

# A NIMROD SZUPERGÉNCSSALÁD EVOLÚCIÓJA

Ph.D. értekezés tézisei

Sipos Botond

Témavezető: Dr. Péntes Zsolt  
Konzulens: Dr. Somogyi Kálmán

Szegedi Tudományegyetem

Biológia Doktori Iskola

Természettudományi és Informatikai Kar

MTA Szegedi Biológiai Központ

Genetikai Intézet

2010

Szeged

# Bevezetés

A nemrégiben leírt Nimrod szupergén család számos tagja rendelkezik a baktérium-kötéshez és a fagocitózishoz társítható funkcióval, ami arra utal, hogy ez a gén csoport a veleszületett immunitás fontos elemét képezheti.

A szupercsalád génei által kódolt fehérjéket egy, az N-terminális régióban található szekvencia-motívum (CCxGY/W) és egy, vagy több speciális EGF-szerű domén, az úgynevezett „NIM domén” jelenléte jellemzi. A fehérjék minden esetben rendelkeznek szignálpeptiddel, ami arra utal, hogy funkciójukat az extracelluláris térben látják el.

A szupercsaládon belül a doménstruktúra nagy változatosságot mutat, ami alapján három fő csoport különíthető el. A Draper típusú fehérjék egy N-terminális EMI domént, egy NIM domént és számos „klasszikus” EGF-szerű domént tartalmaznak. A Nimrod C és Nimrod B típusú („poli-NIM”) fehérjék egy változatos felépítésű N-terminális domént, valamint egy ismétlődő NIM doménekből álló régiót tartalmaznak. A Draper és Nimrod C típusú fehérjék, a Nimrod B típusúakkal ellentétben általában rendelkeznek egy transzmembrán doménnel is.

A Draper típusú doménszerkezet széles körben elterjedt, az ilyen típusú fehérjéket kódoló gének megtalálhatóak például a *Caenorhabditis elegans*, a *Drosophila melanogaster* és az ember genomjában. Ezzel szemben ismétlődő NIM doméneket tartalmazó doménszerkezettel rendelkező fehérjéket kódoló géneket eddig csak rovargenomokban sikerült azonosítani.

Egyes, a Nimrod szupercsaládba tartozó gének, a *centaurin gamma* génnel, valamint az *ance* és a CPF gén családotba tartozó génekkel együtt egy nagy időléptékben (300-350 millió év) konzervált klasztert alkotnak („Nimrod klaszter”).

A Nimrod szupercsaládba tartozó génekről számos olyan irodalmi adat áll rendelkezésre, ami ezen géneknek a veleszületett immunitásában játszott szerepére utal. Több organizmusban azonosítottak a fagocitózisban szerepet játszó Draper típusú doménstruktúrával rendelkező fehérjéket. Ilyen például a *C. elegans* Ced-1, a *Drosophila* Draper, valamint a humán MEGF-10 fehérje.

Számos Nimrod C típusú fehérjét kódoló génről is bebizonyosodott, hogy szerepet játszik a fagocitózisban. Ilyenek például a *D. melanogaster eater* és *nimrodC1*. Mindkét gén expresszáldók a hemocitákban. Az Eater proteinek esetében a baktériumkötés közvetlenül is bizonyított. A *nimrodC4 (SIMU)* génről ismert, hogy az apoptotikus törmelék fagocitózisában vesz részt.

Ugyancsak megtalálható egy Nimrod C típusú, a fagocitózisban szerepet játszó proteint kódoló gén a *Sarcophaga peregrina* húslégy genomjában. Egy, a *Holotrichia diomphalia* bogár fajból azonosított Nimrod B típusú fehérje bakteriális lipopoliszacharid receptorként működik és *in vivo* elősegíti az *Escherichia coli* fagocitózist.

Az „összehangolt” („concerted”), „születés és halál” („Birth-and-Death”, rövidítve „BD”) konceptuális modelleket a kezdetektől fogva használják a géncsaládok evolúciójának jellemzésére. Ezen modellek a gének evolúciójának függetlenségére, illetve az ezt megsértő folyamatok (génkonverzió, egyenlőtlen rekombináció) hatásaira koncentrálnak.

A fehérjék jelentős része tartalmazza ugyanazon domén több példányát („domén ismétlődéseket”, „domén repeateket”), illetve kisebb ismétlődő szekvenciaelemeket. A fentebb ismertetett konceptuális modellek ezen géneknél kisebb egységek evolúciójának jellemzésére is alkalmazhatók, mint amilyenek például a poli-NIM

fehérjék NIM doménjei.

A két modell különböző predikciókat ad a közeli fajokból származó szekvenciákban található domének filogenetikájára nézve. Az összehangolt evolúciós eseményeket megélt gének, illetve domének esetében ugyanis gyakran előfordul, hogy a szekvencia-információkból megbecsült topológia nincs összhangban az őket hordozó fajok filogenetikájával. Erre alapozva lehetőség nyílik az összehangolt evolúciós események detektálása.

## Célkitűzések

- A Nimrod szupergercsalád tagjait az előzőekben prediktált doménstruktúrákra alapozva besorolhatjuk ugyan főbb csoportokba (Nimrod A/Draper, Nimrod C, Nimrod B), de a doménstruktúrák nem informatívak a szupergercsalád részletes evolúciós történetére nézve. Vizsgálataink egyik fő célja ezért a szupergercsalád evolúciójának, vagyis az egyes családok filogenetikájának, valamint családok közötti leszármazási kapcsolatoknak a felderítése volt, a molekuláris filogenetika eszköztárának bevetésével. Vizsgálataink során a rovargenomokban fellelhető, „poli-NIM” típusú doménszerkezettel rendelkező fehérjéket kódoló génekre koncentráltunk.
- Bioinformatikai módszerek segítségével vizsgáltuk a szupergercsalád karakterisztikus doménstruktúrájának eredetét.
- Vizsgálataink másik fő célja a Nimrod B és Nimrod C szekvenciákban található ismétlődő NIM domének evolúciójának jellemzése volt, az összehangolt és a BD evolúciós modellek fényében.
- Céljaink közé tartozott még a Nimrod klaszter konzerváltságának vizsgálata, az irodalomra és a nyilvánosan elérhető adatbázisokban található adatokra alapozva.

## Alkalmazott módszerek

- A NIM domének azonosítását és illesztését az általunk betanított profil Rejtett Markov Modellel (pHMM) végeztük, a HMMER 2.3.2 programcsomag felhasználásával.
- A NIM és EGF-szerű, valamint az EGF-szerű és EMI domének közötti hasonlóságot a PRC által számolt lokális-lokális pHMM illesztéseket megjelenítő páros HMM logó módszerrel vizsgáltuk (Logomat-P webszerver).
- Az doménillesztésekre legjobban illeszkedő aminosav-szubsztitúciós modell kiválasztását és a doménfilogenetikák Maximum Likelihood becslését a ProtTest 1.3 segítségével, a domén-filogenetikák Neighbor-Joining rekonstrukcióját a pedig a MEGA 3.1 szoftverrel végeztük.
- A doméneket kódoló szekvencia-részletek közötti szinonim és nemszinonim távolságokat a módosított Nei-Gojobori módszerrel becsültük meg (MEGA 3.1).
- A Nimrod A, Nimrod B és Nimrod C típusú szekvenciák illesztésére több, különböző heurisztikus stratégiákat alkalmazó illesztőprogramot használtunk (Clustal W 1.3, Muscle 3.6, T-Coffee 4.45, ProbCons 1.1, Dialign 2.2). Bizonyos kritériumok alkalmazásával (pl. a CCxGY/W motívumok és a NIM domének illeszkedése, a T-Coffee által számolt konzisztencia-értékek) megpróbáltuk kiválasztani a biológiai szempontból legrelevánsabb illesztéseket. A feltehetően összehangolt eseményeket megélt NIM doméneket tartalmazó szekvenciákat az illesztésekből és a további filogenetikai elemzésekből kizártuk.

- A ProbCons által számolt szekvencia-illesztésekre legjobban illeszkedő aminosav-szubsztitúciós modellt a ProtTest 1.3 segítségével választottuk ki.
- A ProbCons illesztésekben található filogenetikai szignál mértékét likelihood térképezés segítségével vizsgáltuk (TreePuzzle 5.2).
- Klasszikus (Maximum Likelihood - PhyML 3.0, Neighbor-Joining - MEGA 3.1) és Bayes-statisztikai alapokon nyugvó filogenetikai rekonstrukciós módszerek (MrBayes 3.1) segítségével rekonstruáltuk a Nimrod A, Nimrod B és Nimrod C szekvenciák filogenetikáit. A Bayes MCMC analízisek során bináris karakterekké alakítottuk az illesztésekben található gapeket, és megvizsgáltuk, hogy ezek bevonása milyen hatással van a topológia becslésének bizonytalansága.
- A *D. melanogaster* Nimrod paralógok illesztésének és filogenetikájának egyidejű becslése által (Bali-Phy 2.0.0) következtetni próbáltunk a Nimrod A/Draper, Nimrod C és Nimrod B szekvenciák közötti leszármazási kapcsolatokra.
- A Nimrod klaszter konzerváltságának statisztikai szignifikanciáját egy, a géncsaládok jelenlétét is figyelembe vevő közelítő analitikus módszerrel vizsgáltuk, a *D. melanogaster*, *Anopheles gambiae*, *Apis mellifera* és a *Pediculus humanus* fajok genomjaiban található klaszterek bevonásával.
- A filogenetika fákat az iTOL 1.7 webalkalmazás és az APE R csomag segítségével ábrázoltuk. A poszterior faminták kiértékelése során az konszenzushálózatokat a SplitsTree 4.10 segítségével építettük fel.

## Eredmények és értékelés

- A *D. melanogaster*-hez viszonyított konzerváltság szignifikanciáját felmérő tesztek eredményei arra utalnak, hogy a Nimrod klaszter konzerváltsága nem magyarázható egy semleges, a génsorrend teljes átrendeződését feltételező null modellel. Irodalmi adatokra alapozva feltételezzük, hogy a Nimrod klaszter egységét a gének közelségét igénylő, a kromatin-struktúra szintjén ható közös szabályozási mechanizmusok jelenléte tartja fenn. A konzerváltság és a gének expressziójára, valamint funkciójára vonatkozó adatok alapján valószínűsíthető, hogy ez klaszter a rovarok veleszületett immunitásában fontos szerepet játszó „funkcionális modul”.
- A szupercsalád által kódolt fehérjékben található NIM domének aminosav-szekvenciáiból becsült filogenetikai fák arra utalnak, hogy a szupercsalád legtöbb tagjával ellentétben, az *eater* gének által kódolt egyes domének evolúciója az összehangolt modellel írható le legjobban. Ezen következtetésünket a doméneket kódoló nukleotid-szekvenciák közötti becsült szinonim távolságok is megerősítették.
- A NIM domének összehangolt evolúciójára utaló jeleket találtunk az *A. mellifera* Nimrod CI, a *T. castaneum* Nimrod CI és kisebb mértékben a *T. castaneum* Nimrod CII szekvencia esetében is, azonban közeli ortológok hiányában a homogenizáció tényét nem tudtuk megerősíteni.
- Mivel a gének összehangolt evolúciójának vizsgálata során a leggyakrabban alkalmazott filogenetikai módszerek a domének esetében kevésbé praktikusnak bizonyultak, kifejlesztettünk egy, a domén-filogenetikán alapuló vizuális



módszert, amelynek segítségével egyszerűbben felmérhető, hogy a szekvenciák egyes régióiban található domének evolúcióját melyik konceptuális modell írja le a legjobban. A módszer lényege, hogy egy diagramon ábrázolja az egy adott szekvenciapárban található doméneket, valamint a domén-filogenetikából kinyert, a konceptuális modellekre nézve releváns információkat. A módszert egy Perl nyelven írt alkalmazás formájában implementáltuk, ami elérhető a <http://t2prhd.sf.net> weblapon.

- A vizuális módszer segítségével megállapítottuk, hogy az Eater szekvenciák középső régiójában található NIM domének vesznek részt az összehangolt evolúcióban. Ugyancsak megfigyeltük, hogy a domén-ismétlődés N-terminális régiójában a C-terminális régióhoz képest több domén viselkedik a „születés és halál” konceptuális modellnek megfelelően, amit szelekciós tényezőkkel magyarázunk.
- A páros HMM logók, valamint egyéb megfigyelhető hasonlóságok alapján azt feltételezzük, hogy az EMI és NIM domének két EGF-szerű doménből és egy linker régióból jöttek létre egy olyan strukturális átrendeződés-sorozat révén, amiben az inszerciók kiemelkedő szerepet játszottak. Ennek fényében valószínűsíthető, hogy a Draper típusú doménszerkezet egy transzmembrán régiót és egy EGF-szerű doménekből álló ismétlődést tartalmazó „poli-EGF” architektúra leszarmazottja. A NIM és EMI domének eredete jó példája lehet a doménismétlődések rekrutálásának az új funkciót ellátó domének kialakulása során.
- A *D. melanogaster* Nimrod paralógok illesztésének és filogenetikájának

egyidejű becslése során kapott poszterior mintából felépített konszenzus-hálózat alapján feltételezhető, hogy a Nimrod B típusú szekvenciákat tömörítő klád a Nimrod C2 leszármazási vonal részét képezi.

- A filogenetikai analízisek és a modellillesztések nyomán kapott eredményeinket összegezve kidolgoztunk egy hipotézist, ami a szupercsalád evolúciós történetét magyarázza az első jellegzetes doménstruktúra kialakulásától az egyes családok radiációjáig, ami reményeink szerint hasznosnak bizonyul a szupercsalád génjeire irányuló további vizsgálatok során.

# Közlemények jegyzéke

## Az értekezés alapjául szolgáló közlemények

- Somogyi\*, K., Sipos\*, B., Péntzes, Zs., Kurucz, É., Zsámboki, J., Hultmark, D., Andó, I. (2008): Evolution of genes and repeats in the Nimrod superfamily – *Molecular Biology and Evolution* **25**(11):2337–2347 IF: 7.28.

\* - megosztott első szerzős közlemény.

- Sipos, B., Somogyi, K., Andó, I., Péntzes, Zs. (2008): *t2prhd*: a tool to study the patterns of repeat evolution. – *BMC Bioinformatics* **9**: 27 IF: 3.781.

## Egyéb közlemények

- Szabó, K., Bozsó, M., Sipos, B., Péntzes, Zs.: Genetic diversity of great bustard (*Otis Tarda*) populations in the Carpathian Basin – *Conservation Genetics*, közlésre elfogadva.
- Péntzes, Zs., Melika, G., Bozsóki, Z., Bihari, P., Mikó, I., Tavakoli, M., Pujade-Villar, J., Fehér, B., Fülöp, D., Szabó, K., Bozsó, M., Sipos, B., Somogyi, K., Stone, G. N. (2009): Systematic re-appraisal of the gall-usurping wasp genus *Synophrus* Hartig, 1843 (Hymenoptera: Cynipidae: Synergini) – *Systematic Entomology* **34**(4):688-711 IF (2008): 1.808.
- Márkus, R., Laurinyecz, B., Kurucz, É, Honti, V, Bajusz, I., Sipos, B., Somogyi, K., Kronham, J., Hultmark, D., Andó, I. (2009): Sessile hemocytes as a hematopoietic compartment in *Drosophila melanogaster* – *PNAS* **106**(12):4805-4809 IF (2008): 9.38.

- Álmos, P.Z., Horváth, S., Czibula, Á., Raskó, I., Sipos, B., Bihari, P., Béres, J., Juhász, A., Janka, Z., Kálmán, J. (2008): H1 tau haplotype-related genomic variation at 17q21.3 as an Asian heritage of the European Gypsy population. – *Heredity* **101**(5):416–419 IF: 3.823.
- Markó, B., Sipos, B., Csósz, S., Kiss, K., Boros, I., Gallé, L. (2006): A comprehensive list of the ants of Romania (*Hymenoptera: Formicidae*). – *Myrmecologische Nachrichten* **9**: 65–76.
- Schlick-Steiner, B. C., Steiner, F. M., Konrad, H., Markó, B., Csósz, S., Heller, G., Ferencz, B., Sipos, B., Christian, E., Stauffer, C. (2006): More than one species of *Messor* harvester ants (*Hymenoptera: Formicidae*) in Central Europe. – *European Journal of Entomology* **103**(2):469–476 IF: 0.782.