Ph. D. értekezés tézisei

Génkölcsönhatások és evolúciós újítások vizsgálata anyagcsere-hálózati megközelítéssel

Szappanos Balázs

Témavezető: Dr. Papp Balázs, Ph.D.

Biológia Doktori Iskola

MTA Szegedi Biológiai Kutatóközpont Biokémiai Intézet

Szeged

2016

Bevezetés

Az elmúlt néhány évtizedben a kísérletes technikák fejlődése az általuk szolgáltatott adatmennyiség exponenciális növekedéséhez vezetett. Ez lehetővé tette, hogy a biológiai rendszereket holisztikus megközelítéssel vizsgáljuk és elvezetett a rendszerbiológia kialakulásához. Az anyagcsere az egyik legjobban feltárt alrendszere a sejtnek, és ezáltal kiemelten alkalmas arra, hogy rendszerbiológiai problémákat vizsgáljunk rajta. Az anyagcseréről meglévő ismereteink összegyűjtése lehetővé tette olyan matematikai modellek építését is, amelyek előrejelzéseket tudnak tenni az anyagcserét érő zavarásoknak, például génkiütéseknek a növekedésre gyakorolt hatásáról. A rendszerszintű megközelítés az evolúcióbiológiában is új lehetőségekhez vezet. Az evolúcióbiológia kutatásában ugyanis kulcsszerepe van annak, hogy megértsük, hogyan határozza meg az élőlények genotípusa a fenotípust. Ezt a genotípus-fenotípus leképezést az anyagcseremodellek segítségével akár több száz génre kiterjedően, szisztematikusan tudjuk vizsgálni (Papp et al., 2011).

Doktori munkám során az volt a célom, hogy az anyagcsere-hálózatok felhasználásával az evolúciós genetika és az evolúciós innovációk néhány alapvető problémáját vizsgáljam. Egyrészt megvizsgáltam, hogy az anyagcseregének kiütésének hatásai között van-e kölcsönhatás, ha két gént együttesen ütünk ki, valamint hogy ezek a genetikai kölcsönhatások milyen rendszerszintű mintázatokat mutatnak (Szappanos et al., 2011). Doktori disszertációm második részében pedig az enzimek járulékos aktivitásainak jelentőségét

részletezem új anyagcsereutak evolúciójában (Notebaart et al., 2014).

Genetikai interakciók vizsgálata és modellezése az anyagcserében

Genetikai kölcsönhatásról akkor beszélünk, ha két mutáció együttes hatása eltér attól, amit a mutációk önálló hatása alapján várnánk. A genetikai kölcsönhatás lehet negatív vagy pozitív, attól függően, hogy a kapott kettős mutáció hatása erősebb vagy gyengébb a vártnál. A genetikai interakcióknak fontos szerepe van a genetikai hátterű betegségek kialakulásában (Lehner, 2007). Ugyancsak fontos szerepet kapnak a genetikai interakciók ismeretlen funkciójú gének leírásában, egy új gén funkciója ugyanis megismerhető már ismert génekkel mutatott kölcsönhatásai alapján (Boone et al., 2007). A genetikai interakciók formálják az adaptív tájképet is, ezáltal befolyásolják az adaptáció menetét (Phillips, 2008).

Munkánk során a genetikai kölcsönhatásokat az élesztő anyagcseremodellen számoltuk *in silico* az úgynevezett fluxus balansz analízis (FBA) módszerrel (Segrè et al., 2005). A modell eredményeit egy átfogó kísérletes adatsorral vetettük össze, melyet kollaborátorunk (Boone labor, Torontó) biztosított. Ez az adatsor több, mint 176.000 anyagcseregénpárra tartalmazza a genetikai kölcsönhatás értékét.

Célkitűzések

Kutatásainknak több célja is volt. Egyrészt szisztematikusan tesztelni, hogy milyen hatékonysággal jelzi előre az anyagcseremodell a genetikai interakciós hálózat általános, nagyléptékű tulajdonságait és a konkrét génpárok közötti genetikai kölcsönhatásokat. További célunk volt egy automatizált számítógépes eljárást kifejleszteni, amely segítségével a modell hibás előrejelzései javíthatók és ezáltal a genetikai interakciós adatsor alapján új biológiai hipotéziseket javasolhatunk. Végül pedig az ily módon automatikusan kinyert hipotézisek ellenőrzését is terveztük.

Módszerek

- FBA módszer használata az egyszeres és kettős génkiütések *in silico* fitneszhatásának meghatározására.
- Funkcionális pleiotrópia számolása az anyagcseremodell génjeire az FBA felhasználásával.
- Saját fejlesztésű, genetikai algoritmuson alapuló eljárás a modell előrejelzéseinek feljavítására.

Eredmények

1. A gének genetikai interakciós fokszáma szorosan összefügg a gén fontosságával és pleiotrópiájával

Kísérletesen meghatározott genetikai kölcsönhatásokon megfigyeltek egy érdekes mintázatot, miszerint minél több interakcióban vesz részt egy gén, annál károsabb a kiütése az élőlény számára (Costanzo et al., 2010). Kimutattuk, hogy ezt a mintázatot a modell is mutatja. Ez után a modellt arra akartuk felhasználni, hogy magyarázatot találjunk a gének genetikai interakciós fokszáma és a génkiütés fitnesze közötti összefüggésre. Egy lehetséges magyarázat arra, hogy egy bizonyos génnek sok genetikai kölcsönhatása van, az lehet, hogy a gén nagyszámú biológiai folyamatban vesz részt (erősen pleiotróp). Az elemzésünk kimutatta, hogy erős negatív korreláció van egy gén által érintett bioszintetikus folyamatok száma (pleiotrópia) és a gén kiütésének hatása között. Ezen kívül a pleiotrópia és a genetikai kölcsönhatások száma pedig pozitív korrelációban van. Ez alapján feltételezhető, hogy a több biológiai funkcióban részt vevő (pleiotrópikus) gének kiütésének hatását nagyszámú más gén mutációja befolyásolja, ahol az egyes gének a pleiotrópikus gén más-más funkcióját kompenzálják.

2. Az anyagcseremodell nem tudja előrejelezni a genetikai kölcsönhatások jelentős részét

Az FBA segítségével kiszámoltuk, hogy mely génpárok mutatnak genetikai kölcsönhatást, majd összevetettük a Boone csoportból származó kísérletes adatsorral. Azt találtuk, hogy az *in silico*

kölcsönhatások között erős feldúsulást mutatnak a kísérletes interakciók a véletlenszerűen várthoz képest. Ez alátámasztja, hogy az *in silico* előrejelzett genetikai interakciók biológiailag relevánsak, azonban a kísérletesen látott kölcsönhatásoknak csak egy kis hányadát találjuk meg a szimulációval. A modell ezen hiányosságán automatizált javító eljárással próbáltunk javítani.

3. Az anyagcseremodell predikciós ereje automatizált módszerrel feljavítható

Annak érdekében, hogy javítsunk a modell szimulációs erején, kifejlesztettünk egy gépi tanuláson alapuló eljárást. Módszerünk a modellen különböző módosításokat hajt végre automatikusan, majd a módosítások hatását globálisan teszteli a szimulációs eredmények és a kísérletes adatok összevetésével. Módszerünk több változtatást is javasolt a modellen, amelyek jelentősen javították az előrejelzések sikerességét: az érzékenység 100-267 %-kal, míg a precizitás 44-59 %-kal nőtt.

4. A NAD bioszintézis útvonalak *in silico* kijavítását kísérletesen is igazoltuk

A legjelentősebb javítást az algoritmusunk a NAD bioszintézis módosításával érte el. A szimulációhoz használt élesztő anyagcseremodell három NAD bioszintézis útvonalat tartalmaz, javító algoritmusunk az egyik törlését javasolta. Ennek az útvonalnak az enzimjeiről szekvencia homológia kereséssel kimutattuk, hogy hiányoznak az élesztőből. Ezután olyan előrejelzéseket kerestünk, ahol a javított modell eltér a kiindulásitól

és ezeket az eseteket kísérletesen is teszteltük, igazolva, hogy az algoritmus által kifogásolt harmadik NAD bioszintézis út valóban nincs jelen az élesztőben. A laborkísérleteket Szamecz Béla kollégám végezte. Eredményeink azt mutatják, hogy a gépi tanulásos eljárások nagy áteresztőképességű kísérletes adatokkal egybekötve képesek hipotézisek automatikus előállítására és a hipotézisek tesztelésére is, ez pedig legalább részben automatizált biológiai felfedezésekhez vezet.

Rejtett enzimaktivitások hálózata és evolúciós jelentősége

Régóta ismert, hogy az enzimek képesek olyan biokémiai reakciókat is felgyorsítani melléktermékként, amelyeknek a katalizálására nem specializálódtak. Ezeket a szelekciós nyomás alatt nem álló enzimaktivitásokat rejtett vagy látens aktivitásoknak hívjuk, szemben az enzimek natív vagy főaktivitásával. Jellemző a rejtett aktivitásokra, hogy széles körben elterjedtek a fehérjék között, nagyságrendekkel gyengébbek a natív aktivitásoknál, így közvetlen élettani hatásuk csekély, mutációk révén azonban felerősödhetnek (Khersonsky and Tawfik, 2010).

Annak megértése, hogy hogyan alakulnak ki új útvonalak az anyagcsere-hálózatban, az evolúcióbiológia és a rendszerbiológia egyik kulcskérdése. Az uralkodó elmélet szerint új útvonalak az anyagcserében már meglévő enzimekből épülnek fel, melyek a hálózat különböző pontjairól származnak. Ebben az evolúciós

folyamatban kulcsszerepet játszanak a látens aktivitások, lehetővé téve, hogy az enzimek olyan funkciót is végrehajtsanak (még ha nagyon alacsony aktivitással is), amelyekre nem adaptálódtak. Integrált anyagcsere-hálózatos megközelítésünkkel a látens aktivitások fenotípusos hatását vizsgáltuk új környezetekhez való adaptációban.

Célkitűzések

A látens enzimaktivitások, mint evolúciós nyersanyag tényleges evolúciós potenciáljáról csak nagyon keveset tudunk. Vajon a látens aktivitások milyen gyakran eredményeznek evolúciós újítást (új fenotípust)? Vajon tipikusan a natív hálózattól elszigetelten helyezkednek el, vagy a többségük beköthető abba és ezáltal elvben bekapcsolódhat a sejt anyagcsereforgalmába? És milyen fenotípusos előnnyel járhat a látens aktivitások felerősödése? Ezeknek a kérdéseknek a megválaszolásához összegyűjtöttünk egy átfogó listát az *Escherichia coli* enzimek kísérletesen meghatározott látens aktivitásaiból, majd ezeket a biokémiai reakciókat beépítettük egy genomléptékű anyagcsere-hálózati modellbe.

Módszerek

- *E. coli* látens aktivitások összegyűjtése és a natív anyagcsere-hálózatba építése. A reakciók fő forrása a BRENDA adatbázis és irodalmi adatok voltak.

- FBA módszer alkalmazása a látens reakciók hatásának *in silico* elemzésére.
- Fehérje túltermeléses kísérletek a látens aktivitások hatásának vizsgálatára.

Eredmények

1. A látens reakciók beköthetőek a natív *E. coli* anyagcsere-hálózatba

Az összegyűjtött látens reakciókra megvizsgáltuk, hogy a szubsztrátjaik és termékeik megtalálhatóak-e a natív anyagcseremodellben. Azt találtuk, hogy a látens reakciók 45%-a teljesen beköthető volt a natív hálózatba, és csupán 22%-uk helyezkedik el a natív reakcióktól izoláltan. Azáltal, hogy a látens reakciók többsége beköthető a natív hálózatba, képesek új biokémiai útvonalakat kialakítani a natív reakciókkal közreműködve.

2. A látens reakciók jelenléte fitneszelőnyt biztosít bizonyos környezetekben

A látens reakciók evolúciós előnyének meghatározásához kiszámoltuk a fitnesz-hozzájárulásukat nagyszámú *in silico* környezetben. 19 olyan környezetet találtunk, ahol a látens aktivitások jelenléte elengedhetetlen a növekedéshez. Ez a 19 reakció 2.9%-kal növelte meg az anyagcseremodell által felhasználható nutriensek számát. Figyelembe véve, hogy a natív modell reakcióinak számát a teljesen beköthető látens reakciók hozzáadásával alig 10.8%-kal terjesztettük ki, megállapítható, hogy a

rejtett enzimaktivitásoknak egy tekintélyes hányada evolúciós újítást eredményez. Emellett további 31 környezetben jelentettek a látens reakciók növekedési előnyt.

3. A látens reakciók egyedi hatásának vizsgálata

A következő lépésben megvizsgáltuk, hogy egyetlen látens reakció hozzáadása a natív modellhez milyen előnnyel jár a korábban vizsgált környezetekben. Kiderült, hogy minden olyan környezetben, ahol a látens reakciók hiányában a natív modell nem tud nőni, elegendő egyetlen látens reakció a növekedéshez. Ugyanakkor ritka, hogy ugyanaz a látens reakció több környezetben is fitneszelőnyt nyújtson, tehát a látens reakciók hatása környezetfüggő.

4. Az *in silico* előrejelzések jó egyezést mutatnak a kísérletesen meghatározott metabolikus újításokkal

Eredményeinket egy, az *in silico* szimulációkhoz hasonló rendszerszintű laboratóriumi vizsgálattal támasztottuk alá. A kísérletek alapja, hogy egy adott enzim látens aktivitásait felerősíthetjük, ha megnöveljük az enzim koncentrációját a sejtben, tehát túltermeltetjük. Kísérleteinkben 4269 *E. coli* gén túltermeltetésének hatását vizsgáltuk meg 194, különböző szénforrást tartalmazó környezetben. A laborkísérleteket Kintses Bálint kollégám végezte. A kísérleteket az *in silico* szimulációkkal összevetve megállapítottuk, hogy a modellezés segítségével sikerült előrejelezni a kísérletesen feljött környezetek 44%-át és a kísérletes és *in silico* vizsgálatok során feljött környezet – gén párok közötti átfedés több mint 2 nagyságrenddel felülmúlja a véletlenszerűen

vártat. Ez a nagymértékű átfedés arra utal, hogy a rejtett aktivitások általunk összeállított hálózata már így is jelentős részét lefedi annak az evolúciós nyersanyagnak, mely az új környezetekhez való rövidtávú adaptáció során az élőlény rendelkezésére áll.

Irodalomjegyzék

Boone, C., Bussey, H., and Andrews, B.J. (2007). Exploring genetic interactions and networks with yeast. Nat. Rev. Genet. *8*, 437–449.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. Science *327*, 425–431.

Khersonsky, O., and Tawfik, D.S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. Annu. Rev. Biochem. *79*, 471–505.

Lehner, B. (2007). Modelling genotype–phenotype relationships and human disease with genetic interaction networks. J. Exp. Biol. *210*, 1559–1566.

Notebaart, R.A., Szappanos, B., Kintses, B., Pál, F., Györkei, Á., Bogos, B., Lázár, V., Spohn, R., Csörgő, B., Wagner, A., et al. (2014). Network-level architecture and the evolutionary potential of underground metabolism. Proc. Natl. Acad. Sci. *111*, 11762–11767.

Papp, B., Notebaart, R.A., and Pál, C. (2011). Systems-biology approaches for predicting genomic evolution. Nat. Rev. Genet. *12*, 591–602.

Phillips, P.C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. *9*, 855–867.

Segrè, D., DeLuna, A., Church, G.M., and Kishony, R. (2005). Modular epistasis in yeast metabolism. Nat. Genet. *37*, 77–83.

Szappanos, B., Kovács, K., Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M.J., Jelasity, M., Myers, C.L., et al. (2011). An integrated approach to characterize genetic interaction networks in yeast metabolism. Nat. Genet. *43*, 656–662.

Publikációs lista

Szappanos, B.*, Fritzemeier, J.*, Csörgő, B.*, Lázár, V., Lu, X., Fekete, G., Bálint, B., Herczeg, R., Nagy, I., Notebaart, R.A., et al. (2016). Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat Commun* 7, 11607.
IF: 11.329

Karcagi, I.*, Draskovits, G.*, Umenhoffer, K., Fekete, G., Kovács, K., Méhi, O., Balikó, G., **Szappanos, B.**, Györfy, Z., Fehér, T., et al. (2016). Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. *Mol Biol Evol* 33, 1257–1269.
IF: 13.649

Notebaart, R.A.*, **Szappanos, B.***, Kintses, B.*, Pál, F., Györkei, Á., Bogos, B., Lázár, V., Spohn, R., Csörgő, B., Wagner, A., et al. (2014). Network-level architecture and the evolutionary potential of underground metabolism. *PNAS* 111, 11762–11767.
IF: 9.674

Lázár, V.*, Nagy, I.*, Spohn, R.*, Csörgő, B., Györkei, Á., Nyerges, Á., Horváth, B., Vörös, A., Busa-Fekete, R., Hrtyan, M., Bogos, B., Méhi, O., Fekete, G., **Szappanos, B.**, Balázs, K., Papp, B., Pál, Cs. (2014). Genome-wide analysis captures the determinants of the antibiotic cross-resistance interaction network. *Nat Commun* 5.
IF: 11.470

VanderSluis, B.*, Hess, D.C.*, Pesyna, C., Krumholz, E.W., Syed, T., **Szappanos, B.**, Nislow, C., Papp, B., Troyanskaya, O.G., Myers, C.L., et al. (2014). Broad metabolic sensitivity profiling of a prototrophic yeast deletion collection. *Genome Biol.* 15, R64.
IF: 10.810

Lázár, V.*, Pal Singh, G., Spohn, R., Nagy, I., Horváth, B., Hrtyan, M., Busa-Fekete, R., Bogos, B., Méhi, O., Csörgő, B., Pósfai, G., Fekete, G., **Szappanos, B.**, Kégl, B., Papp, B., Pál, C. (2013). Bacterial evolution of antibiotic hypersensitivity. *Molecular Systems Biology* 9.
IF: 14.099

Szalay-Bekő, M.*, Palotai, R.*, **Szappanos, B.**, Kovács, I.A., Papp, B., and Csermely, P. (2012). ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. *Bioinformatics* 28, 2202–2204.
IF: 5.323

Lázár, V.*, Ecsedi, S., Vízkeleti, L., Rákosy, Z., Boross, G., **Szappanos, B.**, Bégány, A., Emri, G., Adány, R., and Balázs, M. (2012). Marked genetic differences between BRAF and NRAS mutated primary melanomas as revealed by array comparative genomic hybridization. *Melanoma Res.* 22, 202–214.
IF: 2.518

Nagy, L.G.*, Házi, J., **Szappanos, B.**, Kocsubé, S., Bálint, B., Rákhely, G., Vágvölgyi, C., and Papp, T. (2012). The Evolution of Defense Mechanisms Correlate with the Explosive Diversification of Autodigesting Coprinellus Mushrooms (Agaricales, Fungi). *Syst Biol* 61, 595–607.
IF: 12.169

Papp, B.*, **Szappanos, B.**, and Notebaart, R.A. (2011). Use of genome-scale metabolic models in evolutionary systems biology. *Methods Mol. Biol* 759, 483–497.
IF: NA

Szappanos, B.*, Kovács, K.*, Szamecz, B., Honti, F., Costanzo, M., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M.J., Jelasity, M.,

Myers, C.L., et al. (2011). An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat. Genet* 43, 656–662.
IF: 35.532

Angyán, A.F.*, **Szappanos, B.***, Perczel, A., and Gáspári, Z. (2010). CoNSEnsX: an ensemble view of protein structures and NMR-derived experimental data. *BMC Struct Biol* 10, 39.
IF: 2.258

Szappanos, B.*, Süveges, D., Nyitray, L., Perczel, A., and Gáspári, Z. (2010). Folded-unfolded cross-predictions and protein evolution: the case study of coiled-coils. *FEBS Lett* 584, 1623–1627.
IF: 3.601

Gáspári, Z.*, Várnai, P., **Szappanos, B.**, and Perczel, A. (2010). Reconciling the lock-and-key and dynamic views of canonical serine protease inhibitor action. *FEBS Lett* 584, 203–206.
IF: 3.601

* első szerzők