

Ph. D. értekezés tézisei

**Az anyagcsere szerkezetének hatása
a genetikai interakciókra és a genomszerveződésre**

Kovács Károly

Témavezető: Dr. Papp Balázs

Biológiai Doktori Iskola

MTA Szegedi Biológiai Kutatóközpont Biokémiai Intézet

SZTE TTIK

2012

Szeged

Bevezetés

Az utóbbi évekeben elérhetővé vált genomikai és nagy léptékű fenotípusos adatok analízise és matematikai modellbe integrálása révén elsőként vált lehetővé a nukleotidszekvenciától a látható fenotípusos jellegekig vezető út szisztematikus, nagy léptékű feltérképezése. Mivel az anyagcsere talán a legrégebb óta kutatott és legrészletesebben feltérképezett sejtes alrendszer, kézenfekvő választás lehet a fenti cél eléréséhez. Disszertációm két munkát tárgyal (Szappanos et al., 2011; Kovács et al., 2009), melyek mindegyike az anyagcsere szerkezetének, az itt ható génpárok relatív viszonyainak genetikai és evolúciós következményeit vizsgálja. Az első kutatási témában az anyagcserehálózat, mint lehetséges eszköz jelent meg, ami segíthet a gének közötti kölcsönhatások jobb megértésében, a másodikban pedig adaptív hipotéziseket vizsgáltunk meg az anyagcsere működésére vonatkozóan, illetve az anyagcsere ható szelekció lehetséges hatását a genom szerveződésére. Disszertációm kérdésfelvetései a genom anatómiájának vizsgálatával egyrészt a genomikához, másrészt az anyagcserehálózatok rendszerszintű vizsgálatával a rendszerbiológia területéhez kapcsolódnak.

Genetikai interakciók modularitása és prediktálhatósága a sörélesztő anyagcserehálózatában

Munkánk során elsőként vizsgálhattuk az anyagcsere nagyléptékű genetikai interakciós térképét (Szappanos et al., 2011). Két gén genetikai interakcióban (episztázis) áll egymással, ha az egyes gének mutációinak hatása nem független egymástól. A genetikai interakciók fontos szerepet játszanak többek között a gének funkcionális kapcsolatainak feltérképezésében, a sokgénés öröklődésű betegségekben, illetve az evolúció lehetséges útvonalainak meghatározásában (Phillips, 2008). A génkölcsönhatások központi jelentősége ellenére a jelenség mechanisztikus háttere és a kölcsönhatások gének közötti megoszlása még kevésbé ismert.

Vizsgálataink alapja a kísérletes kollaborátorunk (Boone labor, Torontó) által rendelkezésünkre bocsátott genetikai interakció adatsor volt, amelynek segítségével kb. 185 000, a modellünkben szereplő metabolikus génpár kvantitatív genetikai interakcióját elemezhetjük. Ezek között a génpárok között el tudtuk különíteni az

egyszeres génkiütések alapján vártnál magasabb és alacsonyabb rátermettségű kettős mutánsokat, vagyis a pozitív és negatív genetikai interakciókat (multiplikatív modell szerint). A rátermettséget a haploid élesztőtörzsek kolóniamérete alapján becsültük.

Célkitűzések

Munkánk két fő célja: (1.) A *Saccharomyces cerevisiae* genetikai interakciói az anyagcserehálózat funkcionális moduljaihoz való viszonyának jobb megértése. (2.) A genetikai interakciók előrejelezhetőségének vizsgálata funkcionális genomikai és anyagcserehálózati információk valamint genomskálájú anyagcseremodell segítségével.

A modularitás és a genetikai interakciók vizsgálata esetében a kísérletes eredményeket egy korábbi számítógépes modell alapján tett előrejelzésekkel vetettük össze. Ezek szerint i) a genetikai interakciók feldúsulnak az azonos funkciójú csoportban ii) e csoportok között a genetikai interakciók gyakran kizárólag negatívak vagy pozitívak (monokromatikusság) (Segre et al., 2005).

Módszerek

- Randomizációs statisztikai tesztek Perl programnyelv alkalmazásával
- Logisztikus regresszió és random forest statisztikai modellek építése és kiértékelése R statisztikai környezetben (R Development Core Team, 2009), a randomforest R csomagot használva.

1. A legtöbb genetikai interakció funkcionális modulok között van. A kísérletes eredmények részben alátámasztották a anyagcseremodellel prediktált mintázatok létét, de egyben a mdellel tett előrejelzések általánosíthatóságának határait is rámutattak. A hagyományosan definiált funkcionális modulokon belüli feldúsulás negatív és pozitív interakció esetében is szignifikáns, de kis mértékű volt (1,4 és 2,2-szeres). Hasonló eredményt kapunk a funkcionális kapcsolat biokémiaailag pontosan értelmezett definíciójával az ún. fluxus kapcsolttsággal (Papin et al., 2004; Price et al., 2004). A kapcsolt fluxus azt jelenti, hogy az egyik reakció aktivitása minden alkalommal a másik

reakció aktivitásával is jár, ami lehet egyirányú (irányított kapcsoltság) vagy kölcsönös (teljes kapcsoltság). Intuitív módon fluxus kapcsoltság esetén pozitív genetikai interakciót várnánk, hiszen ilyenkor definíció szerint az egyik gén kiütésével lenullázott fluxus a másik génhez tartozó reakcióaktivitás megszűnésével jár, ha az nincs is kiütve. Azonban a hagyományos annotációs csoportokhoz hasonló módon a negatív és pozitív genetikai interakciók feldúsulása itt is kismértékű (1,4 és 2,3-szoros). Az interakciók döntő többsége nem kapcsolt génpárok között fordul elő, míg a fluxus kapcsolt párok mindössze néhány százaléka van genetikai interakcióban. Vagyis míg a negatív és pozitív genetikai interakciók egyaránt feldúsulnak a tradicionális funkcionális modulokban és a biokémiailag definiált kapcsolatokban, a legtöbb genetikai interakció mégis különböző funkciókat köt össze.

2. A monokromatikusság korlátozott mértékben jellemző. A modell második predikcióját tesztelve a modulokkal való kapcsolathoz hasonló eredményt kaptunk: a monokromatikusság mértékét szignifikánsan nagyobbak találtuk a randomizációval kapott nulleloszlás alapján vártnál, de ez mindössze 24-34 %-kal több monokromatikus modulpárt jelentett.

3. A legtöbb genetikai interakció nem jelezhető előre megbízhatóan a génpárjellemzők és anyagcserehálózati modell alapján. Megvizsgáltuk, mennyiben tudjuk prediktálni a genetikai interakciókat az anyagcsere génjeire vonatkozó egyéb ismereteink segítségével. Kétféle módszer előrejelzéseit vizsgáltuk: funkcionális genomikai, és az anyagcserehálózatra vonatkozó ismereteinken alapuló statisztikus modellezést és az anyagcsere nagyléptékű biológiai modelljét (FBA) mely képes egyes és kettes génkiütött törzsek növekedési rátájának becslésére. A statisztikus modellezéshez egyrészt genomszintű génpár jellemzőket használtunk korábbi munkákat követve (pl. koexpresszió, közös transzkripciós faktor) (Wong et al., 2004; Ulitsky et al., 2009), másrészt anyagcserehálózati jellemzőket (pl. kódolt reakciók legrövidebb távolsága). Ezután a fenti jellemzők alapján klasszikus statisztikai (logisztikus regresszió) és egy újabb, döntési fák együttesén alapuló (ensemble) adatbányászati módszert (random forest (Breiman, 2001)) alkalmazva osztályoztuk genetikai interakció adatainkat. Az FBA modell esetében a prediktált genetikai interakciók empirikusan alátámasztott aránya akár 50%, illetve 11% lehet negatív illetve pozitív genetikai interakciók esetében, alátámasztva a legerősebbnek előrejelzett genetikai interakciók fiziológiás jelentőségét. Ugyanakkor azonos küszöbértékeknél a modell az empirikus

genetikai interakció adatoknak csak igen alacsony arányát jelzi előre (2,8%, illetve 12,9%; negatív illetve pozitív genetikai interakció). A genomikai és anyagcserehálózati adatokon alapuló statisztikai modellezés, elsősorban a random forest módszer, az FBA-nál legtöbb esetben jobb predikciót eredményezett, de ez az előrejelzések 10%-os empirikus találati aránya felett még mindig csak a kísérletesen negatív genetikai interakciók 30% -át, illetve a pozitív genetikai interakciók 25%-ának előrejelzését jelenti. A becslés jósága akkor sem változik jelentősen, ha az FBA modell által prediktált adatokat (rátermettség és genetikai interakció értékek) hozzáadjuk a statisztikai modellhez, ugyanakkor negatív genetikai interakció esetében növeli a maximális precision értékét. Ez arra utal, hogy az anyagcseremodellben található olyan komplementer információ, ami a funkcionális genomikai és hálózati tulajdonságokból közvetlenül nem nyerhető ki. Összefoglalva, a genetikai interakciók többségét sem a biokémiai modellel, sem a funkcionális genomikai adatsorok és anyagcserehálózati adatok adatbányászati integrálásával nem tudjuk megbízható pontossággal megjósolni.

Az anyagcsereutak felépítésének hatása az operonális génsorrendre *E. coli*-ban

Jól ismert, hogy a bakteriális génsorrend nem véletlenszerű és az operonok gyakran azonos anyagcsereutak vagy fehérjekomplexek tagjait tartalmazzák. Kutatásunkban azt vizsgáltuk, hogy vajon van-e az operonon belül valamilyen rendezettség a gének sorrendjében és ha igen, annak mi az oka. A kérdés megválaszolásához az *Escherichia coli* operonjait vizsgáltuk a rendelkezésre álló nagyszámú anyagcsereútvonalra vonatkozó és operonszerkezeti adat miatt (Keseler et al., 2009).

Célkitűzések

1. Annak az empirikus tesztelése *E. coli*-ban, hogy az operonok génsorrendje a kódolt enzimek anyagcsereútvonalbeli sorrendjét tükrözik-e (kolinearitás).
2. Általános operonexpresszió és kapcsolt lineáris anyagcsereút matematikai modelljének építése, azt vizsgálendő, hogyan hat az operonbeli génsorrend az útvonal produktivitására.

3. Három különböző adaptív forgatókönyv predikcióinak elméleti tesztelése a modell szimulációival.

4. A három kolinearitás magyarázatára szolgáló hipotézis predikcióinak tesztelése *E. coli*-ban funkcionális genomikai adatok alapján.

Módszerek

- Kinetikai modellezés a Copasi programmal (Hoops et al., 2006).
- Randomizációs statisztikai tesztek Perl programnyelvben.

Eredmények

1. A metabolikus operonok kolinearitásának mértéke nagyobb a véletlenszerűen vártnál *E.coli*-ban

Az operonon belüli génpárok kb. 60%-a mutat az útvonallal megegyező sorrendet, szemben az 50%-kal, amit a randomizáció során kapott nulleloszlás alapján várnánk. Mivel nem ismerünk olyan mutációs hatást, ami a kolinearitás mintázatához vezethetne, különböző adaptív forgatókönyveket kerestünk a mintázat létének magyarázatára. Steady-state állapotban az egyes enzimek mennyisége független a génsorrendtől, ezért összetettebb magyarázatot kerestünk a kolinearitásra. Feltételezve, hogy a kolinearitás az útvonal produktivitásának növelésével növelheti a rátermettséget (növekedési rátát), három lehetséges hipotézist vizsgálunk.

2. Poláris operonok léte nem magyarázza a kolinearitást.

A kolinearitás előnyös lehet, ha a gének expressziós szintje monoton csökkenést mutat az 5' végtől a 3' vég irányába (polaritás) (Nishizaki et al., 2007). E hipotézisnek ellentmond, hogy az *E.coli* operonjai esetében nem találtunk összefüggést a csökkenő expressziós trend és a kolinearitás mértéke között.

3. Nem találtunk kapcsolatot az expressziós variabilitással és a kolinearitás mértéke között

A második hipotézisünk szerint a kolinearitás előnye elsősorban az, hogy az anyagcsereutak gyors aktiválását teszi lehetővé környezeti változás esetén. Megfigyelések szerint egyazon operonon belül a gének expressziós időpontja időben késleltetett az operon transzkripció kezdőpontjától távolodva (Alpers & Tomkins, 1965, 1966) ezért ahogy matematikai modellünkben kimutattuk, a megfelelő sorrend a metabolikus útvonal bekapcsolásakor a végtermék gyorsabb megjelenését eredményezi. Amennyiben így van, azt várnánk, hogy elsősorban a különböző környezetekben különbözőféleképpen expresszálódó operonok mutatnak kolinearitást. Bioinformatikai elemzéseink szerint azonban a kolinearitás mértéke nem függ a különböző környezetek között mutatott expressziós változatosságtól.

4. A kolinearitás a sztochasztikus leállás” hipotézissel magyarázható

A „sztochasztikus leállás” hipotézis szerint alacsony génexpresszió esetén az útvonal kevés példányszámú enzimeit véletlenszerűen elveszhetnek a sejt osztódása ill. a fehérjék lebomlása miatt. Matematikai modellünk segítségével bemutattuk, hogy így a kolinearitás ezekben a génekben folyamatos előnyt jelenthet, ugyanis az útvonal gyors újraindítására gyakran van szükség (t.i. állandó környezeti körülmények között is). Hipotézisünket alátámasztja, hogy valóban csak alacsonyan expresszálódó operonok esetén találunk szignifikáns kolinearitást, ahol az útvonal sztochasztikus leállása valószínűsíthető.

Összefoglalva, elsőként mutattuk ki szisztematikusan az operonon belüli génsorrend nem véletlenszerű mintázatát: az alacsonyan expresszálódó metabolikus operonokban a gének a kódolt enzimek reakciósorrendjét tükrözik (Kovács et al., 2009). Több alternatív adaptív hipotézist matematikai modellel vizsgáltunk, majd azok predikcióit empirikus *E. coli* adatokon teszteltük. Konklúzióink szerint a kolinearitás oka az alacsony expressziójú operonok esetében jelentkező sztochasztikus útvonalleállás idejének minimalizálása lehet.

Irodalomjegyzék

- Alpers, D.H., and Tomkins, G.M. (1965). The order of induction and deinduction of the enzymes of the lactose operon in *E. coli*. *Proc. Natl. Acad. Sci. U.S.A* 53, 797–802.
- Breiman, L. (2001). Random forests. *Machine Learning* 45, 5–32.
- Cai, L., Friedman, N., and Xie, X.S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature* 440, 358–362.
- Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic Gene Expression in a Single Cell. *Science* 297, 1183–1186.
- Heinrich, R., and Klipp, E. (1996). Control Analysis of Unbranched Enzymatic Chains in States of Maximal Activity. *Journal of Theoretical Biology* 182, 243–252.
- Kovács, K., Hurst, L.D., and Papp, B. (2009). Stochasticity in Protein Levels Drives Colinearity of Gene Order in Metabolic Operons of *Escherichia coli*. *PLoS Biol* 7, e1000115.
- Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *Resampling Methods in R: The Boot Package* 18.
- Papin, J.A., Stelling, J., Price, N.D., Klamt, S., Schuster, S., and Palsson, B.O. (2004). Comparison of network-based pathway analysis methods. *Trends in Biotechnology* 22, 400–405.
- Price, N.D., Reed, J.L., and Palsson, B.Ø. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology* 2, 886–897.
- R Development Core Team (2009). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing).
- Segre, 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. *Nature Genetics* 43, 656–662.
- Ulitsky, I., Krogan, N.J., and Shamir, R. (2009). Towards accurate imputation of quantitative genetic interactions. *Genome Biol* 10, R140.
- Wong, S.L., Zhang, L.V., Tong, A.H.Y., Li, Z., Goldberg, D.S., King, O.D., Lesage, G., Vidal, M., Andrews, B., Bussey, H., et al. (2004). Combining biological networks to predict genetic interactions. *Proceedings of the National Academy of Sciences of the United States of America* 101, 15682–15687.

Publikációs lista

Fehér, T. *, Bogos, B. *, Méhi, O. *, Fekete, G., Csörgő, B., **Kovács, K.**, Pósfai, G., Papp, B., Hurst, L.D., Pál, C. (2012) Competition between Transposable Elements and Mutator Genes in Bacteria *Mol Biol Evol* 29: 3153

IF: 5,550

Szappanos, B. *, **Kovács, K.** *, Szamecz, B., Honti, F., Costanzo, F., Baryshnikova, A., Gelius-Dietrich, G., Lercher, M.J., Jelasity, M., Myers, C.L., Andrews, B.J., Boone, C., Oliver, S.G., Pál, C., Papp, B. (2011) An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nature Genetics* 43: 656

IF: 35,532

Kovács, K.*, Hurst, L.D., Papp, B. (2009) Stochasticity in Protein Levels Drives Colinearity of Gene Order in Metabolic Operons of Escherichia coli. *PLoS Biol.* 7: e1000115.

IF: 12,916

* első szerző