

I. Bevezető

Mely gének szükségesek egy mikroorganizmus túléléséhez és meddig egyszerűsíthetők laboratóriumban a genomok? Meg tudjuk-e jósolni több gén együttes eltávolításának fenotípusos hatását és miért nem távolíthatók el bizonyos génpárok egyszerre? Hogyan zajlik a természetben a genomok egyszerűsödése (pl. endoszimbiózis során)? Jósolható-e rendszerbiológiai módszerekkel az evolúciós folyamatok legvalószínűbb kimenetele? Vajon csökken-e a laboratóriumban egyszerűsített genomok evolúciós alkalmazkodóképessége? Kutatásaink során a rendszerbiológiai modellezés, a genommérnökség és a funkcionális genomika eszköztárát felhasználva kerestük a választ a fenti kérdésekre.

II. Eredmények szöveges összefoglalása

1. Projekt: Egyszeres és többszörös génkiütések hatásának előrejelzése rendszerbiológiai modellezéssel. Automatizált tudáskinyerés génkölcsonhatási adatsorokból.

1.1. Korábbról már ismert volt, hogy a sejt anyagcseréjét leíró ún. flux balance analysis modellek nagy (80-90%) pontossággal képesek jósolni az egyszeres génkiütések kvalitatív hatását mikrobákban (vagyis, hogy a génkiütés letális-e vagy sem). A genomok természetben zajló egyszerűsödésének megértéséhez és a laboratóriumi genomegyszerűsítés racionális kivitelezéséhez azonban nem elegendő az egyszeres génkiütések hatásait prediktálni. Két vagy több gén mutációja között ún. génkölcsonhatás léphet föl, azaz fenotípusos hatásuk nem független, erősítheti vagy gyengítheti egymást. Az elmúlt időszakig a génkölcsonhatásokról kevés információ állt birtokunkban, különösen igaz volt ez anyagcserehálózati génekre, amelyek viselkedése rendszerbiológiai modellekkel viszont jól leírható.

Kollaborátorunk, Charles Boone (Toronto), 2009 elején egyedülálló adatsort bocsájtott rendelkezésünkre, amely a *S. cerevisiae* élesztőgombában több százezer (>200 000) anyagcseregépár genetikai kölcsönhatását írja le. Ezen adatok lehetővé teszik, hogy az élesztő anyagcseremodelljét kiterjedten teszteljük, a többszörös génkiütések hatását becsüljük, és módszereket dolgozzunk ki a modell automatikus javítására. Ilyen mennyiségű és minőségű adat nem áll rendelkezésre *E. coli*-ra, így figyelmünk az 1. évtől főleg a *S. cerevisiae* anyagcseréjének vizsgálatára terelődött. Egyik legfontosabb kutatási irányunk a génkölcsonhatásokkal kapcsolatban, hogy egy, az anyagcserét leíró rendszerbiológiai modell képes-e mechanisztikus magyarázatot kínálni az empirikus génkölcsonhatási adatsorban látott összefüggésekre. Meg tudjuk-e magyarázni például, hogy miért mutatnak a fontos gének nagyszámú génkölcsonhatást? Mennyire pontosan képes az anyagcsere számítógépes modellje prediktálni a gének közötti kölcsönhatásokat? És hogyan tudjuk automatizáltan javítani a modellt a génkölcsonhatási adatsorok alapján?

Megmutattuk, hogy az anyagcserét leíró rendszerbiológiai modell képes mechanisztikus magyarázatot kínálni a génkölcsonhatási hálózatok bizonyos általános tulajdonságairól. Az egyik ilyen általános tulajdonság, hogy a génkölcsonhatások száma erősen összefügg a gén fontosságával, azaz ha egy gén kiejtése nagy fitnesscsökkenéssel jár, akkor egy második mutáció (génkjejtés) nagy valószínűséggel erősíti (ún. negatív

génkölcsönhatás) vagy éppen mérsékli (pozitív génkölcsönhatás) az első mutáció hatását. Az anyagcseremodell segítségével *in silico* génkölcsönhatásokat számoltunk, majd megállapítottuk, hogy a génenkénti kölcsönhatások száma a modellben is erősen összefügg a génkiejtés fitnesshatásával. De mit árul el a modell az összefüggés mögött meghúzódó mechanizmusról? Azt találtuk, hogy a nagy fitnesshatású enzimgének egyben erősen pleiotrópák is: az általuk kódolt biokémiai reakciók több különböző bioszintetikus folyamathoz is nélkülözhetetlenek. Ennek eredményeképpen ezek mutációja többféle anyagcserefolyamatot érint és így több gén képes módosítani a mutáció hatását.

A fenti biztató eredményekkel szemben az egyedi génpárok közötti génkölcsönhatások predikciójában kevésbé bizonyult sikeresnek a modell. Kimutattuk, hogy habár az anyagcseremodell által jóslt génkölcsönhatások nagy valószínűséggel helyesek, de a kísérletesen megállapított génkölcsönhatások többségét egyszerűen nem találja meg (azaz nagyon alacsony a modell szenzitivitása). Mindez arra utal, hogy az anyagcserehálózat szerkezetére támaszkodó modell a génkölcsönhatások általános eloszlási mintázatait ugyan jól visszaadja, de az egyedi génkölcsönhatások megbízható predikciójához nem tartalmaz elég információt. Előzetes vizsgálataink szerint ez a konklúzió akkor sem változik, ha a génszabályozásról rendelkezésre álló (egyelőre igencsak töredékes) ismereteinket ötvözni próbáljuk az anyagcseremodellel.

Az a tény, hogy számos olyan kettős géniütés is letális *in vivo* amely az anyagcseremodell alapján életképes két következményt is felvet. Az egyik, hogy az eredeti OTKA pályázatomban megfogalmazott terv miszerint a modell segítségével racionálisan lehet minimalizálni az anyagcsere bizonyos egységeit sokkal problémásabb mint korábban várható volt. Ez alapján a 2008-as pályázatban vázolt 2. Projekt módosult és a genomegyszerűsödés más aspektusait kezdtem el vizsgálni (lásd alább). A másik, biológiai szempontból lényegesebb következmény, hogy a modell és a kísérleti adatok eltérését elvben fel lehetne használni a modell javítására (feltételezve, hogy a modellezési keret fogyatékoságai ellenére is használható és a modell biológiai tartalma javítható). A tudományos felfedezés automatizálása (pl. biokémiai modellek automatikus fejlesztése) alapvető cél a rendszerbiológiában és funkcionális genomikában, és a rendelkezésünkre álló génkölcsönhatási adatok kiváló lehetőséget biztosítanak erre. Ezért kidolgoztunk egy gépi tanulási eljárást amely a ~200 000 génkölcsönhatási adatpont alapján módosításokat (hipotéziseket) javasol az anyagcserehálózati modellben. Módszerünk olyan változtatásokat javasol (pl. biokémiai reakció törlése, reakció reverzibilitásának megváltoztatása), amely javítja a génkölcsönhatások predikciójának sikerét. A módszer segítségével több javítást is javasoltunk az anyagcseremodell NAD bioszintézis útvonalában. A változtatások sikerességét egyrészt keresztvalidálással becsültük, másrészt az egyik NAD útvonalon végzett új genetikai kísérletekkel is ellenőriztük.

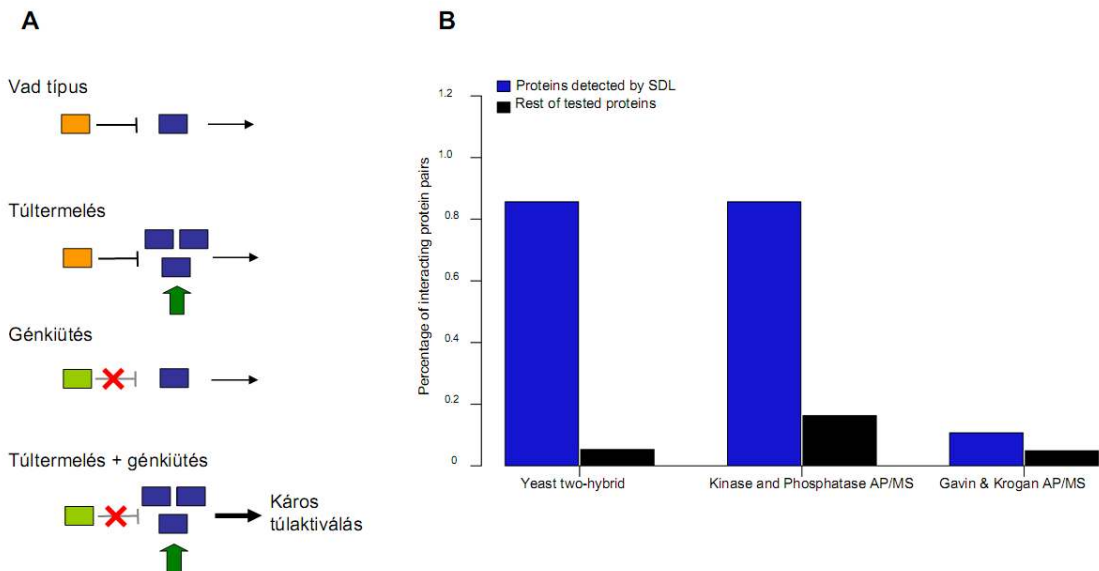
A fenti eredményeket nemrég a Nature Genetics folyóiratban közzétettük kollaborátorainkkal (Szappanos et al., 2011 *Nat Genet* **43**: 656), és cikkünket azóta a *Faculty of 1000* is méltatta („Must Read” besorolással).

1.2. A fenti projekthez kapcsolódóan további vizsgálatokat végeztünk a génkölcsönhatások természetének felderítése érdekében (a Boone, Andrews és Myers csoportokkal együttműködve):

- Duplikált gének evolúciója: Kimutattuk, hogy a duplikált génpárok más génekkel mutatott génkölcsönhatási profiljainak összevetése felfedheti a génpár

funkcionális különbségeit. Azt találtuk, hogy a génpárok kölcsönhatási profiljai aszimmetrikusan különböznek: a pár egyik tagja gyakran nagy számú génkölcsönhatással rendelkezik, míg a testvérpéldány jóval kevesebbel, ráadásul ez evolúciós aszimmetriájukban is tükröződik. Eredményeinket 2010-ben közzétettük a *Molecular Systems Biology* folyóiratban.

- Szintetikus dózis letális génkölcsönhatások: Nemrég szisztematikusan megvizsgáltuk milyen kölcsönhatások lépnek fel egy gén túltermelése és egy másik gén kiütése között. Például ha egy gén túltermelése önmagában nem káros, viszont egy másik gén kiütésével már letális fenotípust okoz, akkor az a két gén között negatív regulációs kapcsolatra utalhat (ún. SDL kölcsönhatás, lásd 1A ábra). A kanadai labor elkészített egy nagyléptékű SDL génkölcsönhatási térképet az élesztő kináz génjeire, ahol sok ilyen típusú regulációs kapcsolatot feltételeznek. A közel 500 000 génpárt tartalmazó adatsor bioinformatikai elemzését mi végeztük, és a következőket találtuk: i) az irodalomból ismert kináz – szubsztrát kölcsönhatások az SDL génkölcsönhatásoknak csak egy csekély hányadát tudják megmagyarázni, ami arra utal, hogy legtöbb SDL kölcsönhatás sokkal közvetettebb és nem közvetlen kináz – szubsztrát negatív regulációt jelent; ii) az SDL kölcsönhatásoknak csak egy része magyarázható közvetlen fizikai fehérje kapcsolattal, de ezek szignifikánsan gyakrabban jelentkeznek tranziens fehérje-fehérje kölcsönhatásoknál mint stabil fehérjekomplexekben (1B ábra); iii) az SDL kölcsönhatások gyakran környezetspecifikusak; iv) az SDL és más típusú génkölcsönhatási térképek integrálásával új génfunkciókat lehet jósolni. Eredményeink közzététel alatt állnak a *Genome Research*-nél (átdolgozás alatt).

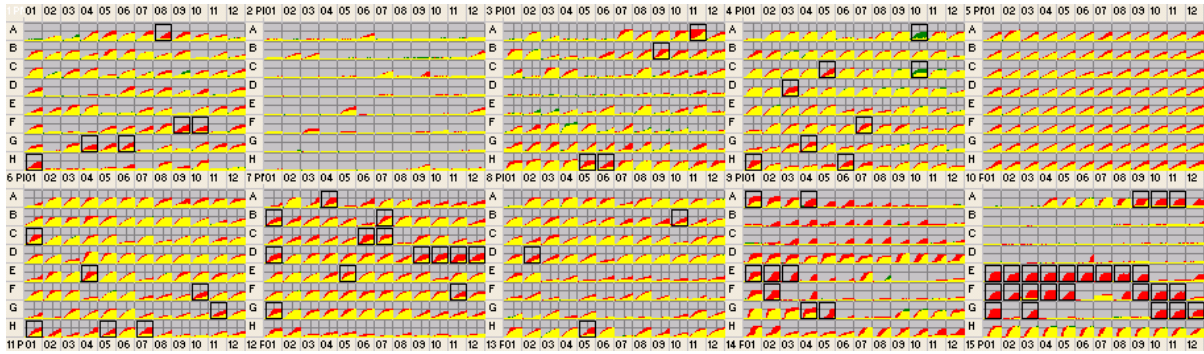


1. Ábra. A) Szintetikus dózis letális (SDL) génkölcsönhatás egy gén túltermelése és egy másik gén kiütése között: önmagában sem az egyik gén kiütése, sem a másik gén túltermelése nem káros, de a kettő együtt már igen (pl. a túltermelt gén gátlás híján túlaktiválódik). B) SDL génkölcsönhatási kapcsolatok átfedése fehérje-fehérje kapcsolatokkal. A „Yeast two-hybrid” és a „Kinase and Phosphatase AP/MS” kategóriák leginkább tranziens interakciókat fednek le, míg a „Gavin & Krogan AP/MS” csoport stabil fehérjekomplexeket jelöl.

2. Projekt: Egyszerűsödött genomok rendszerbiológiája és evolúciója

Eredeti tervünk az volt, hogy egy nagy prediktív erővel bíró *Escherichia coli* anyagcserehálózati modell segítségével megkísérülünk új, egyszerűsített genomú törzseket racionálisan tervezni. Ezt a tervet két okból is módosítani kényszerültünk: i) az 1. Projektben kapott eredmények alapján egy nagyon hasonló élesztőmodell képtelen volt a kettős letális fenotípusok többségét helyesen prediktálni (enélkül pedig nehéz életképes többszörös géntörzseket tervezni), ii) a már rendelkezésre álló (a Pósfai csoport által előállított) egyszerűsített genomú *E. coli* törzsek metabolikus fenotípusát vizsgálva arra az előzetes konklúzióra jutottunk, hogy a modell egyelőre csak részben tudja leírni a kísérleti tapasztalatokat (lásd alább). Ezen okok miatt figyelmünket elsősorban a meglévő egyszerűsített genomok (*MDS42*, *MDS71*) fenotípusos és kísérleti evolúciós vizsgálatára, valamint a természetben lezajló genomegyszerűsödések számítógépes vizsgálatára fordítottuk.

2.1. Az egyik fő törekvésünk, hogy az Intézetben a Pósfai csoport által szintetikus biológiai módszerekkel előállított egyszerűsített genomú *Escherichia coli* törzsek fenotípusos viselkedését megértsük, és a genomegyszerűsödésnek az evolúciós potenciálra gyakorolt hatását felmérjük. E törzsek egyedülállóak abból a szempontból, hogy nagy számú (közel 700 – 1000) gént kiejtettek belőlük, köztük az összes ugráló genetikai elemet, de ennek ellenére a törzsek jó növekedést mutatnak standard laboratóriumon. A 2. évben a Pósfai és Pál csoportokkal kollaborációban megkezdtük ezen redukált kóligenomok fenotípusos viselkedésének kísérletes és számítógépes vizsgálatát. Egyrészt egy új típusú nagy áteresztőképességű eljárás, a Biolog Phenotype Microarrays segítségével több száz tápanyagkörnyezetben megmértük a redukált genomú törzsek metabolikus aktivitását és azonosítottuk azokat a környezeteket, ahol a vad típusú sejthez képest az egyszerűsített törzs gyengébben növekszik (lásd 2. ábra). Másrészt széles körű anyagcserehálózati szimulációkat végeztünk annak felderítésére, hogy az egyszerűsített genomú törzsnek elméletileg mely szimulált tápanyagkörnyezetben kell csökkent növekedést mutatnia. A kísérletes adatok és a rendszerbiológiai szimulációk összevetése feltárta, hogy az *in vivo* megfigyelt fenotípusok egy szignifikáns része megmagyarázható az anyagcsere szerkezetében bekövetkezett változásokkal. Emellett viszont az is kiderült, hogy a kísérletesen tapasztalt növekedési zavaroknak legalább a fele nem prediktálható a rendszerbiológiai modellel, feltehetően egyrészt azért, mert számos anyagcserén kívüli gén is el lett távolítva a genomegyszerűsítés során, másrészt pedig azért, mert a modell helytelenül írja le a gének kölcsönhatások számottevő részét (lásd 1. Projekt).



2. Ábra. Az egyszerűsített genomú *E. coli* törzs növekedése a vad típusú *E. coli*-hoz képest 960 kísérletesen vizsgált tápanyagkörnyezetben (különböző szén, nitrogén, kén és foszfor-források). Bekeretezett piros színű növekedési görbék jelzik azokat a környezeteket, ahol a redukált genomú törzs alulmaradt a vad típusúhoz képest.

De vajon az összes ugráló genetikai elem (ún. IS elem) eltávolítása után rendelkezik-e a kóli genom azzal a képességgel, hogy új környezetekhez evolúciósan adaptálódjon? Az irodalomban gyakran találkozni azzal a nézettel, hogy az ugráló genetikai elemek, variabilitást növelő tényezőként, evolúciós hajtóerejét képviselik. Ezt a kérdéskört a Pál és Pósfai csoportokkal közösen vizsgáltuk laboratóriumi evolúciós kísérletekben (ún. kompetíciós kísérletekben), amelyben különböző genetikai hátterű kólibaktériumokat adaptáltattunk szalicinhez mint új szénforráshoz. Eredményeink azt mutatták, hogy az egyszerűsített genomú, ugráló elemektől mentes törzs nehezebben tudott alkalmazkodni az új szénforráshoz mint egy olyan izogenikus törzs, amelybe egyetlen ugrálóelemet (egy IS1 elemet) visszahelyeztünk. Ráadásul az IS1-et tartalmazó adaptált törzsek átlagos fitnessze magasabb volt, mint a pusztán pontmutációkkal adaptálódott törzseké. Mindez azt mutatja, hogy az ugráló genetikai elemek egyrészt növelik az adaptív mutációk megjelenési gyakoriságát, másrészt nagyobb fenotípusos hatású mutációkat tudnak létrehozni. De vajon mi történik, ha egy másik mutációs gyakoriságot növelő mechanizmussal vetjük össze az IS-elemek hatását? Az ún. mutátor gének a DNS-hibajavításban sérültek, és így akár 100-szor magasabb a mutációs rátájuk. Izogenikus mutátor fenotípusú és IS-t tartalmazó törzsek versenyeztetése azt mutatta, hogy a mutátor gének gátolják az IS-elemek evolúciós adaptációhoz való hozzájárulását. Eredményeink bírálat alatt állnak a *Molecular Biology and Evolution*-ben.

2.2. A másik fő irányvonal a természetben lezajlott genomcsökkenések megértése komparatív genomikai és rendszerbiológiai eszközök integrálásával. Néhány évvel ezelőtt kimutattuk, hogy a *Buchnera aphidicola* endoszimbionta baktérium genomegyszerűsödését (a gének 70 – 75%-a elveszett viszonylag rövid evolúciós idő alatt) meglepően nagy pontossággal lehet prediktálni a mai *E. coli* anyagcserehálózati modelljéből kiindulva (a *Buchnera* és az *E. coli* viszonylag közeli rokonok). Azaz az ősi anyagcserehálózat szerkezete meghatározza, hogy mely géneket lehet elveszteni, és melyeknek kell megmaradni ahhoz, hogy az endoszimbionta saját biomasszáját meg tudja termelni és bizonyos vitaminokkal és aminosavakkal lássa el a gazdasejtet. Munkánk akkor a ma megfigyelhető *Buchnera* genom összetételének predikciójára irányult és először bizonyította, hogy a természetben lezajló hosszú távú evolúciós folyamat

számítógépes rendszerbiológiai módszerekkel előrejelezhető. Az elmúlt évben Eytan Ruppín (Tel Aviv) csoportjával együttműködésben azt vizsgáltuk, hogy vajon megjósolható-e a gének elvesztésének sorrendje is? Azaz, ha ismerjük az ősi anyagcserehálózatot (feltételezzük, hogy ez nagyon hasonló a mai *E. coli* hálózathoz), a felvehető tápanyagokat és a megtermelendő biomassza-összetevőket, akkor egy evolúciós szimulációval meg tudjuk-e jósolni, hogy mely gének vesznek el hamarabb és melyek később? A jósoltok ellenőrzésére a rendelkezésre álló *Buchnera* genomok alapján rekonstruált törzsfát és az ősi endoszimbionta genomok filogenetikai rekonstrukcióját használtuk. Eredményeink várakozásainkat felülmúlóan pozitívak voltak: az evolúciós rendszerbiológiai szimuláció szignifikánsan prediktálja a törzsfá alapján rekonstruált génvesztések sorrendjét, azaz a valóságban lezajlott evolúciós útvonal részben előrejelezhető rendszerbiológiai ismereteink alapján. Eredményeinket nemrég közzétettük a *Molecular Systems Biology* folyóiratban.

3. Összefoglaló közlemények

Az elmúlt 3 évben az alábbi, az OTKA pályázathoz kapcsolódó témákban írtam összefoglaló munkákat:

- 3.1. Anyagcserehálózatok adaptív bélyegei és evolúciójuk (HFSP Journal 2009)
- 3.2. Genomevolúció jóslása rendszerbiológiai eszközökkel (Nature Reviews Genetics 2011)
- 3.3. Anyagcserehálózati modellezés az evolúciós rendszerbiológiában (Methods Mol Biol 2011)
- 3.4. Génkölcsonhatási hálózatok „csomópontjai” és a pleiotrópia kapcsolata (Cell Cycle 2011)

4. Egyéb közlések és teljesítmények

A pályázat 3 éve alatt 15 nemzetközi konferencián ill. nyári iskolán mutattam be eredményeinket, köztük 11 meghívott előadás volt és 1 konferencián szimpóziumszervezőként vettem részt (ESEB 2009). Jelenleg 5 doktorandusz hallgató témavezetését végzem, közülük ketten várhatóan egy éven belül védeni fognak.

III. Költségtervtől történő jelentősebb eltérések indoklása

A kutatások során a korábban tervezettnél több és költségesebb kísérleti munkákat végeztünk el: i) Az 1. projektben a számítógépes jósolások ellenőrzésére genetikai kísérleteket végeztünk. ii) A 2.1. projektben az egyszerűsített genomok növekedését több száz környezeti körülmény között vizsgáltuk. Ezt szolgáltatásként a Biolog cégtől rendeltük, majd laborunkban kisebb léptékű ellenőrző kísérleteket végeztünk. iii) A 2.1 projekt részeként evolúciós kísérleteket végeztünk. A fenti kísérletes munkák olyan tudományos eredményekkel szolgáltak, amelyek tisztán számítógépes módszerekkel elérhetetlenek lettek volna, így kulcsfontosságúak voltak a projektek továbbhaladásához. Mindez azonban 3 év alatt kb. 2 millió forinttal növelte a dologi és készletbeszerzésre fordított kiadásokat. Ezt a konferencia és kollaboratív utazásokra szánt keret csökkentésével tudtam áthidalni (az utazások egy részét más forrásokból sikerült fedezni).