

Bakteriális patogén es ember közötti molekuláris hálózatok vizsgálata

Horváth Balázs

2015

1. Tartalomjegyzék

2. Rövidítésjegyzék

3. Bevezetés

3.1. A bél mikrobióta fontosságának ismertetése

Miért van szükség a bél mikrobióta vizsgálatára?

A humán bél mikrobióta egy komplex ökoszisztéma. A mikrobiomot alkotó sejtek száma nagyjából a humán szomatikus és csírasejtek összegének tízszerese. A bél mikrobiom mind metabolikusan, mind immunológiaiilag komplex kapcsolatban áll az emberrel. Ed- dig több mint három millió nem redundáns mikrobiális gént sikerült kimutatni az em- berben. Ez a nagy genetikai állomány lehetővé teszi, hogy olyan metabolikus folyamatok játszódjanak le a humán bélben, melyeket az emberi sejtek nem képesek végrehajtani. A bél mikrobióta felelős bizonyos glikánok, aminosavak és xenobiotikumok metaboliz- musáért valamint rövid láncú zsírsavak, vitaminok és kofaktorok termeléséért. Emellett kulcsszerepet játszik az immun-homeosztázis fenntartásában. Az immunrendszerrel mind bakteriális mintázatok érzékelő receptorok által, mind a mikro organizmusok által termelt anyagokon keresztül kapcsolatban áll. Ilyen anyag pl. az *Eubacterium spp.* által oligos- zacharidokból képzett vajsav, mely részt vesz az emberi szervezet energia egyensúlyának szabályzásában, valamint karcinogenezis gátló mivel proliferáció gátló és apoptózis indukáló hatása van. Éppen ez okokból a bél mikrobióta tekinthető egy új metabolikus szervnek is. Kapcsolatok mutathatók ki a bél mikrobiom megváltozása és olyan betegségek között mint az IBD (*inflammatory bowel disease*), elhízás vagy a rák. [Karlsson et al., 2011]

A bél mikrobióta vizsgálatának módszerei

Vizsgálatának fő eszközei a „meta omikák”. Ezek közül is a legfőbb eszköz a metage- nomika, de alkalmaznak már metabolomikai, metatranszkriptomikai és metaproteomi- kai technikákat is. A metagenomikai vizsgálatok során a környezetből származó mintát közvetlenül shotgun szekvenálásnak vetik alá. [Karlsson et al., 2011]

Quin és társai 2010-re meghatározták a minimális bél metagenomot. A vizsgálat során Illumina GA short-read alapú technológiával 124 egy kohortba tartozó nordikus és me- diterrán személy székletmintája alapját elemezték. Az ebből kinyert 576,7 gigabázisnyi DNS-ből 3.3 millió nem redundáns mikrobiális gént mutattak ki. Az így kimutatott gének az emberi genom százötvenszeresét teszik ki. A mintákra egészére jellemző, hogy a bennük található gének két fő részre osztható: A legnagyobb csoportba (86%) a sűrűn előforduló

mikrobiális gének, míg a másik fő csoportba pedig a kifejezetten a humán bélflórára jellemző mikrobiális gének tartoznak. Az összes személyből származó vizsgált génhalmaz 99,1%-a *Eubacteria*, 0,8%-a *Archea* és a fennmaradó 0,1%-a pedig vegyesen *Eucaryota* és virális eredetű. A bakteriális eredetű gének összesen 1000-1150 uralkodó baktériumfajhoz tartozhatnak, ami személyenként kb. 160 domináns fajt jelent. A személyekre jellemző nagyjából 160 uralkodó baktériumfaj listái között a személyeket összevetve nagyfokú hasonlóság figyelhető meg. Egy adott személy bél metagenomjának minimálisan 40%-a megtalálható a minták legalább felében. A közelítőleg ezer fajból 75 faj található meg a minták több mint felében és 57 faj van ami a minták nagyobb mint 90%-ban kimutatható. [Qin et al., 2010]

3.2. A szakirodalomban publikált gazda patogén hálózatok

3.3. A Humán-Salmonella kapcsolat ismertetése és hatása az autofágiára

3.4. Ökológiai hálózatok elemzésére használt topológiai mérőszámok

Miért van szükség topológiai mérőszámokra?

A konzervációs biológia az élettudományok azon ága mely a Föld biodiverzitásának megőrzésével foglalkozik. Mivel az összes faj védelme nem megoldható, ezért szükségessé vált olyan fajok kiválogatása melyek kiemelt figyelmet igényelnek konzervációs biológiai szempontból. [Payton et al., 2002] Az 1990-es évek előtt a védelemre való kiválasztás fő szempontja a faj ritkasága volt. A fajok ilyen alapú szelekciója nem veszi figyelembe hogy például az adott taxon kulcsszerepet játszik-e az ökoszisztéma funkciók ellátásában. [Jordán et al., 2007]

Kulcsfajok

1966-ban Robert Paine megalkotta a kulcsfaj koncepciót (*keystone species*). Megfigyelte hogy ha kiesik a Kaliforniai sziklás tengerparti közösségből a *Piaster ochraceus* csúcsragadozó tengeri csillag akkor az egész közösség fajösszetétele összeomlik. A mai legelfogadottabb kulcsfaj definíció szerint ezek olyan fajok, melyek ökológiai hatása aránytalanul nagy az abundanciájukhoz képest. A fogalommal kapcsolatban azonban további kérdések merülnek fel: Milyen hatás számít nagynak? Pontosan mekkora biomassa hányad után

mondható az adott faj ereje aránytalannak? [Payton et al., 2002] Ez utóbbi kérdések megválaszolásához szükség van olyan mérőszámokra, melyek segítségével kvantitatívvá tehető egy adott faj ökológiai fontossága. Másrészt így lehetővé válik a fajkiválasztás során a szubjektivitás csökkentése is. Az ilyen mérőszámok használatával objektív fontossági sorrendet lehet felállítani az adott élőhelyen előforduló taxonok között. [Jordán et al., 2007]

Rangsorolásra használt topológiai mérőszámok az ökológiában

Ma már a kulcsfajok kiválasztása részben ökológiai interakciós hálózatok elemzése alapján történik. A használt hálók kizárólag biotikus-biotikus (faj-faj) kapcsolatokat tartalmaznak. Erre azért van szükség, mert például minden élőlény összekötésben áll a detritusszal és ez eltorzítaná az analízis eredményét. Sőt ilyen esetben a detritusz maga is struktúrális kulcsfajnak számítana. Egy adott fajnak az ökológiai interakciós hálóban betöltött szerepét pozicionális fontossági mérőszámokkal, vagy más néven centralitási indexekkel lehet jellemezni. A konzervációs biológiában sokfajta ilyen mérőszámot használnak, melyeknek közös tulajdonsága, hogy mindegyik valamilyen egyedi tulajdonságra fekteti a hangsúlyt és az alapján rangsorolja a hálózatban szereplő fajokat. Ilyen eltérés lehet két index között például, az hogy az egyik egy adott pont lokális kapcsolati mintázatára, míg a másik az egész hálózatra vonatkozó hatását számszerűsíti. Adott hálóra különböző mérőszámok eltérő fajsorrendeket adnak, de a hasonló tulajdonságok figyelembevételén alapuló mérőszámok között felállíthatók konszenzus fák. [Jordán et al., 2007] **!TODO:** esetleg lehetne még kicsit írni arról, hogy sok ökológiai mérőszám igazából szociológiából jött → Wasserman, S., Faust, K., 1994. Social Network Analysis. Cambridge University Press, Cambridge.

Főbb topológiai mérőszámok

Normalised degree - D

Az adott ponttal kapcsolódó pontok száma elosztva a hálózat összes pontjának számával. [Baranyi et al., 2011]

Closeness centrality - CC vagy C

A pontok száma elosztva az adott pontból eredő azt minden más ponttal összekötő legrövi-

debb topológiai távolságok összegével. [Baranyi et al., 2011] Ez a mérőszám megmutatja, hogy egy adott pontnak mekkora az átlagos távolsága a hálózat összes többi pontjától. Az index kicsi szám olyan pontokra melyek rövid legrövidebb útvonalakon vannak a többi ponttal összekötve. Az ilyen pontok valószínűleg könnyebben elérnek más pontokat vagy nagyobb hatást tudnak gyakorolni más pontokra. Adott i pont átlagos legrövidebb távolságát a többi ponttól a következőképpen lehet kiszámolni: [Newman, 2010]

$$\ell_i = \frac{1}{n-1} \sum_j d_{ij} \quad \text{vagy,} \quad \ell_i = \frac{1}{n} \sum_{j(\neq i)} d_{ij} \quad (1)$$

Ahol:

ℓ_i : Az i pont átlagos legrövidebb távolsága a hálózat többi pontjától.

d_{ij} : Az az i pontot a j ponttal összekötő legrövidebb útvonal (geodézikus útvonal) pontjainak száma.

n : A hálózat pontjainak száma.

A két számítás között stratégiai különbség van. A baloldali egyenlet azt feltételezi, hogy adott pontnak önmagára mért hatása nem releváns a hálózat működésének szempontjából. Azonban még erre az esetre is jellemző, hogy mivel definíció szerint a d_{ii} távolság 0, ezért az összeget ez az érték nem növeli csupán az osztót. [Newman, 2010]

Az ℓ_i érték önmagában még nem centralitási index, mert kis számokat ad a magas központiságú pontokra. Ahhoz, hogy megkapjuk a *Closeness Centrality*-t az ℓ_i inverzét kell vennünk: [Newman, 2010]

$$C_i = \frac{1}{\ell_i} \quad (2)$$

Betweenness centrality - BC

A vizsgálni kívánt ponton áthaladó a hálózat többi pontpárját összekötő legrövidebb utak összege elosztva a hálózat többi pontpárját összekötő összes legrövidebb út összegével. [Baranyi et al., 2011] Ez a mérőszám azt mutatja meg, hogy egy adott pont milyen arányban szerepel a többi pont között futó útvonalakban. A *betweenness centrality* vagy röviden *betweenness* olyan hálózatok jó jellemzője, melyekben valamilyen természetű „áramlás” folyik a pontok között. Ha feltételezzük, hogy egy ilyen hálózat minden kapcsolata között

az áramlás során ugyanannyi kicserélődés történik egy egységnyi idő alatt és a kicserélődés a legrövidebb útvonalakon folyik, akkor az összes geodézikus útvonalon is azonos rátával történik az áramlás. Ez azt jelenti, hogy egy adott ponton átmenő áramlás mennyisége arányos azzal, hogy a hálózat legrövidebb útvonalainak milyen arányában szerepel. [Newman, 2010]

Information Centrality - IC !TODO

Topological importance - TIⁿ

Ez egy teljesen topológiai alapú mérőszám mely összegzi az egy adott pontból kiinduló összes lehetséges n lépéshosszúságú útvonal hatását. A hálózat összes direkt kapcsolatára kiszámítható azok topológiai erőssége:

$$d_{X,Y} = \frac{1}{x} \quad (3)$$

Ahol:

$d_{X,Y}$: Az Y pont hatása X pontra.

x : Az X pont első szomszédainak száma.

Az így kiszámolt közvetlen kapcsolatok hatását egy mátrixban lehet ábrázolni, melynek indexelése a populációdinamika konvencióit követi: d_{ij} jelenti a j pontnak az i pontra gyakorolt hatását. Adott direkt kapcsolat hatásának nagysága a kapcsolat irányától is függ, tehát d_{ij} nem feltétlenül ugyanakkora mint d_{ji} . Egy n lépés hosszú útvonal erejét az ezt alkotó direkt kapcsolatok hatásának szorzataként értelmezzük:

$$d_{p_{XY}}^n = \prod_{i=1}^{n-1} d_{i,i+1}^1 \quad (4)$$

Ahol:

p_{XY} : Útvonal amire igaz hogy $p \in \{X \text{ és } Y \text{ közötti } n \text{ lépés hosszúságú útvonalak}\}$

$d_{p_{XY}}^n$: Az X és Y pontok közötti n lépés hosszú p útvonal ereje.

$d_{n,n+1}^1$: Az útvonal i és $i + 1$ -ik pontja közötti direkt kapcsolat erőssége

Ez alapján egy Y pont hatása X -ra n lépés távolságban:

$$d_{XY}^n = \sum d_{p_{XY}}^n \quad (5)$$

Ahol:

p_{XY} : Útvonal amire igaz hogy $p \in \{X \text{ és } Y \text{ közötti } n \text{ lépés hosszúságú útvonalak}\}$

d_{XY}^n : Az összes Y pontból eredő és X -ben végződő n hosszúságú útvonalak erejének összege.

Mivel a direkt kapcsolatok ereje függ a kapcsolat irányától, így a TI tükrözi a kapcsolat asszimmetrikusságát is. Egy adott pontra TI^n a következő képen számítható ki:

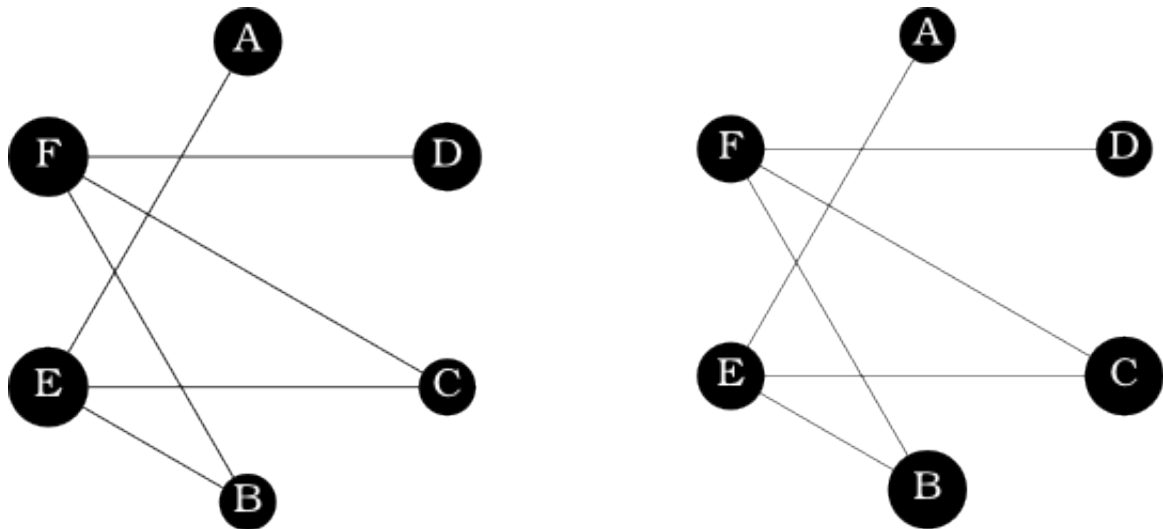
$$TI_A^n = \sum d_{j,A}^n \quad (6)$$

Ahol:

TI_A^n : A pont n lépésre számított topológiai fontossága.

$d_{j,A}$: A és j pont közötti n hosszúságú útvonalak ereje.

A TI^n -t a hálózat összes pontjára ki lehet számítani és ez alapján sorrendet lehet felállítani a nódusok között.



1. ábra. d (bal) és TI^2 (jobb) szemléltetése ugyanazon a példagráfon

A pontok átmérője arányos az adott nódusra kiszámolt d (direkt vagy közvetlen topológiai kölcsönhatás) és TI^2 (topológiai fontosság két lépésre) értékekkel. [Jordán et al., 2003] alapján módosítva.

Az 1. ábrán látható példagráfra rendre felírhatóak a közvetlen kölcsönhatások (d) és a két lépésnyire közvetített indirekt kölcsönhatások (d^2) értékeit tartalmazó mátrixok:

$$\begin{array}{c}
 \begin{array}{c} A \quad B \quad C \quad D \quad E \quad F \\
 \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{3} & 0 \\
 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\
 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\
 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \\
 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\
 0 & \frac{1}{2} & \frac{1}{2} & 1 & 0 & 0 \end{bmatrix} \\
 d \text{ értékek}
 \end{array}
 \quad , \quad
 \begin{array}{c} A \quad B \quad C \quad D \quad E \quad F \\
 \begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 \\
 \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\
 \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\
 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & 0 & 0 \\
 0 & 0 & 0 & 0 & \frac{2}{3} & \frac{1}{3} \\
 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \\
 d^2 \text{ értékek}
 \end{array}
 \end{array}$$

Az ábrázolt mátrixok elrendezése követi a populációdinamikai konvenciókat, tehát például $d_{BF} = \frac{1}{2}$ azt jelenti, hogy F pont a B -re $\frac{1}{2}$ erővel hat. Mindkét mátrixra érvényes az, hogy az adott oszlop értékeinek összege egy. Ez a tulajdonság a d érték definíciójából fakad. d azt mutatja meg, hogy adott pont a cél pont kapcsolatainak hányad részét adja. Ezáltal minden pont egy egységnyi hatást kap ami eloszlik a vele kapcsolatban álló pontok között. [Jordán et al., 2003] Ezt a hatást jól szemlélteti az 1. ábra jobb oldali része amin látható, hogy B , C , D és A pontok kimenő hatása kisebb, mivel célpontjaik sok hatást fogadnak.

Ugyancsak mindkét mátrixra jellemző, hogy a sorok összege azt mutatja meg, hogy egy adott pont mennyire erős kölcsönható, tehát mekkora TI^n értéke. Például B pont két lépés távolságban összesen $\frac{4}{3}$ erővel hat, ez alapján erősebb kölcsönhatónak mondható mint az A pont a maga $\frac{2}{3}$ értékű összesített kimenő két lépés hosszú hatásaival. [Jordán et al., 2003]

Az 1. ábrán az is jól megfigyelhető, hogy C pont a gyengébb közvetlen kölcsönhatók közé tartozik. Ugyanakkor mivel a C -ből eredő két lépéses útvonalak erős elsődleges kölcsönhatókon keresztül érik el végpontjaikat, ezáltal két lépés távolság viszonylatában már C is az erős kölcsönhatók közé tartozik.

Az $n > 1$ lépésszámú d^n értékeket tartalmazó mátrixokban már egy adott pont indirekt hatása önmagára is kiterjedhet. Páros számú lépések esetén viszont mindenképpen

felírhatók olyan útvonalak melyeken a pont eléri önmagát, [Jordán et al., 2003] vagyis $d_{X,X}^n \neq 0$ ha $n \in \{ 2k : k \in \mathbb{Z} \}$. Az 1. ábrán látszik, hogy például az F pont két lépés távolságban a következő útvonalakon hat önmagára: $F \rightarrow B \rightarrow F$, $F \rightarrow C \rightarrow F$ és $F \rightarrow D \rightarrow F$.

Weighted Topological Importance - WIⁿ **!TODO**

4. Források és módszertan

4.1. Adatbázisok

4.1.1. SLK3 Layer 0 összeállítása és ebből a Rho útvonal extrakciója

4.1.2. Arn Core

4.1.3. Salmonet

4.1.4. Predictions

4.2. Az adatbázisok egyesítése

4.2.1. Algoritmusok leírása

4.2.2. Módszerek és technológiák leírása

4.3. Az egyesített hálózat elemzésének módszere

4.3.1. Topológiai mérőszámok jellemzői

4.3.2. Algoritmusok, alkalmazott technológiák

5. Eredmények (A hálózat elemzése)

5.1. Főbb statisztikák

5.2. Kapott topológiai adatok és jelentésük

6. Diskusszió

7. Összefoglalás

8. Summary

9. Hivatkozások jegyzéke

10. Köszönetnyilvánítás

11. Nyilatkozat

Hivatkozások

- [Baranyi et al., 2011] Baranyi, G., Saura, S., Podani, J., and Jordán, F. (2011). Contribution of habitat patches to network connectivity: Redundancy and uniqueness of topological indices. *Ecological Indicators*.
- [Jordán et al., 2007] Jordán, F., Benedek, Z., and Podani, J. (2007). Quantifying positional importance in food webs: A comparison of centrality indices. *Ecological modelling*.
- [Jordán et al., 2003] Jordán, F., Liu, W.-C., and van Veen, F. (2003). Quantifying the importance of species and their interactions in a host-parasitoid community. *Community ecology*.
- [Karlsson et al., 2011] Karlsson, F. H., Nookaew, I., Petranovic, D., and Nielsen, J. (2011). Prospects for systems biology and modeling of the gut microbiome. *Cell*.
- [Newman, 2010] Newman, M. E. J. (2010). *Networks: An Introduction*. Oxford University Press.
- [Payton et al., 2002] Payton, I. J., Fenner, M., and Lee, W. G. (2002). Keystone species: the concept and its relevance for conservation management in New Zealand. *Science for conservation*.
- [Qin et al., 2010] Qin, J., Li, R., Raes, J., Arumugam, M., Bork, P., and Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*.