

Bakteriális patogén és ember közötti molekuláris hálózatok vizsgálata

Diplomamunka
biológus mesterszak
Molekuláris Genetika, Sejt- és Fejlődésbiológia szakirány

készítette:
Horváth Balázs

Témavezetők:
KADLECSIK TAMÁS, PhD hallgató
ELTE Biológiai Intézet, Genetikai Tanszék
KORCSMÁROS TAMÁS, beosztás
TGAC, The Genome Analysis Centre, Norwich, Egyesült Királyság

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
TERMÉSZETTUDOMÁNYI KAR
BIOLÓGIAI INTÉZET



Budapest, 2016

Tartalomjegyzék

1. Rövidítésjegyzék	3
2. Bevezetés	4
2.1. A bél mikrobióta fontosságának ismertetése	4
2.2. A Humán- <i>Salmonella</i> kapcsolat ismertetése és hatása az autofágiára	5
2.3. Ökológiai hálózatok elemzésére használt topológiai indexek	8
3. Célkitűzések	15
4. Források és módszertan	17
4.1. Informatikai módszerek	17
4.2. Az források feldolgozásának eszközei	20
4.3. A források	22
4.3.1. Autophagy Regulatory Network (ARN)	22
4.3.2. Az ARN feldolgozása	22
4.3.3. Salmonet	23
4.3.4. A Salmonet feldolgozása	24
4.3.5. A predikciók	24
4.3.6. A predikciók feldolgozása	24
4.3.7. Az adatbázisok egyesítése	24
4.4. A hálózatok topológiai elemzése	25
4.4.1. A TopologyAnalyser osztály	25
5. Eredmények	28
5.1. Topológiai indexek az alhálózatokban	28
5.1.1. ARN	28
5.1.2. Salmonet	31
5.2. Topológiai indexek az integrált hálózatban	33
5.3. A hármas típusú szekréciós rendszer kapcsolatai	36
5.3.1. A <i>Salmonella</i> kapcsolatai az autofágia fehérjékkel	36
5.3.2. A T3SS kapcsolatai és a <i>Salmonella</i> belső kapcsolatai	38

6. Diskusszió	39
7. Összefoglalás	42
8. Summary	43
9. Köszönetnyilvánítás	44
10.Nyilatkozat	44

1. Rövidítésjegyzék

<i>Rövidítés</i>	<i>Magyarázat</i>	<i>Rövidítés</i>	<i>Magyarázat</i>
ATG3	ubikvitin-szerű-konjugációs enzim ATG3	mTORC1	mammalian target of rapamycin complex 1
ATG8	Autophagy-related protein 8	<i>NCBI</i>	National Center for Biotechnology Information
BC	Betweenness Centrality	NBR1	Next to BRCA1 gén
BECN1	Beclin-1	<i>NDP52</i>	nuclear dot protein 52 kDa
CC	Closeness Centrality	OPTN	Optineurin génje
E1	Ubikvitin aktiváló enzimek	p62 (SQSTM1)	ubikvitin-kötő fehérje p62, Sequestosome-1
E2	Ubikvitin konjugáló enzimek	PI3K	foszfatidilinozitol-4,5-biszfosfát 3-kináz
E3	Ubikvitin ligáz	PPI	fehérje-fehérje kapcsolat (Protein-Protein Interaction)
GABRAP	Gamma-aminobutyric acid receptor-associated protein	PSI	Proteomics Standards Initiative
GPCR	G-fehérje kapcsolt receptor (G-protein coupled receptor)	SCFA	rövid láncú zsírsav (Short Chain Fatty Acid)
GPR41	Free fatty acid receptor 3	SCV	Salmonella Containing Vacuole
GPR43	Free fatty acid receptor 2	SIF	Salmonella által kiváltott filamentumok (Salmonella induced filaments)
IBD	gyulladásos bélbetegség (Inflammatory Bowel Disease)	SPI-I és II	Salmonella patogenitási sziget 1 és 2
IFR	Institute of Food Research	SQL	Structured Query Language
LC3	light chain 3 (ATG8 homológ)	T3SS	Hármas típusú szekréciós rendszer (Type three secretion system)
LC3	Light Chain 3	TrEMBL	Translated EMBL Nucleotide Sequence Data Library
M cell	Microfold sejt	ULK-1	szerin-threonin-protein-kináz ULK1
MAP	Microtubule Associated Protein	VPS34	foszfatidilinozitol-3-kináz VPS34

2. Bevezetés

2.1. 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 Backhed és mtsai., 2005. A bél mikrobiom mind metabolikusan, mind immunológiai komplex kapcsolatban áll az emberrel Karlsson és mtsai., 2011. Az emberi béltraktusban eddig több mint három millió nem redundáns mikrobiális gént sikerült kimutatni Qin és mtsai., 2010. Ez a nagy genetikai állomány lehetővé teszi, hogy olyan metabolikus folyamatok játszódjanak le az emberi bélben, melyeket az sejtjeink nem képesek végrehajtani Karlsson és mtsai., 2011. A bél mikrobióta felelős bizonyos glikánok, aminosavak és xenobiotikumok metabolizmusáért valamint rövid láncú zsírsavak (*short chained fatty acids* - SCFA-k), vitaminok és kofaktorok termeléséért. A gazda által meg nem emésztett poliszacharidok bontását is a bél mikrobióta végzi. Ez utóbbi folyamat eredményeképpen olyan rövid láncú zsírsavak keletkeznek mint az acetát, propionát és vajsav Backhed és mtsai., 2005.

A bélflóra kulcsszerepet játszik az immun-homeosztázis fenntartásában. A mikrobióta az immunrendszerrel bakteriális mintázatokat észlelő receptorokon és G-fehérje kapcsolt receptorokon (GPCR) keresztül van kapcsolatban. A mikroorganizmusok által termelt SCFA-k képesek GPCR-eken révén a sejtszignalizáció indítására. Például a veleszületett immunrendszer nagy részét alkotó monociták és neutrofil granulociták rendelkeznek GPR43 receptorral. A GPR43 egy SCFA érzékeny receptor. Tehát az immunrendszer a patogén mintázatot felismerő receptorok mellett, a metabolitokra érzékeny GPR43-on keresztül is kapcsolatban a bélflórával Brown és mtsai., 2003.

A bélflóra hatással van még a gazda metabolizmusára is. Az *Eubacterium spp.* által oligoszacharidokból képzett vajsav részt vesz az emberi szervezet energia egyensúlyának szabályzásában Karlsson és mtsai., 2011. Az enteroendokrin sejtek és az adipociták is rendelkeznek GPR41 receptorral mely vajsavra és propionátra is érzékeny. Adipocitáknál ez a GPR41 szignalizáció *leptin* elválasztást eredményez Brown és mtsai., 2003. A vajsav segít a karcinogenezis kivédésében mivel apoptózis indukáló és proliferáció gátló hatása van. Éppen ezen okokból a bél mikrobióta tekinthető egy új metabolikus szervnek is Backhed

és mtsai., 2005. Kapcsolatok mutathatók ki a bél mikrobiom megváltozása és olyan betegségek között mint az IBD (*inflammatory bowel disease*), elhízás vagy a különböző rák típusok Karlsson és mtsai., 2011.

A bél mikrobióta vizsgálatának módszerei

A mikrobióta vizsgálatát elsősorban a különböző meta-omikák eszköztárával közelítik meg. Ezek közül is a legfontosabb eszköztár a metagenomika, de alkalmaznak már metabolomikai, metatranszkriptomikai és metaproteomikai megközelítést is. A metagenomikai vizsgálatok során a környezetből származó mintát megfelelő előkészítés után közvetlenül *shotgun* szekvenálásnak vetik alá Karlsson és mtsai., 2011.

Qin és társai 2010-re meghatározták a minimális bél metagenomot. A vizsgálat során *Illumina GA short-read* alapú technológiával 124 egy kohortba tartozó nordikus és mediterrán személy széketmintáját elemezték. Az ebből kinyert 576,7 gigabázisnyi DNS-ből mintegy 3,3 millió nem redundáns mikrobiális gént mutattak ki. Az így kimutatott gének az emberi genom százötvenszeresét teszik ki. A minták 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 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 és mtsai., 2010.

2.2. A Humán-*Salmonella* kapcsolat ismertetése és hatása az autofágiára

Salmonella spp.

A *Salmonella* olyan Gram-negatív patogén mely az állatok széles skáláját képes fertőzni. A tudomány jelenleg több ezer szerotípust ismer, melyek két fő kategóriára oszthatók. Az

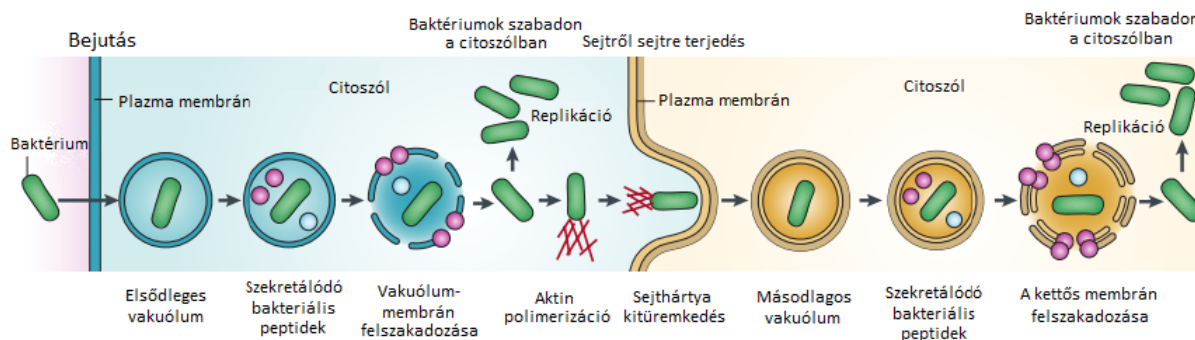
egyik fő típus a *Typhoid*, ebbe a csoportba tartozik a *Typhi* és *Paratyphi* melyek kifejezetten embert fertőznek. A másik fő csoport a *Non-typhoid* amelybe tartozó baktériumok már széleskörű gazdaspecificitással rendelkeznek Haraga, Ohlson és Miller, 2008.

A fertőzés kontaminált étel vagy folyadék fogyasztásával történik. A *Salmonella* az alacsony pH és oxidatív stressz ellen adaptív toleranciával rendelkezik, így képes eltűrni a gyomor savasságát és a veleszületett immunrendszer egyéb hatásait. A vékonybélbe jutva az epithélium sejtjeit fertőzik. Fő célpontjaik a *microfold* (*M cells*) sejtek, melyek fő feladata, hogy pinocitózissal mintákat vegyenek a középbel tartalmából. Amennyiben a felvett anyag károsnak bizonyul, azt antigén prezentáló sejtekhez juttatják Haraga, Ohlson és Miller, 2008. A *Salmonella* másik célpontja a nem fagocita típusú enterociták. Ezekbe a sejtekbe úgynevezett baktérium-közvetített endocitózissal képesek bejutni Haraga, Ohlson és Miller, 2008.

A *Salmonella* életciklusa

Az intracelluláris baktériumok életciklusa általánosan három stádiumra osztható: A bejutáshoz használt vakólum elhagyása, replikáció a citoszólban és a citoszolikus veleszületett immunitás elemeinek manipulációja. A *Salmonella* az úgynevezett *trigger* mechanizmussal jut be a sejtbe. A mechanizmus során a baktérium olyan fehérjéket juttat be az eukarióta sejtbe, melyek képesek a sejtvázzal kölcsönhatni. Ezek a bakteriális effektorfehérjék nagyfokú sejt váz-átrendeződést váltanak ki az eukarióta gazdában. A folyamat végén a baktérium egy vakólummal határolva a sejt belsejébe kerül Ray és mtsai., 2009. Ezt a struktúrát a szakirodalomban SCV-nek nevezik (*Salmonella containing vacuole*) Haraga, Ohlson és Miller, 2008.

A fagocitózis végeztével a *Salmonella* átesik egy úgynevezett bakteriális felszín átformázáson (*bacterial surface remodeling*). A folyamat során gátlódik az olyan bakteriális gének kifejeződése amelyeket a gazda könnyen fertőzési jelnek tekinthet. Ilyen gének például a SPI1, a T3SS és a flagellin. Mindezek mellett megváltozik a baktériumok felszíni lipopoliszacharid mintázata is Haraga, Ohlson és Miller, 2008.



1. ábra. Az intracelluláris baktériumok életciklusa

Bejutáskor a baktériumok egy elsődleges vakólumba érkeznek. A sejt a belsejében a mikrobák olyan fehérjéket szekretálnak, melyek felbontják az őket határoló elsődleges vakólum membránját. A legtöbb intracelluláris baktériumra jellemző, hogy befolyásolni tudja az aktin polimerizációt és ezáltal képes az intra- és intercelluláris mozgásra. A szomszédos sejtbe átjutott baktériumok egy másodlagos membránburokba kerülnek, melyet ugyancsak felbontanak.

Normális körülmények között a vakólum pH-ja mindaddig fokozatosan csökken amíg érett degradatív fagolizoszómává nem válik. A baktériumok kétféleképpen képesek életben maradni ebben a környezetben: A vakólum-lizoszóma fúzió gátlásával, vagy a fagolizoszóma összetételének aktív módosításával Ray és mtsai., 2009. A szakirodalomban még nincs kialakult álláspont arról, hogy a *Salmonella* melyik mechanizmust használják. Bizonyítottan képesek életben maradni, olyan SCV-ben mely már fuzionált a lizoszómával, viszont a fő útvonal valószínűleg a vakólum savanyítási folyamatának késleltetése lehet Haraga, Ohlson és Miller, 2008.

Az SCV-n belül a *S. Typhimurium* képes a replikációra. A hármas típusú szekréciós rendszer segítségével a baktériumsejtek olyan anyagokat tudnak kibocsájtani, melyek lehetővé teszik az SCV-ből kijutást és citoplazma invázióját Jo és mtsai., 2013.

A hármas típusú szekréciós rendszer (T3SS vagy TTSS)

A T3SS evolúciósan a flagelláris export rendszerrel mutat rokonságot. Jelenléte esszenciális ahhoz, hogy a *Salmonella* képes legyen a fertőzésre és gazda sejtjeinek kolonizálására. A T3SS felelős a baktérium virulencia vagy effektor fehérjéinek átviteléért. Az effektorok az eukarióta sejtbe jutva megváltoztatják annak sejtfuncióit. Az virulenciafehérjék átalakítják a gazda citoszkeleton architektúráját, membrán anyagáramlását, szignál transzdukcióját és citokin expresszióját, ezzel segítve a baktériumok túlélését és további kolonizációját Haraga, Ohlson és Miller, 2008.

A *Salmonella* és az autofágia kapcsolata

Az autofágia egy intracelluláris katabolikus folyamat melynek szerepe van a fehérje aggregátumok és károsodott sejtorganellek eltávolításában és a veleszületett immunrendszer működésében. A *xenofágia* az autofágiának azon formája mely során az intracelluláris baktériumok és vírusok szelektív felismerése és lebontása történik. A szelektív felismerésért az autofágia adaptor fehérjéi felelősek. Ilyen receptor fehérje például a p62 (SQSTM1), a NDP52, optineurin (OPTN) és az NBR1. Az előbb felsorolt receptorok a szubsztrátjuk megkötése után kargo adaptorként viselkednek az LC3 (ATG8) számára. *Salmonella* fertőzéskor a sérült SCV-ből kilépett baktériumok sejtfelszíni fehérjéi poliubiquitin borítást kapnak amit a kargo adaptor fehérjék érzékelnek. *S. Typhimurium* fertőzéskor a poliubiquitinált baktériumokat NDP52 és a p62 is felismeri. Az így megkötött baktériumok xenofágia útján eltávolítódnak Jo és mtsai., 2013.

2.3. Ökológiai hálózatok elemzésére használt topológiai indexek

Miért van szükség topológiai mérőszámokra?

A konzervációbiológia 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, Fenner és Lee, 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, Benedek és Podani, 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 nagyknak? Pontosan mekkora biomassa hányad után mondható az adott faj ereje aránytalanul nagy? Payton, Fenner és Lee, 2002? Ez utóbbi kérdések megválaszolásához szükség van olyan mérőszámokra, melyek segítségével kvantitatív

tehető egy adott faj ökológiai fontossága. Másrészt így lehetővé válik a fajkiválasztás során fellépő szubjektivitás csökkentése. Az ilyen mérőszámok használatával objektív fontossági sorrendet lehet felállítani az adott élőhelyen előforduló taxonok között Jordán, Benedek és Podani, 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, Benedek és Podani, 2007.

Főbb topológiai mérőszámok

Normalised degree - D

Az adott ponttal kapcsolódó pontok száma elosztva a hálózat összes pontjának számával Baranyi és mtsai., 2011.

Closeness centrality - CC vagy C

A pontok száma elosztva az adott pontból eredő azt minden más ponttal összekötő legrövidebb topológiai távolságok összegével Baranyi és mtsai., 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 és mtsai., 2011. Ez a mérőszám azt mutatja meg, hogy egy adott pont milyen arányban szerepel a többi pont között futó útvonalakon. 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.

Topological importance - TIⁿ

A diplomamunkám során ezt a mérőszámot alkalmazom az általam összeállított molekuláris hálózatokra. Ezen okból a következő néhány bekezdésben, egy jóval részletesebb leírásban mutatom be a TI tulajdonságait.

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:

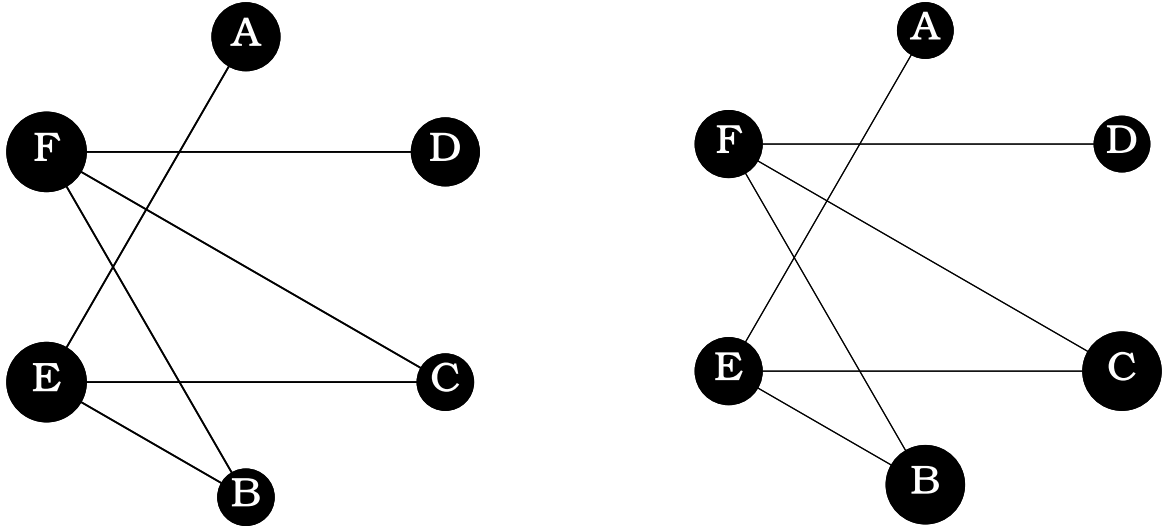
$$TI_A^n = \sum d_{j,A}^n \quad (6)$$

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.



2. ábra. TI^1 (bal) és TI^2 (jobb) szemléltetése ugyanazon a példagráfon
A pontok átmérője arányos az adott nódusra kiszámolt TI^1 (bal) és TI^2 (jobb). Jordán,
Liu és Veen, 2003 alapján módosítva.

A 2. 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, Liu és Veen, 2003). Ezt a hatást jól szemlélteti a 2. ábra bal oldali része amin látható, hogy az A , B , C és D 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, Liu és Veen, 2003).

Az 2. 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, Liu és Veen, 2003 vagyis $d_{X,X}^n \neq 0$ ha $n \in \{ 2k : k \in \mathbb{Z} \}$. Az 2. 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 $F \rightarrow D \rightarrow F$.

3. Célkitűzések

A diplomamunka célja

A diplomamunkám célja egy több adatbázisból integrált fehérje-fehérje kapcsolatokat tartalmazó humán-*Salmonella* gazda-patogén hálózat létrehozása és az így elkészült hálózat topológiai elemzése.

Az elkészítendő hálózatnak a következőket kell tartalmaznia:

1. kézzel gyűjtött *H. sapiens* fehérje-fehérje kapcsolatok
2. kézzel gyűjtött *Salmonella* fehérje-fehérje kapcsolatok
3. *H. sapiens* és *Salmonella* közti prediktált fehérje-fehérje kapcsolatok

A topológiai elemzés során kapott adatok alapján véleményt szeretnék alkotni arról, hogy felhasználhatók-e az ökológiában fajok közti kapcsolatok vizsgálatára használt tisztán topológiai adatokon alapuló mérőszámok a molekuláris kapcsolati hálók elemzésére. Valamint, hogy az így előállított rangsorok mennyire korrelálnak a jelenleg használt *Salmonella* és humán belsejteket vizsgáló módszerek eredményeivel.

A célok eléréséhez tervezett feladatok

1. Program írása mely képes az *ARN (Autophagy Regulatory Network)* adatbázis kézzel gyűjtött humán autofágia specifikus fehérje-fehérje kapcsolati rétegének („ARN core”) *MiTab SQL* formátumra átalakítására.
2. Program írása mely képes a *Salmonet* adatbázis kézzel gyűjtött *Salmonella* fehérje-fehérje kapcsolati hálózatának *MiTab SQL* formátumra átalakítására.
3. Programok írása melyek képesek a Krishnadev és Skhirshagar féle humán-*Salmonella* fehérje-fehérje kapcsolati predikciók *MiTab SQL* formátumra alakítására.
4. Program írása mely képes a létrehozott *MiTab SQL* fájlokban a fehérjék azonosítójának *Uniprot* azonosítóra fordítására.
5. Program írása mely képes a már csak *Uniprot* azonosítókat tartalmazó adatbázisok összeállítására.

6. Program írása mely képes hálózatokban a megadott útvonalhosszra kiszámolni a TI-t.
7. A kapott adatsorok értékelése, hálózatok ábrázolása, biológiai relevancia keresése.

4. Források és módszertan

4.1. Informatikai módszerek

A problémák megoldására használt programnyelvek

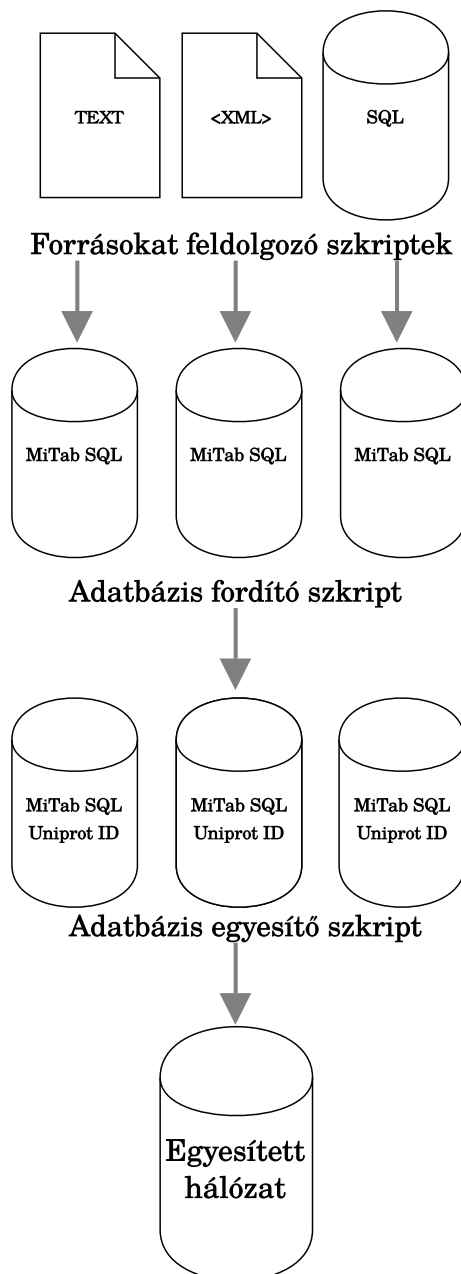
A teljes adatbázisok feldolgozására valamint az adatbázisokból származó adatok rendszerezésére és megfelelő formátumúra alakítására *Python* programnyelven írtam szkripteket. A diplomamunkám során a *Python* 2.7-en és 3.4-en futtatható szkripteket alkalmaztam. A fehérjék azonosítójának fordítását végző szkriptek egyike témavezetőm Kadlecsek Tamás *Javascript*-ben írt fordítószkriptjének kismértékű módosítása.

A diplomamunkám során alkalmazott szkriptek egy részét a *Signalink 3 (SLK 3)* szignalizációs adatbázis kézzel gyűjtött és külső adatokat tartalmazó rétegeinek létrehozásakor készítettem. Mivel a *Signalink* rétegei is több adatbázisból integrálnak fehérje-fehérje interakciókat, így az ott alkalmazott munkafolyamat felhasználható volt a diplomamunkám gazda-patogén hálózatának létrehozásakor is. (3. ábra) A humán-*Salmonella* hálózat szerkezete azonban különbözik a *Signalink 3*-étól. A *Signalink* adatbázisának készítésekor kifejezetten szűrtük például az interspecifikus kapcsolatokat. A *Signalink*kel szemben a humán-*Salmonella* hálózat prediktált éleket is tartalmaz. A két hálózat különbségei miatt, a diplomamunkámban az adatokat kezelő algoritmusok bár hasonlítanak a *Signalink*et létrehozókra, de azokkal nem azonosak.

Az adatok tárolása

Az adatok ideiglenes tárolására, már a *Signalink 3* készítése óta témavezetőm Kadlecsek Tamás javaslatára *SQLite 3* adatbázis fájlokat alkalmazunk. Az *SQLite 3* egy nyílt forráskódú, *C* nyelven írt API-val rendelkező, beágyazott relációs adatbázis motor. Az *SQL* sztenderd szintaxisának nagy részét tartalmazza. Sok népszerű programnyelv rendelkezik már beépített *SQLite* támogatással, ilyen például a *Python* is Owens, 2006.

Az adatok ilyen módú tárolása lehetővé teszi azok gyors szűrését, kategorizálását és átalakítását *SQL* parancsok segítségével. Ilyen módon még azelőtt gyorsan információkat nyerhetünk nagy méretű hálózatokról, mielőtt azokat olyan jóval lassabb működésű hálózat kezelő programokkal elemezni kezdenénk mint például a *Cytoscape*.



3. ábra. A hálózat létrehozásának folyamata

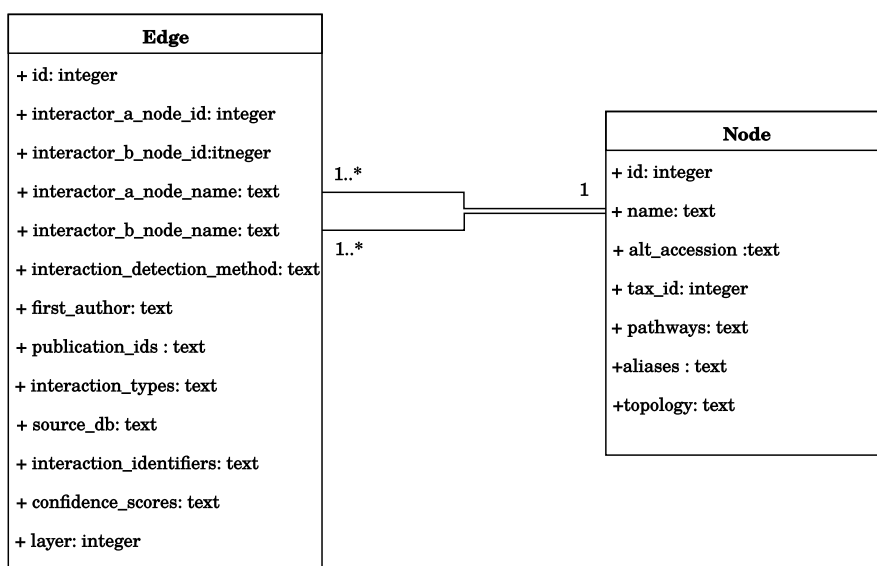
A különböző forrásokból származó adatok esetén először a forrás formátumokat feldolgozni képes szkriptek átalakítják azokat a belső szabványként használt *MiTab SQL* formátumra. Általában a különböző adatforrások különféle azonosítókkal illetik a komponenseiket. Ahhoz, hogy több hálózatot egyesíteni tudjunk, szükség van arra, hogy egy adott biológiai entitás csak egyfajta azonosítóval szerepeljen. A fordító szkript *MiTab SQL* fájlból olyan *MiTab SQL* fájlt készít, amiben az elsődleges azonosító már a kívánt, esetemben *Uniprot* azonosító. Legvégül az adatbázis egyesítő szkript úgy „összefűzi” a különböző hálózatok pontjait és éleit, hogy ne legyen benne redundáns információ.

Az *SQLite* adatbázisfájlok másik előnye, hogy rendelkezésünkre áll az *SQL* nyelv. Mivel *SQL* parancsok segítségével gyorsan kezelhetőek a feldolgozott adatok, így csak ritkán

van szükség adatmanipulálási célból egy újabb szkript írására. Amennyiben mégis szükséges újabb szkript írása, a legtöbb szkriptnyelv rendelkezik valamilyen *SQLite* adatbázis kezelési opcióval. Nagy méretű és mennyiségű biológiai adatot tároló *SQLite* fájlban a keresés is igen gyorsan megoldható az adatbázis beindexelésével, sőt még gyorsabb keresés is megvalósítható az indexelt táblák memóriába csatolásával. Az *SQLite* segítségével könnyen lehet importálni és exportálni a legtöbb népszerű adattárolási formátumba. Az előbb felsorolt okokból a csoportban a szöveges fájlok helyett az *SQLite* adatbázisokat használjuk az adatok köztes tárolására.

Hálózatok tárolására a csoport által létrehozott *MiTab SQL* formátumot használtam. A *MiTab SQL* egy *SQLite* 3-ban tárolt a *PSI-MI Tab* formátummal közel megegyező adatstruktúra. A *PSI-MI Tab* egy *HUPO Proteomics Standards Initiative* (PSI) szervezet által meghatározott proteomikai adatok tárolására használt formátum. A *PSI-MI Tab* formátum specifikációja a szervezet honlapján elérhető. Hermjakob és mtsai., 2004

A pontok és az élek külön táblában vannak letárolva az adatbázisban, így a *PSI-MI Tab* specifikáció pontra és az élre vonatkozó tulajdonságai a megfelelő táblába kerülnek. A *MiTab SQL* táblák oszlopai azonban nem teljesen egyeznek a *PSI-MI Tab* kategóriákkal. Ilyen különbség például, hogy a *MiTab SQL* nem használ néhány opcionális *PSI-MI* kategóriát viszont tartalmaz a *PSI-MI*-re nem jellemző tulajdonságokat is mint a topológia. Az éleket tartalmazó táblában a forrás (*interactor_a_node_name*) és a cél pont név oszlopa a *node* tábla azonosító oszlopának idegen kulcsai. (4. ábra)



4. ábra. A *MiTab SQL* sémája

Verziókövetés

Diplomamunkám készítése során a *Git* verziókövető rendszert használtam, melynek tartalmát a web-alapú *GitHub* tárhely szolgáltatásra töltöttem fel. A diplomamunkám *GitHub* tárhelyén Horváth, 2015 a következő általam írt kódok érhetők el:

- A 3. ábrán ábrázolt munkafolyamatot lebonyolító szkriptek
- A fordításhoz használt adatbázist megépítő szkript (Kadlecsek Tamás szkriptje alapján)
- A fordítást végző szkriptek
- Az adatbázisokat összeépítő szkriptek
- A MiTab SQL formátumot kezelő osztály
- A topológiai elemzést végző szkript
- Az adatszűrésre használt SQL szkriptek

Tesztelés

A bonyolultabb algoritmusok esetén *egységteszteket* (*unit test*) alkalmaztam. A TI-t kiszámító *Python* osztály összes metódusának működését ilyen módon ellenőriztem. A teszteléshez a *python* saját *unittest* nevű csomagját használtam.

Adatok ábrázolása A szakdolgozatomban látható képek és gráfok végleges formáját az *InkScape* nevű ingyenes vektorgrafikus programmal állítottam elő. A gráfokat a *Cytoscape* nevű hálózatelemző programmal készítettem.

4.2. Az források feldolgozásának eszközei

MiTab SQLite adatbázis API

A *PsimiSQL* egy *Python* 2.7-es szintaxisban írt osztály, melyet még a *Signalink 3* összeállításához készítettem, de azóta más projektekben is használtam és továbbfejlesztettem. A *PsimiSQL* segítségével a molekuláris biológiai hálózatok könnyen átalakíthatók *MiTab SQLite* adatbázisokká. Az osztály számos függvényével megkönnyíti a *MiTab SQLite* adatbázisok kezelését *Python* alól. Ilyen függvény például a redundáns adatok adatbázisba

illesztését gátló *insert_unique_node()* mely ellenőrzi, hogy az adott hálózatban szerepel-e már az importálni kívánt pont. Az osztály példányosításakor a memóriában létrejön egy példányhoz kötött *MiTab SQL* sémával rendelkező *SQLite 3* adatbázis. Az adatbázis benépesítése és az adatok keresése tehát nagy sebességgel történik. Az osztálynak vannak olyan függvényei melyekkel könnyen importálni és exportálni lehet *MiTab SQL* adatbázis fájlokat.

A szótárak építése és a fordítás

Ahhoz hogy a feldolgozott forrásokat össze lehessen fűzni egy nagy adatszettbe, szükség van arra, hogy a hálózatokban ne szerepeljen ugyanaz a biológiai entitás más azonosítóval. Ennek érdekében mindegyik hálózat fehérjéit a legfrissebb *Uniprot* adatbázis azonosítókra fordítottam amihez két szkriptet kellett írnom.

A *Salmonet* és az *ARN* már eleve *Uniprot* azonosítókat használ. Azonban az *Uniprot* adatbázis állandó frissítései miatt, fenn áll a lehetőség, hogy nem egy időben készült fájlok ugyanarra a fehérjére más *Uniprot* azonosítót használnak. Egy másik hibaforrás az lehet, hogy a *Uniprot* adatbázis egy fehérjét több azonosítóval is tárol. Amikor egy fehérjét beletesznek a *Uniprot* adatbázisba, akkor kap egy elsődleges azonosítót. Primer azonosítót kapnak még olyan fehérjék is, melyek már benne voltak az adatbázisban de később külön izoformákra lettek szétválasztva. Új elsődleges azonosítót kapnak olyan fehérjék is melyeket több vélt fehérjéből egyesítettek. Minden ilyen művelet után, a legfrissebb elsődleges azonosító marad az új primer azonosító, az összes többi pedig másodlagos azonosítók lesznek. A *Uniprot* azonosítókat még csoportosítani lehet az alapján is, hogy a fehérje manuálisan vagy automatikusan lett annotálva. Az első típusba az úgynevezett *Swissprot* az utóbbiba pedig a *trEMBL* azonosítók tartoznak. A szkriptem az összes pont azonosítójára, ha az nem *Swissprot*, kikeresi a *Swissprot* azonosítót ha létezik, vagy az elsődleges *trEMBL* azonosítót. Az *ARN* és a *Salmonet* fordításához szükség volt egy *Salmonella-Salmonella* és egy humán-humán szótárra, amiket a Kadlecik Tamás szótárépítő szkriptjével állítottam elő. Az azonosítókat a szótárak alapján saját készítésű szkripttel fordítottam.

A predikciókhoz egy olyan szótárat kellett létrehozni, mely *Salmonella* génazonosítókhoz rendel *Salmonella Uniprot* azonosítókat. Több okból is lehetséges ebben az esetben a génről fehérjére fordítás. A *Salmonella* adatbázisok általában csak génazonosítókat tartalmaznak, sőt vannak olyan publikált fehérje-fehérje kapcsolatokat leíró szövegfájlok is

melyek csak génneveket tartalmaznak. Másik fő ok, az hogy *Salmonellára* még nincs megfelelő mennyiségű és felbontású adat. Valamint prokariótákban nincs alternatív splicing, így egy génről készülő fehérjék is jobban megfeleltethetők. Ezt szintén Kadlecsek Tamás szkriptjével állítottam elő. Egy általam írt másik fordítószkript segítségével pedig az előzőhöz hasonló módon fordítottam a predikciókat.

4.3. A források

4.3.1. Autophagy Regulatory Network (ARN)

Az *ARN* egy széles terjedelmű autofágia adatbázis. Az adatbázis az irodalomból kézi gyűjtéssel kapott élek mellett tartalmaz még 19 más adatbázisból importált valamint 4 féle predikcióval készült feltételezett kapcsolatokat is. Az *ARN*-ben található 1485 darab fehérje között 4013 kapcsolat van. Az adatbázis komponensei között vannak az autofágia mechanizmusában szerepet játszó fehérjék és ezek regulátorai valamint transzkripciós faktorai. Az adatbázisban 413 transzkripciós faktor valamint 386 olyan miRNS melyek képesek lehetnek autofágia komponensek szabályzására Turei és mtsai., 2015.

Az ARN hat rétegből épül fel:

1. Autofágia fehérjék.
2. Az első réteg fehérjeinek autofágia specifikus forrásokból származó regulátorai.
3. Olyan poszt-transzlációs regulátorok melyek közvetlenül hatnak az első két réteg fehérjeire.
4. Az első három réteg transzkripciós szabályzói.
5. Az első négy réteg poszt-transzlációs regulátorai.
6. Olyan jelátviteli útvonalak és fehérje-fehérje interakciók melyek különböző útvonalakat az autofágia szabályzóikhoz kötnek Turei és mtsai., 2015.

4.3.2. Az *ARN* feldolgozása

A gazda patogén hálózat összeállításához az *ARN* adatbázisnak csak az első, autofágia fehérjéket tartalmazó rétegét használtam fel. A hálózat letöltését követően azt egy általam írt *Python* scripttel *MiTab SQL* formátumba alakítottam. A fordító szkripttel az akkor

legfrissebb *Uniprot* adatbázis azonosítókra fordítottam. Mindezekre azért volt szükség, mert az *ARN* létrehozásakor még nem használtuk a *MiTab SQL* formátumot.

4.3.3. Salmonet

A *Salmonet* a csoportunk által készített de még nem publikált molekuláris hálózat. A hálózat adatai kézi adatgyűjtésből, nagy áteresztőképességű módszerekből valamint predikciókból származnak. Az integrált hálózat transzkripcionális szabályzási, metabolikus és fehérje-fehérje kölcsönhatási szinteket tartalmaz. A *Salmonet* összesen öt-öt gasztrointesztinális és extraintesztinális *Salmonella* törzs hálózatából lett egyesítve. Egy törzs hálózata a metabolikus, szabályzási és fehérje-fehérje alhálózatok egyesítéséből készült. A *Salmonet* ezen tíz törzs hálózatainak összeállításával jött létre Métris és mtsai., 2015.

A metabolikus hálózatok

Ha egy metabolit egy bizonyos reakció képződménye és egy másik szubsztrátja, akkor a két reakciót katalizáló fehérje kapcsoltnak tekinthető. A reakciók adatai két forrásból származnak. Az adatok egy része a *Salmonella* Typhimurium LT2 kézi gyűjtésű metabolikus modelljéből származik, a másik része pedig a *BioModels* adatbázisból származó predikciókból Métris és mtsai., 2015.

A szabályzási hálózatok

Szabályzási kapcsolatnak tekintjük a transzkripció faktor promóterhez való kötődését. Az ilyen típusú kapcsolatok egy része a szakirodalomból kézzel gyűjtött kísérletesen bizonyított forrásból érkezik. A másik része pedig nagy áteresztőképességű kísérletek adataiból jön Métris és mtsai., 2015.

A fehérje-fehérje kapcsolati hálók

A fehérje-fehérje kapcsolatok is több forrásból kerültek a *Salmonet*-be. Kézi gyűjtéshez a csoportunk által a *Signalink* adatbázis létrehozásához használt módszert alkalmaztuk. A kézi gyűjtés mellett kísérletesen igazolt kapcsolatokat az *IntAct* adatbázisból is integráltuk. Az előző két módszer mellett ez a réteg is tartalmaz predikciókból származó kapcsolatokat Métris és mtsai., 2015.

4.3.4. A Salmonet feldolgozása

Az *Salmonet* átalakítása az *ARN*-hez hasonló módon történt, azzal a különbséggel, hogy fordításkor *Salmonella-Salmonella* szótárat használtam.

4.3.5. A predikciók

Az eddig ismertetett források csak fajon belüli kapcsolatokról felépülő hálózatokat tartalmaztak. Ahhoz, hogy szakdolgozatomban tudjam tanulmányozni a humán-*Salmonella* kapcsolatot szükségem van még interspecifikus élekre is. A predikciós forrásokból származó interspecifikus kapcsolatok fogják összekapcsolni a gazda hálózatát a patogénével. Szakdolgozatomban Krishnadev és Srinivasan, 2011 és Kshirsagar, Carbonell és Klein-Seetharaman, 2012 humán-*Salmonella* predikcióit használtam.

4.3.6. A predikciók feldolgozása

A két predikció feldolgozására külön *Python* szkripteket írtam. Csakúgy mint az előző forrásokat, az így elkészült adatbázisokat a legújabb *Uniprot* azonosítóra fordítottam.

4.3.7. Az adatbázisok egyesítése

Az adatbázisok egyesítésekor a fő szempont az, hogy a végleges hálózatban ne legyenek redundáns pontok vagy élek. Az adatbázis egyesítő szkript beolvassa az összes adatbázisfájlt és egy *has-map*-ben eltárolja azok pontjait és éleit, valamint ezek tulajdonságait. A *hasm-map* adatstruktúra használatával sokkal gyorsabban kikereshető, hogy egy adott él vagy pont benne van-e a hálózatban, mint *SQLite* adatbázisból. Amíg a szkript végigmegy az összes adatbázisfájlon a *has-map*-ek tartalmából létrehoz egy *MySQL* fájlt mely már nem redundánsan tartalmazza az összes forrás adatbázis tartalmát.

Az egyesített adatbázis tartalmazza a:

- A *Salmonet* pontjait és fajon belüli kapcsolatait.
- Az *ARN* pontjait és fajon belüli kapcsolatait.
- A predikciókból származó interspecifikus éleket

A predikciós élek közül is csak azokat tartottam meg, melynek mindkét nódusa megtalálható valamelyik kézi gyűjtésű forrásban. Tehát a predikciókból új pont nem került

bele a hálózatba.

Ezt az adatbázist már újabb *Python* szkript írása nélkül csupán *SQL* lekérdezésekkel hoztam létre.

4.4. A hálózatok topológiai elemzése

4.4.1. A *TopologyAnalyser* osztály

A *TopologyAnalyser* osztály *Python 3* szintaxist használ. Az osztály segítségével kiszámítható a Jordán Ferenc féle TI^n . A *TopologyAnalyser* egyetlen külső függősége a *NetworkX* csomag.

A *NetworkX* egy teljesen *Python*ban írt, komplex hálózatok létrehozására, manipulálására és elemzésére használható ingyenes csomag. Vannak még más ingyenes hálózatkezelő *Python* modulok, mint például a *graph-tool* és az *Igraph*. Az utóbbi kettő csomag már részben *C/C++*-ban írták, így kisebb futásidőre képesek. Választásom mégis azért esett a kisebb teljesítményű *NetworkX*-re mert így felhasználhattam régebben megírt szkriptjeimet. A *Signalink* adatbázisának építésekor ugyanis szükségem volt például XGMML fájlok gyors feldolgozására, és erre a célra a *NetworkX* egy alosztály használtam.

Az osztály konstruktorának egyetlen paramétere egy éllista. Példányosítás után a *TopologyAnalyser* típusú objektum, egy éllistát, egy *Graph* típusú objektumot és egy egység-élerősségi mátrixot tartalmaz.

A *Graph* osztály a *NetworkX* csomag része és irányítatlan gráfok tárolására alkalmas. A hálózat példányhoz kötött *Graph* típusú objektum (*self.graph*) tárolása előnyös, mert így elérhetők a *NetworkX* csomag különböző gráf elemzésre használható metódusai. Ilyen függvény például a *Graph.neighbors_iter(node)* mely egy adott pont szomszédjainak iterálható listáját adja vissza.

Az élerősség mátrix (*self.edgeStrength*) egy példányhoz kötött *Python Dictionary*. A *Dictionary* osztály *Python*ban *hashmap* adatstruktúrával van implementálva, ezáltal a benne tárolt adatok gyorsan elérhetők. A szótár kulcsai a mátrix indexei, az értékei pedig az élerősségek *Fraction* típusú objektumokként letárolva. Az egységerő mátrix celláinak indexei maguk az élt alkotó pontok *Uniprot* azonosítói. Az élerősség mátrix 0 értékkel rendelkező cellái nincsenek letárolva a *hasmap*-ben.

A *Fraction* objektumban a racionális számok két szám hányadosaként tárolódnak. Az

törtszámok ilyen módú tárolása előnyös mert így elég csak akkor tizedestört formátumra hozni a számokat, ha már elvégeztük az összes műveletet. Ilyen módon sokkal pontosabb számításokat végezhetünk. Hátránya, az osztály a matematikából ismert tört műveleteket használja, így több számítást kell végezni.

Az osztály fontosabb metódusai és működésük

- *getNthNeighbors()*

A függvény rekurzívan kikeresi az adott pont n . szomszédját.

- *pathFinder()*

Ez a függvény megkeresi az összes n hosszúságú útvonalat egy forrás és egy cél nódus között:

```
def pathfinder(mélység, forrás_nódus, cél_nódus, útvonal, összegyűjtött_útvonalak):
    if útvonalhossz == 0 then
        forrás_nódus hozzáadása az útvonalhoz ;
        mélység -= 1
    else if mélység == 0 és forrás_nódus == cél_nódus then
        | return útvonal
    else if mélység == 0 és forrás_nódus ≠ cél_nódus then
        | return
    for forrás_nódus első szomszédai do
        | szomszéd_nódus hozzáadása az útvonalhoz;
        | pathfinder( mélység - 1, szomszéd_nódus, cél_nódus, útvonal, összegyűjtött_útvonalak) ;
        | az utolsó pont eltávolítása az útvonalból;
```

A pathFinder metódus algoritmusának pszeudokódja

- *buildEdgeUnitStrengthDict()*

Ez a függvény egyszer hívódik meg a konstruktor lefutása során. A metódus imperatív módon egy **for** ciklussal végigiterál a kapott éllistán. A függvény visszatérési értéke egy egységerő mátrix.

- *getNthPathwaysForNodes()*

A függvény először minden pontnak megkeresi az n . szomszédját. Mint az irodalmi bevezetőben ismertettem, a TI számolásánál páros lépésszámoknál az adott pont

legalább egy hurkon keresztül eléri önmagát. A metódus két egymásba ágyazott **for** ciklust használ. A külső ciklus minden pontra előállítja a nódus n . szomszédjainak halmazát. A belső ciklus pedig minden cél és forrás pont párra meghívja a *pathFinder()* függvényt, majd a kapott útvonalakat imperatívan hozzáadja egy listához, amit még a ciklusokon kívül definiáltam. A metódus visszatérési értéke egy lista amelyben minden lehetséges forrás és célpont közötti n hosszúságú útvonal benne van.

- *getEdgesFromPathway()*

A metódus egy útvonalat fogad és az azt alkotó élek listájával tér vissza.

- *countPathwayStrength()*

A metódus egy útvonalat kap, amire meghívja a *getEdgesFromPathway* függvényt. Végigmegy a kapott éllistán és minden élre kikeresi a megfelelő értéket az egységerő mátrixból. Az összes él erejének összeszorozása után az út erősségével tér vissza.

- *getPathWaysToNthNeighbours()*

Ez a függvény kikeresi egy adott pont összes n . szomszédját és egy listában letárolja. Amikor elkészült a szomszédok listájával, akkor ezen egy **for** ciklus segítségével végigmegy. Ekkor minden periódusban meghívja a forrás és az aktuális cél nódusra a *pathfinder* függvényt, majd a kapott útvonalakat imperatíven hozzáfűzi egy a cikluson kívül definiált listához. A ciklus után a függvény visszatér az elkészült listával.

- *countTI()*

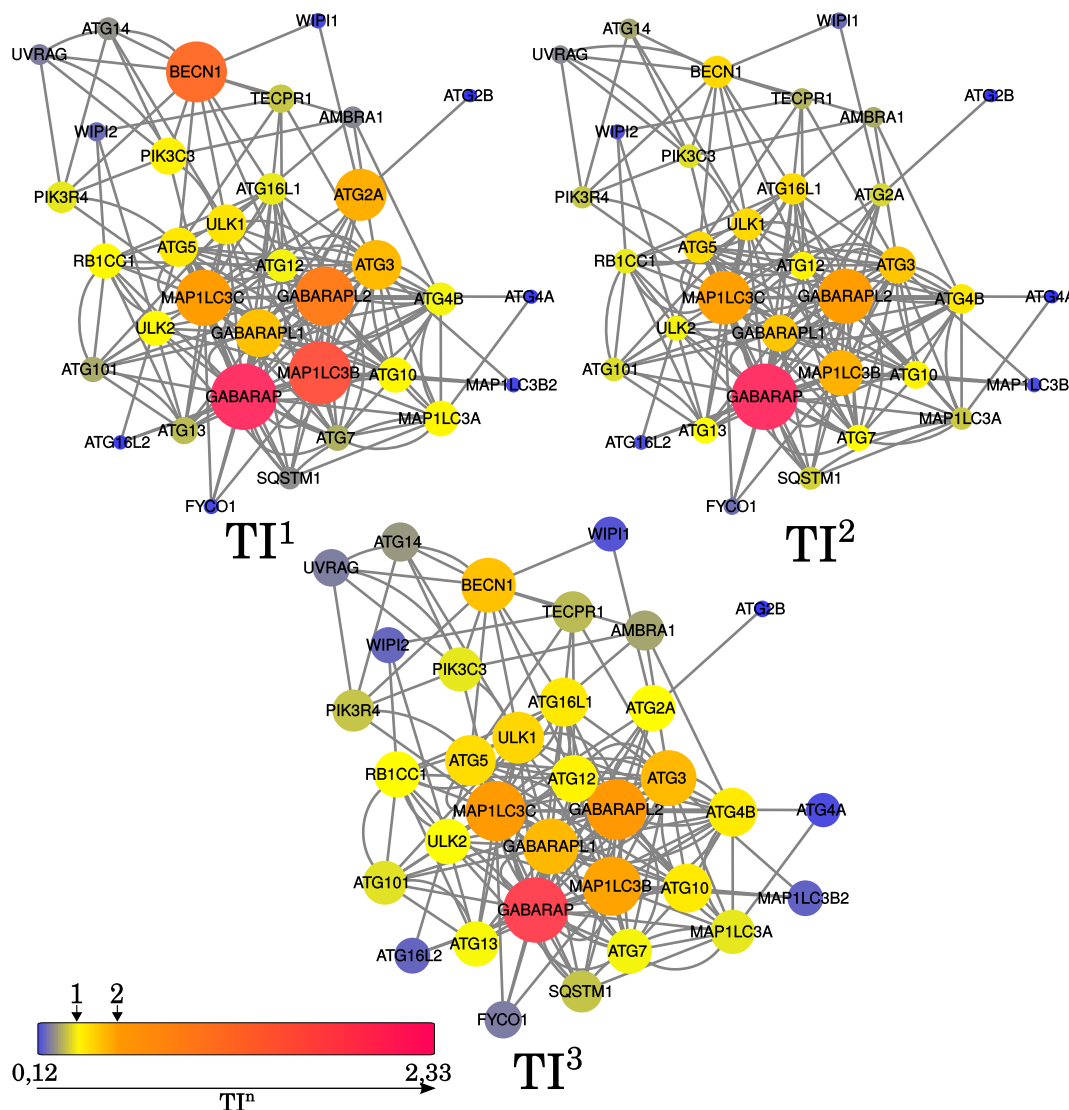
A függvény egy adott pontra meghatározza annak TI értékét. Első lépésként létrehoz egy útvonal listát a *getPathWaysToNthNeighbours* metódus segítségével, majd végigmegy az útvonalak listáján egy **for** ciklusban. Minden periódusban meghívja a *countPathwayStrength(pathway)* függvényt az aktuális útvonalra, és annak eredményét egy a cikluson kívül definiált változóba akkumulálja.

5. Eredmények

Az elemzéseimet a *Salmonet*, az *ARN* és az egyesített hálózatra is elvégeztem. A TI-t mind a három hálóra kiszámoltam 1, 2 és 3 lépés távolságra.

5.1. Topológiai indexek az alhálózatokban

5.1.1. ARN



5. ábra. Az *ARN* 1, 2 és 3 lépésre számított topológiai fontosságai

A pontok színmélységét leíró skála a bal oldalon látható. Az ábra a különböző lépéshosszokra kiszámolt TI értékek eltéréseit szemlélteti. A többfajta lépéshosszra meghatározott indexek információt szolgáltatnak arról, hogy egy adott pont TI értéke hogyan változik egy hálózaton belül.

1. táblázat. Az *ARN* legnagyobb TI értékű pontjai különböző lépésszámoknál

Rang	Azonosító	TI ¹	Azonosító	TI ²	Azonosító	TI ³
1.	GABARAP	2,41	GABARAP	2,21	GABARAP	2,33
2.	MAP1LC3B	2,27	GABARAPL2	1,96	GABARAPL2	2,00
3.	BECN1	2,17	MAP1LC3C	1,93	MAP1LC3C	1,99
4.	MAP1LC3C	2,11	MAP1LC3B	1,75	MAP1LC3B	1,68
5.	ATG2A	1,99	ATG3	1,56	GABARAPL1	1,68

Az egyesített hálózat az *ARN*-nél jóval kompaktabb, az *ARN*-ben pontok a kevésbé gazdagon kapcsolnak egymással. Az egyesített hálózat denzitása 0,00136 %, míg az *ARN*-é 0,00099%. A denzitási értékekkel ellentétesen az *ARN*-ben nagyobb az olyan nódusok aránya melyeknek nagy a TI értéke. Az egyesített hálózatban csupán a pontok 31,9%-a rendelkezik egynél nagyobb TI³ értékkel, ugyanakkor ez az arány az *ARN* esetén már 44,1%.

Az 1. táblázatban ábrázoltam az *ARN*-re kiszámolt TI¹, TI² és TI³ kategória első öt legfontosabb fehérjét. A táblázatból látható, hogy egy adott fehérje különböző lépéshosszoknál számolt topológiai indexe nagyjából hasonló értékeket vesz fel. A TI² adatok mindenképp magukba foglalják azt, hogy a pontok hogyan hatnak vissza önmagukra. A TI³-as értékek is tartalmazhatnak olyan útvonalakat melyeken egy pont önmagára visszahat, de csak akkor ha a két nódusnak van közös első szomszédja. A TI³ tehát magába foglalja az adott pont önhatását de csak közvetett módon. A TI számításánál az egy pontba érő hatás ereje attól függ hogy a pontot hány darab útvonal éri el. Tehát TI számításánál az adott pontba érkező hatások összege mindig egy. Fordítva, egy pont egységnyi erőt képes befogadni, ami a pontot elérő kapcsolatok között oszlik meg. A páros n úthosszoknál az n . szomszédoktól érkező hatások mindenképp gyengébbek lesznek mert a pontra ható egység hatás még az önható útvonalak hatásával is megoszlik.

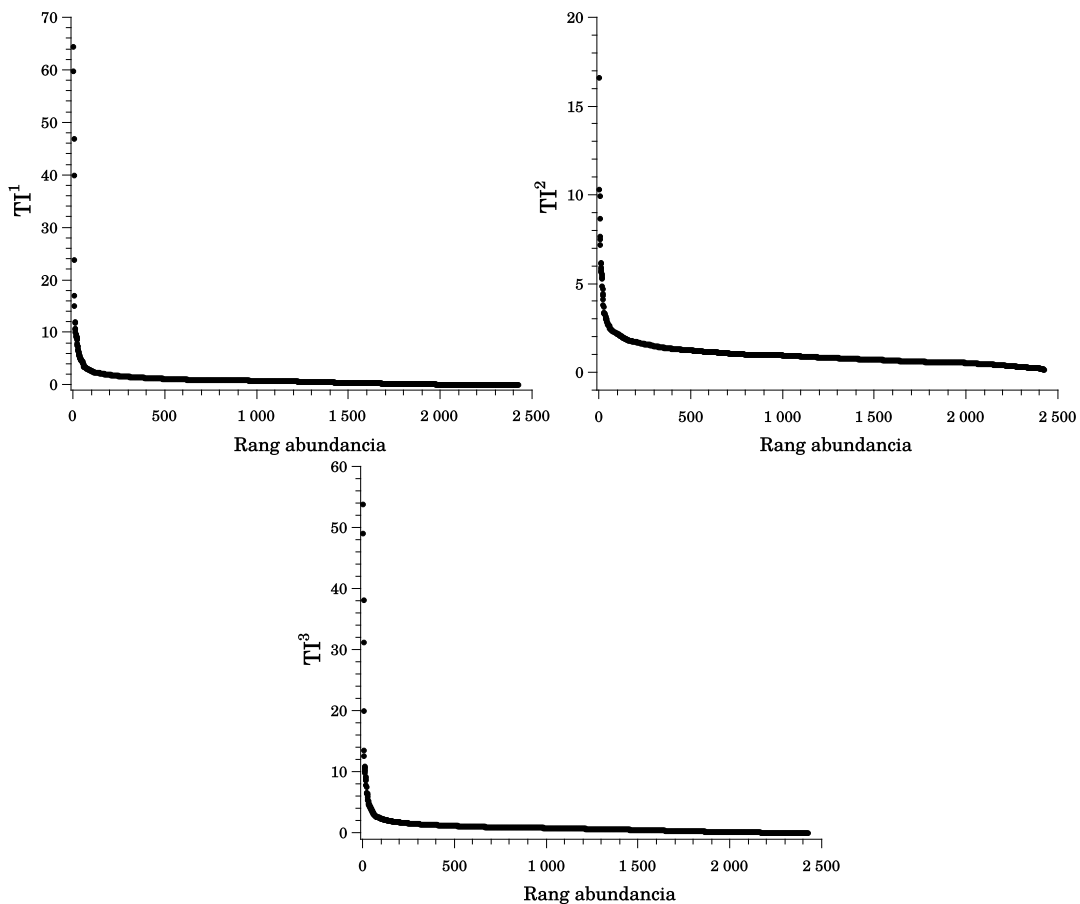
Az *ARN* legnagyobb TI értékű fehérjéi

A következő felsorolásba összegyűjtöttem azokat a fehérjéket amelyek mind a három TI kategóriában megtalálhatók a tíz legfontosabb fehérje halmazában. A TI¹ a közvetlen topológiai hatást veszi legerősebben figyelembe. A három index közül a TI² fekteti a legnagyobb hangsúlyt az önhatásra. A TI³ távolabbi szomszédságot is elérő közvetett hatás szempontjából fontos. A következő nyolc fehérje tehát mindegyik szempontból a benne van a legrangosabb pontok listáján:

- **ATG3** – Más néven *ubiquitine-like-conjugating enzyme ATG3*, egy konjugáló enzim, mely az autofágiában és a mitokondrium homeosztázisban tölt be szerepet. Autofágia során ez az enzim katalizálja a foszfatidiletanolamin konjugációját az ATG8 szerű fehérjék C-terminális glicinjére. Ilyen ATG8 szerű fehérje a listán szereplő GABRAP, GABARAPL1, GABARAPL2 valamint ezen a listán nem szereplő MAP1LC3A is Glick, Barth és Macleod, 2010.
- **BECN1** – Más néven *Beclin-1*. Az autofágia egyik kulcs fehérjéje. Bizonyos jelátviteli útvonalak hatására a VPS34-el együtt a fagofór kialakulását szabályozza. A *Beclin-1* egy PI3K komplex alegysége, VPS34-el való kapcsolódása fokozza annak katalitikus hatását Glick, Barth és Macleod, 2010.
- **GABRAP, GABARAPL1 és a GABARAPL2** – A GABRAP a nevét onnan kapta, hogy neuronokban a GABA_A receptorokhoz kapcsolódik. A GABRAP egy 14kDa-os citoplazmatikus molekuláris *chaperone*. Ma már tudjuk, hogy a fehérje ennél sokkal szélesebb körű kapcsolatrendszerrel rendelkezik. Autofágiában fagofór formálódásánál csakúgy mint az LC3-ak, a GABRAP-ok is liopidálódnak. Amikor egy sejtben az autofágia felpörög akkor a GABRAP-ek foszfatidiletanolaminra konjugált állapotban találhatók, ezzel ellentétes helyzetben viszont konstitutív proteolitikus bomlásra vannak ítélve Szalai és mtsai., 2015.
- **MAP1LC3B** – Ez a fehérje a MAP-ok családjába tartozik és emberben két majdnem identikus izoformája van. Első izoláláskor a MAP-1 alegységének vélték és ezért kapta LC3 (*light chain 3*) nevet. Az LC3 és a MAP1LC3B név tehát ugyanarra a fehérjére vonatkozik. Ma már tudjuk, hogy a mikrotubulusok mellett, az autofágiában is kulcsszerepet játszik Szalai és mtsai., 2015. Az autofágiában két ubikvitin rendszer szerű enzim-komplex dolgozik. Az LC3 autofágia indukciójakor proteolitikus hasítást szenved a citoplazmában és így jön létre az LC3B-I-es izoforma. Az LC3B-I-et az Atg7 egy E1-szerű enzim aktiválja, majd az E2 szerű ATG3-ra kerül. Ezután LC3B-I C-terminálisán helyezkedő glicinje foszfatidiletanolaminnal konjugálódik. Ez utóbbi formát nevezzük LC3B-II-nek. Az LC3B-II a növekvő fagofór külső és belső felszínén is megtalálható. Jelenlegi ismereteink alapján az LC3B-II két fő szerepei a membránok hemifúziója és a lebontásra irányított fehérjék szelekciója az autofagoszómához Glick, Barth és Macleod, 2010.

- **MAP1LC3C** – Ez a fehérje egy LC3C paralóg, az *LC3C* génről íródik át melynek szabályozása független az LC3-tól. Az LC3-hoz hasonlóan ez a fehérje is a makroautofágia szelektivitását növeli Szalai és mtsai., 2015.
- **ULK1** – Az ULK-1 egy szerin-threonin-protein-kináz, az ATG1 emlős homológja. Az ULK-1 a fagofór kéződés egyik kulcs faktora. Aktivitását az az mTORC1 szabályozza. Az mTORC1 egyik legfontosabb feladata emlős sejtekben az energiaszint érzékelés. Az mTORC1 magas tápanyagtartalom vagy aktiváló jelátvitel hatására foszforillálja az ULK-1 szubsztrátját, így az ULK-1 nem tudja beteljesíteni a fagofór kéződésében játszott szerepét és az autofágia gátlódik Glick, Barth és Macleod, 2010.

5.1.2. Salmonet



6. ábra. A *Salmonet* pontjainak rang abundancia diagramjai

Az ábrán látható grafikonokon a *Salmonet* pontjai a különböző TI értékek szerint sorba vannak rendezve.

A *Salmonet* 2425 nódust és 7973 élt tartalmaz. A *Salmonet* jobban hasonlít az egyesített hálózatra. Az *ARN*-el ellentétben de az integrált hálózathoz hasonlóan, jóval magasabbak a legnagyobb TI-vel rendelkező pontjainak értékei. Sőt a *Salmonet* különböző lépésszámokhoz rendelt öt legfontosabb fehérjének listája megegyezik az egyesített hálózatával. (2. és 3. táblázat)

2. táblázat. A *Salmonet* legnagyobb TI értékű pontjai különböző lépésszámoknál

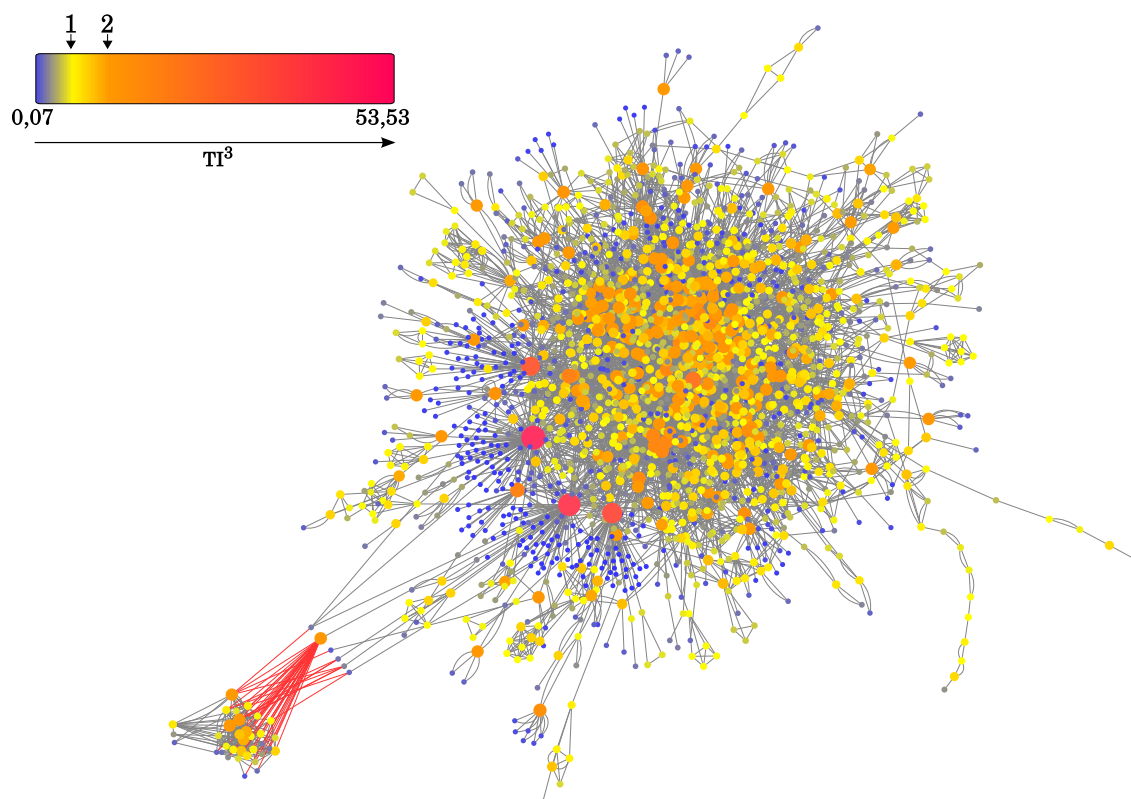
Rang	Azonosító	TI ¹	Azonosító	TI ²	Azonosító	TI ³
1.	phoP	64,46	trxA	16,61	phoP	53,84
2.	ssrB	59,77	phoP	10,30	ssrB	49,06
3.	rpoN	46,89	yajL	9,95	rpoN	38,11
4.	hilD	39,87	hilD	8,70	hilD	31,20
5.	trxA	23,77	ssrB	7,70	trxA	20,11

A *Salmonet*-ben négy olyan fehérje van mely mind a három topológiai index szempontjából a legerősebb tíz helyezettben van:

- **hilD** – *Salmonella* Typhimuriumban a patogénitási sziget 1 (SPI-1) kódolja a T3SS fehérjéit amik az invázióhoz szükségesek. Az SPI-1 génjei többnyire a szigeten kívüli transzkripció faktorok által szabályozódnak. A sziget mester transzkripció faktora a hilA, melyet két másik fehérje mellett a hilD szabályoz. A hilD-nek eddig 17 kötési helyét írták le a *Salmonella* genomban Petrone, Stringer és Wade, 2014.
- **phoP** – A T3SS komponenseit kódoló patogénitási sziget mellett egy másik fontos virulencia rendszer is szerepet játszik a *Salmonella* fertőzés sikerességében. A phoP-phoQ egy kétkomponensű rendszer mely intracelluláris szignálokra aktiválódik, és nélkülözhetetlen a baktériumsejtek túléléséhez és replikációjához. A phoQ egy kináz mely a megfelelő jelátvitel hatására foszforilálja a phoP-t ami ekkor ki tudja fejezni transzkripció aktivátor hatását Lathrop és mtsai., 2015.
- **ssrB** – Az ssrA-ssrB is egy kétkomponensű rendszer mely szintén egy szenzor kinázból és egy regulátor transzkripció faktorból áll. Az ssrA-ssrB rendszer szükséges a baktériumsejtek makrofágokon belüli túléléséhez és replikációjához. A PSI-2 (patogénitási sziget 2) több ssrB promótert is tartalmaz. Az ssrB több gén mellett az SPI-2-ben kódolt T3SS fehérjéinek átírásáért is felelős Walthers és mtsai., 2007.
- **trxA** – A trxA is egy több komponensű rendszer része. A trxA egyike a *Salmonella* genom két ismert thioredoxinjának. A trxA által kódolt fehérje a trx1 egy kicsi

oldható diszulfid reduktáz enzim. A *trx1* sok specifikus célfehérjéinn thiol-diszulfid redox folyamatot hajt végre, így fontos szerepet játszik az oxidatív stressz elleni védelemben. Az oxidált *trx1* a redukált *trxB*-ből regenerálódik, mely NADPH-ból szerzi a redukálóerőt Peters és mtsai., 2010.

5.2. Topológiai indexek az integrált hálózatban



7. ábra. Az egyesített hálózat

A hálózaton piros élek jelölik az *ARN*-t (*baloldalt alul*) és a *Salmonet*-et elválasztó prediktált éleket. Az ábrán a színátmenetek és méretek arányosak a pontokra kiszámolt TI^3 -al. A színátmenetet szemléltető nem lineáris skála a bal felső sarokban látható. A topológiai fontos pontok kiemelése érdekében három csoportot állítottam fel: Az első csoportba azok a pontok tartoznak, melyekből távozó hatás kisebb vagy egyenlő mint a beérkező ($TI^3 \leq 1$). A hálózatban 1674 ilyen nódus található, ami a hálózat 68,1%-át teszi ki. Az utóbbi pontok színe a sötétkéktől a citromsárgáig terjed. A középső a színskálán citromsárgától narancsig terjedő kitöltésű pontok azok melyeknek már van nettó kimenő hatásuk ($1 < TI^3 < 2$). Ebbe a halmazba 624 pont azaz a hálózat 25,4%-a tartozik. A harmadik narancstól vörösig terjedő csoportot azok a nódusok alkotják melyeknek a legnagyobb a TI értékük. Ez utóbbi 161 pont a hálózat 6,5%-át teszi ki.

3. táblázat. Az egyesített hálózat legnagyobb TI értékkel rendelkező pontjai különböző lépésszámoknál

Rang	Azonosító	TI ¹	Azonosító	TI ²	Azonosító	TI ³
1.	phoP	64,38	trxA	16,61	phoP	53,53
2.	ssrB	59,17	phoP	10,31	ssrB	46,15
3.	rpoN	46,21	yajL	9,95	rpoN	37,36
4.	hilD	39,79	hilD	8,71	hilD	31,07
5.	trxA	23,77	ssrB	7,93	trxA	20,10

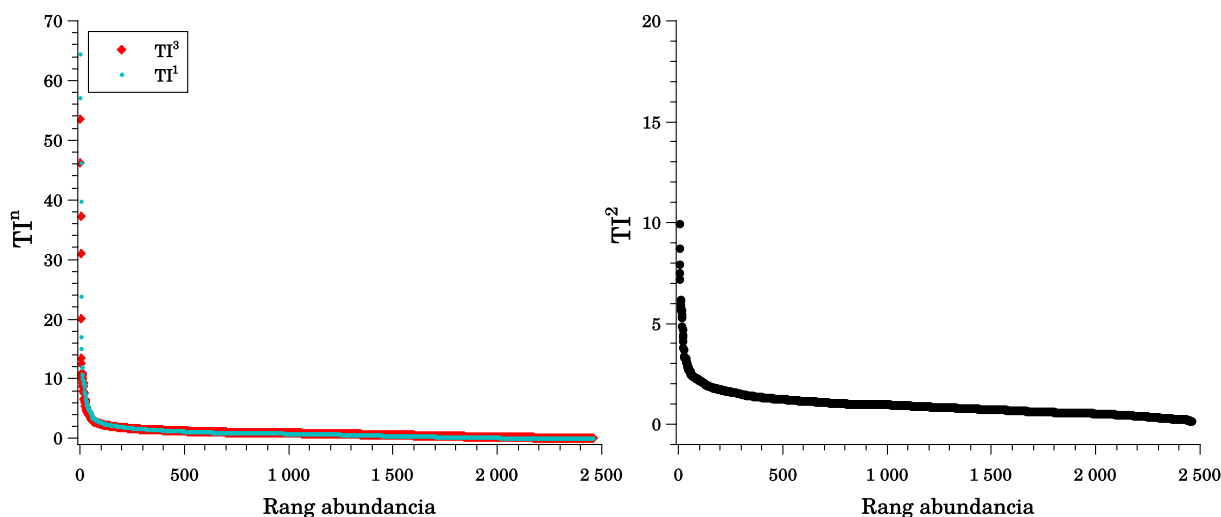
Eddig csak adott hálózaton belül hasonlítottuk össze a pontok topológiai fontosságát. Azonban fontos információt nyújt az is, hogy egy pont TI értéke hogyan változik azzal, hogy az eredeti vagy az egyesített hálózatban vizsgáljuk.

A feldolgozott *ARN* hálózat 34 pontot tartalmaz amik között 186 él van. A *Salmonet*-ben pedig 2425 pont van amik között 7973 él húzódik. A predikciók az *ARN*-ből és *Salmonet*-ből származó 2459 nódus közé még 40 kapcsolatot adnak. A végleges hálózat így 2459 pontot és 8199 élt tartalmaz.

Az integrált hálózatban a *Salmonet* tulajdonságai érvényesülnek. Az első öt nódus mind a három TI kategóriájában *Salmonet* eredetű. (3. táblázat) A hálózatban a legerősebb hatással rendelkező humán fehérje a GABARAP, ami TI¹ (2,54) és TI³ (2,12) esetén a 88., TI²-nél (2,61) pedig a 112. legerősebb hatású pontnak számít.

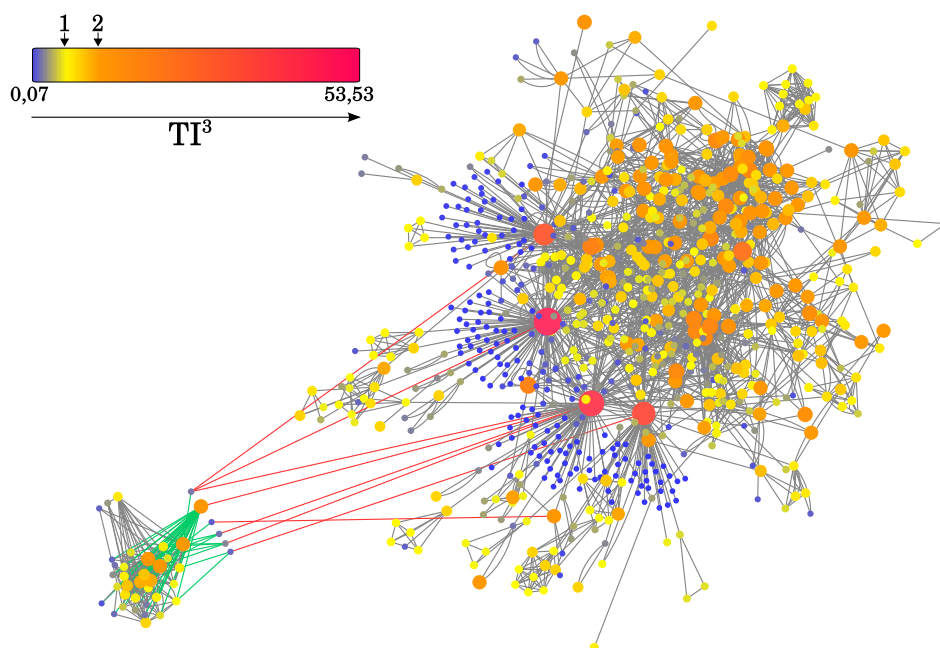
A hálózat fontossági rangsorára felrajzolhatók az ökológiában alkalmazott rang abundancia diagramok. (8. ábra) A rang abundancia diagramok x értékei a pontok rangsor-számai. Az értékkészletük pedig a pontra vonatkoztatott valamilyen mérőszám értékeinek halmaza. Ebben az esetben az y értékek a különböző lépéshosszhoz számított TI-k.

A 8. ábrán megfigyelhető, hogy a hálózatban csupán néhány kiugróan magas TI értékű pont van. Az is jól látszik, hogy a visszahatást figyelmen kívül hagyó TI¹ és a csak közvetett útvonalakon közvetített önhatást figyelő TI³ közel azonos lefutású. TI²-nél minden pont eléri önmagát az összes első szomszédján keresztül. Egy pont hatása a cél pontja fokszámaival fordítottan arányos. A megnövekedett bemenetszámok miatt a TI² értékek is kisebbek, viszont jól szemléltetik azt, hogy az adott pontok mennyire hatnak vissza önmagukra



8. ábra. Az egyesített hálózat rang abundancia görbéi

Az ábra bal oldalán a TI^1 pontjai pirossal, a TI^3 pedig kékkel. Jól látható, hogy a TI^1 és a TI^3 lefutása nagyjából azonos. A jobb oldalon a TI^2 pontjai láthatóak. A két grafikon skálázása alapján leolvasható, hogy a TI^2 értékei kisebbek a TI^1 és TI^3 értékeitől.



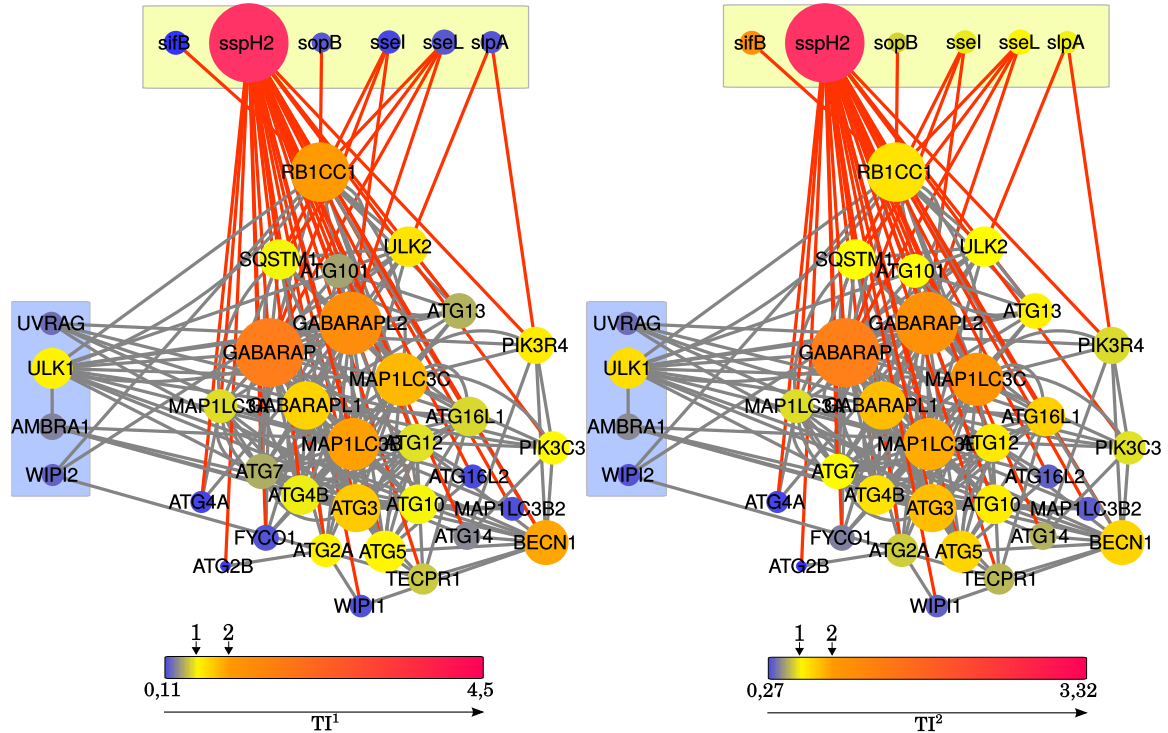
9. ábra. Az interfész fehérjéktől három lépésre elérhető pontok

Az ábra az egyesített hálózat azon részét tartalmazza mely elérhető három lépésre a *Salmonella* *Salmonella* hat darab interfész fehérjéjétől. A zöld színű élek a predikciókból származnak és a *Salmonella* (jobb) kötik össze az *ARN*-el (bal). A piros színű élek ez esetben a *Salmonella* interfész fehérjéi és ezek *Salmonellán* belüli első szomszédai között húzódnak. Látványos, hogy ez utóbbi élekből a legtöbb a legerősebb TI értékű pontokhoz tart. A pontok színezése megegyezik az 7. ábránál ismertetett sémával, a színskála a bal felső sarokban látható.

5.3. A hármas típusú szekréciós rendszer kapcsolatai

5.3.1. A *Salmonella* kapcsolatai az autofágia fehérjékkel

Az *ARN* pontjaiból 30 létesít összesen 40 kapcsolatot valamilyen *Salmonella* fehérjével. A predikciók alapján tehát hálózatban szereplő humán autofágia fehérjék 88%-a létesíthet kapcsolatot *Salmonella* fehérjével (10. ábra).



10. ábra. Az *ARN* és kapcsolatai az egyesített hálózattal

Az ábrán látható hálózatok az *ARN*-t és a hat *Salmonella* interfész fehérjét ábrázolják. A hálók a TI^1 (bal) és a TI^2 (jobb) értékeket szemléltetik. A hálózatok bal oldalán elhelyezkedő kék háttérű pontok jelképezik az *ARN*-nek azon részét, mely csak belső kapcsolatokkal rendelkezik. A hálózatok felső részén elhelyezkedő sárga téglalap tartalmazza a *Salmonella* interfész fehérjéit. A szűkre színű élek a emberi fehérjék közötti kapcsolatok, a narancs színűek pedig a predikciókból származó interspecifikus kapcsolatok. A pontok nagysága arányos a TI értékükkel. A pontok színének megfeleltetése a bal alsó sarokban látható színskálán látható.

A 40 darab interspecifikus kapcsolatért *Salmonella* oldalról csupán a következő hat *Salmonella* fehérje a felelős: sifB, slpA, sopB, sseI, sseL, sspH2 (10. ábra). A predikciók alapján tehát a *Salmonella* fehérjék mindössze 0,2%-a létesít csak kapcsolatot a gazda autofágia fehérjéivel. Az ábrán az is megfigyelhető, hogy a négy csak humán kapcsolattal rendelkező fehérje viszonylag kisebb TI -vel rendelkezik, pedig valóban központi szerepűek.

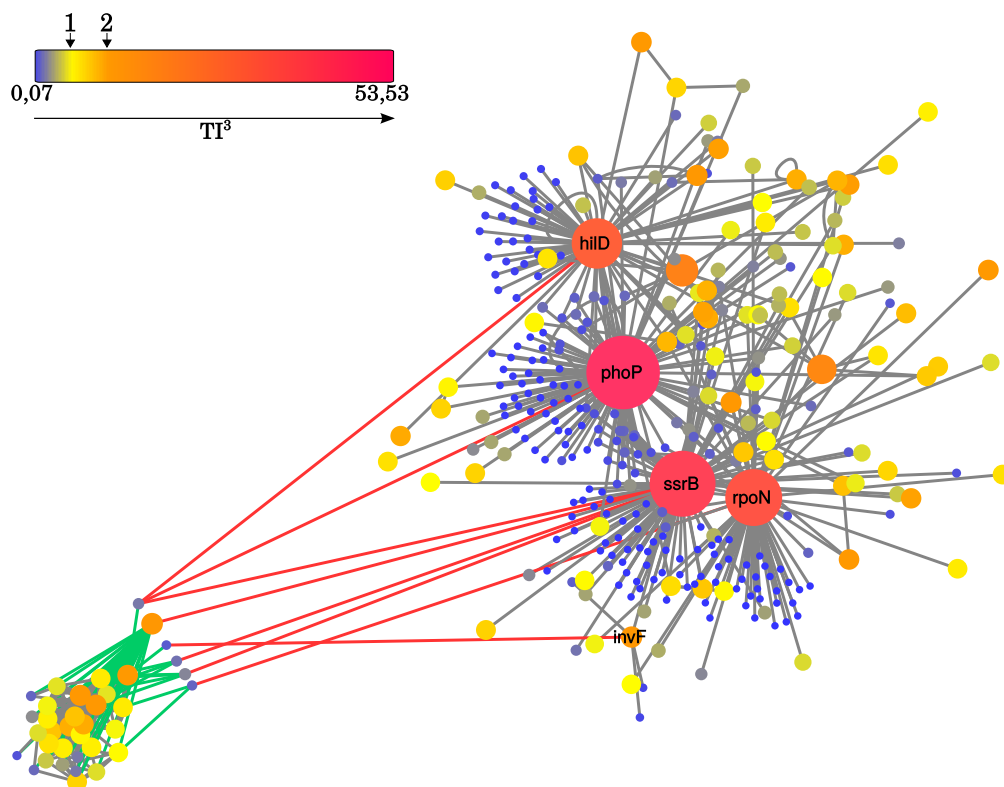
A 10. ábra azt is szemlélteti, hogy az interfész fehérjék többsége egy lépésre még

nem számít topológiaiilag erős kölcsönhatónak, viszont két lépésnél már igen. Ennek a jelenségnek az az oka, hogy az interfész fehérjékből viszonylag kevés kapcsolat ered, azonban ezek a kapcsolatok erősen kölcsönható fehérjék. Ezáltal az interfészfehérjékből kiinduló kétlépéses útvonalak erős elsődleges kölcsönhatókon keresztül érik el célpontjukat. A hat fehérje közül a következő öt bizonyítottan a T3SS-el jut ki a baktériumsejtből, míg az utolsóról kevés információnk van:

- **sifB** – A virulencia vagy más néven effektor fehérjék közé tartozik. Fertőzéskor *Salmonella* által kiváltott filamentumok (*Sif*) mentén eltávolodnak az SCV-től Freeman, Ohl és Miller, 2003.
- **sspH2** – Szintén effektor fehérje. Egy olyan E3 ubikvitin ligáz mely a gazda ubikvitinációs szignalizációját befolyásolja. A szubsztrát hiányában öngátló állapotban van. Szubsztrát kötésekor bekapcsolódik Quezada és mtsai., 2009. Az sspH2 gátló hatással van az aktin polimerizációra Haraga, Ohlson és Miller, 2008.
- **sopB** – Egy olyan esszenciális virulencia fehérje mely inozitol-foszfát-foszfataz mechanizmusával közvetetten aktiválja a gazda Rho-GTP-ázait. Ilyen kis-G fehérje a Cdc42 és a Rho-G. Az aktiválások hatásának eredője az aktin sejtváza átrendeződése, ami elősegíti a *Salmonella* bejutását a gazdasejtbe Haraga, Ohlson és Miller, 2008.
- **sseI** – Egy olyan effektor fehérje melynek aminotermínusa nagy hasonlóságot mutat a SspH2-vel. Filaminon keresztül képes a polimerizált aktin kötésére Haraga, Ohlson és Miller, 2008.
- **sseL** – Egy deubikvitináz enzim, mely mono és polyubikvitin láncok hidrolizálását végzi. Az sseL-nek *in vitro* a lizin-63 kapcsolt ubikvitin láncok sokkal jobb szubsztrátjai, melyből arra lehet következtetni, hogy a proteoszómalis degradáció befolyásolása helyett inkább az ubikvitin szignalizáció módosításáért felelős Rytkonen és mtsai., 2007.
- **slpA** – Ez a fehérje egy peptidil-prolil cisz-transz izomeráz. A szakirodalomban a *Salmonella* ezen fehérjéjéhez nem találtam olyan cikket, mely részletesen leírta volna a funkcióját. Az *Uniprot* adatbázis segítségével azonban találtam a legalább 90%-os hasonlóságot mutató fehérjék listájában egy *Swissprot* bejegyzést. Ez a hasonló de annotált fehérje az *E. coli* FKBP típusú 16 kDa-os peptidil-prolil izomeráza, mely

fehérjék hajtogatási sebességét növeli Uniprot, 2015. A két fehérje szekvenciáját az *NCBI BLAST* algoritmusán alapuló *blastp suite-2sequences* és az *ExPASy* fehérjeszekvencia összehasonlító programjával is összehasonlítottam. Az első 92%-os egyezést adott ($E=5 \cdot 10^{-104}$) a második pedig 91,9%-t.

5.3.2. A T3SS kapcsolatai és a *Salmonella* belső kapcsolatai



11. ábra. A T3SS fehérjéinek első és második szomszédai

Az ábra a 9. ábrán látható hálózathoz az interfész fehérjéket valamint ezek első és második szomszédjait tartalmazza. A feltüntetett nevű fehérjék az interfész fehérjék *Salmonella* oldali első szomszédai. A predikciókból származó interspecifikus kapcsolatok zölddel vannak jelölve.

Az interfész fehérjéket és *Salmonella*-án belüli első szomszédjaikat piros élek kötik össze. A pontok nagysága és színmélysége a TI^3 értékükkel arányos. A jobb felső sarokban a színeket magyarázó skála látható.

A 11. ábrán az öt feltüntetett fehérje közül csupán egy az *invF* olyan amelyik nincs benne se a 3. se a 2. táblázatban. Így elmondható, hogy a T3SS fehérjék *Salmonella* oldali első szomszédjai a mind a három TI érték szempontból a legmagasabb rangú pontok közé tartoznak.

6. Diszkusszió

Diplomamunkámban egy tisztán topológiai ökológiai mérőszámot, a topológiai fontosságot (TI) alkalmaztam molekuláris hálózatokra. Az index segítségével megpróbáltam felderíteni, hogy mik a topológiailag fontos fehérjék a humán-*Salmonella* hálózatban és az ezt felépítő két alhálózatban.

Szakdolgozatomban humán oldalról csupán a *Salmonella* autofágiára gyakorolt hatását vizsgáltam. A *Salmonella* azonban befolyásolja még a sejtvázszerkezetet is például a Rho útvonalon keresztül Haraga, Ohlson és Miller, 2008 Ray és mtsai., 2009. Mivel a *Signalink* tartalmazza a Rho útvonalat is, így egy ilyen hálózatra is fel lehetne írni topológiai sorrendet .

Utóbbiak mellett számtalan forrás Haraga, Ohlson és Miller, 2008 Ray és mtsai., 2009 Petrone, Stringer és Wade, 2014 említette, hogy a *Salmonella* a bélhámsejtek mellett immunsejtekben, köztük makrofágokban is tud szaporodni. A *Signalink* készítésekor abba jó minőségű fehérje-fehérje kapcsolatokat integráltam az *InnateDB*-ből. Az *InnateDB* egy kifejezetten immunológiai útvonalakkal foglalkozó publikus rendszer-biológiai adatbázis mely kísérletesen bizonyított jól annotált fehérje-fehérje kapcsolatokat tartalmaz. Az *InnateDB* főleg a veleszületett immunrendszer molekuláris kapcsolataival foglalkozik, melyből már több mint 18 700 található az a adatbázisban InnateDB, 2015. Az adatbázis felhasználásával és a TI alkalmazásával meglehetne vizsgálni, hogy a melyek a topológiailag fontos fehérjék az immunrendszer és *Salmonella* kölcsönhatásakor. Egy ilyen hálózat akár új gyógyszer-célpontokra is rámutathat.

Mint a bevezetőben említettem, a bél mikrobióta metabolitokkal kommunikál egymással és a humán sejtekkel is. Ilyen mechanizmus a leptin rendszer befolyásolása, vagy a különböző GPCR-ek által például a sejtciklus befolyásolása Brown és mtsai., 2003. A bél mikrobióta ilyen módon olyan komoly népegészségügyi betegségekre is hatással van mint az IBD vagy különböző rák típusok Karlsson és mtsai., 2011. A bél mikrobióta mindezekre a saját maguk által termelt metabolitokkal is képes hatni Backhed és mtsai., 2005. Mivel a *Salmonet* tartalmaz metabolikus kapcsolatokat, így egy jó kiindulási pont lehetne egy bél mikrobióta metabolikus hálózat készítéséhez. Egy ilyen hálózat TI-vel való elemzésével ki lehetne használni a TI azon ökológiai aspektusát, hogy a pontok páros lépésszámoknál erősen visszahatnak magukra. Ez életszerű mivel a bél mikrobió-

tában az egyes baktériumok metabolitjaikkal hatnak más baktériumok anyagcseréjére és így valószínűleg azok biomasszájára is Karlsson és mtsai., 2011. Egy ilyen a TI-n alapuló rangsorban ki lehetne jelölni hogy a meta-omikák által megállapított személyre jutó körülbelül 160 domináns baktérium fajból Qin és mtsai., 2010 melyek azok amelyek úgymond kulcsfajként viselkednek. Gyógyászati szempontból meg lehetne nézni a különbséget a topológiailag legfontosabb kulcs baktérium fajok között egészséges és például Crohn betegségben, vagy rákban szenvedő személyek mikrobiótája között. Hasonló vizsgálatokkal a gazda-patogén szempont mellett akár a mutualisztikus kapcsolatokat is vizsgálni lehetne humán-bélbaktérium, vagy bélmikrobióta baktérium-baktérium hálózatokban.

A szakirodalomban már csináltak másik humán-baktérium gazda-patogén hálózatot, ami inkább a metabolikus megközelítésre fektette a hangsúlyt. A vizsgálatban tüdő makrofágok *Mycobacterium tuberculosis* általi fertőzését modellezték különböző omikai adatok integrálásával készült hálózatban. A teljes genom szintű integrált hoszt-patogén hálózat segítségével metabolikus változásokat modelleztek, melyek segítségével a fertőzést három különböző patológiai fázisra tudták bontani Bordbar és mtsai., 2010. Mivel a *Salmonet* is tartalmaz metabolikai adatokat, így az általam összeállított integrált adatbázisban is lehetne hasonló modelleket alkalmazni és ezáltal a fertőzés folyamatának patológiai aspektusaira rávilágítani.

A diplomamunkámban csak a TI aspektusában vizsgáltam a gazda-patogén hálózat fehérjéinek fontosságát. Mint a bevezetésben említettem, az ökológiában a kulcs fajok kijelölésére több mérőszámot is használnak egyszerre, melyek más és más aspektusban vizsgálják a fajok fontosságát Jordán, Benedek és Podani, 2007 Jordán, Liu és Veen, 2003. Érdeemes lenne, megvizsgálni azt, hogy más mérőszámok hogyan viselkednek az általam felállított gazda-patogén hálózaton, vagy hogy korrelálnak-e a TI-vel megállapított rangsorral.

Az eddig felsorolt kitekintések és alkalmazási lehetőséget többnyire új elméleti módszerekre világítottak rá, melyek kísérletes ellenőrzése meglehetősen nehézkes. Azonban organoidok segítségével lehetőség lenne a legtöbb eddig felsorolt elméleti vizsgálat *in vitro* ellenőrzésére. Az organoidokon végzett vizsgálatok már eddig is sok információval szolgáltak az organogenezis, a regeneratív orvoslás a tumorigenezis és az emésztőrendszer működésével kapcsolatban. Az organoidok egyetlen izolált szöveti őssejtből előállított, a valódi *in vivo* szövet karakterisztikáit tükröző mesterségesen *in vitro* struktúrák Date

és Sato, 2015. Egér bélből állítottak elő olyan organoidokat melyek tükrözték a bél lumen szerkezetét, ezekben kriptákat határoló kitüremkedések voltak megfigyelhetők. Az ilyen módon előállított szérumból, csupán szöveti niche faktorokkal kezelt organoidok önfenntartók és akár több mint egy évig életben maradnak Date és Sato, 2015. Ilyen rendszerekben lehetőség nyílik a rendszer-biológiai modellekből és fontossági listákból nyert elméleti eredmények, jóslatok kísérleti ellenőrzésére.

Az eddig említett elméleti és gyakorlati lehetőségek mind valamilyen szinten humán relevanciával rendelkeztek. Azonban bél mikrobiomhoz hasonlóan a növények is komplex mikrobiológiai rendszerben élnek, melyben mikrobák széles skálája fordul elő. A hasonlóan az állati immunrendszerhez a növények is felismerik a patogén mintázatot és specifikus válaszokat adnak rá. A növényi fertőzések kis molekulák széles skálája vesz részt a védekezésben. Ezek a molekulák között egymás erősítő és gátló hatások figyelhetők meg. A növényi patogének hasonló molekulákkal próbálják a fitohormon homeosztázist módosítani Pieterse és Tsai, 2009 . Az ilyen adatokból épített molekuláris hálózatok elemzése jó alapot szolgáltat a mezőgazdaságilag fontos növények és patogénjeik között fellépő mechanizmusok feltérképezéséhez.

7. Összefoglalás

Diplomamunkám során egy humán-*Salmonella* gazda-patogén hálózatot készítettem és elemeztem. A hálózat előállításánál olyan eszközöket fejlesztettem melyek felhasználhatók hasonló hálózatok építésére és elemzésére. Ezek eszközök segítségével könnyen integrálni lehet a bioinformatikában és rendszer-biológiában használt hálózati formátumokat. A csoportunk által alkalmazott MiTab SQL formátum megkönnyíti a molekuláris hálózati adatok tárolását, átalakítását és elemzését. Az általam írt *PsimiSQL Python* osztály megkönnyíti a különböző adatformátumok MiTab SQL-be alakítását. Az általam készített *TopologyAnalyser Python* osztály képes a bármilyen hálózat formátumból könnyen előállítható éllista fájl topológiai elemzésére. A csoport által felépített protokoll egy hatékony módszert biztosít molekuláris hálózatok létrehozására. A protokoll jól használhatónak bizonyult *Signalink* és humán-*Salmonella* adatbázis készítésénél.

A TI bár ökológiai eredetű, de képes volt gazda-patogén molekuláris hálózaton a központi molekulák kijelölésére. Az ilyen topológiai mérőszámokat többnyire nem használják molekuláris hálók vizsgálatára. Diplomamunkám rávilágít arra is, hogy biológiailag releváns eredményeket lehet kapni az ilyen mérőszámok molekuláris hálózatokra alkalmazásával.

A TI paramétertől függően több szempontból is le tudja írni a hálózatot. Az egy lépésre számolt index megmutatja az erős közvetlen kölcsönhatókat. A páros lépésre számolt esetben az önhatás alakítja a rangsort. Az egynél hosszabb páratlan útvonalakon alapuló TI pedig a pontok közvetett hatását jellemzi.

Ebben az adatszettben a predikciós adatbázisok és a TI együttesen rámutatott egy olyan fehérjére aminek eddig nem írták le a szerepét a *Salmonella* virulenciájában. A TI² rávilágított a *Salmonella* interfész fehérjéinek fontosságára, amit az irodalom is alátámaszt. Ez abból adódott, hogy a belőlük eredő viszonylag kevés kapcsolat erős elsődleges kölcsönhatókon éri el célpontjaikat. A TI megmutatta azt is hogy az interfész fehérjék első szomszédai is nagyon fontosak mind az autofágia, mind a *Salmonella* virulencia és életciklus szabályzásában. Ezzel rávilágítva arra, hogy a *Salmonella* életciklusa erősen függ a gazdával való kölcsönhatásoktól.

8. Summary

In my thesis work I built and analyzed a human-*Salmonella* host-pathogen network. While constructing the network I created tools that can be used for building and analyzing similar networks. This tools can be used to integrate data formats used in bioinformatics and systems-biology. Our group's MiTab SQLite format makes it easier to store, transform and analyze molecular network data. I built the *PsimiSQL Python class* that can be used to convert several data formats to MiTab SQL easily. I also built another *Python class TopologyAnalyser* that can analyze edge-list files. Edge-list files can be easily obtained from any network. Our research group created a powerful protocol for constructing molecular networks. The protocol proved to be well usable for the establishment of the Signalink and the human-*Salmonella* database.

Although the topological importance is an ecological index, it is proven that it can be used to mark the central proteins of a host-pathogen network. These kind of topological indices are rarely used for the analysis of molecular networks. My thesis reveals that biologically relevant results can be obtained by using indices like this.

Based on the parameter, the topological importance can describe the analyzed network in several aspects. The index calculated for one step paths shows nodes that have strong direct topological effect. The ranking produced with the formula used with even parameters emphasizes nodes with heavy effects on themselves. Using longer odd step paths marks nodes that are important indirect inter-actors.

In this data set, the predictions and the topological importance marked a protein that haven't been described in *Salmonella* virulence yet.

TI² emphasized the importance of the *Salmonella* interface proteins. The importance of these proteins was already described in the literature. The reason behind that effect is that the interface proteins have just a few edges in the direction of *ARN*, but those edges pass through topologically very powerful proteins. The TI also showed that the first neighbors of the interface proteins are very important in autophagy and in the regulation of *Salmonella* life cycle. This shines light on the fact that the life cycle of *Salmonella* is dependent on interaction with the host cell.

9. Köszönetnyilvánítás

10. Nyilatkozat

Felhasznált irodalom

- Backhed F. Ley R. E. Sonnenburg J. L. Peterson D. A. (2005). “Host-bacterial mutualism in the human intestine”. *Science* 307.5717, pp. 1915–1920.
- Baranyi Gabriella, Saura Santiago, Podani János, Jordán Ferenc (2011). “Contribution of habitat patches to network connectivity: Redundancy and uniqueness of topological indices”. *Ecological Indicators*.
- Bordbar A. Lewis N. E. Schellenberger J. Palsson B. ?. Jamshidi N. (2010). “Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions”. *Mol. Syst. Biol.* 6, p. 422.
- Brown A. J. Goldsworthy S. M. Barnes A. A. Eilert M. M. Tcheang L. Daniels D. Muir A. I. Wigglesworth M. J. Kinghorn I. Fraser N. J. Pike N. B. Strum J. C. Steplewski K. M. Murdock P. R. Holder J. C. Marshall F. H. Szekeres P. G. Wilson S. Ignar D. M. Foord S. M. Wise A. Dowell S. J. (2003). “The Orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids”. *J. Biol. Chem.* 278.13, pp. 11312–11319.
- Date S. Sato T. (2015). “Mini-gut organoids: reconstitution of the stem cell niche”. *Annu. Rev. Cell Dev. Biol.* 31, pp. 269–289.
- Freeman J. A. Ohl M. E. Miller S. I. (2003). “The Salmonella enterica serovar typhimurium translocated effectors SseJ and SifB are targeted to the Salmonella-containing vacuole”. *Infect. Immun.* 71.1, pp. 418–427.
- Glick D. Barth S. Macleod K. F. (2010). “Autophagy: cellular and molecular mechanisms”. *J. Pathol.* 221.1, pp. 3–12.
- Haraga A. Ohlson M. B. Miller S. I. (2008). “Salmonellae interplay with host cells”. *Nat. Rev. Microbiol.* 6.1, pp. 53–66.
- Hermjakob H. Montecchi-Palazzi L. Bader G. Wojcik J. Salwinski L. Ceol A. Moore S. Orchard S. Sarkans U. Mering C. Roechert B. Poux S. Jung E. Mersch H. Kersey P. Lappe M. Li Y. Zeng R. Rana D. Nikolski M. Husi H. Brun C. Shanker K. Grant S. G. Sander C. Bork P. Zhu W. Pandey A. Brazma A. Jacq B. Vidal M. Sherman D. Legrain P. Cesareni G. Xenarios I. Eisenberg D. Steipe B. Hogue C. Apweiler R. (2004). “The HUPO PSI’s molecular interaction format—a community standard for the representation of protein interaction data”. *Nat. Biotechnol.* 22.2, pp. 177–183.

- Horváth Balázs (2015). *Host Pathogen Networks Repository*. <https://github.com/blaisehorvath/host-pathogen-networks>.
- InnateDB (2015). *InnateDB*. <http://innatedb.com/redirect.do?go=aboutIDB>.
- Jo E. K. Yuk J. M. Shin D. M. Sasakawa C. (2013). “Roles of autophagy in elimination of intracellular bacterial pathogens”. *Front Immunol* 4, p. 97.
- Jordán Ferenc, Benedek Zsófia, Podani János (2007). “Quantifying positional importance in food webs: A comparison of centrality indices”. *Ecological modelling*.
- Jordán Ferenc, Liu W.-C. Veen F.J.F. (2003). “Quantifying the importance of species and their interactions in a host-parasitoid community”. *Community ecology*.
- Karlsson F. H. Nookaew I. Petranovic D. Nielsen J. (2011). “Prospects for systems biology and modeling of the gut microbiome”. *Trends Biotechnol.* 29.6, pp. 251–258.
- Krishnadev O. Srinivasan N. (2011). “Prediction of protein-protein interactions between human host and a pathogen and its application to three pathogenic bacteria”. *Int. J. Biol. Macromol.* 48.4, pp. 613–619.
- Kshirsagar M. Carbonell J. Klein-Seetharaman J. (2012). “Techniques to cope with missing data in host-pathogen protein interaction prediction”. *Bioinformatics* 28.18, pp. i466–i472.
- Lathrop S. K. Binder K. A. Starr T. Cooper K. G. Chong A. Carmody A. B. Steele-Mortimer O. (2015). “Replication of *Salmonella enterica* Serovar Typhimurium in Human Monocyte-Derived Macrophages”. *Infect. Immun.* 83.7, pp. 2661–2671.
- Métris A. Sudhakar P. Fazekas D. Demeter A. Ari E. Branchu P, Kingsley R. Baranyi J. Korcsmáros T. (2015). “SalmoNet, an integrated network of ten *Salmonella enterica* strains reveals common and distinct pathways to host adaptation”. *Kézirat*.
- Newman Mark (2010). *Networks: An Introduction*. Oxford University Press. ISBN: 0199206651. URL: <http://www.amazon.com/Networks-An-Introduction-Mark-Newman/dp/0199206651%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0199206651>.
- Owens Mike (2006). *The Definitive Guide to SQLite*. Apress. ISBN: 1590596730. URL: <http://www.amazon.com/Definitive-Guide-SQLite-Mike-Owens/dp/1590596730%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D1590596730>.

- Payton Ian J. Fenner Michael, Lee William G. (2002). “Keystone species: the concept and its relevance for conservation management in New Zealand”. *Science for conservation*.
- Peters S. E. Paterson G. K. Bandularatne E. S. Northen H. C. Pleasance S. Willers C. Wang J. Foote A. K. Constantino-Casas F. Scase T. J. Blacklaws B. A. Bryant C. E. Mastroeni P. Charles I. G. Maskell D. J. (2010). “Salmonella enterica serovar typhimurium *trxA* mutants are protective against virulent challenge and induce less inflammation than the live-attenuated vaccine strain SL3261”. *Infect. Immun.* 78.1, pp. 326–336.
- Petrone B. L. Stringer A. M. Wade J. T. (2014). “Identification of HilD-regulated genes in Salmonella enterica serovar Typhimurium”. *J. Bacteriol.* 196.5, pp. 1094–1101.
- Pieterse C. M. Leon-Reyes A. Ent S. Van Wees S. C. (2009). “Networking by small-molecule hormones in plant immunity”. *Nat. Chem. Biol.* 5.5, pp. 308–316.
- Qin J. Li R. Raes J. Arumugam M. Burgdorf K. S. Manichanh C. Nielsen T. Pons N. Levenez F. Yamada T. Mende D. R. Li J. Xu J. Li S. Li D. Cao J. Wang B. Liang H. Zheng H. Xie Y. Tap J. Lepage P. Bertalan M. Batto J. M. Hansen T. Le Paslier D. Linneberg A. Nielsen H. B. Pelletier E. Renault P. Sicheritz-Ponten T. Turner K. Zhu H. Yu C. Li S. Jian M. Zhou Y. Li Y. Zhang X. Li S. Qin N. Yang H. Wang J. Brunak S. Dore J. Guarner F. Kristiansen K. Pedersen O. Parkhill J. Weissenbach J. Bork P. Ehrlich S. D. Wang J. Antolin M. Artiguenave F. Blottiere H. Borruel N. Bruls T. Casellas F. Chervaux C. Cultrone A. Delorme C. Denariáz G. Dervyn R. Forte M. Friss C. Guchte M. Guedon E. Haimet F. Jamet A. Juste C. Kaci G. Kleerebezem M. Knol J. Kristensen M. Layec S. Le Roux K. Leclerc M. Maguin E. Minardi R. M. Oozeer R. Rescigno M. Sanchez N. Tims S. Torrejon T. Varela E. Vos W. Winogradsky Y. Zoetendal E. (2010). “A human gut microbial gene catalogue established by metagenomic sequencing”. *Nature* 464.7285, pp. 59–65.
- Quezada C. M. Hicks S. W. Galan J. E. Stebbins C. E. (2009). “A family of Salmonella virulence factors functions as a distinct class of autoregulated E3 ubiquitin ligases”. *Proc. Natl. Acad. Sci. U.S.A.* 106.12, pp. 4864–4869.
- Ray K. Marteyn B. Sansonetti P. J. Tang C. M. (2009). “Life on the inside: the intracellular lifestyle of cytosolic bacteria”. *Nat. Rev. Microbiol.* 7.5, pp. 333–340.

- Rytkenon A. Poh J. Garmendia J. Boyle C. Thompson A. Liu M. Freemont P. Hinton J. C. Holden D. W. (2007). "SseL, a Salmonella deubiquitinase required for macrophage killing and virulence". *Proc. Natl. Acad. Sci. U.S.A.* 104.9, pp. 3502–3507.
- Szalai P. Hagen L. K. S?tre F. Luhr M. Sponheim M. ?verbye A. Mills I. G. Seglen P. O. Engedal N. (2015). "Autophagic bulk sequestration of cytosolic cargo is independent of LC3, but requires GABARAPs". *Exp. Cell Res.* 333.1, pp. 21–38.
- Turei D. Foldvari-Nagy L. Fazekas D. Modos D. Kubisch J. Kadlecsek T. Demeter A. Lenti K. Csermely P. Vellai T. Korcsmaros T. (2015). "Autophagy Regulatory Network - a systems-level bioinformatics resource for studying the mechanism and regulation of autophagy". *Autophagy* 11.1, pp. 155–165.
- Uniprot (2015). *FKBP-type 16 kDa peptidyl-prolyl cis-trans isomerase*. <http://www.uniprot.org/uniprot/POAEM2>.
- Walthers D. Carroll R. K. Navarre W. W. Libby S. J. Fang F. C. Kenney L. J. (2007). "The response regulator SsrB activates expression of diverse Salmonella pathogenicity island 2 promoters and counters silencing by the nucleoid-associated protein H-NS". *Mol. Microbiol.* 65.2, pp. 477–493.