkedése tranziens, normál körülmények között visszaáll

alacsonyabb szintre.

• Mindezek a megállapítások pontmutációk tanulmányozásából erednek. A mutációs spektrumnak azonban más összetevői is vannak: a dolgozat tárgyát képező IS elemek jelentős tényezői a bakteriális genom változékonyságának. A legtöbb genom nagy számban tartalmaz IS elemeket, ezek inaktiváló inszerciókat, deléciókat, továbbá átrendeződéseket, kriptikus operonok aktiválását okozhatják, és elősegíthetik a horizontális géntranszfert.

2. 3. Az inszerciós szekvenciák szerkezete és működése

Az IS-ek a legegyszerűbb DNS transzpozonok, [21], [22], felfedezőjük Barbara McClintock [23], aki az 1950-es években írta le először az önálló áthelyeződésre képes DNS szakaszok meglétét kukoricában. Azóta már több mint 2400 IS-t írtak le és soroltak be 20 nagycsaládba (IS Finder, http://www-is.biotoul.fr/ [24]), genetikai felépítésük valamint transzpozázuk aminosav-szekvenciája alapján. A tipikus bakteriális IS hossza 0.7-2.5 kilobázispár (kb); ez a rövid DNS-darab magában hordozza az IS autonóm transzpozíciójához szükséges összes fontos elemet. Egyik ilyen fontos elem a terminális szekvencia, mely 10-40 bázispárnyi tökéletes vagy majdnem tökéletes ismétlődő szekvencia (inverted repeat sequences = IR). Két funkcionális részre osztható: a terminális doménre, amely 2-3 bázispárnyi, és a vágási valamint a száláthelyeződési folyamatok helye, és a transzpozáz kötéséért felelős doménre. A terminális szekvenciák között helyezkedik el a transzpozícióért felelős gén (2. ábra). A transzpozáz általában homodimer formában aktív fehérje, ahol az egyes monomereken több egymástól topológiailag és funkcióban is elkülöníthető motívum található. Egy-egy monomer esetében legalább két domént különböztetünk meg: a katalitikus részt, mely a C-terminálison, valamint a transzpozícióért felelős részt, mely az N-terminálison található. Egyes esetekben (pl. IS1 [25]) a két domént egymástól eltérő leolvasási keretben találjuk meg, melyek a transzláció során a leolvasási keret eltolódásával íródnak át egy fehérjévé [26].

Az IS-ek további fontos jellemzője az "tandem ismétlődő szekvenciák", vagy DR (direct repeat) szekvenciák. Ezek a transzpozíció során keletkeznek az IS két szélénél a cél (target) DNS-szekvenciák duplikációjával. Hosszuk 2-14 bázispár között mozog, és szekvenciájuk általában az adott IS-re jellemző. Előfordulhatnak IS elemek, melyek hosszabb vagy akár variábilis DR-tel rendelkeznek [27], [28].



2. ábra: Egy tipikus IS (nyitott doboz) szerveződése. Sötétszürke színnel az IR szekvenciák láthatóak, közöttük a transzpozáz helyezkedik el. *WXY* részek: az inszerció során keletkező DR szekvenciák. A transzpozáz promótere az IRL régióban található [29].

Az IS-ek transzpozáza a foszforiltranszferázok családjába tartozik Egy savas aminósav hármas felelős a katalítikus folyamatokért. Számos IS családnál és a retrovirális integrázoknál ez a hármas aminósavcsoport egy erősen konzervált "DDE" szakasz. Ez a motivum más további csoportokkal felelős a terminális bázispárok megfelelő pozícionálásáért[30].

A transzpozíciós aktivitás normál körülmények között alacsony szinten mozog, ez a transzpozáz nagyfokú szabályozottságának köszönhető. Ez mind a transzkripciós aktivitás kontrollálásával, mind poszt-transzlációs szabályozással megtörténhet, de megfigyelhetőek az adott gazda sejt faktorai által kontrollált folyamatok is [31].

Az IS-ek transzpozázának transzkripciós szabályozása történhet klasszikus módon, azaz endogén vagy alternatív promóterek használatával, transzkripciós terminációval, az mRNS stabilitásának befolyásolásával. Több IS esetében is leírtak olyan speciális mechanizmust, mely tompítja az aktivitást a gazdában történő transzpozíció (impinging-transcription) után. Transzlációs szabályozásukban is megtalálhatjuk a klasszikus példákat, valamint különleges, csak IS–ekre jellemző szabályozási módokat. Gyakori transzlációs szabályzásuk a programozott transzlációs frameshift, mely egy igen hatékony eszköz a transzpozáz aktivitásának irányítására, egyesítve két egymást követő nyitott leolvasási keretet (open reading frame vagy ORF), s ezzel lehetőséget teremtve a különböző fehérjék alternatív expressziójára. A transzpozíció szabályozására szolgálhat a cisz aktivitás is, ugyanis a transzpozáz hatékonyabban működik a saját elemén, mint más kópiáján ugyanannak az elemnek. Ugyanakkor maga a transzpozáz stabilitása is gátat szabhat a transzpozíciós Lon proteáz) szenzitív transzpozázok is. A gazdasejtben nem csak a proteázok hatnak a transzpozázra, hanem más különböző hatások is befolyásoló tényezők lehetnek (pl. hisztonszerű DNS chaperonok, girázok, metilázok, SOS-válaszban résztvevő szabályozó elemek). Nemcsak endogén, hanem külső környezeti tényezők is befolyásolhatják a transzpozíciós gyakoriságot, mint például a hőmérséklet, UV sugárzás, mágneses mező [32], [33], [34].

Maga a transzpozíció mechanizmusa több, jól elkülöníthető biokémiai lépésre osztható. Első lépésben a transzpozáz az IS elem mindkét végén megtalálható, fordítottan ismétlődő elméhez köt, és létrehoz egy ún. szinaptikus komplexet. Ebben komplexben transzpozáz és más, az IS fajtájától függő asszociált fehérjék is megtalálhatóak. A végek kivágódását, újrarendeződését követően létrejön egy speciális fehérje-DNS komplex, melyet transzposzómának nevezünk [35].

Megkülönböztetünk non-replikatív és replikatív transzpozíciót. Az előbbi esetben nem készül másolat az áthelyeződés során, míg utóbbi esetben az áthelyeződés egyben az IS duplikációjával jár.

Az áthelyeződére képes genetikai elemek egyik fontos tulajdonsága a transzpozon immunitás: ha már a célmolekulán van egy meghatározott transzpozíciós elem, akkor jelentősen csökkenti egy másik kópia beépülését. Ezen tulajdonság érdekes módon nem jellemző az IS-ekre. Eddig egyetlen bizonyított esetet találtak, ahol ez megfigyelhető volt (IS21) [36].

Az IS-ek inszerciója során általában megfigyelhető egy változó, az adott elemre adott célszekvenciával jellemző preferencia egy-egy vagy régióval szemben (targetspecificitás) [37]. Ennek a célszekvencia specificitásnak különböző fokai vannak az egészen véletlenszerűtől a konzervált helyspecifikus preferenciáig. Például míg a Mu fág szinte bármely szekvenciába inszertálódhat [38], egyes IS-ek hosszabb, 100 bázispárnyi szakaszokba, mások jól definiálható, rövidebb konszenzus szekvenciákba képesek beépülni [39]. Megfigyelhető teljesen helyspecifikus inszerció is (IS91-nél GAAC/CAAG a célszekvencia [40]). Egyes elemek régióspecificitást mutatnak például a GC- (IS186 [41]) vagy AT-gazdag DNS részeknél (IS1 [42]). Esetenként olyan általánosabb jellemzőkkel is összefügghet ez a helypreferencia, mint a helyi DNS struktúra, a supercoiling mértéke, vagy a DNS görbülete.

2. 4. Az inszerciós szekvenciák szerepe a genom változékonyságában

A transzpozíciós elemek általánosan előfordulnak. Ritka kivétel pl. az egyik elsőnek teljesen megszekvenált genom - a fontos modellorganizmus, a *Bacillus subtilis* egyik törzse – melyben nem találtak sem IS-t, sem egyéb transzpozíciós elemet [43]. Később két másik *B. subtilis* törzset szekvenálva megtalálták a rájuk jellemző IS-t [44]. Napjainkban már jelentős mennyiségű teljes genomszekvencia áll rendelkezésre, ezek adataiból kitűnik, hogy más, többnyire kisméretű genomok esetében is hiányozhatnak ezek a szekvenciák [45]. Ugyanakkor találhatunk törzseket, felhalmozódhattak (**3. ábra**). Meglepő módon még a nagyon közeli rokon fajok összehasonlításakor is találhatunk eltérő IS számot és variabilitást. A *Shigella flexneri* 2a szerotípusának kromoszómáján 247 teljes és 67 töredék IS található, míg a vele közeli rokon, patogén *E. coli* O157 EDL933 kromoszómáján 21 ép és 19 hiányos IS-et azonosítottak.



3. ábra: IS családok elterjedése az Eubaktériumokban és az Archaeákban. A különböző családok különböző színnel jelöltek [45].

Ezek a mobilis szekvenciák a genetikai változásokban a spontán mutációs ráta jelentős részéért felelősek [46], [47], [48]], annak - a genetikai valamint fiziológiai környezettől függően - igen széles spektrumát képviselhetik. Egy adott gén funkcióváltozásának 3.9%-tól

[49] akár 98%-ig [50] terjedő mértékben is okozói lehetnek. Nemcsak a spontán mutációs eseményekben, de a különböző endogén valamint környezeti stresszre adott válaszokban is jelentős szerepük lehet. A genom különböző helyeire történő transzpozíció által a bakteriális sejtek képesek lehetnek olyan megváltozott körülményekhez alkalmazkodni, mint a szénforrások megváltozása [51], [52], antibiotikum gátlás [53], ozmotikus stressz [54]. Az IS-ek többféleképpen változtathatják meg az adott sejt genomját. Áthelyeződésük során operonok vagy gének aktivációját okozhatják [52], [55], inaktiváló inszerciókat, deléciókat, továbbá kromoszóma átrendeződéseket hozhatnak létre [56], [57]. Poláris hatással erőteljesen befolyásolhatják a szomszédos gének működését [58], valamint elősegíthetik a horizontális géntranszfert [14], [59].

Sokrétű hatásuk azt sejteti, hogy evolúciós szerepük nem elhanyagolható. Az általuk okozott legjellemzőbb hatás az inszercióval létrejövő géninaktiváció. Ezt a jelenséget használja ki a mai molekuláris biológiában is egy igen gyakori technika, a transzpozon mutagenezis [60].

Az ellenkező jelenségre, az inszerciós aktiválásra is van alaposan tanulmányozott példa: a kriptikus *bgl* operoné. A *bgl* operon kódol minden funkciót, mely szükséges a β -glükozid szénforrások (szalicin, arbutin) transzportjához és hidrolíziséhez [61], de ahhoz hogy a növekedéshez megfelelő szinten fejeződjön ki, szükséges egy aktiváló mutáció. Ez a mutáció az esetek nagy részében (98%) egy IS – túlnyomó többségben IS1 vagy IS5 – inszerciója a promóter egyik meghatározó régiójába, mely a transzkripciós startponttól számított –132 és +91 bázispár között helyezkedik el [50]. Hasonló példát nemcsak operonoknál, de egyedi géneknél is megfigyelhetünk. Így az *ade* kriptikus gén esetében is, mely egy mangánfüggő adenin deamináz, és erős IS-közvetítette aktivációja figyelhető meg [55].

Nemcsak az anyagcserével kapcsolatos változásokban találhatunk IS elem okozta aktivációt. Gyakran megfigyelhető inszerciós aktiváció egy, a flagelláris transzkripciós szabályozásért felelős operon, az *flhD* esetében, mely az így megnövekedett expressziója révén fokozza a mozgásképességet [62].

A génkifejeződés befolyásolását poláris hatással is elérheti egy IS. Kísérletes adatokból tudjuk, hogy többüknél (IS1, IS2 és IS5) megtalálhatóak a terminális szekvenciában kifelé mutató -35 promóter hexamerek, melyek ha pontos távolságra kerülnek a transzpozíció során egyes, helyben előforduló -10 hexamerektől, egy új promóter képződhet, mely képes elindítani a szomszédos gének átíródását [39].



4. ábra: Horizontális transzfer által szerzett DNS eloszlása az MG1655 kromoszómáján. A kapott régiók tulajdonságai (koruk, illetve repetitív és mobilis szekvenciák jelenléte) színkódolva vannak. Fizikai elhelyezkedésük az replikációs origóhoz és terminátorhoz viszonyítva a függőleges oszlopon, míg nagyságuk a vízszintes tengelyen olvasható le [59]

Megvizsgálva a horizontális transzfer által szerzett DNS régiók szekvenciáját az MG1655 kromoszómáján, kimutatható (**4. ábra**), hogy a törzs IS elemeinek 68%-a (37-ből 25) szorosan összekapcsolható velük [59]. Ez az eloszlás több feltételezést indukál. Egyrészt jelentheti azt, hogy az IS elemek a horizontális génátviteli események során bejutott részekkel kerülnek a kromoszómára. Az is fennállhat, hogy a genomi szigetek kevéssé fontos génjei jobban tolerálják, s így megőrzik az utólag odaugrott IS-eket. Gyakori előfordulásuk a saját és az átvitt DNS közötti kapcsoló régióban azonban inkább arra utalhat, hogy az IS-ek elősegítik az idegen DNS inszertálódását a replikációjuk során keletkező kointegrátum képzéssel, s nem később helyeződnek át a már horizontálisan átvitt szakaszokba.

Kérdés, hogy ezek a genetikai változások csak melléktermékei egy molekuláris parazita működésének, vagy összehangolt tényezői a genomevolúciónak. Evolúciós nézőpontból két, egymásnak ellentmondó elmélet van az IS elemek (valamint a mobilis elemek) genomi jelenlétének és szerepének magyarázatára. Egyik nézet szerint az IS-ek genomi paraziták, kimagaslóan hatékony önző gének, melyek a gazdát kihasználva önmaguk fenntartására törekednek [63], [64], [65]. Másik nézet szerint a gazdasejttel együtt fejlődtek, azzal összehangoltan működnek és a sejtnek hasznos részei [66], [56]. A mobilis genetikai elemek gazdag szakirodalmában mindkét feltételezés alátámasztására találhatunk bizonyítékot és ellentmondást is. Egyrészt úgy tűnik, önző paraziták: hiszen az IS elemek adott gazdában elszaporodhatnak, de ki is veszhetnek, majd újra fertőzhetnek [14], [65]. Hatásaik sokszor károsak a gazdasejt számára (deléciós mutánsok, géninaktiválás, új fehérjevariánsok létrehozása). Másrészt a gazdasejttel összhangban működnek és nem elhanyagolható szerepük van az evolúciós adaptációkban, sőt bizonyos esetekben (pl. kriptikus operonok-gének aktiválásával) a túlélést is biztosíthatják a sejt számára.

2. 5. IS-ek okozta hátrányok biotechnológiai munkákban: klónozási műtermékek

A bakteriális sejtek köztük az *E. coli* biotechnológiai alkalmazását nagymértékben nehezíthetik az IS-ek előidézte DNS újrarendeződési események, valamint a transzpozíciók során bekövetkező mutációk, mivel ezek műtermékekhez, a kívánt klón instabilitásához, a fáradságosan előállított termelő sejtek leromlásához vezethetnek. A szakirodalomban több esetben olvashatunk ilyen klónozási és fenntartási hibákról, annak ellenére, hogy a sikertelen kísérleteket ritkán közlik. A *Proteus vulgaris* restrikciós-modifikációs rendszerének

klónozásakor három, egymástól független esetben kaptak IS1 inszerciót a klónozott génekben [67]. A *Serratia marcescens lpp* génjének klónozásakor is találtak IS1 és IS5 inszerciókat [68]. A *Streptococcus equisimilis* H46A sztreptokinázának tanulmányozása során szintén IS1 inszerció volt tapasztalható, mely ugyan épen hagyta magát a gént, de az expresszióját csökkentette [69]. Az ilyen klónozási hibák *E. coli*-n kívül más organizmusban is előfordulnak, *Agrobacterium tumefaciens* gazdasejtnél is egy transzgént inaktivált a beleugró IS136 [70].

Az inszerciós elemek gondot okozhatnak DNS vakcinák előállításakor, fermentáció során. Ilyen jelenséget a Merck Kutatási Laboratóriumában figyeltek meg HIV vakcina előállítása közben. *E. coliban* fermentálva a vektoron bekövetkező IS1 inszerció okozott alacsony termelési hatékonyságot [71]. A probléma fontossága szintén nem elhanyagolható a "shotgun" szekvenáló kutatásoknál, ahol az adatbázisok tisztaságát szavatolni kell. Már az emberi genom "cosmid" alapú könyvtárának tanulmányozásakor észrevették a szekvencia IS-szennyeződését [72]. Az adatbázisok *in silico* vizsgálatával derült fény a különböző genomszekvenciákban felhalmozódott nem kívánt mobilis szekvenciák gyakori előfordulására [73], [74], [75].

Eklatáns példa az IS-ek nem várt felbukkanására a C. Venter munkacsoportja által elkészített első szintetikus genom esete [76]. A sok éves munkával összerakott *Mycoplasma mycoides* genom újraszekvenálása egy nem várt IS1 elemet mutatott ki a genomban. Az IS1 a köztes lépések során – mikor a szintetikus DNS-darabokat *E. coli* gazdában tartották fenn – ugorhatott a *M. mycoides* DNS-be.

2. 6. Közvetlen előzmények: IS-ek az E. coli K-12 genomban

A megszekvenált MG1655 törzs tízféle, összesen 44 IS-t, nyolc kriptikus vagy hibás profágot tartalmaz genomjának különböző helyein; ezek az összes DNS tartalom 2 % -át teszik ki [3]. Az azonosított mobil genetikai elemek különböző transzpozíciós aktivitást mutatnak [77]. Különösen aktív az IS150, mely a hosszú távú evolúciós - 10000 generációs, [57] valamint a szúrt kulturákat vizsgáló [78] kísérletekben és a váltakozó fagyasztási-olvasztási ciklusokkal dolgozó - kísérletben [79] a genom újrarendeződések nagy részéért felelt. Magas transzpozíciós aktivitással rendelkezik az IS1 melyet nemcsak a hosszú távú evolúciós kísérletekben de egyéb egyedi géneket mint az ebgR [51], bglR [50], cycA [53], tonB [80] vizsgálva is rendkívül gyakran felbukkan.

A csoportunkban végzett genomredukciós munka során a sorozatos deléciókkal

fokozatosan csökkent a genomban az IS-ek száma. A munka közben derült fény arra, hogy a törzsben a szekvenálás óta újabb IS transzpozíciós események történtek, és 5 újabb genomi pozícióban is található IS. Ezek feltérképezését és delécióját követően a 42 deléciót hordozó, összességében 14%-kal redukált genomú MDS42 lett a törzssorozat első olyan tagja, mely nem tartalmaz egyetlen IS-et sem [81], [82]. A vad típusú MG1655, a deléciós törzssorozat különböző számú IS-et hordozó tagjai és az IS-mentes MDS42 egyedi lehetőséget nyújtott az IS-ek genomstabilitásra gyakorolt hatásának felméréséhez.

A munka során felhasználtunk egy különösen instabil, IS elemeket "begyűjteni" látszó plazmidot. A konstrukció együttműködő kollégánktól, Günther Keiltől származik. A koleratoxin (CTX) B alegységét próbálta klónozni a nyúl hemorrhágiás vírus [83] VP60 fehérjéjét kódoló génnel kapcsolva. A fúziós gén állati vakcinagyártáshoz készült, és kodonhasználatát tekintve növényi expresszióra optimalizálták [84]. Az előzetes tapasztalatok szerint a pBR322 típusú plazmidba klónozott fúziós gént nem lehetett stabilan fenntartani. Tekintet nélkül az *E. coli* klónozó gazdasejt típusára, a plazmidba 12 órás növesztés alatt is IS-ek épültek, illetve az esetek kisebb részében deléciók és egyéb mutációk módosították a szekvenciát.

3. Célkitűzések

Munkánk célja, hogy megvizsgáljuk az IS elemek genomstabilitásra gyakorolt hatását az *E. coli* baktériumban, illetve megmutassuk az IS-mentes, redukált genomú törzsek klónozási munkákban érvényesülő előnyeit.

A vizsgálódás közelebb vihet számos, nyitott kérdés megválaszolásához. Az IS elemek a genomevolúció integráns elemeit képezik-e, vagy elsősorban önálló, molekuláris paraziták, amelyek csak járulékosan, véletlenszerűen okoznak előnyös változásokat? A változékonyság generálásában milyen speciális, csak az IS elemekre jellemző szerepük van? Gyakorlati szempontból: elhanyagolható-e az IS elemek szerepe a sejt biotechnológiai / szintetikus biológiai alkalmazásakor, vagy ellenkezőleg, jelentős hibaforrást jelenthetnek az IS transzpozíció okozta műtermékek klónozási munkákban?

A vizsgálódáshoz néhány egyedi eszközzel rendelkezünk: i) soklépcsős deléciós munkával előállított, különböző számú IS-t tartalmazó ill. teljesen IS-mentes *E. coli* törzs, ii) egy olyan – feltételezhetően a sejt számára toxikus - gént hordozó plazmid, mely látszólag "IS-csapdaként" működik, azaz normál gazdasejtben igen gyorsan IS hordozóvá válik.

A következőket kívántuk vizsgálni:

• Mekkora az IS-ek átlagos transzpozíciós gyakorisága *E. coli*ban? Nagyságrendileg azonos-e a pontmutációs rátával?

Történik-e stressz hatására transzpozíciók indukciója, hasonlóan a pontmutációkhoz?

• Hogyan működik az esetünkben talált "IS-csapda"? DNS-, RNS- vagy fehérje-szinten funkcionál az IS-ugrást kiváltó tényező?

• Hogyan lehetséges, hogy a ritka eseménynek számító IS inszerció eredményeként létrejött mutáns "IS-csapda" plazmid rövid idő alatt domináns lesz a populációban? Milyen az IS inszerciót szenvedett plazmid szelekciójának dinamikája?

• Az "IS-csapdaként" működő plazmidnál tapasztaltak mennyire általánosíthatók? Milyen gyakori probléma a klónozási munkák során az IS-transzpozíció miatt bekövetkező klónozhatatlanság? Találunk-e analóg eseteket nyers shotgun genomszekvenálási adatbankokban?

4. Anyagok és módszerek:

4. 1. Felhasznált törzsek, táptalajok és plazmidok

Kísérleteinket *E. coli* K-12 MG1655-ös törzsön [3], valamint számos módosított változatán végeztük. Az MDS12, MDS30, és MDS42 (valamint más MDS) törzsek sorszámuknak megfelelő genomi deléciót tartalmazó, csökkentett genomú változatai az MG1655-nek [81][82]. Az MG1655-T7, valamint az MDS42-T7 olyan T7 polimeráz expresszáló törzsek, melyek tartalmaznak egy IPTG-vel indukálható, lac operátor/T7 polimeráz kazettát az eredeti törzs *yahA-yaiL* genomi regiójában. Az MDS42+IS1 és MDS42+IS2 törzsek egy-egy kópia IS1 (*yeaJ*) ill. IS2 (*yddA*) inszerciós elemet tartalmaznak.

A kísérletek során általános gazdag táptalajt, folyékony és szilárd Luria-Bertani (LB) táptalajt használtunk [85]. Mutációs ráta méréseknél, D-cikloszerin antibiotikum alkalmazásakor MS minimál táptalajt használtunk [50]. Antibiotikumokat a következő végkoncentrációban alkalmaztuk: kloramfenikol (Cm) 25 μg/ml, ampicillin (Amp) 100 μg/ml, D-cikloszerin (Cyc) 0.04 mM. A D-cikloszerin törzsoldata (4 mg/ml) mindig frissen, használat előtt készült.

Az eredeti pCTX, pVP60 valamint pCTXVP60 plazmidokat Günther M. Keil készítette [67]. A pSG1144 plazmid a Scarab Genomics (Madison, WI, USA) tulajdona; F-típusú replikációs origót, R2 plazmidról származó *trfA*-oriV replikont [86], T7 promótert, valamint kloramfenikol rezisztencia gént hordoz. Az *orf238* és a *ctxvp60* IPTG-vel indukálható változata a pSG1144 klónozó helyébe illesztve készült. A mesterséges *ctxvp60opt* és *ctxvp60dezopt* szekvenciákat a GenScript Corporation-nal (Piscataway, NJ, USA) szintetizáltattuk; ezeket szintén a pSG1144 plazmidba klónoztunk. A pProEX HT-CAT (Invitrogen, Cat. No. 10711) plazmid a Scarab Genomics-tól (Madison, WI, USA) származik. A plazmidokat kizárólag IS mentes (MDS42) gazdatörzsből, hagyományos alkáli-lízis módszerével tisztítottuk. A toxikus pCTXVP60 plazmidot külön eljárással tisztítottuk: frissen elektroporált sejteket agar lemezre szélesztettünk, majd 14-15 óra növesztés után a normál, egészséges fenotípust mutató (mutáns) kolóniákat kivágtuk és eltávolítottuk, majd a toxikus fenotípust mutató, gyönge növekedésű kolóniákat TE-vel lemostuk. Az így kapott sejtszuszpenziót Sigma GenEluteTM Plasmid Miniprep Kit-el dolgoztuk fel.

4. 2. Kompetens sejtek és transzformáció

Az elektrokompetens sejteket standard protokoll alapján készítettük. 100 ml megfelelő antibiotikumot tartalmazó LB-tápoldatba 1 ml starter kultúrát oltottunk át, ezt $OD_{590}=0,5-0,6$ értékig növesztettük, majd 4 °C-ra hűtöttük. A sejteket 15 percig centrifugálva ülepítettük (4000 rpm), majd 100 ml jéghideg desztillált vízben szuszpendáltuk. Újabb centrifugálás után a sejteket 50 ml desztillált vízben vettük fel. A harmadik centrifugálást követően 4 ml jéghideg, 20%-os glicerinben szuszpendáltuk a sejteket. A következő centrifugálás (15 perc, 5000 rpm) után 0,2 ml hideg, 20%-os glicerinben vettük fel a sejteket, majd 40 μ l-enként Eppendorf-csövekbe osztottuk szét. Az így elkészített sejtek azonnal használhatók elektroporálásra, vagy –80 °C-on tárolhatók.

A kompetens sejtek elektrosokkolását Invitrogen Elektroporator II vel, 0.1 cm-es küvettákban 1800 V maximális feszültség és 150 Ω maximális ellenállás mellett végeztük. Ezután 1 ml LB tápoldatot adtunk a sejtekhez, majd egy órán át inkubáltuk a megfelelő hőmérsékleten, végül a megfelelő agarlemezre szélesztettük a kultúrát.

4. 3. Növekedési görbék automatizált meghatározása

A különféle plazmid-gazdasejt kombinációk növekedési dinamikáját 100-well-es Honeycomb 2 lemezekben (Oy Growth Curves Ab, Helsinki, Finland), a megfelelő antibiotikummal kiegészített, folyékony LB táptalajban vizsgáltuk. A sejtek növekedését 540nm-es hullámhosszon mért optikai denzitáson (O.D.) követtük Bioscreen C Automated Microbiology Growth Analysis System (Oy Growth Curves Ab, Helsinki, Finland) használatával. A kísérletekhez frissen elektroporált (14-15 órát növesztett) lemezekről szuszpendáltunk fel egy-egy kolóniát (200 μ l LB táptalajban, melyet 25 μ g/ml kloramfenikolal egészítettünk ki), minden gazda-vektor párosításból 10 párhuzamossal dolgozva. Ezeket a Honeycomb lemezekben 37°C-on folyamatos rázatás mellett növesztettünk fel.

4.4. Indukálható toxikus és kontroll plazmid konstrukciók növekedésgátló hatásának vizsgálata folyadékkultúrában

MDS42-T7 sejteket elektroporáltunk a megfelelő plazmid konstrukciókkal. Telítettségig felnövesztett starter kultúrákból 1-1 ml-t oltottunk tovább (2 párhuzamossal dolgozva) 50 ml folyékony LB táptalajt tartalmazó lombikokba, majd 220 rpm-en 37°C-on növesztettük tovább. A sejtek növekedési görbéjét OD 550 nm-en, 30 perces mintavételekkel, WPA Colourmax CO75000 koloriméter segítségével végzett mérésekkel állapítottuk meg. Az indukálandó mintákhoz 1 mM IPTG inducert adtunk 90 perces növesztés után, kivétel az SG1144-ORF238 konstrukció, ahol az időpont 120 percre tolódott.

4. 5. Mutációs ráta vizsgálatok

A mutációs ráta vizsgálata során D-cikloszerin [87][88] rezisztens sejtek megjelenési gyakoriságát követtük fluktuációs teszt segítségével [41]. A kísérlet során frissen elektroporált sejteket oltottunk 20-20 cső 1 ml, glükózzal kiegészített MS oldatba (10⁴ CFU/ml), melyet szükség szerint antibiotikummal egészítettünk ki, majd a kultúrákat 37°C-on korai stacioner fázisig növesztettük. Minden egyes csőből 50 µl kultúrát D-cikloszerin (0.04 mM) tartalmú minimál táptalajra szélesztettünk. A kapott telepszámból a becsült mutáció per cső számot (m) Ma-Sandri-Sarkar maximum-likelihood (MSS-MLL) eljárással [89] határoztuk meg. Az 50 µl-e vonatkozó m számból a teljes 1 ml-es kultúrára vonatkoztatható m számot Stewart nyomán számoltuk ki [90]. Csak azokat a kísérleteket vettük figyelembe, ahol az összehasonlítani kívánt kultúrák összsejtszámainak eltérése elhanyagolható (<3%, p \geq 0.6 kétmintás t-teszttel) volt. Az összsejtszámokat 4-4 csőből vett minták higításainak nem lemezekre történő kiszélesztésével kaptuk. szelektív agar А mutációs rátát (mutáció/sejt/generáció) a mutáció per cső és az átlag összsejtszám per cső hányadosából számoltuk.

A fehérje túltermelés mutációs rátára gyakorolt hatását [82] pProEX HT-CAT plazmidot tartalmazó MG1655 sejtekkel vizsgáltuk. Mind IPTG indukcióval, mind anélkül fluktuációs teszteket végeztünk. A sejtekhez az IPTG-t 0.6 mM-os koncentrációban, exponenciális növekedési fázisban adtuk hozzá (A590 = 0.7), majd a kultúrát 37°C-on telítettségig növesztettük, és a fent leírt módszerrel D-cikloszerin rezisztens kolóniákra

szelektáltunk. A mutációs ráta méréseket minimum háromszor ismételtük meg.

4. 6. Mutációk típusainak vizsgálata

A D-cikloszerin rezisztenciáért felelős *cycA* gén esetében meghatároztuk annak mutációs spektrumát [53]. 1877 bázispárnyi genomi régiót - mely tartalmazta a teljes gént - amplifikáltunk a mutáns sejtekből cycA1/cycA2 primer párokkal, kolónia PCR segítségével. Reprezentatív mintának 20 kolóniát használtunk minden parallel lemezről, így kísérletenként 400 mintával dolgoztunk. Az amplifikált DNS fragmenteket agaróz gélen elektroforézissel elválasztottuk, és egy vad típusú fragmenthez hasonlítva elemeztük. Azonos nagyság egy vagy néhány nukleotid változását jelzi, csökkent méret vagy fragment hiánya deléciót mutat, a megnövekedett méret pedig IS elem jelenlétére utal. A kapott mutációs változatok arányait a paralell kísérletekben kétmintás t-teszttel hasonlítottuk.

A *ctxvp60* gén mutációs spektrumát a fent leírt módszerrel határoztuk meg: a mutáns sejtekből egy 2481 bázispárnyi szakasz T7/chi3 primerpárral történő amplifikálásával, kolónia PCR segítségével. Inszerciós mutánsok esetében az egyes IS elemeket IS- és plazmidspecifikus primerekkel, PCR–el azonosítottuk. A mutáns szekvenciák közül 12-t szekvenálásnak vetettünk alá [91].

4. 7. Plazmid konstrukciók létrehozása

Az eredeti pCTXVP60 plazmid frameshift-tel rendelkező változatát helyspecifikus mutagenezissel, többlépcsős PCR-el készítettük el. A frameshift-et létrehozó (egy extra T bázis hozzáadását biztosító), egymással ellentétes irányba mutató, átfedő primer párt (vpf1/vpf2) terveztünk; ezek magukban hordozták az inszerciót. További két, külső (T7/CVEK3) primer segítségével, rekombináns PCR módszerrel elkészítettük a gén frameshift-et tartalmazó változatát, melyet *NcoI- Hind*III helyre, az eredeti plazmidba klónoztunk. A szekvenciák pontosságát szekvenálással ellenőriztük.

Az SG1144 vektor indukálható konstrukcióit a plazmid *Eco*RI-*Bam*HI valamint *NdeI-HindIII* klónozóhelyébe inszertálva készítettük el. A különböző vektorokból származó szakaszokat minden esetben PCR segítségével emeltük ki. A felhasznált plazmidok létrehozásában a részletesen nem tárgyalt rekombinációs lépéseket közismert molekuláris

biológiai módszerekkel végeztük [85]. A plazmidok térképe a Függelékben található.

4. 8. Reverz-transzkripcióhoz kapcsolt, szemikvantitatív PCR (RT-PCR)

Az RT-PCR során használt teljes RNS mintákat RNS tisztító készlet segítségével preparáltuk (Sigma-Aldrich, GenEluteTM Total RNA purification kit, Cat. No.: BTR7). Egyedi kolóniákról "overnight" kultúrákat indítottunk 3 ml LB tápoldatban. A pCTXVP60 ismert IS mutánsaiból nyert (pCTXVP60 IS/2, 3, 7, 8) starterekből 3 ml-es kultúrákat indítottunk, majd az exponenciális növekedési fázisban teljes RNS-t tisztítottunk. A pCTXVP60 plazmid friss transzformációjából származó kis és nagyméretű telepekből annyit gyűjtöttünk össze, hogy az 1 ml-re higított szuszpenzió elérje az $OD_{600}=1.0$ töménységet. A teljes RNS-t ebből a szuszpenzióból tisztítottuk a friss kultúráknál alkalmazott módszerrel. A tisztított RNS minőségét agaróz gélelektroforézissel ellenőriztük. A nyers RNS-ből 1 µg-nyi mennyiséget a gyártó protokollja szerint DNáz (Fermentas, Cat. No.: EN0521) kezeléssel tisztítottuk meg a szennyező, RT-PCR-t zavaró genomi DNS-től. A DNase-kezelt RNS mintákat PCR-el ellenőriztük, hogy sikerült-e eliminálni a szennyező DNS-t, majd cDNS szintéziséhez használtuk templátként (Applied Biosystems, GeneAmp^R RNA PCR, Part No.: N808-0017). A cDNS szintéziséhez random hexamer primereket használtunk a gyártó protokollja leírásának megfelelően. A keletkező cDNS templát vizsgálatát ORFstart/ORFstop primerekkel, PCR-rel végeztük. Belső kontrollként a recA transzkriptum mennyiségét detektáltuk. A PCR-rel kapott fragmenteket agaróz gélelektroforézissel vizsgáltuk.

4.9. Mikroszkópos sejtanalízis

A sejteket 37°C-on 30 percig 10 µM nonyl acridine orange (NAO) és 10µg/ml ethidiumbromid (EtBr) oldatban inkubáltuk. A DNS festéséhez 70%-os etanolos fixáció (30 perc) és PBS mosás után 500ng/ml 4',6-diamidino-2-phenylindole (DAPI) oldatot alkalmaztunk (10 perc). A mikroszkópos megfigyeléshez a fedőlemezeken 2 százalékos agarblokkokkal (1cm x 1cm) immobilizáltuk a mintákat. A konfokális lézer pásztázó mikroszkópiát Olympus Fluoview FV1000 konfokális lézer pásztázó mikroszkóppal végeztük (Olympus Life Science Europa GmbH, Hamburg, Germany), UPLSAPO 60x (olaj, NA:1.35) objektívvel. A képek kiértékeléséhez Olympus Fluoview software (version 1.7.2.2) programot használtuk.

4. 10. Az *E. coli* IS elemek előfordulásának bioinformatikai analízise a shotgun szekvenálási adatbázisokban

A bioinformatikai elemzések technikai részét – konzultációk után – Pósfai János végezte. A szekvencia adatokat az "NCBI Trace Archive" adatbázisból töltöttük le. A bakteriális és az archea genomokat akkor használtuk fel, ha az egyedi részszekvenciák, az összeállított szekvenciák, az annotációk és a kísérletes mellékletek is hozzáférhetőek voltak. Az újabb generációs szekvenáló technológiákkal (primerwalk, 454 stb.) felépített adatbázisokat kizártuk a keresés során, mivel azok előállításához nincs szükség klónozási lépésekre. Két BLAST adatbázist építettünk fel minden vizsgált genom esetében, egyet az egyedi "shotgun" klónok szekvenciáiból, egyet pedig ugyanazon genom összeállított, annotált szekvenciájából. Összegyűjtöttük az E. coli K-12 MG1655 törzs (U00096) IS elemeit, kiegészítve a DH10B törzs IS10 szekvenciájával (NC_010473). Az IS-ekből összeállított szekvenciakészlet nyomait kerestük az általunk létrehozott BLAST adatbázisokban nukleotid szintű homológiák után kutatva. Az inszerciós események számának megbecsüléséhez szigorú paramétereket használtuk (expectation value 0.1, minimum score 75, minimum run length 40 nucleotides). Az első keresési körben az IS-ek homológjait kerestük a "shotgun" klón-könyvtárakban. Ez a keresés olyan találatokat szolgáltatott, melyek feltehetően E. coli IS-szerű szekvenciákat tartalmazó klónokra mutatnak. Az első keresési körben megtalált klónokat – ha tartalmaztak genomi eredetű szekvenciát is - betérképeztük az összeállított genomszekvenciákra.

Következő lépésként az IS homológjait kerestük immár az összeállított szekvenciákon. Ezek a találatok nagy valószínűséggel endogén IS-re mutatnak (és nem a klónozási lépések során beugrott, szennyező IS szekvenciákra), és a további vizsgálatokból kizártuk őket. A maradék "shotgun" klónok (az első keresés találatai közül kizártuk a második keresés találatait) genomi szakaszát használtuk a következő keresésben, és a nyers "shotgun" könyvtárakban kerestük meg homológjaikat. Végül a találatokat újra megkerestük az összeállított genomokon is, és csak a kölcsönösen legjobb találatokat tartottuk meg.

Ezzel az "előre-hátra" BLAST módszerrel azonosítottunk olyan szekvenciákat, melyek a "shotgun" könyvárak létrehozása közbeni klónozási lépések során szereztek *E. coli* eredetű IS elemeket. A módszer alapján azonosíthatóak voltak olyan genomi szekvenciák, melyekben megemelkedett gyakorisággal fordultak elő a klónozások során szerzett IS-ek. Ezeken a genomi szakaszokon azonosítottuk a kódoló géneket és azok feltételezett funkcióját (GenBank annotációk alapján).

5. Eredmények

5. 1. A pCTXVP60 klónok heterogén fenotípusának és genotípusának vizsgálata

Előzetes kísérletek alapján a fentebb ismertetett, kiméragént (nyúl hemorrhágiás vírus kapszid fehérje gén (VP60) + koleratoxin B alegység (CTX)[84]) hordozó pCTXVP60 plazmid instabilnak bizonyult különböző hagyományosan használt *E. coli* törzsekben (HB101, C600, DH10B és MG1655). A fúziós gént tartalmazó plazmidokról fehérjetermelés nem mutatható ki, és a DNS kinyerés hozama is alacsony volt. Ugyanakkor külön-külön klónozva a fúziós gén elemeit, a plazmidok megfelelő hatékonysággal, változatlan szekvenciával fenntarthatóak. A pCTXVP60 plazmiddal történő transzformálás után a kapott kolóniák heterogének voltak LB-agar lemezeken. A transzformáns telepek többsége kicsi, lassú növekedésű, áttetsző volt. Kevés egészséges kinézetű, normál növekedésű telep jelent meg a kontroll pCTX klónokhoz viszonyítva. Hosszú (>24 h) inkubáció hatására végül minden kicsi telep normális sebességű szektoros növekedésnek indult. Ezzel ellenétben IS-mentes MDS42 törzset használva a transzformánsok egységesebb morfológiával rendelkeztek LB-agar lemezen. Telepeiket ugyancsak gátolt növekedés jellemezte, ám köztük csak < 0,1 % mutatott nagyobb, normális növekedést éjszakán át inkubált lemezeken (**5. ábra**).



5. ábra (A) pCTXVP60/MG1655 és (B) pCTXVP60/MDS42 transzformánsok LB+ kloramfenikol agar lemezen. A nyilak a valószínűsíthetően már valamilyen mutációval inaktivált fúziós gént hordozó plazmiddal rendelkező telepeket mutatják [91].

A normál (nem IS-mentes) gazdából visszaizolált plazmidok *NcoI/Eco*RI restrikciós mintázata többségében eltérést mutatott az elméletileg várthoz képest, és arra utalt, hogy a klónozott gén mérete a legtöbb esetben megnövekedett. A különböző törzsekre jellemző

hasítási mintázatból a jellegzetes DNS fragmentumokat izoláltuk és megszekvenáltuk. A kapott szekvenciákban minden esetben IS szekvenciákat találtunk a *ctxvp60* génben. Az MDS42 gazdasejtből kinyert pCTXVP60 plazmid viszont a várt restrikciós mintázatot mutatta (**6. ábra**).



6. ábra A pCTXVP60 plazmid jellegzetes restrikciós mintázata. Különböző törzsekből izolált pCTXVP60 NcoI-EcoRI hasítási képe. (1) MDS41-ből; normál hasítási kép (2) MDS42-ből; normál hasítási kép (3) DH10B-ből (4) DH10B-ből; normál hasítási kép – szekvenálás alapján átrendeződött szekvencia (5) C600-ból (6) C600-ból (7) C600-ból [82]

További elemzéshez az MG1655-ös törzset használtuk fel. 100 nagyméretű, egészséges pCTXVP60 transzformáns kolónia PCR analízise alapján megállapítottuk, hogy 92 %-uk IS inszerciót hordozott, míg 8 % mutatott deléciót vagy egyéb átrendeződést. Az IOS-inszerciók esetében IS1, IS3, és IS5 transzlokációkat mutattunk ki. Az inszerciók pontos helyét 12 klónban szekvenálással határoztuk meg (**1. táblázat**). Minden inszerciós esemény a fúziós gén 5' végén, a *ctx* régióban vagy a *vp60* 5'végénél következett be (**7. ábra**).



7. ábra A ctxvp60 térképe az IS inszerciók helyével. [91]

A kapott eredmények arra utaltak, hogy a CTXVP60 fehérje toxikus hatású a gazdasejt számára. A véletlenszerűen, méghozzá nagy százalékban specifikusan IS inszerció következtében létrejövő, inaktivált toxikus gént hordozó sejt viszont már normálisan képes növekedni. Felmerült a kérdés, hogy a gyors, IS-ek általi inaktiválódás i) erősen megnövekedett transzpozíciós aktivitásnak, ii) a gátolt növekedés miatt előnyt élvező inaktivált, normál növekedésű mutánsok gyors szelekciójának, iii) esetleg mindkét folyamatnak az eredménye. További kérdés, hogy miért elsősorban IS-inszerciók (és nem pontmutációk) képesek inaktiválni a gént.

	week and a	2 0.902 es
Vektor	IS típusa	inszerció helye
pCTXVP60IS/1	IS5	350
pCTXVP60IS/2	IS5	350
pCTXVP60IS/3	IS5	350
pCTXVP60IS/4	IS1	-5
pCTXVP60IS/5	IS5	350
pCTXVP60IS/6	IS1	250
pCTXVP60IS/7	IS3	384
pCTXVP60IS/8	IS3	397
pCTXVP60IS/9	IS5	350
pCTXVP60IS/10	IS5	350
pCTXVP60IS/11	IS1	336
pCTXVP60IS/12	IS1	751

táblázat A pCTXVP60 plazmid megszekvenált, IS inszerciót hordozó mutánsainak adatai.
Az inszerciók helye a *ctxvp60* gén start kodonjától mért távolsággal van megadva (bázispárban)

5. 2. A transzpozíciós gyakoriság összehasonlítása normál és fehérjetermelés okozta stressz-körülmények között

5. 2. 1. Toxikus fehérje termelése okozta stressz hatása a mutációs rátára.

Annak eldöntésére, hogy a toxikus gén okozta stressz megnöveli-e a transzpozíciós gyakoriságot, megvizsgáltuk a pCTXVP60 (toxikus fehérjét expresszáló) és a pCTX (nemtoxikus *ctx* gént hordozó) plazmidokat tartalmazó gazdasejtek mutációs rátáját és spektrumát. A mutációs ráta méréséhez a csoportunk által kidolgozott, *cycA* inaktiváción alapuló fluktuációs teszt módszerét alkalmaztuk [53] A módszer egy olyan, bizonyos aminosavak és a D-cikloszerin antibiotikum felvételéért egyaránt felelős gént (*cycA*) inaktiváló mutációk kimutatásán alapul, amelyek rezisztenssé teszik a sejtet erre az antibiotikumra. A géntermék nem játszik szerepet minimál táptalajon való növekedés során, mutációinak száma és aránya tehát nem torzul esetleges szelekciós hatások miatt [53].

Hogy elkerüljük a kísérlethez felhasznált pCTXVP60 plazmidpopuláció mutációk okozta heterogenitását, speciális módon tisztított, homogén plazmidokat használtunk (ld. Anyagok és módszerek fejezet). A gazdasejteket (MG1655 és MDS42) elektroporáltuk, majd közvetlenül az előinkubációs periódus után fluktuációs tesztben telítettségig növesztettük, végül D-cikloszerin rezisztens sejtekre szelektáltunk MS/glükóz agar lemezeken. A kapott telepszámokkal fluktuációs analízist végezve meghatároztuk a mutációs rátát, a cycA gént PCR-el ellenőrizve pedig típizáltuk az egyes mutánsokat. Az MG1655/pCTXVP60 sejtek magasabb mutációs rátát mutattak (1.19×10^{-7}) , mint a kontrollként alkalmazott MG1655/pCTX (5.7 $\times 10^{-8}$). Az eltérést elsősorban az IS transzpozíciós események négyszeres emelkedése okozta. IS specifikus PCR segítségével kimutattuk, hogy legalább négy különböző IS is részt vesz a magasabb mutációs ráta létrehozásában (IS1, IS2, IS5, IS150). A kapott rezisztens telepeken nemcsak a cycA mutációit, de a plazmidot is ellenőriztük IS mutációkra nézve. A fúziós gént PCR-el vizsgálva a sejtek 32 százaléka tartalmazott mind a plazmidon, mind a kromoszómán IS inszerciót (8. ábra). A várakozásoknak megfelelően az MDS42-es gazdatörzsek IS-mentesek, így bennük IS okozta mutációt nem találtunk. A pontmutációs és deléciós rátájuk, valamint összsejtszámuk mind a kontroll, mind a vizsgált plazmid esetében nem tért el szignifikánsan az MG1655-ös törzsben mértekétől.



8. ábra A pCTXVP60 és pCTX plazmidok hatása a *cycA* gén mutációs spektrumára MG1655 és MDS42 törzsekben [82].

5. 2. 2. Nem-toxikus fehérjetermelés okozta stressz hatása a mutációs rátára

Megvizsgáltuk egy a sejt által jól tolerált, elterjedten használt fehérjének, a kloramfenikol acetil transzferáz (CAT) túltermelésének hatását a mutációs rátára, ezen belül az IS transzpozícióra. MG1655 törzsbe elektroporáltunk CAT túltermelő plazmidot (pProEX HT-CAT/ Invitrogen), majd ezután fluktuációs mérést végeztünk a fehérjetermelés indukciójával ill. indukció nélkül. D-cikloszerin rezisztens sejtekre szelektáltunk és a cycA gént PCR-el ellenőriztük. A kapott rezisztens sejtek pontmutációs rátája az indukált és indukálatlan mintákban egymáshoz hasonlítva nem változott meg szignifikánsan. Ugyanakkor az IS inszerció gyakorisága a cycA génen mérve 136%-al magasabb volt az indukált mintákban, mint a nem indukáltakban (9. ábra). IS-specifikus primerekkel végzett PCR-el IS1, IS2, IS5 és IS150 szerepe is kimutatható volt. Duplamutánsok számának meghatározása céljából a telepekről plazmid preparátumokat készítettünk, majd ellenőriztük a restrikciós mintázatukat. Fehérjetúltermelés hatására a pProEX HT-CAT plazmidon a CAT génjében inszerciót nem találtunk. További kontrollkísérletekként, az elektroporáció és az IPTG indukció esetleges mutagén hatásának ellenőrzése érdekében plazmid nélkül is elektroporáltunk MG1655 kompetens sejteket, ill. MG1655 kultúrákhoz IPTG-t adtunk. Ezekkel a mintákkal mutációs ráta mérést végeztünk: a kapott ráta nem tért el szignifikánsan az MG1655 mutációs rátájától.



9. ábra CAT fehérjetúltermelés hatása a *cycA* gén mutációs spektrumára MG1655/ pProEX HT-CAT sejtekben.[82]

Összefoglalva: Kismértékű stressz (nem-toxikus fehérje túltermeltetése) nem okozott változást a gazdasejt pontmutációs rátájában, a transzpozíciós gyakoriságot viszont 2.5-szeresére emelte. Erős stressz (toxikus fehérje termeltetése) viszont mind a pontmutációs rátát (1.5x), mind a transzpozíciós gyakoriságot (4x) – ez utóbbit különösen markánsan – megemelte. A kísérleti rendszerben (*cycA*) négyféle IS mutagén hatását tudtuk kimutatni.

A semleges génen mért mutációs adatok azt mutatják, hogy az IS-ek normál körülmények között is szignifikáns mértékben járulnak hozzá a sejt mutációs terheléséhez. Stresszkörülmények között szerepük relatíve tovább növekszik: a vizsgált körülmények között nagyobb arányban indukálódnak, mint a pontmutációk. Ez a növekedés azonban nem olyan mértékű, hogy a pCTXVP60 esetében tapasztalható gyors IS általi inaktivációt megmagyarázza.

5. 3. A pCTXVP60 plazmid toxicitásának vizsgálata

5. 3. 1. A pCTXVP60 plazmid toxicitása folyadékkultúrában

Az agar lemezeken észlelt, pCTXVP60 okozta növekedésgátlás pontosabb detektálásához a *ctxvp60* gént indukálható konstrukcióban, a pSG1144 plazmidba klónoztuk, majd MDS42-T7 sejtekbe transzformáltuk. Az IPTG-indukált (az ábrán "+"-al jelölt) és indukálatlan, 50 ml-es folyadékkultúrák növekedési görbéje jól mutatja a *ctxvp60* gén/konstrukt növekedésgátló hatását (**10. ábra** A, A+, B, B+; az ábra egyéb görbéinek magyarázata a további alfejezetekben). (A méréseket saját előzetes kísérletek után Balikó Gabriella végezte.).



10. ábra. Különböző *ctxvp60* variánsok expressziójának hatása MDS42-T7 sejtek növekedésére. (A, A+) pSG1144 plazmid (B, B+) pSG1144-*ctxvp60* (C, C+) pSG1144-*ctxvp60opt* (D, D+) pSG1144-*ctxvp60dezopt* (E, E+) pSG1144-*orf238*. A nyilak az IPTG indukció kezdetét jelölik.

5. 3. 2. Ritka kodon használat és toxikusság kapcsolatának vizsgálata

A *ctxvp60* szintetikus gén nagy számban tartalmaz ritka Arg kodonokat (1 AGG és 21 AGA). Feltételezésünk az volt, hogy a gén transzlációja kimeríti a sejt ritka arginin tRNS készletét, a csökkenő Arg tRNS mennyisége pedig gátolja az alapvető fehérjék megfelelő szintű szintézisét, s ez lehet a toxikus hatás magyarázata [92] [93]. Ennek bizonyításához két mesterséges *ctxvp60* gént szintetizáltattunk: a *ctxvp60opt* verzió az *E. coli* számára csak optimális Arg kodonokat tartalmazott, a *ctxvp60dezopt* viszont fokozott mértékben tartalmazott ritka Arg kodont (22 AGG). A kapott szintetikus géneket az SG1144 vektor *NdeI-HindIII* helyére klónoztuk, majd MDS42-T7 sejtekbe elektroporáltuk. A kapott transzformánsokból növekedési görbét mértünk inducer jelenlétében és indukció nélkül (**10. ábra**). Az előzetes várakozással ellentétben az optimalizált és dezoptimalizált verziók növekedési retardációt. A hipotézis tehát nem igazolódott – nem a ritka kodonok fokozott használata okozza a pCTXVP60 plazmid toxikusságát.

5. 3. 3. A ctxvp60 gén és az azon belüli, mesterséges ORF toxikusságának vizsgálata

A toxikus gén (részlet) azonosításához egy olyan variánst készítettünk, amely frameshift mutációt (T inszerció a 3227-s pozícióban) hordoz a *vp60* szakasz 5' végén (**11. ábra**). A várakozással ellentétben a leolvasási keret eltolódása nem szüntette meg a toxikus hatást (**16. ábra E**). Ez arra utalt, hogy nem maga a *ctxvp60* gén terméke okozza a növekedésgátlást.



11. ábra A *ctxvp60/orf238* térképe az IS inszerciók helyével. A függőleges vonalkák az *orf238*-on található lehetséges transzlációs reiniciációs helyeket jelölik. F: frameshift mutáció helye a vp60 5' régiójában.

Az eredeti ctxvp60 gén szekvenciájának vizsgálatakor felmerült, hogy a génen belül más, nem

tervezett leolvasási keret is átíródhat (ORF238), melyről egy 238 aminosavat tartalmazó fehérje képződhet. Ez az ORF a *ctx* és a *vp60* rész fúziójánál kezdődik és 3' vége belenyúlik a *vp60* gén szekvenciájába (**11. ábra**). Az ORF238 transzlációja során egy extrém leucin gazdag fehérje képződik (102 db Leu azaz 42.9%) (**12 ábra**). HMMTOP és DAS-TMfilter [94], [95], [96] szerkezetpredikciós programok szerint a fehérje erősen hidrofób jellegű, és 4 feltehető transzmembrán domént tartalmaz. Ezek alapján feltételezhető, hogy a fehérje a sejtek membránjába épül be (Függelék).

>238 aminosav hosszúságú ORF238 peptidszekvencia

MLLLLFLWLMDQDPWKERLELLHKLELLELLLLLLFQELLLMEWTQELLLLLL LLLKILLLLLELEDHHNKLINKKLGELIFIIMMFLLGLLLMLQDLFFILFNILHK IIHSLLFFLKCMLDGLEECNSDSLLLDLEFLEEDLLLLLFHQELKLDQVLKLDNFH MLLLMLGLLNQLLLLCQTLDQICIIQLEIQDLFQLLFFLFIIILLIHLEDLLLLFKLLL KLDHLKILNLL

12. ábra Az ORF238 peptidszekvenciája

Megvizsgáltuk, hogy az *orf238* kifejeződése okozza-e a növekedési gátlást. Ennek bizonyításául PCR segítségével kiemeltük az *orf238* szekvenciát, és indukálható konstrukcióként a pSG1144 plazmidba klónoztuk. Ezt a plazmidot MDS42-T7 és MG1655-T7 sejtekbe transzformáltuk. Indukció hiányában normális fenotípusú telepeket kaptunk. A telepekből indított növekedési görbe mérések alapján inducer hiányában nem mutatkozik toxikus fenotípus. Ezzel szemben, ha IPTG-vel indukáltuk az *orf238* átíródását, a pCTXVP60-éhoz hasonló, erős növekedésgátlást észleltünk (**10. ábra**). Megállapítható tehát, hogy nem maga a *ctxvp60* fúziós gén, hanem egy nem tervezett leolvasási keretben képződő hidrofób fehérje okozza a toxikus hatást.

A toxikus hatás fenotípusos vizsgálatához CTXVP60 és ORF238 expresszáló sejteket konfokális lézer pásztázó mikroszkópos vizsgálatnak vetettük alá. A plazmidmentes MG1655 sejtet hasonlítottunk össze az MG1655 pCTXVP60 plazmidot tartalmazó (konstitutívan expresszáló) sejtjeivel. Továbbá MG1655-T7 törzsbe transzformáltunk SG1144-ORF238 plazmidot, és megvizsgáltuk a sejteket indukálatlan és IPTG-indukált állapotban. Nukleinsav festést (etídium-bromid vagy DAPI) kombináltunk különféle interferencia kontrasztos képfeldolgozásokkal (DIC), valamint membránfestést (nonyl acridine orange) is alkalmaztunk. A kísérletek során az MG1655 törzs, valamint az ORF238-at nem indukált formában tartalmazó sejtek normál fiziológiás fenotípust és sejtméretet mutattak (**13. ábra A**, **C, D).** Ezzel szemben mind a pCTXVP60 plazmiddal rendelkező, mind az indukált ORF238at tartalmazó minták rendellenes hosszúságú, lassú növekedésű és sejtosztódásukban gátolt sejteket tartalmaztak (**13. ábra B, E, F).**



13. ábra Toxikus és kontroll plazmidot tartalmazó sejtek fluoreszcens mikroszkópikus képei. Toxikus konstrukció pCTXVP60 (B) és indukált ORF238 (E,F) minták, kontroll minták MG1655 (A) és nem indukált ORF238 (C,D). (A) és (B) ábrán a membránokat nonyl acridine orange festéssel (zöld) tettük láthatóvá, ethidium bromid festékkel (piros) nukleotidokat festettük. (C) és (E) ábrán DAPI-val festett nukleotidok (piros) láthatóak (B) és (E) ábrán nyilakkal jelöltük a rendellenesen megnyúlt sejteket. Mérce: 2µm. Az ábrát Ferhan Ayaydin készítette

5. 4. IS inszerció hatása a leucingazdag ORF238 átíródására.

Megállapítottuk tehát, hogy a toxikus, sejtnövekedést gátló hatásért az *orf238* expressziója felelős. A pCTXVP60 plazmidon térképezett IS-inszerciók pozíciója azt sugallta, hogy az IS elemek a toxikus *orf238* 5' régiójába ugorva annak transzkripcióját blokkolja, megszüntetve ezzel a plazmid toxikus hatását.

Ennek bizonyítása érdekében MG1655 sejtbe transzformáltunk pCTXVP60 plazmidot, majd a kapott agar lemezekről mintát gyűjtöttünk az eltérő fenotípusú kolóniákból (kis és nagy telepekről). Az agar lemezekről levett és felszuszpendált mintákból RNS izoláltunk és szemikvantitatív RT-PCR segítségével próbáltuk detektálni a toxikus *orf238* transzkriptumait. Belső kontrollként a *recA* gén átíródását követtük. A még toxikus fenotípust mutató kisebb telepekről vett mintában kimutatható volt az *orf238* transzkriptuma. Ezzel ellentétben az egészséges méretű sejtekből vett mintákból már nem volt kimutatható. (**14. ábra**)





14. ábra A transzkripciós gátlás bizonyítása RT-PCR segítségével telepekről. (A) MG1655 pCTXVP60 heterogén telepeiről izolált RNS minták RT-PCR elektroforetikus képe. (1) lassú növekedésű telep (2) gyors növekedésű telep ORFstart és ORFstop primerekkel. (3) lassú növekedésű telep (4) gyors növekedésű telep *recA* kontroll primerekkel (B) a *ctxvp60* fúziós gén térképe ábrázolja az *orf238* és a PCR primerek elhelyezkedését, valamint egy IS elem tipikus előfordulási pozícióját.

A következő kísérlethez kiválasztottunk négy különböző, az IS elemet a ctxvp60 5' végének

eltérő helyein tartalmazó pCTXVP60 vektort (pCTXVP60 IS/ 2, 4, 7, 12). Ezekből exponenciális növekedési szakaszban RNS-t izoláltunk, majd RT-PCR segítségével kerestük az *orf238*-ről származó transzkriptumot, viszonyítva az előzőekben már detektált pCTXVP60 kis telepeiről kapott transzkriptumhoz. Egyik plazmidból származó mintából sem tudtuk kimutatni a toxikus gén termékét, igazolva, hogy az IS inszerciók a gén 5' végén meggátolták annak átíródását (**15. ábra**).



15. ábra A toxikus gén transzkripciója gátlásának bizonyítása RT-PCR segítségével, különböző pozíciókban IS-t hordozó pCTXVP60 plazmidok esetében. (A) A pCTXVP60-IS/ 2, 4, 7, 12 plazmidokat tartalmazó sejtekből (sorrendben 1-től 4-ig), valamint az egészséges növekedést mutató telepekből (5) nyert RNS minták RT-PCR elektroforetikus gélképe ORFstart és ORFstop primerekkel. (B) Ugyanezen minták kontroll gélképe (*recA* amplifikációja).

5. 5. pCTXVP60 plazmidot tartalmazó törzsek növekedési/szelekciós dinamikája

Az eddigi eredmények alapján tisztázódott, hogy a pCTXVP60 plazmid egy toxikus mellékterméke gátolja a sejtek növekedését. A stresszelt sejtekben megemelkedik az IS-ek transzpozíciós aktivitása. A plazmid meghatározott régiójába ugró IS elemek meggátolják a toxikus szakasz átíródását, lehetővé téve a sejt normális növekedését. Mindezeknek a folyamatoknak a dinamikáját, a gazdasejt IS-készletétől való függését különféle gazdasejt-plazmid kombinációk növekedési görbéjének kimérésével vizsgáltuk.

MG1655 és MDS42 gazdatörzseket elektroporáltunk a toxikus fúziós gént tartalmazó pCTXVP60, valamint kontrollként a pVP60 és a pCTX plazmidokkal (utóbbiak nem

mutatnak toxikus hatást). Minden gazda-vektor kombinációból tíz párhuzamos, külön telepről származó kultúrát növesztettünk, és Bioscreen C használatával, automata leolvasásokkal követtük növekedésüket. Mivel az MG1655 törzsben a pCTXVP60 plazmid transzformációja heterogén telepeket eredményezett LB agar lemezeken, ebből a törzsből külön kis és nagy (még toxikus ill. már inaktiváló mutációs plazmiddal rendelkező) kolóniákból is vettünk mintát. A kísérletek során a kis telepekről leoltott, intakt pCTXVP60 toxikus plazmidokat tartalmazó kultúrák bizonyos sejtsűrűség elérése után (OD₅₄₀=0.2-0.5 között) nem növekedtek tovább (**16. ábra B**). Azonban ha az inkubációt folytattuk, a minták változatos időpontokban újra növekedésnek indultak, és elérték a kontroll kultúrák (MG1655/pCTX) végső sűrűségét (**16. ábra A**). Az ezekből a kinőtt mintákból izolált plazmidok minden esetben IS inszerciót tartalmaztak a fúziós génben. Ezekből a variábilis növekedést mutató, telítettségig növesztett kultúrákból újraoltást követően normális, megszakítás nélküli növekedési görbéket figyeltünk meg, csakúgy, mint a nagy telepekről leoltott kultúrák esetén.

Ezeknek a folyadékultúrás kísérleteknek az eredményei összhangban vannak az agar lemezeken megfigyeltekkel [82]: i) pCTXVP60 növekedésgátlást okoz, ii) a sejtek egy része véletlenszerű módon IS-inszerciót szenved a *ctxvp60* génben a genomi IS-ek (fokozott) mobilitásának köszönhetően, iii) az IS-ek által inaktivált toxikus gént hordozó sejtek visszanyerik normális növekedési ütemüket.

A vadtípusú gazdasejthez képest az MDS42 törzsben a pCTXVP60 jelenléte mellett mért növekedési görbék egységesebb lefutásúak voltak. A kezdeti növekedés $OD_{540}=0.3$ közelében megállt, és a tíz párhuzamos mintából csupán egy nőtt a kontrollhoz mérhető szaturációig (**16. ábra D**). Ebből az egy felnőtt mintából plazmidot izolálva deléciót találtunk a *ctxvp60* génben. Az IS-mentes MDS42-ben tehát kisebb eséllyel inaktiválódik a toxikus gén, ami arra utal, hogy az inaktiválódás elsősorban IS-inszerció révén mehet végbe.


16. ábra A pCTXVP60 hatása MG1655 és MDS42 sejtek növekedésére. (A) és (B): MG1655 sejtek növekedési görbéje kontroll pCTX ill. pCTXVP60 plazmid jelenlétében. (C) és (D): ugyanezen plazmidok hatása MDS42 sejteken. (E): pCTXVP60 frameshift-tel rendelkező változatának hatása.

A következőkben a pCTXVP60 plazmidot hordozó, genomi IS számukban különböző gazdatörzsek növekedését vizsgáltuk.

MG1655 (44 IS), MDS12 (25 IS), MDS30 (5 IS), valamint MDS42 (IS-mentes) sejteket transzformáltuk pCTXVP60 plazmiddal, és meghatároztuk tíz párhuzamos kultúra növekedési görbéjét. A kapott adatokból az adott gazdasejtre vonatkozó növekedési görbék középértékét ábrázoltuk. (17. ábra) Megállapítható, hogy a nem-toxikus, mutáns plazmidvariáns kialakulásához szükséges idő korrelál a gazdasejt IS-számával: minél több IS található a genomban, annál gyorsabb a normál növekedésű mutánsok evolúciója.



17. ábra (A): 4 különböző deléciós fázisból származó törzs növekedési görbéje pCTXVP60 jelenlétében. (B): az adott törzs IS elem tartalma.

Egy további kísérletben izogenikus gazdatörzspárok használatával (MDS42 és MDS42+IS1 ill. MDS42+IS2) megállapítottuk, hogy akár egyetlen IS jelenléte is drámai hatással lehet a növekedési képességre (**18. ábra**).



18. ábra A pCTXVP60 hatása egyetlen IS-t tartalmazó MDS42 sejtek növekedésére. (A): MDS42+IS1/ pCTXVP60, (B): MDS42+IS2 /pCTXVP60

5. 6. Az *E. coli* genomi IS-ek és exogén szekvenciák interferenciájának bioinformatikai tanulmányozása a shotgun szekvenálások adatbázisaiban

A következőkben arra a kérdésre kerestük a választ, hogy a toxikus pCTXVP60 plazmid ISek általi semlegesítését vizsgáló kísérletek tanulságai mennyiben általánosíthatók. Egyedi jelenségről van-e szó, vagy ellenkezőleg, szignifikáns mértékű-e a gazda IS-ek és az exogén szekvenciák interferenciája?

A vizsgálatra alkalmas, nagy volumenű adatbázist szolgáltatnak a nyers shotgun genomszekvenálási szekvenciaadatok. A Sanger-módszer szerinti shotgun szekvenálási protokollok első lépései lényegében nagyléptékű klónozási kísérletnek tekinthetők. A (nem *E. coli*) célgenom darabjait plazmidba ligálják, majd *E. coli* gazdába transzformálják. A transzformánsokat a lemezekről folyadékkultúrákba oltják, majd az innen kinyert DNS-t szekvenálják. A sejtek növesztése (lemezen majd folyadékban) tulajdonképpen egy biológiai teszt, mely megmutatja a gazdasejt reagálását az adott klónozott DNS-sel szemben. A szekvenciaadatok analízise felvilágosítást adhat arról, vajon vannak-e a fentiekhez hasonló esetek, azaz előfordul-e toxikus klónozott szegmensek gazda IS általi inaktiválása.

Egy tipikus bakteriális shotgun könyvtár, melyet Sanger féle szekvenáláshoz készítettek, 2- $4x10^4$ klónt tartalmaz; ezeket általában 20 generáción át szilárd táptalajon, majd további 10 generáción át folyékony kultúrákban növesztették feldolgozás előtt. Tekintettel arra, hogy az átlagos spontán transzpozíciós ráta 10^{-8} /gén/generáció [53], az esély, hogy egy spontán

kialakult inszerciós mutánst szekvenáljunk, $1.2x10^{-2}$ (= $30x4x10^{4}x10^{-8}$). Ez azt jelenti, hogy megközelítőleg 100 szekvenálási projekt közül mindössze 1 esetében találnánk olyan génszekvenciát, amelybe a klónozó *E. coli* gazda IS eleme inszertálódott. Ennél gyakoribb előfordulás már az ilyen IS-hordozó klónok erős szelekció okozta feldúsulását jelezné.

295 shotgun genom szekvenálás adatait vizsgáltuk. A sorozatok átlagosan 50-70 ezer leolvasott szekvenciából (read) állnak. Megközelítőleg 18 millió felhasznált szekvenciából 22 ezerben találtunk – szigorú kritériumok szerinti - *E. coli* IS szekvenciarészletet. Míg 166 genomi könyvtár egy vagy több *E. coli* IS-es részszekvenciát tartalmazott, addig 129 könyvtár IS szennyeződés mentesnek bizonyult. A 166 könyvtár közül csak 30 esetében állt rendelkezésre befejezett, összeállított genom – csupán ezeket vettük figyelembe. További rostálást eredményezett, hogy számos esetben az IS-szennyezés a leolvasott szekvencia olyan részére esett, amely a klónozó vektort reprezentálta (pl. *E. coli lacZ* génrészlet). Olyan könyvtárakat sem vettünk figyelembe, amelyek esetében a célgenom az *E. coli*-éhoz hasonló endogén IS elemeket tartalmazott. A szűrések után fennmaradó 14 génkönyvtárban 109 olyan esetet találtunk, ahol a klónozott darabba *E. coli* gazda IS épült (**19. ábra**) Az esetek többségében IS10 és IS1 dominált (68x és 30x), de 6 másik típusú IS is megtalálható volt (IS2, IS3, IS4, IS5, IS150, IS186).



19. ábra Példa egy jellegzetes IS inszerció találatra a shotgun adatbázis szekvenciákban (*Streptococcus agalactiae* 986965..987994 genomi szegmensének átfedő ""nyers" szekvenálási adatai). Zöld ill. kék vonalak jelzik a felső ill. alsó szállal egyező szekvenciákat. Piros vonal mutatja az egyik olvasatban a talált IS5 szekvenciát, míg a szürke sáv a régióban kódolt ABC transzporter gén elhelyezkedését mutatja.

6. Az eredmények megvitatása.

A genetikai változásokért felelős molekuláris folyamatok és a változó környezet általi automatikus szelekció együttes dinamizmusa alakítja egy bakteriális populáció genotípusát és fenotípusát. Míg hosszú távon az alkalmazkodás képessége szükséges a túléléshez, a gyakorlati, biotechnológiai/szintetikus biológiai alkalmazásokban az alkalmazkodás folyamat inkább nemkívánatos tényező [97]. A szintetikus biológiai elképzelés alapja éppen az, hogy pontos tervezéssel, szabványosított alkatrészekből, mérnöki módszerek alkalmazásával lehessen összeállítani biológiai eszközöket, lehessen gyártani biomolekulákat. A termék előállítása közben fontos, hogy az alkotóelemek interakciói tökéletesen megjósolhatóak legyenek [98], [99]. Ennek a célnak elérése érdekében hatalmas kihívás, hogy az élő sejtek belső genetikai instabilitását legyőzzük, vagy legalábbis kordában tartsuk. A biológiai termékek gyártására tervezett sejtek szinte mindig fitnesz csökkenést mutatnak a kiinduló sejtekhez képest. Ez a szelekciós nyomás mutációk felhalmozódásához vezethet, melyekkel a sejt a számára idegen, rekombináns technológiával bejuttatott exogén szekvenciák funkcionális degradációját okozhatják. E probléma egyik lehetséges megoldása a kívánt termék előállítási lépéseinek beépítése esszenciális anyagcsereutakba, így a sejt nem nélkülözheti a terméket [100], [101]. Másik megoldás lehet a sejt evolúciós potenciáljának csökkentése a mutációk létrehozásáért felelős mechanizmusok gátlásával. Ebben a dolgozatban ezzel az utóbbi lehetőséggel foglalkoztunk, különös tekintettel a mobilis IS elemek genomstabilitásban játszott szerepére. Összetett jelenségről van szó; itt csak bizonyos - elsősorban gyakorlati indíttatású - szempontokból vizsgáltuk az IS-ek meglétének és hiányának következményeit.

6. 1. Az IS-ek részesedése a mutációs folyamatokban

Az *E. coli*-ban a mutációs spektrum jelentős hányadát az IS elemek okozta genomi újrarendeződések teszik ki. Csoportunk korábbi munkái során – többek között - meghatároztuk egy szelekciós hatás alatt nem álló génben (*cycA*) mutatkozó, normál (nem stresszes) körülmények közötti mutációs eloszlást. A mérések szerint az összes mutáció 20-25 %-áért IS elemek felelősek [82], [53]. Ez az arány stressz hatására megváltozik, és a transzpozíciós események gyakorisága megnövekedhet. Ilyen stresszkörülményt akár egy rekombináns fehérje túltermelése is okozhat [102]. Munkám során a CAT gén túltermelésével okoztunk enyhe stresszt a sejtek számára. A fehérje jelenlétét ugyan jól tolerálták a sejtek,

mégis, túltermelése mintegy 2,5-szörösére emelte a *cycA* gén IS elemek okozta inaktivációjának gyakoriságát (**9. ábra**). A nem szelektált génben megnövekedett IS transzpozíció jelzi, hogy általános IS-indukcióról van szó.

Még jelentősebb volt a transzpozíciós gyakoriság megnövekedése toxikus fehérje (ORF238) termeltetése okozta stressz esetében. Míg kísérleti körülményeink között a pontmutációs ráta 1.5-szörösre növekedett, az IS-ugrás mértéke négyszeres emelkedést mutatott.

Ismeretesek olyan szituációk (pl. a szalicil szénforráshoz alkalmazkodás [50]), amikor domináns szerepe van az IS-ek generálta mutációknak a túlélés szempontjából. Az általunk bizonyított, fehérjetermeltetésre bekövetkező általános IS-transzpozíció növekedés arra utal, hogy speciális eseteken túl is jelentős szerepe van az IS-eknek a sejt evolúciójában.

A megnövekedett IS-aktivitást jelző kísérleteinkben - MG1655 sejtekben - IS specifikus PCR segítségével legalább négy különböző IS elemet mutattunk ki a cycA génben, valamint legalább hármat a *ctxvp60* génben (1. táblázat.). A különféle IS elemek eloszlása a detektált transzpozíciós eseményekben nem feltétlenül tükrözi a viszonylagos transzpozíciós aktivitásukat. Így például a szakirodalomban leírt (egy 10,000 generációs evolúciós kísérletben [57], valamint a váltakozó fagyasztás-olvasztási ciklusokkal dolgozó kísérletben [79]) genomi újrarendeződések nagy többségét az IS150 okozta. Az általunk elvégzett kísérletek során az ORF238-ban nem tapasztaltuk az IS150 dominanciáját, feltételezésünk szerint az IS150 erős célszekvencia specificitása miatt. Másrészről viszont az IS1 magas transzpozíciós rátája és lazább specifitású célszekvenciái magyarázhatják igen magas előfordulási arányát az izolált pCTXVP60 inszerciós mutánsok között. Hasonló, magas IS1 aktivitás mutatkozott az ebgR [51], bglR [50], cycA [53], tonB [80] gének vizsgálata esetében és egyes evolúciós kísérletekben [57], [79], továbbá az általunk elvégzett bioinformatikai analízisben. Az IS10 gyakori megjelenése meglepő lehet, hiszen ez nem természetes összetevője a K-12 törzseknek, az 1970-es években véletlenül került be egyes gyakorta alkalmazott klónozó gazdákba (DH10B) a Tn10 használata során [103], [104].

6. 2. A ctxvp60 toxikus gén analízise

Az állati vakcinának kifejlesztett *ctxvp60* fúziós gént saját promóter nélkül, a vektor kloramfenikol rezisztencia génjétől 3' irányban klónozták. Átíródása a CAT promóteréről

lehetséges. Ez a konstrukció rendkívül instabilnak bizonyult hagyományos, IS elemeket tartalmazó *E. coli* törzsekben. Az esetek döntő többségében a laboratóriumban használt sejtekből transzformálás után agar lemezen heterogén kolóniákat kaptunk (**5. ábra**.). A különböző fenotípusú telepek restrikciós hasítási mintázatának vizsgálata rámutatott, hogy az általuk tartalmazott plazmid is heterogén (**6. ábra**.), továbbá elhúzódó inkubáció mellett változik is. Az egészséges, gyors növekedésű kolóniákról, valamint a hosszabb inkubáció során a lassú növekedésű telepekből szektorosan kinövő, már egészséges fenotípust mutató teleprészekből izolált plazmidokon a *ctxvp60 5*' végén IS inszerciókat mutattunk ki (**7. ábra**.) Három különféle IS elemet azonosítottunk. Míg IS1 és IS3 inszercióit különböző pozíciókban és mindkét orientációban is megtaláltuk, addig IS5 csak egy, az integrációs forró pontjának "CWAR" [105] megfelelő helyen, ellenkező orientációban inszertálódott.

A pCTXVP60 plazmid változatainak részletes analízise kimutatta, hogy nem maga a fúziós gén, hanem annak egy belső, nem a tervezett leolvasási keretben expresszálódó szakasza (ORF238) jelenti a toxikus komponenst. Érdekesség, hogy ez a másodlagos leolvasási keret mesterséges manipuláció eredményeként jött létre melynek során az eredeti *ctx* és *vp60* szekvenciák kodonhasználatát növényi expresszióhoz igazították.

Az ORF238 hidrofób fehérje extrém leucin-gazdag, és predikció alapján transzmembrán doméneket hordoz. A sejtek fluoreszcens mikroszkópos vizsgálata indirekt bizonyítékot szolgáltatott a toxikus termék és a sejtmembrán kölcsönhatására: a sejtek megnyúlt alakot és aberráns osztódási képet mutattak.

IS-inszerciós mutánsok transzkripciójának RT-PCR-rel történő ellenőrzésével kimutattuk, hogy a gén upstream, 5' szakaszába ugró IS-ek blokkolják a downstream szakasz transzkripcióját, felszabadítva ezzel a sejtet a növekedésgátlás alól.

6.3. Az IS-ek közvetítette mutációk létrejöttének és érvényre jutásának dinamikája

Az IS-ek általános transzpozíciós gyakorisága *E. coli*-ban 10⁻⁸/gén/generáció [53]. Kísérleteink szerint ez az érték stresszhatásra többszörösére (4x) növekedhet [82]. Ez a gyakoriság azonban önmagában messze nem elég, hogy a populáció genotípusában/fenotípusában érzékelhető változást okozzon. Az "IS-csapda"-szerű viselkedés – az IS-mutáns plazmidokat tartalmazó sejtek gyors elszaporodása – a folyadékkultúrás kísérletekben demonstrált szelekciós folyamatokkal magyarázható: i) a plazmid terméke (ORF238) lelassítja a gazdasejt növekedését, ii) a stresszre megemelkedett, de még így is

relatíve alacsony rátájú transzpozíciós események során a megfelelő helyre véletlenszerűen beugrott IS-ek révén adott esetben gátlódik a toxikus termék képződése, iii) a gazdasejt normális növekedésre vált, és túlnövi a nem-mutáns, lassan növő sejteket.

A különböző számú és összetételű IS elemeket tartalmazó törzsek vizsgálatal azt sugallja hogy az idő, ami szükséges egy nem toxikus plazmid kialakulásához hozzávetőlegesen arányos az adott törzs IS elem tartalmával (**17. ábra**.) A csak egy IS elemet tartalmazó, de izogenikus MDS42 törzsek növekedési görbéjéből kitűnik, hogy még egy IS-kópia jelenléte is jelentős hatással van a sejt evolúciójára (**18. ábra**).

6. 4. Az IS-ek okozta mutációk speciális szerepe

Mi lehet annak oka, hogy szinte kizárólag IS-inszercióval "elrontott" plazmidokat kapunk, noha a pontmutációk általánosságban gyakoribbak, és deléciók is okozhatnának hasonló hatást? A magyarázat feltehetőleg a plazmid és az ORF238 sajátságaiban keresendő. A gén környezetében nincsenek ismétlődő szekvenciák, amelyek deléciók létrejöttét elősegítik [106], [107]. Egyes pontmutációk pedig nem képesek alapvetően megváltoztatni az ORF238 erősen hidrofób jellegét, és valószínűleg nem befolyásolnák a membrán-beépülési hajlandóságot. Frameshift mutációk elvileg elronthatnák az ORF238-at. A szekvencia azonban számos lehetséges in-frame reiniciációs helyet mutat (Shine-Dalgarno-szerű szekvenciát és startkodont), tehát egy frame-shift mutáció csak egy rövidebb szakaszt képes "elrontani".

Mindez arra utal, hogy vannak olyan genomi konfigurációk, amikor az IS elemek molekuláris "sorompóként" különösen hatásos, más mutációfajtáknál erőteljesebb hatást képesek kifejteni.

6. 5. Az *E. coli* IS elemek előfordulása a shotgun szekvenálások adatbázisaiban

A shotgun szekvenálási könyvtárakból származó nyers, feldolgozatlan szekvenciák ("read"ek) megfelelően nagy adatbázist biztosítanak annak ellenőrzésére, hogy mennyire széles körben elterjedt az IS szekvenciák megjelenése a klónozott szekvenciákban. Számításaink szerint, figyelembe véve egy tipikus shotgun könyvtár létrehozásának és feldolgozásának lépéseit és a háttér klónozó gazda IS transzpozíciós rátáját, gyakorlatilag nem várhatunk gazda IS elem megjelenést a génméretű részszekvenciákban. Ennek ellenére munkánk során a shotgun szekvenciák nagy készletének áttanulmányozása számos gazda IS elemet fedett fel a klónozott részletekben. Ez az emelkedett arány magyarázható a *ctxvp60* klónozása során is megfigyelt folyamatokkal. Azokban az esetekben, mikor a klónozott szakasz növekedési hátrányt okozott a klónozó gazda számára, a szekvenáláshoz szükséges DNS preparációhoz történő egyedi növesztés során túlsúlyba kerülhettek az IS inszerciót tartalmazó, gyorsabban növekedő mutáns változatok. Így aztán a klónozó gazda IS elemei megjelenhettek a szekvenálási adatokban.

A munkánk során szigorú kritériumokat alkalmaztunk a szennyező IS elemek azonosítására, így valószínűsíthetően alulbecsültük a valódi IS transzpozíciók számát. Az IS elemek deléciókat és újrarendeződéseket is okozhatnak – ezek az esetek meg sem jelennek az analízisben. Az IS elemmel rendelkező szakaszok előfordulási gyakorisága ill. azonosítása valószínűleg a kísérletes eljárásoktól és az adatgyűjtési protokolloktól is függhet. A rendszertanilag távol eső organizmusokból származó gének nem valószínű, hogy átíródnak E. coli gazdában, így feltehetően kisebb eséllyel okoznak növekedési rendellenességet. Ennek megfelelően az IS-inszertált mutánsaik nem jelentenek szelekciós előnyt a sejt számára. A klónok kezelésére, valamint az adatbázisok létrehozására vonatkozó információk sokszor nem voltak hozzáférhetők, vagy más esetekben inkább ellentmondásosak (nem egységes annotáció, azonos adattípusok különböző tartalomra vonatkozhatnak). Valószínűnek látszik, hogy a legtöbb adatkészlet már egy bioinformatikai előszűrésen (pl. gazda szekvenciák kiszűrése) ment keresztül, így bennük az IS transzpozíciók valódi száma rejtve maradhat. Mindent összevetve, még ezeknek a tényezőknek az elhanyagolása után is jelentős számú IS transzpozíciós esemény mutatható ki, s ez arra figyelmeztet, hogy klónozással járó kísérletekben feldúsulhatnak az IS-ek okozta műtermékek.

6. 6. IS elem: a gazdasejt védekező rendszere?

Az eredmények fényében hipotézisként felmerülhet az az elképzelés, hogy az IS elemek – túl az önző genetikai elem szerepen és az alkalmankénti, a gazdasejt számára előnyös közreműködésen - egyfajta, a horizontális géntranszfer negatív következményeivel szembeni védelmi rendszer részei lehetnek. Az attraktív elképzelés ellen szól ugyanakkor, egy ilyen központi, IS-alapú, a gazdasejt integritására is lehetséges veszélyt jelentő rendszer fenntartása valószínűleg nem gazdaságos a sejt számára. Továbbá, a mennyiségi/populációs viszonyokat

figyelembe véve (rendszerint nagyszámú sejt van a populációban, közöttük csak egy kis töredék szenved egy adott szituációban horizontális géntranszfert) egyszerűbb "megoldásnak" tűnik, hogy a horizontális géntranszfer során hátrányba került sejtek egyszerűen kivesznek a populációból.

6. 7. Gyakorlati következtetések

Az extrém instabil pCTXVP60 plazmid klónozhatatlanságát vizsgálva megállapítottuk, hogy egy, a gazdasejtben expresszálódó, toxikus klón igen gyorsan, standard laboratóriumi növesztési intervallumban is mutációt szenvedhet és ez a mutáns változat dominánssá válhat. Arra is fény derült, hogy vannak olyan konstrukciók, ahol az IS elemek speciális, robusztus hatásuk révén elsődleges szerepet játszanak az inaktiváló mutációk létrehozásában. Egy másik, részletes elemzés kimutatta, hogy a probléma nem elhanyagolható masszív klónozási kísérletekben. Irodalmi és anekdotikus információk alapján hasznos molekulák *E. coli*-val történő termeltetése során is gondot okoz az IS elemek mutagén hatása. Megállapítható, hogy biotechnológiai/szintetikus biológiai alkalmazásokban jelentős hibaforrást képeznek a gazda IS elemei.

Mindez rámutat a redukált genomú, egyszerűsített *E. coli* sejtek gyakorlati alkalmazásának egyik lényeges előnyére: az MDS42 törzs mentes az IS elemektől [108]. Az MDS42 törzset és különféle változatait együttműködő partnerünk, a Scarab Genomics forgalmazza. Egyik jelentős felhasználója a vezető állati vakcinagyártó Merial cég, mely egyes DNS-vakcinái stabilitását nem tudta megoldani más módon, csak az MDS42 alkalmazásával

Köszönetnyilvánítás:

Szeretném megköszönni témavezetőmnek, Dr. Pósfai Györgynek szakmai irányítását és hasznos tanácsait, a szövegszerkesztésben nyújtott segítségéért és a kézirat lelkiismeretes átolvasásáért. Ötleteivel és értékes hozzászólásaival nagyban hozzájárult ahhoz, hogy a dolgozat a jelenlegi formában létrejöjjön.

Hálás köszönet illeti Szalkanovics Gábornét, Ágit, aki mindvégig önzetlenül segített a labormunkában.

Köszönettel tartozom a Genommérnöki Csoport minden jelenlegi és volt tagjának (Balikó Gabriella, Csörgő Bálint, Draskovits Gábor, Fehér Tamás, Győrfy Zsuzsa, Karcagi Ildikó, Tímár Edit, Vitaly Kolisnycenko) munkám során nyújtott segítségért. Külön köszönetet mondok Fehér Tamásnak, aki mindig türelmet tanúsított irántam, és mindenben a segítségemre volt. Hálás vagyok Balikó Gabriellának, munkámhoz nyújtott óriási segítségéért és szakmai tanácsaiért.

Köszönet illeti Dr Ferhan Ayaydint a konfokális lézer pásztázó mikroszkópos munkákban nyújtott segítségért.

Köszönettel tartozom Dr Pósfai Jánosnak a shotgun szekvenálások adatbázisainak bioinformatikai tanulmányozásában nyújtott elengedhetetlen segítségéért.

Köszönetet mondok Frederik R Blattnernek, Guy Plunkettnek III, David Frischnek, a Scarab Genomicstól (Madison, WI, USA) a munkánkban nyújtot elengedhetetlen segítségükért.

Külön köszönet a pCTXVP60 plazmidnak a folyamatos szakmai és technikai kihívásért.

Szakirodalomi hivatkozások

[1] F.C. Neidhardt, *Escherichia Coli* and *Salmonella:* Cellular and Molecular, American Society Microbiology, 1999.

[2] J LEDERBERG, *Genetic studies with bacteria*, Genetics in the 20th century, 1951.

[3] F.R. Blattner, G. Plunkett, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, and Y. Shao, "The Complete Genome Sequence of *Escherichia coli* K-12," *Science*, vol. 277, Sep. 1997, pp. 1453-1462.

[4] J.L. Ingraham and F.C. Neidhardt, *Escherichia Coli & Salmonella Typhimurium:* Cellular & Molecular Biology, ASM Press, 1987.

[5] J.L. Reed, T.D. Vo, C.H. Schilling, and B.O. Palsson, "An expanded genomescale model of *Escherichia coli* K-12 (iJR904 GSM/GPR)," *Genome Biology*, vol. 4, 2003, p. R54.

[6] A.M. Feist, C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, and B.Ø. Palsson, "A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information," *Molecular Systems Biology*, vol. 3, 2007.

[7] A. Matsuyama, H. Yamamoto, N. Kawada, and Y. Kobayashi, "Industrial production of (R)-1,3-butanediol by new biocatalysts," *Journal of Molecular Catalysis B: Enzymatic*, vol. 11, Jan. 2001, pp. 513-521.

[8] R.A. Welch, V. Burland, G. Plunkett, P. Redford, P. Roesch, D. Rasko, E.L. Buckles, S. Liou, A. Boutin, J. Hackett, D. Stroud, G.F. Mayhew, D.J. Rose, S. Zhou, D.C. Schwartz, N.T. Perna, H.L.T. Mobley, M.S. Donnenberg, and F.R. Blattner, "Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, Dec. 2002, pp. 17020 -17024.

[9] N.T. Perna, G. Plunkett, V. Burland, B. Mau, J.D. Glasner, D.J. Rose, G.F. Mayhew, P.S. Evans, J. Gregor, H.A. Kirkpatrick, G. Pósfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E.J. Grotbeck, N.W. Davis, A. Lim, E.T. Dimalanta, K.D. Potamousis, J. Apodaca, T.S. Anantharaman, J. Lin, G. Yen, D.C. Schwartz, R.A. Welch, and F.R. Blattner, "Genome sequence of enterohaemorrhagic *Escherichia coli*

O157:H7," Nature, vol. 409, 2001, pp. 529-533.

[10] D.A. Rasko, M.J. Rosovitz, G.S.A. Myers, E.F. Mongodin, W.F. Fricke, P. Gajer, J. Crabtree, M. Sebaihia, N.R. Thomson, R. Chaudhuri, I.R. Henderson, V. Sperandio, and J. Ravel, "The Pangenome Structure of *Escherichia coli*: Comparative Genomic Analysis of *E. coli* Commensal and Pathogenic Isolates," *J. Bacteriol.*, vol. 190, Oct. 2008, pp. 6881-6893.

[11] J. Hacker and E. Carniel, "Ecological fitness, genomic islands and bacterial pathogenicity," *EMBO Reports*, vol. 2, May. 2001, pp. 376-381.

[12] M. Juhas, J.R. Van Der Meer, M. Gaillard, R.M. Harding, D.W. Hood, and D.W. Crook, "Genomic islands: tools of bacterial horizontal gene transfer and evolution," *FEMS Microbiology Reviews*, vol. 33, 2009, pp. 376-393.

[13] M. Touchon, C. Hoede, O. Tenaillon, V. Barbe, S. Baeriswyl, P. Bidet, E. Bingen, S. Bonacorsi, C. Bouchier, O. Bouvet, A. Calteau, H. Chiapello, O. Clermont, S. Cruveiller, A. Danchin, M. Diard, C. Dossat, M.E. Karoui, E. Frapy, L. Garry, J.M. Ghigo, A.M. Gilles, J. Johnson, C. Le Bouguénec, M. Lescat, S. Mangenot, V. Martinez-Jéhanne, I. Matic, X. Nassif, S. Oztas, M.A. Petit, C. Pichon, Z. Rouy, C.S. Ruf, D. Schneider, J. Tourret, B. Vacherie, D. Vallenet, C. Médigue, E.P.C. Rocha, and E. Denamur, "Organised Genome Dynamics in the *Escherichia coli* Species Results in Highly Diverse Adaptive Paths," *PLoS Genetics*, vol. 5, Jan. 2009.

[14] U. Dobrindt, M. Chowdary, G. Krumbholz, and J. Hacker, "Genome dynamics and its impact on evolution of *Escherichia coli*," *Medical Microbiology and Immunology*, Jun. 2010.

[15] H. Tettelin, V. Masignani, M.J. Cieslewicz, C. Donati, D. Medini, N.L. Ward, S.V. Angiuoli, J. Crabtree, A.L. Jones, A.S. Durkin, R.T. DeBoy, T.M. Davidsen, M. Mora, M. Scarselli, I. Margarit y Ros, J.D. Peterson, C.R. Hauser, J.P. Sundaram, W.C. Nelson, R. Madupu, L.M. Brinkac, R.J. Dodson, M.J. Rosovitz, S.A. Sullivan, S.C. Daugherty, D.H. Haft, J. Selengut, M.L. Gwinn, L. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K.J.B. O'Connor, S. Smith, T.R. Utterback, O. White, C.E. Rubens, G. Grandi, L.C. Madoff, D.L. Kasper, J.L. Telford, M.R. Wessels, R. Rappuoli, and C.M. Fraser, "Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome"," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, 2005, pp. 13950 -13955.

[16] O. Lukjancenko, T. Wassenaar, and D. Ussery, "Comparison of 61 Sequenced

Escherichia coli Genomes," Microbial Ecology.

[17] H. Jeong, V. Barbe, C.H. Lee, D. Vallenet, D.S. Yu, S. Choi, A. Couloux, S. Lee, S.H. Yoon, L. Cattolico, C. Hur, H. Park, B. Ségurens, S.C. Kim, T.K. Oh, R.E. Lenski, F.W. Studier, P. Daegelen, and J.F. Kim, "Genome Sequences of *Escherichia coli* B strains REL606 and BL21(DE3)," *Journal of Molecular Biology*, vol. 394, Dec. 2009, pp. 644-652.

[18] K.B. Zeldovich, P. Chen, and E.I. Shakhnovich, "Protein stability imposes limits on organism complexity and speed of molecular evolution," *Proceedings of the National Academy of Sciences*, vol. 104, Oct. 2007, pp. 16152-16157.

[19] M. Lynch, "Evolution of the mutation rate," *Trends in Genetics*, vol. 26, Aug. 2010, pp. 345-352.

[20] B.A. Bridges, "Hypermutation in bacteria and other cellular systems," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 356, Jan. 2001, pp. 29-39.

[21] D.E. Berg and M.M. Howe, *Mobile DNA*, ASM Press, 1989.

[22] E. Galun, Transposable elements: a guide to the perplexed and the novice : with appendices on RNAi, chromatin remodeling and gene tagging, Springer, 2003.

[23] B. McClintock, "The Origin and Behavior of Mutable Loci in Maize," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 36, Jun. 1950, pp. 344-355.

[24] P. Siguier, J. Perochon, L. Lestrade, J. Mahillon, and M. Chandler, "ISfinder: the reference centre for bacterial insertion sequences," *Nucl. Acids Res.*, vol. 34, Jan. 2006, pp. D32-36.

[25] Y. Sekine and E. Ohtsubo, "DNA sequences required for translational frameshifting in production of the transposase encoded by IS 1," *Molecular and General Genetics MGG*, vol. 235, Nov. 1992, pp. 325-332.

[26] P.J. Farabaugh, "Programmed Translational Frameshifting," *Annual Review of Genetics*, vol. 30, 1996, pp. 507-528.

[27] M.J. Calcutt, J.L. Lavrrar, and K.S. Wise, "IS1630 of *Mycoplasma fermentans*, a Novel IS30-Type Insertion Element That Targets and Duplicates Inverted Repeats of Variable Length and Sequence during Insertion," *J. Bacteriol.*, vol. 181, Dec. 1999, pp. 7597-7607.

[28] B.B. Plikaytis, J.T. Crawford, and T.M. Shinnick, "IS1549 from *Mycobacterium smegmatis* Forms Long Direct Repeats upon Insertion," *J. Bacteriol.*, vol. 180, Mar.

1998, pp. 1037-1043.

[29] T. Binnewies, Y. Motro, P. Hallin, O. Lund, D. Dunn, T. La, D. Hampson, M. Bellgard, T. Wassenaar, and D. Ussery, "Ten years of bacterial genome sequencing: comparative-genomics-based discoveries," *Functional & Integrative Genomics*, vol. 6, Jul. 2006, pp. 165-185.

[30] J. Kulkosky, K.S. Jones, R.A. Katz, J.P. Mack, and A.M. Skalka, "Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases.," *Mol. Cell. Biol.*, vol. 12, May. 1992, pp. 2331-2338.

[31] Z. Nagy and M. Chandler, "Regulation of transposition in bacteria," *Research in Microbiology*, vol. 155, Jun. 2004, pp. 387-398.

[32] H.J. Reif and H. Saedler, "IS1 is involved in deletion formation in the gal region of *E. coli* K12," *Molecular and General Genetics MGG*, vol. 137, Mar. 1975, pp. 17-28.

[33] B. Del Re, F. Garoia, P. Mesirca, C. Agostini, F. Bersani, and G. Giorgi, "Extremely low frequency magnetic fields affect transposition activity in *Escherichia coli*," *Radiation and Environmental Biophysics*, vol. 42, Jul. 2003, pp. 113-118.

[34] Z. Eichenbaum and Z. Livneh, "UV light induces IS10 transposition in *Escherichia coli.*," *Genetics*, vol. 149, Jul. 1998, pp. 1173-1181.

[35] N.L. Craig, Mobile DNA II, ASM Press, 2002.

[36] V.N. Danilevich and D.A. Kostiuchenko, "[Immunity to repeated transposition of the insertion sequence IS21]," *Molekuliarnaia Biologiia*, vol. 19, Oct. 1985, pp. 1242-1250.

[37] N.L. Craig, "TARGET SITE SELECTION IN TRANSPOSITION," *Annual Review of Biochemistry*, vol. 66, 1997, pp. 437-474.

[38] B.A. Castilho and M.J. Casadaban, "Specificity of mini-Mu bacteriophage insertions in a small plasmid.," *J. Bacteriol.*, vol. 173, Feb. 1991, pp. 1339-1343.

[39] J. Mahillon and M. Chandler, "Insertion Sequences," *Microbiol. Mol. Biol. Rev.*, vol. 62, Sep. 1998, pp. 725-774.

[40] M.V. Mendiola and F. de la Cruz, "Specificity of insertion of IS91, an insertion sequence present in alpha-haemolysin plasmids of *Escherichia coli*," *Molecular Microbiology*, vol. 3, Jul. 1989, pp. 979-984.

[41] C. Sengstag, S. Iida, R. Hiestand-Nauer, and W. Arber, "Terminal inverted repeats of prokaryotic transposable element IS186 which can generate duplications of

variable length at an identical target sequence," Gene, vol. 49, 1986, pp. 153-156.

[42] J. Meyer, S. Iida, and W. Arber, "Does the insertion element IS1 transpose preferentially into A+T-rich DNA segments?," *Molecular and General Genetics MGG*, vol. 178, May. 1980, pp. 471-473.

[43] F. Kunst, N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo,
M.G. Bertero, P. Bessières, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans,
M. Braun, S.C. Brignell, S. Bron, S. Brouillet, C.V. Bruschi, B. Caldwell, V. Capuano,
N.M. Carter, S.K. Choi, J.J. Codani, I.F. Connerton, and A. Danchin, "The complete
genome sequence of the gram-positive bacterium *Bacillus subtilis*," *Nature*, vol. 390,
Nov. 1997, pp. 249-256.

[44] T. Nagai, L.S. Phan Tran, Y. Inatsu, and Y. Itoh, "A new IS4 family insertion sequence, IS4Bsu1, responsible for genetic instability of poly-gamma-glutamic acid production in *Bacillus subtilis*," *Journal of Bacteriology*, vol. 182, May. 2000, pp. 2387-2392.

[45] P. Siguier, J. Filée, and M. Chandler, "Insertion sequences in prokaryotic genomes," *Current Opinion in Microbiology*, vol. 9, Oct. 2006, pp. 526-531.

[46] T. Naas, M. Blot, W. Fitch, and W. Arber, "Dynamics of IS-related genetic rearrangements in resting *Escherichia coli* K-12," *Mol Biol Evol*, vol. 12, Mar. 1995, pp. 198-207.

[47] K. Kanbashi, X. Wang, J. Komura, T. Ono, and K. Yamamoto, "Frameshifts, base substitutions and minute deletions constitute X-ray-induced mutations in the endogenous *tonB* gene of *Escherichia coli* K12," *Mutation Research/DNA Repair*, vol. 385, Dec. 1997, pp. 259-267.

[48] H. Rodriguez, E.T. Snow, U. Bhat, and E.L. Loechler, "An *Escherichia coli* plasmid-based, mutational system in which *supF* mutants are selectable: Insertion elements dominate the spontaneous spectra," *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 270, Nov. 1992, pp. 219-231.

[49] J.A. Halliday and B.W. Glickman, "Mechanisms of spontaneous mutation in DNA repair-proficient *Escherichia coli*," *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 250, Sep. , pp. 55-71.

[50] B. Hall, "Activation of the bgl operon by adaptive mutation," *Mol Biol Evol*, vol. 15, Jan. 1998, pp. 1-5.

[51] B.G. Hall, "Spectra of Spontaneous Growth-Dependent and Adaptive Mutations at *ebgR*," *J. Bacteriol.*, vol. 181, Feb. 1999, pp. 1149-1155.

[52] A.E. Reynolds, J. Felton, and A. Wright, "Insertion of DNA activates the cryptic bgl operon in *E. coli* K12," *Nature*, vol. 293, 1981, pp. 625-629.

[53] T. Fehér, B. Cseh, K. Umenhoffer, I. Karcagi, and G. Pósfai, "Characterization of *cycA* mutants of *Escherichia coli*: An assay for measuring in vivo mutation rates," *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, vol. 595, Mar. 2006, pp. 184-190.

[54] D.M. Stoebel, K. Hokamp, M.S. Last, and C.J. Dorman, "Compensatory Evolution of Gene Regulation in Response to Stress by *Escherichia coli* Lacking RpoS," *PLoS Genet*, vol. 5, Oct. 2009, p. e1000671.

[55] C. Petersen, L.B. Møller, and P. Valentin-Hansen, "The Cryptic Adenine Deaminase Gene of *Escherichia coli:* SILENCING BY THE NUCLEOID-ASSOCIATED DNA-BINDING PROTEIN, H-NS, AND ACTIVATION BY INSERTION ELEMENTS," *Journal of Biological Chemistry*, vol. 277, 2002, pp. 31373-31380.

[56] D. Schneider and R.E. Lenski, "Dynamics of insertion sequence elements during experimental evolution of bacteria," *Research in Microbiology*, vol. 155, Jun. 2004, pp. 319-327.

[57] D. Schneider, E. Duperchy, E. Coursange, R.E. Lenski, and M. Blot, "Longterm experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements.," *Genetics*, vol. 156, Oct. 2000, pp. 477-488.

[58] H. Saedler, H.J. Reif, S. Hu, and N. Davidson, "IS2, a genetic element for turnoff and turn-on of gene activity in *E. coli*," *Molecular and General Genetics MGG*, vol. 132, Dec. 1974, pp. 265-289.

[59] J.G. Lawrence and H. Ochman, "Molecular archaeology of the *Escherichia coli* genome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, 1998, pp. 9413-9417.

[60] D. Botstein and D. Shortle, "Strategies and applications of in vitro mutagenesis," *Science*, vol. 229, 1985, pp. 1193-1201.

[61] K. Schnetz, C. Toloczyki, and B. Rak, "Beta-glucoside (bgl) operon of *Escherichia coli* K-12: nucleotide sequence, genetic organization, and possible evolutionary relationship to regulatory components of two *Bacillus subtilis* genes.," *J. Bacteriol.*, vol. 169, Jun. 1987, pp. 2579-2590.

[62] C.S. Barker, B.M. Pruss, and P. Matsumura, "Increased Motility of Escherichia

coli by Insertion Sequence Element Integration into the Regulatory Region of the flhD Operon," *J. Bacteriol.*, vol. 186, Nov. 2004, pp. 7529-7537.

[63] W.F. Doolittle and C. Sapienza, "Selfish genes, the phenotype paradigm and genome evolution," *Nature*, vol. 284, Apr. 1980, pp. 601-603.

[64] L.E. Orgel and F.H. Crick, "Selfish DNA: the ultimate parasite," *Nature*, vol. 284, Apr. 1980, pp. 604-607.

[65] A. Wagner, C. Lewis, and M. Bichsel, "A survey of bacterial insertion sequences using IScan," *Nucl. Acids Res.*, vol. 35, Aug. 2007, pp. 5284-5293.

[66] M. Blot, "Transposable elements and adaptation of host bacteria," *Genetica*, vol. 93, Feb. 1994, pp. 5-12.

[67] R.M. Blumenthal, S.A. Gregory, and J.S. Cooperider, "Cloning of a restrictionmodification system from *Proteus vulgaris* and its use in analyzing a methylasesensitive phenotype in *Escherichia coli.*," *J. Bacteriol.*, vol. 164, Nov. 1985, pp. 501-509.

[68] K. Nakamura and M. Inouye, "Inactivation of the *Serratia marcescens* gene for the lipoprotein in *Escherichia coli* by insertion sequences, IS1 and IS5; sequence analysis of junction points," *Molecular and General Genetics MGG*, vol. 183, 1981, pp. 107-114.

[69] J. Muller, H. Reinert, and H. Malke, "Streptokinase mutations relieving *Escherichia coli* K-12 (prlA4) of detriments caused by the wild-type *skc* gene.," *J. Bacteriol.*, vol. 171, Apr. 1989, pp. 2202-2208.

[70] P. Rawat, S. Kumar, D. Pental, and P. Burma, "Inactivation of a transgene due to transposition of insertion sequence (IS136) of *Agrobacterium tumefaciens*," *Journal of Biosciences*, vol. 34, Jun. 2009, pp. 199-202.

[71] K. Prather, M. Edmonds, and J. Herod, "Identification and characterization of IS1 transposition in plasmid amplification mutants of *E. coli* clones producing DNA vaccines," *Applied Microbiology and Biotechnology*, vol. 73, Dec. 2006, pp. 815-826.

[72] C. Fernandez, D. Larhammar, B. Servenius, L. Rask, and P.A. Peterson, "Spontaneous insertions into cosmid vector: a warning," *Gene*, vol. 42, 1986, pp. 215-219.

[73] M. Binns, "Contamination of DNA database sequence entries with *Escherichia coli* insertion sequences.," *Nucleic Acids Research*, vol. 21, Feb. 1993, p. 779.

[74] Kovařík, Matzke, Matzke, and B. Koukalová, "Transposition of IS10 from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts," *Molecular*

Genetics and Genomics, vol. 266, Oct. 2001, pp. 216-222.

[75] G. Astua-Monge, A. Lyznik, V. Jones, S.A. Mackenzie, and C.E. Vallejos, "Evidence for a prokaryotic insertion-sequence contamination in eukaryotic sequences registered in different databases," *TAG Theoretical and Applied Genetics*, vol. 104, Jan. 2002, pp. 48-53.

[76] D.G. Gibson, J.I. Glass, C. Lartigue, V.N. Noskov, R. Chuang, M.A. Algire,
G.A. Benders, M.G. Montague, L. Ma, M.M. Moodie, C. Merryman, S. Vashee, R.
Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E.A. Denisova, L. Young, Z.
Qi, T.H. Segall-Shapiro, C.H. Calvey, P.P. Parmar, C.A. Hutchison, H.O. Smith, and
J.C. Venter, "Creation of a Bacterial Cell Controlled by a Chemically Synthesized
Genome," *Science*, vol. 329, Jul. 2010, pp. 52-56.

[77] I. Szeverényi, A. Hodel, W. Arber, and F. Olasz, "Vector for IS element entrapment and functional characterization based on turning on expression of distal promoterless genes," *Gene*, vol. 174, 1996, pp. 103–110.

[78] T. Naas, M. Blot, W.M. Fitch, and W. Arber, "Insertion Sequence-Related Genetic Variation in Resting *Escherichia coli* K-12," *Genetics*, vol. 136, Mar. 1994, pp. 721-730.

[79] S.C. Sleight, C. Orlic, D. Schneider, and R.E. Lenski, "Genetic Basis of Evolutionary Adaptation by *Escherichia coli* to Stressful Cycles of Freezing, Thawing and Growth," *Genetics*, vol. 180, Sep. 2008, pp. 431-443.

[80] K. Kitamura, Y. Torii, C. Matsuoka, and K. Yamamoto, "DNA sequence changes in mutations in the *tonB* gene on the chromosome of *Escherichia coli* K12: insertion elements dominate the spontaneous spectra," *Idengaku Zasshi*, vol. 70, Feb. 1995, pp. 35-46.

[81] V. Kolisnychenko, G. Plunkett, C.D. Herring, T. Fehér, J. Pósfai, F.R. Blattner, and G. Pósfai, "Engineering a Reduced *Escherichia coli* Genome," *Genome Research*, vol. 12, Apr. 2002, pp. 640-647.

[82] G. Posfai, G. Plunkett, T. Feher, D. Frisch, G.M. Keil, K. Umenhoffer, V. Kolisnychenko, B. Stahl, S.S. Sharma, M. de Arruda, V. Burland, S.W. Harcum, and F.R. Blattner, "Emergent Properties of Reduced-Genome *Escherichia coli*," *Science*, vol. 312, May. 2006, pp. 1044-1046.

[83] G. Meyers, C. Wirblich, and H. Thiel, "Rabbit hemorrhagic disease virus-molecular cloning and nucleotide sequencing of a calicivirus genome," *Virology*, vol. 184, Oct. 1991, pp. 664-676. [84] H. Mikschofsky, H. Schirrmeier, G.M. Keil, B. Lange, P.L. Polowick, W. Keller, and I. Broer, "Pea-derived vaccines demonstrate high immunogenicity and protection in rabbits against rabbit haemorrhagic disease virus," *Plant Biotechnology Journal*, vol. 7, 2009, pp. 537-549.

[85] Sambrook J, Fritsch, EF, Maniatis, T, *Molecular cloning: a laboratory manual.*, Cold Spring Harbor Laboratory Press, New York, 1989.

[86] J. Wild, Z. Hradecná, G. Pósfai, and W. Szybalski, "A broad-host-range in vivo pop-out and amplification system for generating large quantities of 50-to 100-kb genomic fragments for direct DNA sequencing," *Gene*, vol. 179, 1996, pp. 181-188.

[87] R.J. Wargel, C.A. Shadur, and F.C. Neuhaus, "Mechanism of D-Cycloserine Action: Transport Mutants for D-Alanine, D-Cycloserine, and Glycine," *J. Bacteriol.*, vol. 105, Mar. 1971, pp. 1028-1035.

[88] R.P. Holmes and R.R.B. Russell, "Mutations Affecting Amino Sugar Metabolism in *Escherichia coli* K-12," *J. Bacteriol.*, vol. 111, Jul. 1972, pp. 290-291.

[89] S. Sarkar, W.T. Ma, and G.V.H. Sandri, "On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants," *Genetica*, vol. 85, Jan. 1992, pp. 173-179.

[90] F.M. Stewart, D.M. Gordon, and B.R. Levin, "Fluctuation Analysis: The Probability Distribution of the Number of Mutants Under Different Conditions," *Genetics*, vol. 124, Jan. 1990, pp. 175-185.

[91] K. Umenhoffer, T. Feher, G. Baliko, F. Ayaydin, J. Posfai, F. Blattner, and G. Posfai, "Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications," *Microbial Cell Factories*, vol. 9, 2010, p. 38.

[92] N.A. Burgess-Brown, S. Sharma, F. Sobott, C. Loenarz, U. Oppermann, and O. Gileadi, "Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study," *Protein Expression and Purification*, vol. 59, May. 2008, pp. 94-102.

[93] J.F. Kane, "Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*," *Current Opinion in Biotechnology*, vol. 6, 1995, pp. 494-500.

[94] G.E. Tusnady and I. Simon, "The HMMTOP transmembrane topology prediction server," *Bioinformatics*, vol. 17, Sep. 2001, pp. 849-850.

[95] M. Cserzo, F. Eisenhaber, B. Eisenhaber, and I. Simon, "On filtering false

55

positive transmembrane protein predictions," *Protein Eng.*, vol. 15, Sep. 2002, pp. 745-752.

[96] G.E. Tusnady, L. Kalmar, H. Hegyi, P. Tompa, and I. Simon, "TOPDOM: database of domains and motifs with conservative location in transmembrane proteins," *Bioinformatics*, vol. 24, Jun. 2008, pp. 1469-1470.

[97] J.D. Keasling, "Synthetic Biology for Synthetic Chemistry," ACS Chemical Biology, vol. 3, Jan. 2008, pp. 64-76.

[98] M. Heinemann and S. Panke, "Synthetic biology--putting engineering into biology," *Bioinformatics*, vol. 22, Nov. 2006, pp. 2790-2799.

[99] B. Canton, A. Labno, and D. Endy, "Refinement and standardization of synthetic biological parts and devices," *Nat Biotech*, vol. 26, Jul. 2008, pp. 787-793.

[100] P. Pharkya, A.P. Burgard, and C.D. Maranas, "Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock," *Biotechnology and Bioengineering*, vol. 84, 2003, pp. 887-899.

[101] C.T. Trinh, P. Unrean, and F. Srienc, "Minimal *Escherichia coli* Cell for the Most Efficient Production of Ethanol from Hexoses and Pentoses," *Appl. Environ. Microbiol.*, vol. 74, Jun. 2008, pp. 3634-3643.

[102] F.T. Haddadin and S.W. Harcum, "Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*," *Biotechnology and Bioengineering*, vol. 90, 2005, pp. 127-153.

[103] M. Ayele, R.A. Gibbs, B. Csorgo, G. Posfai, G.M. Weinstock, F.R. Blattner, T. Durfee, R. Nelson, S. Baldwin, G. Plunkett, V. Burland, B. Mau, J.F. Petrosino, X. Qin, and D.M. Muzny, "The Complete Genome Sequence of *Escherichia coli* DH10B: Insights into the Biology of a Laboratory Workhorse," *J. Bacteriol.*, vol. 190, Apr. 2008, pp. 2597-2606.

[104] J. Bender, J. Kuo, and N. Kleckner, "Genetic Evidence Against Intramolecular Rejoining of the Donor DNA Molecule Following IS10 Transposition," *Genetics*, vol. 128, Aug. 1991, pp. 687-694.

[105] J.A. Engler and M.P. van Bree, "The nucleotide sequence and protein-coding capability of the transposable element IS5," *Gene*, vol. 14, Aug. 1981, pp. 155-163.

[106] G.L. Dianov, A.V. Kuzminov, A.V. Mazin, and R.I. Salganik, "Molecular mechanisms of deletion formation in *Escherichia coli* plasmids," *Molecular and General Genetics MGG*, vol. 228, 1991, pp. 153-159.

[107] A. Albertini, "On the formation of spontaneous deletions: The importance of

short sequence homologies in the generation of large deletions," *Cell*, vol. 29, 1982, pp. 319-328.

[108] C. Chakiath and D. Esposito, "Improved recombinational stability of lentiviral expression vectors using reduced-genome *Escherichia coli*," *BioTechniques*, vol. 43, 2007, pp. 466-470.

Saját publikációk jegyzéke

Umenhoffer K, Feher T, Baliko G, Ayaydin F, Posfai J, Blattner FR, Posfai G Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. MICROB CELL FACT 9: 10.1186/1475-2859-9-38- p. (2010) IF: 3.432

 Feher T, Karcagi I, Gyorfy Z, Umenhoffer K, Csorgo B, Posfai G

 Scarless engineering of the *Escherichia coli* genome

 In:
 Gerdes
 S,
 Osterman
 AL
 (szerk.)

 Microbial Gene Essentiality:
 Protocols and Bioinformatics:
 Methods in Molecular Biology

 series.
 Vol. 416..

TOTOWA, USA: HUMANA PRESS INC., 2008. pp. 251-259.

Posfai G, Plunkett G, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR Emergent properties of reduced-genome *Escherichia coli* SCIENCE 312: 1044-1046 (2006) IF: 30.927

Feher T, Cseh B, Umenhoffer K, Karcagi I, Posfai G
Characterization of cycA mutants of *Escherichia coli* - An assay for measuring in vivo mutation rates
MUTAT RES-FUND MOL M 595: 184-190 (2006)
IF: 3.340

Summary

Escherichia coli K-12 is one of the best understood and most thoroughly analyzed organisms. *Escherichia coli* K-12 and O157:H7 were amongst the first bacteria chosen for whole genome sequencing. By far the highest number of individual genome sequences is available for this species, providing us an extraordinarily useful tool for genomic comparisons. Moreover, *Escherichia coli* is an ideal model for studies of processes involved in bacterial genome evolution.

Genomes of *Enterobatericeae* in general, and of *Escherichia coli* strains in particular are now recognized to be mosaics in which a backbone of conserved genes in conserved order is interspersed with strain-specific horizontally transmitted "islands". Our group`s earlier goal was to construct an improved *Escherichia coli* by using scarless genomic reduction of the sequenced K-12 strain MG1655 to physically realize a bacterium whose genome consists of the backbone elements common to most *E. coli*.

E. coli is littered with genes that mediate horizontal gene transfer, including IS (insertion sequence) elements, transposases, defective phages, integrases, and site specific recombinases. Amongst them IS elements are the small, genetically compact units, which generally encode no functions other than those involved in their mobility. These are "selfish" DNA elements, which can translocate, duplicate, and be maintained in the genome even if detrimental to the host, behaving in some sense like an infectious agent. In a typical bacterial cell, IS elements generate a significant share of genetic variation. IS elements, when present in more than one copy, also provide repeats that lead to inversions, duplications, and deletions mediated by homologous recombination. Genome alterations due to IS translocation occur surprisingly frequently and many commonly used laboratory strains have unrecognized genome alterations because of this. IS hopping from chromosome to plasmids is also common, leading to difficulties in cloning, expression, and library construction. IS elements can also pass from strain to strain during laboratory manipulations. Having a unique IS free Escherichia coli (MDS42) strain in our hand, we aimed to investigate the role of IS elements on genome evolution and their impact during synthetic biological-biotechnical applications. Our plan was to test whether IS-mediated instability of cloned sequences could be more widespread than anticipated.

During our study we used an extremely mutable "IS magnetlike" plasmid. A recombinant ectopic gene, constructed for vaccine development, and composed of a synthetic

gene of the structural capsid protein VP60 of rabbit haemorrhagic disease (RHD) virus fused to the gene of the B subunit of cholera toxin (CTX) was very unstable in conventional *E. coli* hosts, while *ctx* and *vp60* individually, both were stable.

Upon transformation by plasmid pCTXVP60, heterogeneous colonies of transformants were obtained on agar plates (small, slow growing, and normal growth). A vast majority, 92 %, of the large colonies carried plasmids with IS insertions, while 8 % displayed deletions or no alteration detectable by PCR. In the case of IS insertions, we detected IS1, IS3, and IS5 translocations, all of them occurring in the 5' end of *ctxvp60*. In contrast, the recombinant plasmid was stable in IS-less MDS42. Transformants in MDS42 showed homogenic colonies, plasmid DNA recovered yielded unaltered restriction digestion pattern and nucleotide sequence. To understand the causes of the host-dependent stability of the plasmid, we've planned to investigate the molecular details of this behaviour.

To determine whether cloning of pCTXVP60 increases the IS-mediated mutation rate, bystander mutation tests were performed, detecting mutations in a neutral, chromosomal gene (*cycA*). The appearance of *cycA* mutant (d-cycloserine resistant) colonies was counted in MG1655 cells carrying pCTX (control, stable plasmid), and in cells carrying the pCTXVP60 plasmid. The overall mutation rate increased 2-fold in MG1655 carrying pCTXVP60, compared to MG1655 carrying pCTX. However, the IS transposition rate was elevated 4-fold. Four different IS elements contributed to this elevated transposition rate. Furthermore, 32 % of the chromosomal *cycA* IS mutants also carried an IS in pCTXVP60. This significantly high proportion of double IS mutants can not be caused by independent single transposition events. Our hypothesis is that stress-induced IS transposition, combined with fast selection of IS-inserted mutants could provide an explanation.

To test whether protein overexpression *per se* increases the frequency of IS transposition in growing cells, a culture of MG1655 carrying pPROEX HT-CAT (expressing a well-tolerated protein, chloramphenicol-acetyltramsferase /CAT/) was divided into two portions, and the frequency and types of mutations were studied in 20-tube fluctuation tests with or without induction of CAT expression. We found that the number of transpositions into *cycA* was 136% higher in expression-induced cells. Point mutation frequencies remained virtually unchanged. We conclude that overexpression of even a well-tolerated protein leads to elevated IS transposition.

For better understanding of the toxic phenomenon of *ctxvp60*, we cloned it as an inducible construct in plasmid pSG1144. Upon induction, a strong growth retardation effect was observed, possibly caused by the chimeric gene. The gene *ctxvp60* contains a large

number of rare Arg codons. We assumed that translation of CTXVP60 exhausted the rare tRNA_{Arg} pool of the cells resulting in the toxic effect. We tested this hypothesis by synthesizing two different versions of *ctxvp60*, that were different only in codon usage. Comparing the induced expression of both constructs (optimized and dezoptimized codon usage for *E. coli*) with the original *ctxvp60* showed that the extreme codon usage had no negative effect on the growth of the cell. Additionally, a frameshift version of *ctxvp60* was constructed, but this modification did not eliminate the toxic effect either. All together, this indicated that neither the *ctxvp60* gene/CTXVP60 protein *per se*, nor the supposed rare tRNA depletion phenomenon were the cause of the toxic effect of pCTXVP60.

Bioinformatical analysis of the fusion gene revealed an artificially formed ORF out of the original reading frame (ORF238). Translation of ORF238 results in an extremely Leu-rich (102 Leu residues) protein, which is predicted to have four putative transmembrane domains. To determine whether this small artificial ORF is responsible for the abnormal phenotype, we cloned it as an inducible construct. Upon induction, ORF238 produced the same growth retardation as seen with the original pCTXVP60 plasmid. We concluded that ORF238 was the culprit of the growth defect. Confocal laser scanning microscopy imaging of the cells expressing either CTXVP60 or ORF238 supported our findings. Both construct showed abnormally long cells with multiple nucleoids.

Based on sequencing 12 IS mutant pCTXVP60, we defined that insertions occurred in the 5' third of the fusion gene, in the *ctx* part or near the 5' end of *vp60*. Since the 5' end of ORF238 is located in the joint between *ctx* and *vp60*, we concluded that IS insertions in this region might block the transcription, relieving the growth retardation. RT-PCR experiments from IS mutant and original pCTXVP60 plasmids confirmed our conception.

Growth characteristics of strains transformed with toxic or mutant plasmids were monitored in liquid medium. The results correlated well with the orserved growth characteristics of transformants on agar plates. The original pCTXVP60 i) causes slow growth ii) cells pick up an IS insertion in *ctxvp60*, due IS transposition from the chromosome iii) cells harboring an IS-inactivated ORF238 resume normal growth and quickly become dominant in the culture. By using strains that possess various numbers of IS-es, we observed that the time needed for evolution of a non-toxic plasmid variant in the culture inversely correlated with the number of IS-es in the host genome. Additionally, even a single copy of an IS present in the host genome had a marked effect on plasmid stability.

In silico analysis of sequence libraries supported our theory that the case of *ctxvp60* is not unique, but far more widespread mechanism of self-defence against not tolerated, ectopic

recombinant genes. Data from 295 shotgun genome sequencing projects were analyzed with stringent criteria searching for IS contamination from cloning host genome. We've found a total of 109 IS-containing reads with 8 different kinds of IS elements. This elevated rate could be explained by processes analogous to those, which have been seen when cloning *ctxvp60*.

Recombined products that are toxic for *E. coli* may accumulate neutralizing, ISmediated mutant clones, which can become dominant by their faster growth and may cause undesired genotypic and phenotypic changes. We've shown that this phenomenon is widespread and leads to inactivation of biotechnological/synthetic-biological products and can diminish productivity. Removing all IS elements from the genome of a bacterial cell results in a significant increase in genomic stability and phenotypic uniformity, yielding an improved cellular chassis with reduced evolutionary potential.

Összefoglalás

Az *Escherichia coli* baktérium az egyik legalaposabban tanulmányozott és megismert élőlény. Napjainkra számos törzsének ismertté vált a teljes genomszekvenciája; ez is hozzájárul ahhoz, hogy kiváló modellorganizmusként szolgáljon a baktériumok genomi szintű evolúciójának tanulmányozásához.

Az Enterobaktériumok, és köztük is különösen az *Escherichia coli* törzsek genomja mozaikos szerkezetet mutat. Az alapvető, minden törzsben megtalálható gének sorát törzsspecifikus, horizontálisan transzfer révén szerzett génszigetek szakítják meg. Csoportunk korábbi munkája során egy már ismert genomú K-12 törzs, az MG1655 genomredukciójával létrehoztunk egy módosított törzset, melynek csökkentett méretű genomja közelít az *E. coli* törzsekre jellemző alapgénkészlettel leírhatóhoz.

Az E. coli kromoszómáján szétszórva található, horizontális géntranszfer által szerzett genomszakaszok jellemzően tartalmazhatnak inszerciós szekvenciákat (IS), transzpozázokat, hibás fágokat, integrázokat és helyspecifikus rekombinázokat. Az IS-ek olyan kisméretű, genetikailag kompakt elemek, melyek általánosságban csak a saját áthelyeződésükért felelős faktorokat hordoznak. Ezek az "önző" DNS szakaszok képesek az autonóm áthelyeződésre, valamint replikatív duplikációra. Egy átlagos bakteriális sejtben az IS elemek a genetikai változások szignifikáns részéért felelősek. Ha több kópiában megtalálhatóak a genomban, az általuk képzett szekvenciaismétlődések révén inverziókat, duplikációkat, valamint homológ rekombinációval kisebb-nagyobb deléciókat is okozhatnak. Az IS-ek okozta genomi változások meglepően gyakoriak, emiatt akár a laboratóriumokban általánosan használatos, elvileg azonos identitású törzsek is eltérő genotípusúak lehetnek. Az IS-ek molekuláris biológiai munkák során is problémákat okozhatnak. Klónozáskor vagy expressziós könyvtár létrehozáskor a genomi IS elemek plazmidokba helyeződhetnek át. Nemcsak természetes közegükben, hanem a laboratóriumi munka során is átfertőződhetnek *E. coli* törzsek egymás IS elemeivel.

A csoportunk által kifejlesztett egyedi, IS-mentes MDS42-es törzs egyedülálló lehetőséget kínált az IS elemek szerepének tanulmányozására mind a genom evolúciós folyamataiban, mind a különféle szintetikus-biológiai/biotechnológiai alkalmazásokban.

Kísérleteinkhez rendelkezésünkre állt egy IS transzpozíciót kiváltani, és IS elemeket igen hatékonyan begyűjteni látszó, extrém változékonyságot mutató pCTXVP60 plazmid ("IS csapda"). Ez a konstrukció növényekben történő vakcina előállítása céljából készült a nyúl

hemorrhágiás vírus (VP60) kapszidfehérjéjét és a koleratoxin (CTX) B alegységét kódoló gének fúzionálásával. Habár külön-külön mindkét gén tartósan fenntartható volt, kiméra változatuk minden hagyományos *E. coli* törzsben instabilnak mutatkozott.

A pCTXVP60 plazmiddal történő transzformálás után a kapott kolóniák változékony morfológiát mutattak LB-agar lemezeken (apró, lassú növekedésű és nagyobb, normális fenotípusú). Restrikciós enzimes emésztési mintázat és PCR ellenőrzések alapján az egészséges telepek 92 %-a inszerciót, 8 %-a deléciót hordozott a plazmidon lévő kiméra génben. Az IS inszerciók esetében specifikus primerekkel IS1, IS3 és IS5 elemeket mutattunk ki a *ctxvp60* gén 5' végén. Az IS-mentes MDS42 törzsben viszont tartósan fenntartható volt a plazmid. Transzformánsai egynemű kolóniákat adtak, a kinyert pCTXVP60 plazmid a normál restrikciós mintázatot mutatta.

Annak eldöntésére, hogy a toxikus gén okozta stressz megnöveli-e a transzpozíciós gyakoriságot, és ez hozzájárul-e a magas IS inszerciós gyakorisághoz, megvizsgáltuk a pCTXVP60 és a pCTX (kontroll, nem toxikus) plazmidot tartalmazó MG1655 gazdasejtek *cycA* génen mért mutációs spektrumát. A pCTXVP60-at tartalmazó sejtek magasabb mutációs rátát mutattak: az IS transzpozíció ráta négyszeresére emelkedett (4 különböző IS elem közreműködésével), míg a pontmutációk gyakorisága 1,5-szeresre növekedett. Továbbá, a sejtek 32 százaléka tartalmazott együttesen a plazmidon is és a kromoszómán is IS inszerciót. A kettős mutánsok ilyen magas aránya nem magyarázható csupán az IS transzpozíciós események gyakoriságának emelkedésével, hanem az IS mutánsok erős szelekciós előnyét is feltételezi.

Vizsgáltuk, hogy a fehérje túltermelés önmagában hogyan hat a mutációs rátára, ezen belül az IS transzpozícióra. pPROEX HT-CAT (a sejt által jól tolerált kloramfenikolacetiltranszferázt /CAT/ termelő) plazmidot tartalmazó MG1655 sejteken mutációs ráta mérést végeztünk a CAT fehérje indukciója során, illetve indukció nélkül. Az IS transzpozíció gyakorisága a *cycA* génen mérve 136 %-ra emelkedett az indukció hatására, ugyanakkor a pontmutációs ráta nem változott. A mért adatok arra utalnak, hogy még jól tolerált fehérje túltermeltetésekor is jelentős szerepe van az IS elemeknek az adott populáció mutációs változékonyságában.

A kiméra gén toxikus hatásának pontosabb vizsgálatához a *ctxvp60* gént indukálható konstrukcióban, a pSG1144 plazmidba klónoztuk. Indukció során kimutatható volt a kiméragén erős növekedésgátló hatása. Maga a *ctxvp60* gén igen nagy számban tartalmaz ritka Arg kodonokat. Feltételeztük, hogy a CTXVP60 transzlációja kimeríti a sejt ritka Arg tRNS készletét, ezzel okozva toxikus hatást. Ennek vizsgálatára két eltérő kodonhasználatú

változatát hoztunk létre. Az eredeti *ctxvp60* konstrukcióval szemben egyik sem okozott növekedésgátlást. Ez arra utalt, hogy nem a ritka kodonok nagymértékű használata okozza a toxikus hatást. Elkészítettünk továbbá egy "frameshift" mutációt hordozó *ctxvp60* változatot is, de a leolvasási keret eltolása – a várakozással ellentétben - sem szüntette meg a toxikus hatást. Mindezek az adatok arra utaltak, hogy nem a *ctxvp60* gén/CTXVP60 fehérje okozza a sejtekre gyakorolt negatív hatást.

Az eredeti *ctxvp60* gén szekvenciájának *in silico* vizsgálata feltárta egy nem a tervezett leolvasási keretben található ORF jelenlétét (ORF238). Transzlációja során egy olyan 238 aminosavat tartalmazó, extrém módon leucin gazdag fehérje (102) képződik, mely szerkezeti modellje alapján 4 transzmembrán, hidrofób doménnel rendelkezik. Annak eldöntésére, hogy ez a mesterségesen létrejött ORF felelős-e a rendellenes növekedési fenotípus kialakulásáért, kiemeltük és indukálható konstrukcióként klónoztuk. Indukált expressziója során az eredeti pCTXVP60 plazmidéval azonos hatást mutattunk ki. A hatás vizuális detektálásához CTXVP60- és ORF238-expresszáló sejteket konfokális lézer pásztázó mikroszkópos vizsgálatnak vetettünk alá. Mindkét konstrukció expressziójának eredményeként a sejtek rendellenes hosszúságúak, lassú növekedésűek és sejtosztódásukban gátoltak lettek. Megállapítható tehát, hogy a toxikus hatásért maga az ORF238 a felelős.

Az általunk megszekvenált 12 IS-mutáns pCTXVP60 plazmid segítségével meghatároztuk az inszerciók pontos pozícióit. Az inszerciós esemény szinte kivétel nélkül a fúziós gén 5' végén, a *ctx* régióban vagy a *vp60* 5'végénél következett be. Mivel az ORF238 a két gén kapcsolódási pontján helyezkedik el, valószínűsíthető, hogy az IS elemek a toxikus *orf238* 5' régiójába ugorva annak transzkripcióját blokkolják, megszüntetve ezzel a plazmid toxikus hatását. Az eredeti pCTXVP60 és annak IS mutáns változatai transzkripciójának RT-PCR-rel történő ellenőrzésével ezt igazolni is tudtuk.

A pCTXVP60 plazmidot tartalmazó törzsek növekedési/szelekciós dinamikáját különféle gazdasejt-plazmid kombinációk növekedési görbéjének kimérésével vizsgáltuk. Ezeknek a folyadékultúrás kísérleteknek az eredményei összhangban vannak az agar lemezeken megfigyeltekkel. Az eredeti pCTXVP60 i) növekedésgátlást okoz, ii) a sejtek egy része véletlenszerű módon IS-inszerciót szenved a *ctxvp60* génben a genomi IS-ek (fokozott) mobilitásának köszönhetően, iii) az IS-ek által inaktivált toxikus gént hordozó sejtek visszanyerik normális növekedési ütemüket és dominánssá válnak a kultúrában. A különböző számú és összetételű IS elemeket tartalmazó törzsek (MDS sorozat) vizsgálata azt sugallja, hogy a nem toxikus plazmid kialakulásához szükséges idő fordítottan arányos az adott törzs IS elem tartalmával. Megállapítottuk, hogy akár egyetlen IS-kópia jelenléte is jelentős

hatással van a plazmid stabilitására, ezzel a sejt evolúciójára.

hozzáférhető "shotgun" szekvenálások adatbázisainak bioinformatikai А tanulmányozása alátámasztotta elképzelésünket, miszerint az általunk megfigyelt ctxvp60 esete nem egyedi. Sokkal inkább egy feltehetően széles körben előforduló, automatikus folyamat, amelynek eredményeként a sejtekben a nem tolerált, külső, rekombináns gének 295 inaktiválódnak. Szigorú szekvenciakritériumok alkalmazásával "shotgun" klónkönyvtárból 14 esetében 109 olyan esetet találtunk, ahol a klónozott darabba E. coli gazda IS épült (8 különböző fajtájú IS-t azonosítottunk). Ez a magas arány nem magyarázható az IS-ek normál transzpozíciós gyakoriságával, hanem a *ctxvp60* klónozása során megfigyelt folyamatokat is feltételez.

Az eredmények demonstrálják, hogy az *E. coli* gazdasejtben expresszált, toxikus termékekben igen gyorsan, általános laboratóriumi körülmények között is felhalmozódhatnak IS okozta neutralizált, mutáns változatok, és ezek rövid idő alatt dominánssá válhatnak a kultúrákban. Ez a folyamat biotechnológiai szempontból előnytelen genotípusos és fenotípusos változatok felhalmozódásához vezethet. Az összes IS elem eltávolítása a bakteriális genomból szignifikánsan növeli a genom állandóságát, fenotípusos egységességét, és csökkenti az evolúciós potenciált.

Táblázatok és Függelékek

```
DAS-TMfilter prediction results 2/8/2010
Calculating prediction for the following proteins
with reference library 08:
>238aa Leu-rich pCTXVP60
... Done.
*** List of predicted non-TM-protein codes ***
none
*** List of predicted TM-protein codes ***
>238aa Leu-rich pCTXVP60
=== Result of the prediction ===
>238aa Leu-rich pCTXVP60
# TMH: 6 Q: trusted
@ 33 4.915 core: 26 .. 41 1.003e-05
@ 57 9.401 core: 46 .. 68 1.332e-12
@ 92 6.834 core: 83 .. 121 1.145e-08 Twin peaks - two TMH with a
short linker
          3.572 core: 83 .. 121 1.148e-03
@ 109

        @
        173
        4.243 core:
        168
        .
        183
        1.073e-04

        @
        210
        7.788 core:
        200
        .
        227
        3.957e-10

<----> end of list ---->
```

F1. AZ ORF238 fehérje DAS-TMfilter szerkezetpredikciója

HMMTOP version 2.0 prediction results

```
Protein: 238aa Leu-rich pCTXVP60
Length: 238
N-terminus: IN
Number of transmembrane helices: 4
Transmembrane helices: 48-67 84-107 167-184 203-222
Total entropy of the model: 17.0012
Entropy of the best path: 17.0026
The best path:
   seq MLLLLFLWLM DQDPWKERLE LLHKLELLEL LLLLLFQELL LMEWTQELLL
                                              50
   seq LLLLLLKIL LLLLLELE DHHNKLINKK LGELIFIIMM FLLGLLLMLQ
                                              100
   seq DLFFILFNIL HKIIHSLLFF LKCMLDGLEE CNSDSLLDL EFLEEDLLLL
                                              150
   seg LFHQELKLDQ VLKLDNFHML LLMLGLLNQL LLLCQTLDQI CIIQLEIQDL
                                              200
   seg FQLLFFLFII ILLIHLEDLL LLFKLLLKLD HLKILNLL 238
   pred ooHHHHHHHH HHHHHHHHHH HHiiiiiii iiiiiiI
```

F2. AZ ORF238 fehérje HMMTOP szerkezetpredikciója



Molecule Features:

Start	End	Name	Name	
219	1926	C CAT		
1786	1430	C CTX		

F3. pCTX plazmid térképe



F4. pCTXVP60 plazmid térképe



A COLUMN TO A	Concernence of the second second		
NO	0011	0	Hastinge.
1.10 7			reacures.

Start	End		Name
219	3690	С	CAT
3227	3227		extraT
3551	1459	C	CTXVP60frameshift

F5. pCTXVP60Frameshift plazmid térképe


Mo	1001	110	FO	2 t	1120	
1.10	エレしい		20		UL C	-0.

Start	End	Name
781 1898 3232 4407 6390	122 0 2653 4407 5378 6440	CAT RepE parA parB T7
0290	0440	1 /

F6. pSG1144 plazmid térképe



Molecule Features:

Start	End	Name
781	122 (C CAT
1898	2653	RepE
3232	4407	parA
4407	5378	parB
6390	6440	T 7
6482	7198	ORF238

F7. pSG1144-ORF238 plazmid térképe



Molecule Features:

Start	End		Name
781	122	C	CAT
1898	2653		RepE
3232	4407		parA
4407	5378		parB
6390	6440		T7
6496	8586		CTXVP60
6788	7504		ORF238
3232 4407 6390 6496 6788	4407 5378 6440 8586 7504		parA parB T7 CTXVP6 ORF238

F8. pSG1144-CTXVP60 plazmid térképe



Mo]	lecul	e	Features:

t

F9. pSG1144-CTXVP60OPTt plazmid térképe



Molecule Features:

Start	End	Name
781	122 C	CAT
1898	2653	RepE
3232	4407	parA
4407	5378	parB
6390	6440	T 7
6470	8560	ctxvp60dezopt

F10. pSG1144-CTXVP60DEZOPT plazmid térképe

Név	szekvencia (5'-3')	felhasználása
ORFstart	cgggatccatgctattgctgctatttc	ORF238 klónozása
ORFstop	cggaattctcataacaaattcaaaatcttcag	ORF238 klónozása
ctxstarteco	ccggaattccgtcgactttaaagttcatc	ctxvp60 klónozása
ctxstopbam	ggatccggccgcccggggcaatggct	ctxvp60 klónozása
CTXVP2	cctcctggaactggttga	szekvenáló primer ctxvp60
СТХУРЗ	ggaggatctacttctgct	szekvenáló primer ctxvp60
CVek3	gcctggttgtacgcctgaa	szekvenáló primer ctxvp60
т7	taatacgactcactataggg	szekvenáló primer/ szekvenciaellenőrző ctxvp60
chi3	gtgagtttcaccagttttga	szekvenáló primer/ szekvenciaellenőrző ctxvp60
vpvege	gttcgctccaagctggactg	szekvenáló primer/ szekvenciaellenőrző ctxvp60
pctxkozep	tgcccttaaacgcctggttg	szekvenáló primer/ szekvenciaellenőrző ctxvp60
vpf1	cctgggtcctattccatcagtagtagttaat	ctxvp60 frame-shift létrehozó primer
vpf2	tgatggaataggacccaggagttgttgct	ctxvp60 frame-shift létrehozó primer
IS1-1	tcgctgtcgttctca	IS1-specifikus primer
IS1-2	aagccactggagcac	IS1- specifikus primer
IS2-1	tcgcaggcataccatcaa	IS2- specifikus primer
IS2-2	cagacgggttaacggca	IS2- specifikus primer
IS3-1	agcggctggtatacgtggt	IS3- specifikus primer
IS3-2	tcatgcgtggcgacattga	IS3- specifikus primer
IS5-1	gacagttcggcttcgtga	IS5- specifikus primer
IS5-2	gctcgatgacttccacca	IS5- specifikus primer
IS150-1	acgtgccgagatgatcct	IS150- specifikus primer
IS150-2	cagacctatatgcctcgt	IS150- specifikus primer
cycA1	ctgatgccggtaggttct	cycA gén ellenőrzése
cvcA2	gcgccatccagcatgata	cycA gén ellenőrzése
RecA Up1	attacccoocatoacaooao	RecA belső kontrollhoz
RecA_Down1	cgacccttgtgtatcaaacaagac	RecA belső kontrollhoz
IS3-1 IS3-2 IS5-1 IS5-2 IS150-1 IS150-2 cycA1 cycA2 RecA_Up1 RecA_Down1	agcggctggtatacgtggt tcatgcgtggcgacattga gacagttcggcttcgtga gctcgatgacttccacca acgtgccgagatgatcct cagacctatatgcctcgt ctgatgccggtaggttct gcgccatccagcatgata attacccggcatgacaggag cgacccttgtgtatcaaacaagac	IS3- specifikus primer IS3- specifikus primer IS5- specifikus primer IS5- specifikus primer IS150- specifikus primer IS150- specifikus primer <i>cycA</i> gén ellenőrzése <i>cycA</i> gén ellenőrzése RecA belső kontrollhoz RecA belső kontrollhoz

F11. Felhasznált primerek szekvenciái