

A szulfanilsav-bontó *Novosphingobium
resinovorum* SA1 törzs strukturális
genomikai és transzkriptomikai vizsgálata

Ph.D. értekezés

Hegedüs Botond

Biológia Doktori Iskola

Témavezető:

Dr. Rákhely Gábor

Szegedi Tudományegyetem, Biotechnológiai Tanszék
Magyar Tudományos Akadémia, Szegedi Biológiai Kutatóközpont,
Biofizikai Intézet



Szeged

2018

Tartalomjegyzék

1	Rövidítések jegyzéke	5
2	Bevezetés.....	7
3	Irodalmi áttekintés	8
3.1	Aromás vegyületek előfordulása a természetben és az iparban.....	8
3.2	A szulfanilsav előállítása, ipari felhasználása	10
3.3	A szulfanilsav ipari felhasználása, útja a természetbe	10
3.4	A szulfanilsav biológiai úton történő bontása	12
3.5	A szulfanilsav bontási útvonalának központi elemei.....	14
3.5.1	A szulfanilsav bontás perifériás útvonala:.....	16
3.5.2	A szulfanilsav bontás centrális útvonala.....	17
3.6	A szulfanilsav bontási útvonalának kiegészítő elemei.....	20
3.6.1	SA dioxygenáz elektron transzport láncának lehetséges elemei	20
3.6.2	Az aromás savak transzporterei	22
3.6.3	Szulfid oxidáló enzimek	23
3.6.4	Szulfid transzporterek	25
3.7	A <i>Sphingomonadaceae</i> család tagjainak változatos genomi összetétele.....	25
3.8	Genomszekvenálás	26
3.9	Transzkriptom vizsgálat	27
4	Célkitűzések.....	30
5	Anyagok és módszerek	32
5.1	A munka során használt baktérium törzsek, plazmidok és oligonukleotidok	32
5.1.1	Baktérium törzsek megnevezése és leírása.....	32
5.1.2	Plazmidok megnevezése és leírása.....	32
5.1.3	Az intergénikus régiók PCR segítségével történő amplifikációjához használt oligonukleotidok megnevezése és szekvenciája	33
5.1.4	Az RT-qPCR során használt oligonukleotidok megnevezése és szekvenciája.....	33

5.2	Növesztési körülmények	34
5.3	Felhasznált tápanyagok.....	34
5.3.1	Luria-Bertain (LB) tápoldat	34
5.3.2	Minimál tápoldat.....	34
5.4	Molekuláris biológiai munkák	35
5.4.1	Genomi DNS izolálása N. resinovorum SA1 törzsekből.....	35
5.4.2	Az N. resinovorum SA1 genomszekvenálása.....	35
5.4.3	Kémiai kompetens E. coli sejtek készítése és transzformálása	36
5.4.4	Elektrokompetens N. resinovorum SA1 törzsek létrehozása és transzformálása.....	36
5.5	Promóter régió azonosítása	37
5.5.1	Plazmid DNS izolálása E. coli és N. resinovorum SA1 törzsekből	37
5.5.2	Restrikciós enzimekkel történő hasítás.....	37
5.5.3	DNS polinukleotidkináz kezelés.....	37
5.5.4	Alkalikus foszfatáz kezelés.....	37
5.5.5	Ligálás	37
5.5.6	Agaróz gélelektroforézis.....	38
5.5.7	DNS fragment izolálás.....	38
5.5.8	Nukleotidsorrend meghatározása.....	38
5.5.9	A promóter régió meghatározásához szükséges promóter próba vektor (pPROBE) változatok elkészítése.....	38
5.5.10	Promóter régió azonosításának menete	39
5.5.11	Polimeráz-láncreakció (PCR).....	39
5.6	Analitikai módszerek	40
5.6.1	A GFP kifejeződésének meghatározása.....	40
5.6.2	A szulfanilsav fogyasztás mérése	40
5.6.3	Glükóz koncentráció mérése	40
5.7	Transzkriptomikai munkák	41
5.7.1	Teljes RNS tisztítása N. resinovorum SA1-ből.....	41

5.7.2	Az <i>N. resinovorum</i> SA1 teljes transzkriptomjának a meghatározása RNA-seq módszerével	41
5.7.3	Az RT-qPCR folyamata	42
5.8	Bioinformatikai vizsgálatok	42
5.8.1	Az <i>N. resinovorum</i> SA1 genomjának összeillesztése és annotációja.	42
5.8.2	NCBI BLAST keresés	43
5.8.3	Filogenetikai vizsgálatok.....	43
5.8.4	Az <i>N. resinovorum</i> SA1 RNA-Seq során kapott leolvasások informatikai feldolgozása	43
5.8.5	Megváltozott kifejeződést mutató gének vizsgálata.....	44
5.8.6	Egyéb statisztikai vizsgálatok.....	44
6	Eredmények és tárgyalásuk.....	45
6.1	Az <i>N. resinovorum</i> SA1 törzs genomszekvenciájának meghatározása.....	45
6.2	A szulfanilsav bontási útvonalának komparatív vizsgálata	47
6.3	A szulfanilsav bontási útvonal evolúciós eredetének tanulmányozása	50
6.4	A perifériás és centrális útvonal génjeinek transzkripció szabályozó régiói.....	56
6.5	Teljes RNA-Seq alapú transzkriptom vizsgálat	62
6.6	A szulfanilsav bontásában résztvevő gének kifejeződése	70
6.7	A szulfanilsav bontást kiegészítő elemek azonosítása.....	72
6.7.1	A szulfanilsav gyúróhidroxiláló dioxigenáz elektrontranszport láncának elemei.....	72
6.7.2	Szulfanilsav transzporter	73
6.7.3	A szulfid sorsa a sejtben	73
6.8	Az RNA-seq eredményeinek alátámasztása RT-qPCR használatával.....	76
7	Összefoglalás	78
8	Summary.....	82
9	Hivatkozások jegyzéke.....	86
10	Köszönetnyilvánítás.....	97
11	Függelék.....	98

1 Rövidítések jegyzéke

2-MA:	2-maleilacetát
3-CM:	3-karboxi-cisz-cisz-mukonát
3-SM:	3-szulfo-cisz-cisz-mukonát, 3-sulfo-cis-cis-muconate
4-CML:	4-karboxi-mukonolakton
4-SC:	4-szulfokatekol, 4-sulfo-catechol
4-SL:	4-szulfolakton
AAHS:	aromatic acid/H ⁺ symporter
AMP:	adenozin-monofoszfát
APS:	adenozin-5'-foszfo-szulfáton
BTEX:	benzol, toluol, etil-benzol, xilol
CA:	korrespondencia analízis, correspondence analysis
CCR:	karbon katabolit represszió, carbon catabolic repression
COG:	clusters of orthologous groups of proteins
DEG:	változó kifejeződést mutató gének, differentially expressed genes
ETC:	elektrontranszport lánc, electron transport chain
FDR:	false discovery rate
GFP:	zöld fluoreszkáló fehérje, green fluorescent protein
Glc:	glükóz, glucose
GSEA:	géncsoport dúsulási vizsgálat, Gene Set Enrichment Analysis
GSOA:	géncsoport túltreprezentációs elemzés, Gene set overrepresentation analysis
HB:	hidroxibenzoát
kbp:	kilo bázispár (1 000 bp)
Mbp:	mega bázispár (1 000 000 bp)
MFS:	major facilitator superfamily
OD:	optikai denzitás
P34O:	protokatekol-3,4-dioxigenáz
PABA:	p-amino-benzoésav, p-aminobenzoic acid
PAH:	poliaromás szénhidrogén, polyaromatic hydrocarbon
PC:	protokatekol, protocatechol
PCA:	főkomponens analízis, principal component analysis

PCR:	polimeráz-lánreakció, polymerase chain reaction
RHO:	gyűrű hidroxiláló oxigenáz, ring hydroxylating oxygenase
RO:	Rieske-típusú oxigenáz
SA:	szulfanilsav, sulfanilic acid, 4-aminobenzol szulfonát (4-ABS),
SOE:	szulfit-oxidáló enzim
TCA:	trikarbonsav ciklus, Szent-Györgyi–Krebs-ciklus, tricarboxylic acid cycle

2 Bevezetés

Az ipar az utóbbi évszázadban rohamos változásokon ment keresztül. Ennek egyik eredménye, hogy a természetes alapanyagok használata napjainkra már kezd egyre inkább háttérbe szorulni, helyüket a mesterségesen előállított vegyületek veszik át. Ezek iránt a vegyületek iránt mutatott megnövekedett igény és sokrétű felhasználási módjuk elkerülhetetlenné tette, hogy végül a természetes környezetben is egyre nagyobb mennyiségben találkozzunk velük. Ezen vegyületek egyik közös tulajdonsága, hogy olyan mesterségesen kialakított szerkezettel rendelkeznek, amelyek csak nagyon ritkán vagy egyáltalán nem fordulnak elő a környezetben. Emiatt a szerves anyagok biológiai körforgásáért felelős mikroorganizmusok sok esetben nem képesek őket hatékonyan bontani, ami felhalmozódásukhoz, esetleg a táplálékláncba történő bejutásukhoz vezet.

A szulfanilsav (SA) is egy ilyen mesterségesen előállított vegyület, amelyet az ipar nagy mennyiségben állít elő és használ fel főleg festékek és gyógyszerek alapanyagaként. Aromás kémiai szerkezete nagyfokú stabilitást kölcsönöz a molekulának, amit tovább fokoz a természetben előforduló aromás vegyületekre alig jellemző szulfonsav funkciós csoport megléte.

Az SA ipari felhasználása során jelentős mennyiségben kerül a szennyvízhálózatba, ahonnan a legtöbb esetben átalakítás nélkül jut tovább a környezetbe. A szennyvíz magas SA tartalmának csökkentése már régóta komoly kihívást jelent a szakemberek számára, azonban hatékony és megfizethető megoldást eddig még nem sikerült találni rá.

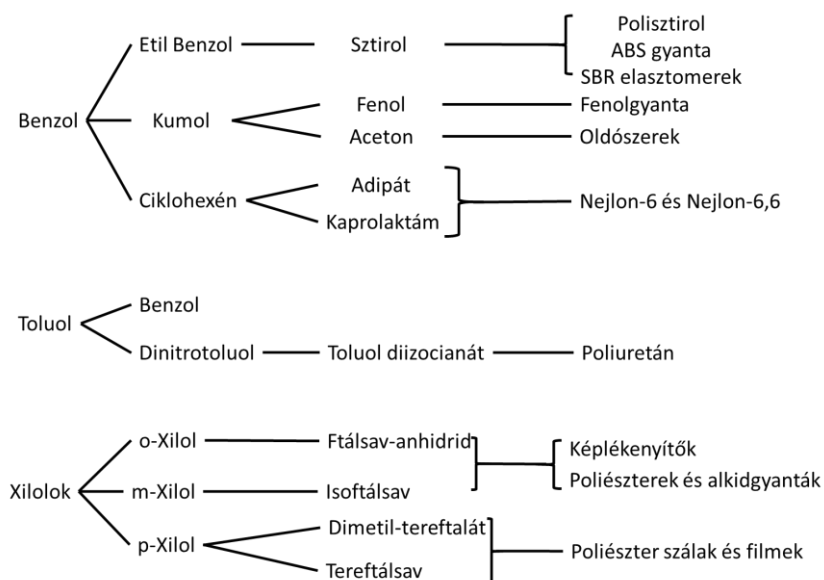
Lehetséges megoldás a már sok esetben bizonyított biológiai úton történő kármentesítés. Számos példa mutatja, hogy a mikroorganizmusok változatos anyagcsere útvonalai segítségével képesek tápanyagforrásként felhasználni olyan nehezen bontható vegyületeket is, mint az SA. Azonban ahhoz hogy a bioremediáció az ipar számára alternatív megoldássá válhasson a kevésbé környezetbarát kémiai és fizikai eljárásokkal szemben, biztosítani kell megbízható működésüket. Ez csak úgy érhető el, ha minél pontosabban feltárjuk a bontás folyamatának elemeit és megértjük működésük szabályozását.

3 Irodalmi áttekintés

3.1 Aromás vegyületek előfordulása a természetben és az iparban

Az aromás vegyületek a természetben előforduló vegyületek egy fontos és változatos csoportját alkotják. Általánosan minden élőlényben megtalálhatóak pl. fenil-alanin, tirozin és triptofán aminosavak formájában. Legnagyobb mennyiségben és változatosságban azonban a növények állítják elő őket. Elsősorban a szerkezeti vázukat alkotó lignin építőelemeiként (fenil-propán alegységek) jelentősek, de számos illatanyag esetén is megtalálhatóak. Az aromás elnevezésük is onnan származik, hogy a legkorábban azonosított aromás vegyületek jellegzetes illattal rendelkeztek. Az aromás vegyületek jelentőségét növeli, hogy a szárazföldi növények biomasszájának akár egynegyedét is kitehetik ¹. Nemcsak az élővilág, hanem a nyers kőolaj és földgáz is jelentős mennyiségben tartalmaz aromás vegyületeket. Ezek egy része monoaromás: benzol, toluol, etil-benzol és xilol, amelyeket együtt BTEX-nek neveznek ², illetve lehetnek poliaromások, amelyeket általában PAH-nak (PolyAromatic Hydrocarbons) neveznek.

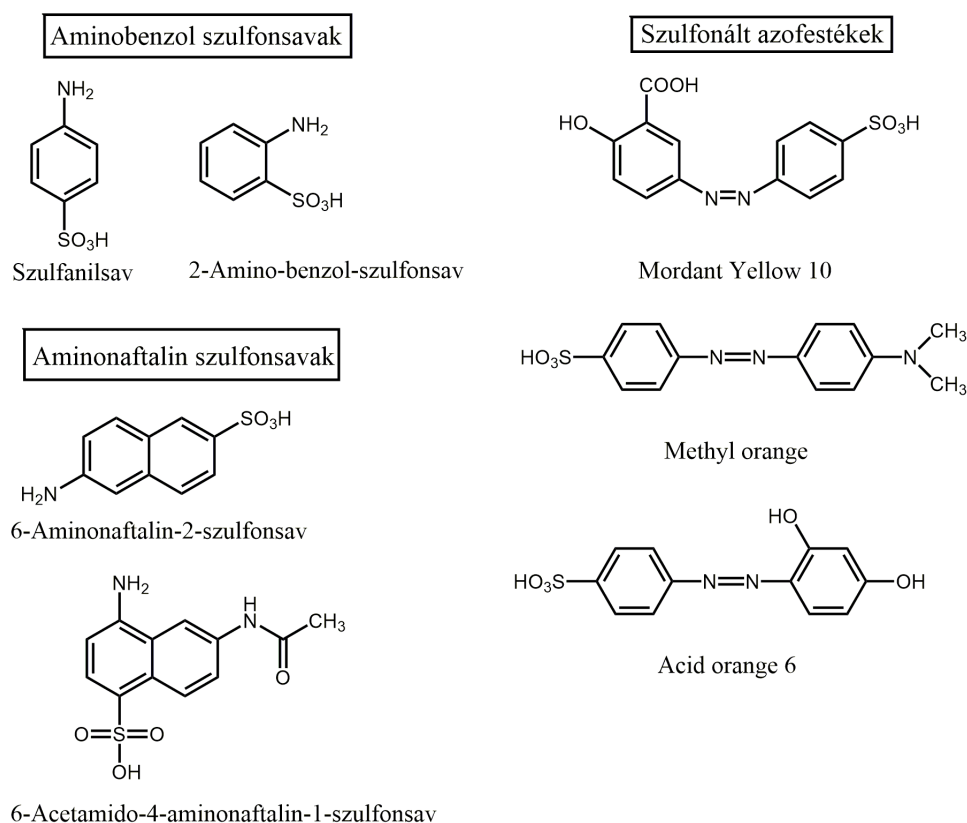
Az aromás vegyületek már régóta fontos szerepet játszanak az iparban. Legnagyobb mennyiségben a BTEX vegyületcsalád elemeit hasznosítják közvetlenül az üzemanyagok oktánszámának növelésére vagy alapanyagként különböző típusú műanyagok, gumik, kenőanyagok, festékek, tisztítószer, gyógyszerek, növényvédőszer és robbanószer előállítására (1. ábra).



1. ábra. Néhány példa a BTX vegyületek ipari felhasználására ³

Az aromás vegyületek szerkezeti sajátosságuk miatt relatíve nehezen lépnek kémiai reakcióba. Azonban elektrofil szubsztitúcióval az ipar számos kémiai módosítást képes végrehajtani rajtuk. Az elektrofil szubsztitúció során az aromás gyűrű egyik atomját, amely a legtöbb esetben egy hidrogén, egy elektrofil tulajdonsággal rendelkező másik funkció csoportra cserélik le. A legfontosabb ilyen típusú reakció közé tartozik az aromás vegyületek nitrálása, halogénezése, szulfonálása és Friedel–Crafts-acilezése.

Az így létrehozható szubsztituált aromás vegyületek közül kiemelkedik a szulfonált aromás vegyületek csoportja, különösen a szulfonált aromás aminoké. Ezeknek a vegyületeknek közös tulajdonsága, hogy a szerkezetük alapját képző egy vagy több aromás gyűrűhöz a szulfonált aromás vegyületekre jellemző szulfonsav funkció csoport (-SO₃H) mellett amino (-NH₂) csoport is kapcsolódik (2. ábra).



2. ábra. Néhány példa a szulfonált aromás aminokra és szulfonált azofestékekre

A szulfonált aromás aminok kiemelt jelentőségüket elsősorban a festékiparnak köszönhetik. Ennen a fő oka az amino csoportjaik között kialakítható azo kötés (-N=N-), ami az azo festékek előállításához nélkülözhetetlen. A szulfonsav csoport pedig az azofestékek kémiai tulajdonságát képes pozitívan befolyásolni: az általa biztosított negatív töltés miatt a szulfonált

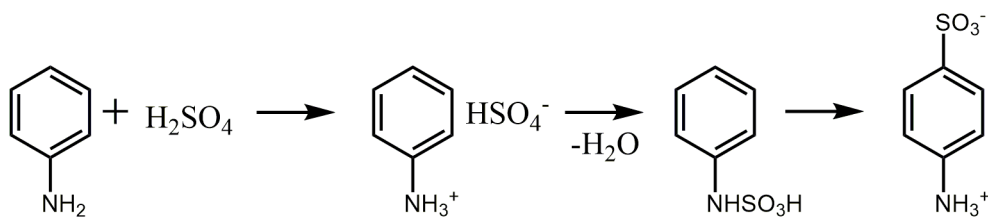
aromás vegyületek erős savakként viselkednek és vizes oldatokban jól oldódnak. A nukleofil szubsztitúciós reakció során a szulfonsav könnyen kilépő csoportként viselkedik, emellett csak minimális hatással van a festék színére.

Amíg a szulfonált aromás vegyületek többnyire negatív töltésűek a szulfonsav csoport miatt, addig a szulfonált aromás aminok ikerionos jellegűnek (zwitter ionos) tekinthetők és gyenge savakként viselkednek. Ennek következtében vízben és szerves oldószerekben a szulfonált vegyületektől eltérően rosszul oldódnak. Azonban lúgokkal képzett alkáli sójuk vízben jól oldódik ⁴.

A szulfonált aromás aminok természetes folyamatok útján csak nagyon ritkán jönnek létre. A környezetben megtalálható ilyen típusú vegyületek többsége mesterséges eredetű. Az egyik legfontosabb szulfonált aromás amin: a szulfanilsav (SA).

3.2 A szulfanilsav előállítása, ipari felhasználása

Ipari mértékben a szulfonált aromás vegyületeket aromás szerkezettel rendelkező szerves vegyületek szulfonálásával állítják elő különböző szulfonáló anyagok felhasználásával mint például kén-trioxid (SO_3), kénsav (H_2SO_4), klór-kénsav (HSO_3Cl), amido-kénsav (H_3NSO_3) vagy szulfid ionok (SO_3^{2-}) ⁵. Az SA előállítása esetén a szerves aromás vegyület szerepét az anilin tölti be, míg a szulfonáláshoz kénsavat használnak. A reakció $190\text{ }^\circ\text{C}$ fölött játszódik le, víz kiválása mellett, a két vegyület 1:1 arányú elegyének használatával (3. ábra). Egyszerű és olcsó előállíthatósága az ipari elterjedését nagymértékben elősegítette.

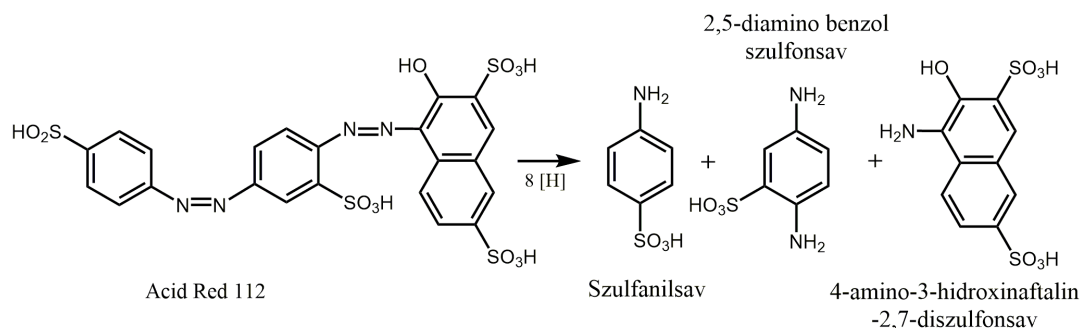


3. ábra. Szulfanilsav ipari előállításának lépései

3.3 A szulfanilsav ipari felhasználása, útja a természetbe

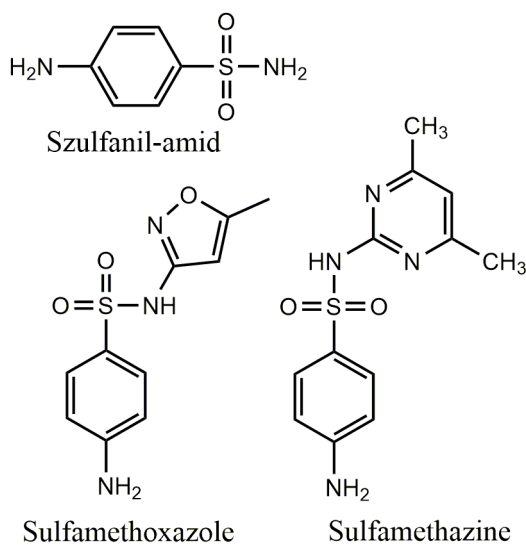
Az SA viszonylag olcsón előállítható vegyület, amelyet különböző iparágak főleg alapanyagként hasznosítanak, ezek közül is kiemelkedik a festékipar. Az azofestékeknek igen sok változata van és a legnagyobb mennyiségben előállított és használt festéktípus. Világszerte több mint 800,000 tonna festéket állítanak elő évente, amelynek 60-70% tartozik az azo

festékek közé. Becslések szerint az előállításuk és felhasználásuk során megközelítőleg 10-15%-uk juthat ki a természetbe ^{6,7}, ami jelentős környezeti terhelést jelent. Az azofestékek abiotikus úton történő bomlása a természetben megjelenő szulfanilsav elsődleges forrása ⁷⁻¹⁰ (4. ábra).



4. ábra. Az Acid Red 112 azo festék redukciója számos szulfonált aromás aminoszulfonát aromás aminoszulfonátok képződéséhez vezet köztük a szulfanilsavéhoz is

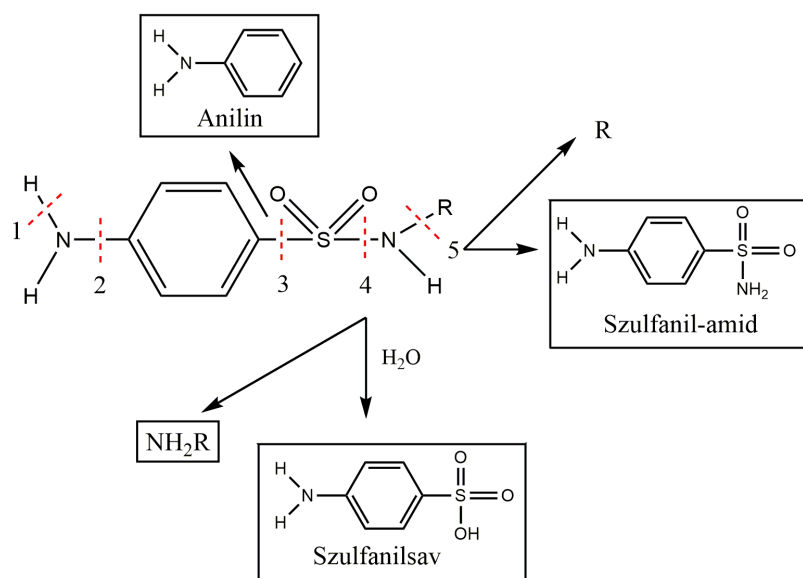
Az SA másik jelentős forrása a gyógyszeripar, amely többnyire szulfonamid típusú antibakteriális szerek előállítására használja ezt a vegyületet (5. ábra). Ezt a gyógyszer családot eredetileg Gram-negatív baktériumok ellen fejlesztették ki, melyek között több humán patogén törzs is megtalálható ¹¹. Hatásmechanizmusuk alapján a szulfonamidok bakteriosztatikus antimikrobiális szerek, amelyek a baktériumok folsavszintézisét kompetitív módon gátolják. A gátlást az teszi lehetővé, hogy a folyamat prekursorának, a p-amino-benzoésavnak (PABA) az SA szerkezeti analógja. A folsav nélkülözhetetlen a baktériumok számára, mivel számos anyagcsere folyamat mellett a DNS szintézisében is szerepet játszik.



5. ábra. Fontosabb szulfonamid származékok

A szulfonamid gyógyszerek történelmi jelentősége óriási. Ezek voltak az első hatékony antibakteriális szerek a penicillin megjelenéséig. Azonban az új, hatékonyabb antibiotikumok elterjedése miatt humán gyógyászati jelentőségük napjainkra csökkent. Ennek ellenére a gyógyszeripar továbbra is nagy mennyiségben állítja elő őket főleg állatgyógyászati célokra ¹².

A szulfonamidok elsősorban a vizes környezetben halmozódnak fel, ahol jellemzően fotolízis útján bomlanak. A bomlásuk során legnagyobb mértékben SA képződik (6. ábra) ¹³.



6. ábra. A szulfonamid származékok lehetséges direkt fotolítikus módjai. A bekeretezett vegyületek valóban megjelentek a fotolízis során

Az SA önmagában is rendelkezik a szulfonamidokra jellemző bakteriosztatikus hatással, azonban kémiai sajátosságai miatt a sejtek számára nehezen felvehető, ami hatékonyságát jelentősen csökkenti. Főleg emiatt közvetlen használata nem terjedt el, legfeljebb tartósítószer formájában.

A festékeken és a gyógyszereken túl az SA-t még optikai fehérítőszer és növényvédő szerek alapanyagaként is használják.

3.4 A szulfanilsav biológiai úton történő bontása

A mikroorganizmusok számos aromás vegyület bontására képesek anaerob körülmények között ¹⁴, azonban az aromás aminok általában kivételt képeznek ez alól. Nagyon kevés példa van arra, amikor ilyen reakciót sikerült megfigyelni. Nincs egyértelmű bizonyíték arra sem, hogy a szulfonált aromás vegyületeket lehetséges lenne anaerob körülmények között bontani. Ugyanez igaz a szulfonált aromás aminokra is, ahová az SA is tartozik ⁹.

Szulfonált aromás vegyületek aerob bontására már viszonylag korán sikerült bizonyítékot találni. Erősen szennyezett területekről származó iszapból azonosítottak olyan mikroorganizmusokat, amelyek képesek voltak az SA bontására ^{15,16}. Későbbiekben az azonosított baktériumok száma csak kis mértékben növekedett jelezve azt, hogy az SA bontás képessége nem általános a baktériumok között. Ezeknek a baktériumoknak nagy része csak ko-kulturában, mutualista kapcsolat kialakításán keresztül képes az SA bontására ^{17,18}. Csak kevés baktérium képes önállóan is bontani ezt a vegyületípust ¹⁹⁻²¹.

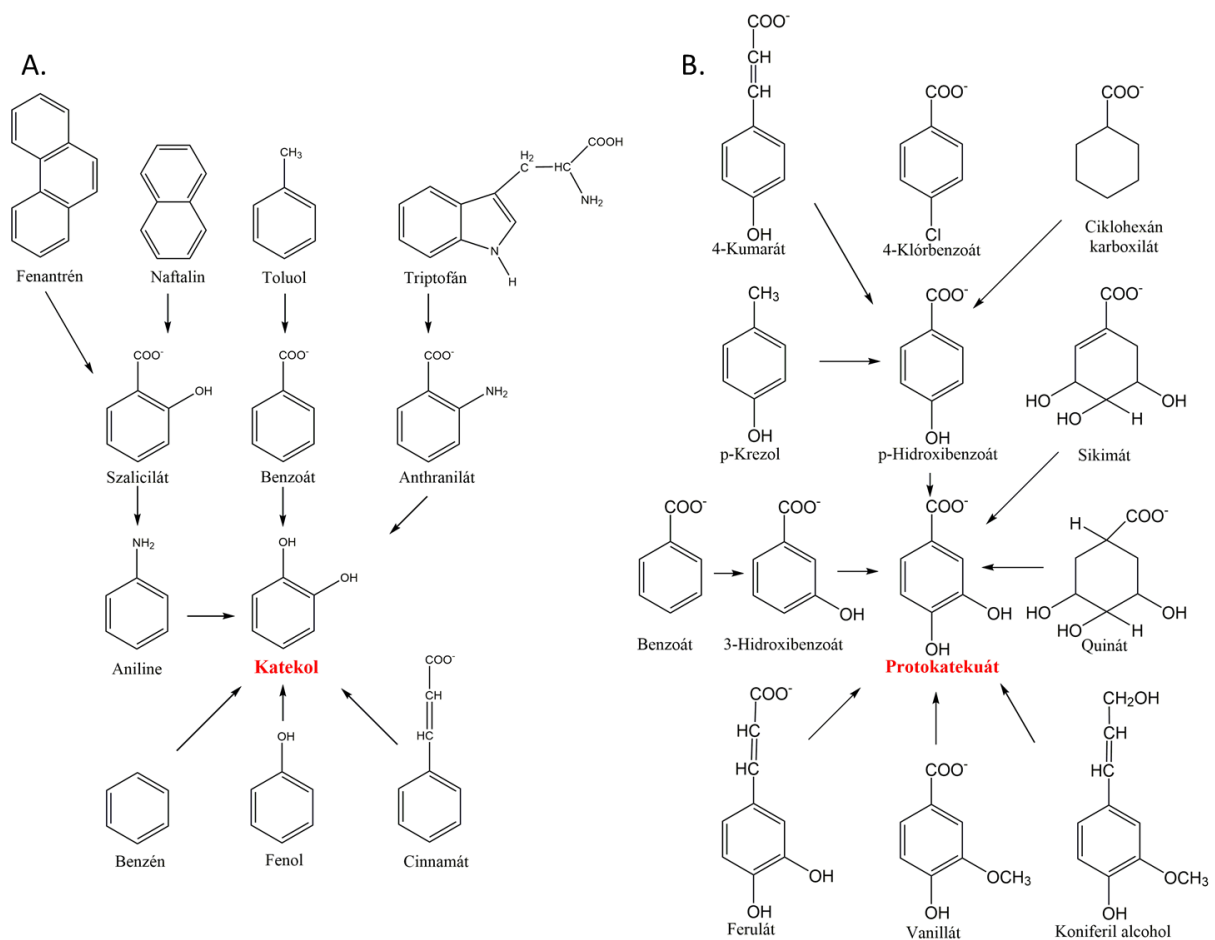
Az SA mikrobiális úton történő bontásának modellje főleg egy két baktériumból álló közösség vizsgálatával alakult ki. Az első olyan baktérium közösséget, amely képes volt a szulfanilsavat felhasználni egyedüli szén-, nitrogén-, kén- és energiaforrásként már 1988-ban sikerült azonosítani ¹⁷. A ko-kultúra legfontosabb tagja, az akkor még ismeretlen fajba tartozó, majd *Hydrogenophaga palleronii* S1-ként, végül *Hydrogenophaga intermedia* S1-ként azonosított baktérium törzs, amely különlegessége, hogy képes az SA-t 4-szulfokatekolon keresztül teljesen lebontani. A kultúra másik tagja, amelyet *Agrobacterium radiobacter* S2-ként soroltak be, ezzel szemben csak a 4-szulfokatekol bontására volt képes. A két baktérium között, SA-n növesztve, kölcsönösen pozitív kapcsolat jön létre. Csak közösen, kevert kultúrát alkotva képesek SA-n stabilan fennmaradni. Feltételezések szerint az *Agrobacterium radiobacter* S2 az SA *Hydrogenophaga intermedia* S1 általi bontása során felszabaduló szulfokatekolt használja tápanyagforrásként és cserébe vitaminokkal (p-amino-benzoosav (PABA), biotin) látja el a *Hydrogenophaga* törzset ²². A másik baktériumközösséget - amely esetén szintén kölcsönös pozitív kapcsolat létrejöttén keresztül történik az SA bontása - csak jóval később, 2011-ben azonosították ¹⁸. Ezt a közösséget is két baktérium faj egy-egy törzse alkotja, a *Hydrogenophaga* sp. PBC, újabb nevén *Hydrogenophaga intermedia* PBC és a *Ralstonia* sp. PBA. Érdekes módon ebben az esetben csak a közösség egyik tagja, a *H. intermedia* PBC képes az SA konverziójára, a *Ralstonia* sp. PBA se az SA-t se a belőle képződő szulfokatekolt nem képes hasznosítani. Feltételezhetően ebben az esetben is a *Hydrogenophaga* sp. PBC hasznosítja a kultúra másik eleme által termelt vitaminokat, azonban a *Ralstonia* sp. PBA szerepe a mutualista kapcsolatban még nem tisztázott ²³.

Az első olyan baktériumot, amely önállóan is képes volt az SA bontására 2001-ben izolálták Perei és munkatársai ¹⁹. Az izolált baktériumot először *Pseudomonas paucimobilis*-ként azonosították, majd az elnevezése későbbiekben *Sphingomonas subarctica* SA1-re majd végül *Novosphingobium resinovorum* SA1-re módosult. Napjainkra az azonosított, az SA-at önállóan

is bontani képes, baktérium fajok száma kis mértékben tovább növekedett az *Agrobacterium sp.* PNS-1²⁰ és *Pannonibacter sp.* W1²¹ törzsekkel.

3.5 A szulfanilsav bontási útvonalának központi elemei

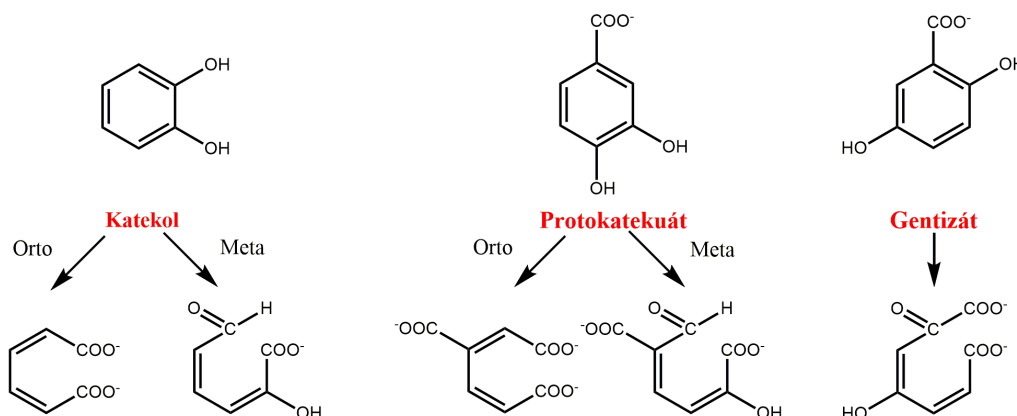
Nagy általánosságban az aromás vegyületek bontási folyamata két szakaszra bontható. Az első, perifériás útvonalnak nevezett szakasz fő célja az aromás gyűrű előkészítése a későbbi hasítására. Ez a szakasz számos párhuzamos enzimátikus kaszkádot tartalmaz, amelyek az aromás vegyületek széles spektrumát képesek átalakítani kisszámú köztiterméké. A sokszínű enzimátikus reakciók egy közös, gyűrű-hidroxiláló oxigenázok (RHO) által katalizált mono- vagy dioxigenációs lépésben megegyeznek, amely dihidroxilált benzol gyűrű, leggyakrabban katekol vagy protokatekuát képződéséhez vezet (7. ábra).



7. ábra. A baktériumok változatos perifériás útvonalhálózattal rendelkeznek, amelyek segítségével képesek az aromás vegyületek széles skáláját kis számú köztiterméké alakítani

A folyamat második, úgynevezett centrális szakaszában történik meg a gyűrű hasítása és továbbalakítása. A folyamat legvégül a trikarbonsav ciklusba (TCA) torkollik. A gyűrű

hasítását speciális dioxigenázok katalizálják. A gyűrű felnyitása történhet a két hidroxil csoport között, ebben az esetben orto-hasítási útvonalról (más néven intradiol hasításról) beszélünk. Történhet valamelyik hidroxil csoport mellett, amely esetben meta-hasítási útvonalról (más néven extradiol hasításról) beszélünk. A harmadik típusú útvonal, a gentizát útvonal, abban az esetben játszódik le, amikor az aromás gyűrűn jelen lévő hidroxil csoportok egymáshoz képest para helyzetben helyezkednek el. Ilyenkor a hasítás egy karbonsav csoportot tartalmazó szénatom és az egyik hidroxil csoportot tartalmazó szénatom között történik meg (8. ábra).



8. ábra. Az aromás gyűrű hasításának lehetséges módjai

A három útvonal közül a meta-hasítási illetve az orto-hasítási, vagy más néven β-ketoadipát útvonalak a legismertebbek. A β-ketoadipát elnevezés onnan származik, hogy a β-ketoadipát az orto-hasítási útvonal fő köztiterméke.

Az eddig vizsgált baktériumok esetén az SA bontása egy módosult β-ketoadipát útvonalon keresztül történt meg. Az útvonal perifériás szakasza során játszódik le az SA dihidroxilációja, amely 4-szulfokatekol (4-SC) képződéséhez vezet. A keletkező köztitermék orto-hasítása az útvonal centrális szakaszában történik 3-szulfomukonát (3-SM) képződése mellett, amely további lépéseken keresztül végül szukcinil-CoA és acetil-CoA formájában lép be a TCA ciklusba ²⁴.

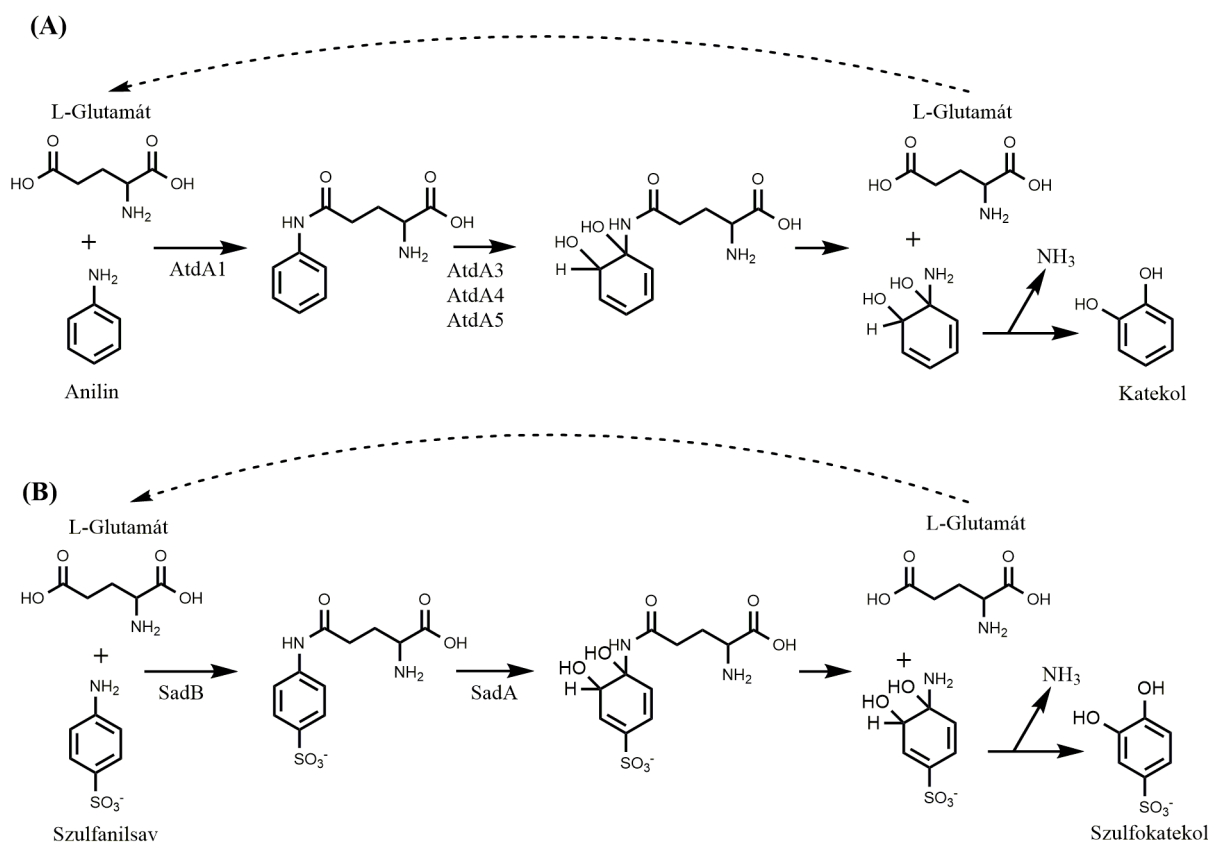
Az SA bontásához alkalmazkodott β-ketoadipát útvonala annyiban tér el a klasszikus β-ketoadipát útvonalaktól, hogy az aromás gyűrű szulfonsav csoportja csak a bontás későbbi szakaszában válik le hidrogénszulfít (HSO_3^-) formájában és alakul szinte azonnal tovább szulfáttá (SO_4^{2-}). Ez azt eredményezi, hogy az SA bontása során olyan szulfonált köztitermékek jönnek létre, amelyek továbbalakítása az enzimek bizonyos szintű specializációját igényli.

3.5.1 A szulfanilsav bontás perifériás útvonala:

Az SA bontására specializálódott útvonal perifériás szakaszát eddig csak egyetlen baktérium esetén sikerült megbízhatóan feltérképezni ²⁵. A *H. intermedia* PBC törzsön végzett transzpozonos mutagenézis segítségével azonosítottak egy olyan géncsoportot, amelynek elemei nélkülözhetetlen szerepet játszanak a bontás első lépései során. Ez a *sadABD* génekből álló csoport egy - az SA által jól szabályozott - operont alkot.

Az operon első tagja, *sadA*, kódolja az útvonal oxigenáz elemét. A gén által kódolt 4-aminobenzolszulfonát (4-ABS)-3,4-dioxigenáz fehérje szekvencia hasonlósága alapján a Rieske típusú oxigenázok (RO) I-es csoportjába sorolható be ²⁶. Az operon második tagja, a *sadB*, egy glutamin szintetázhoz hasonló fehérjét kódol. Szerepe nem teljesen egyértelmű. Az ismert bontási útvonalak esetén a glutamin szintetáz és a dioxigenáz együttes előfordulása elég ritka. Csak az anilin oxidatív deaminálása esetén írtak le hasonló enzimeket. Filogenetikai vizsgálatok is azt erősítették meg, hogy a *sadB* által kódolt fehérje az anilin bontásában szerepet játszó glutamin szintetázokkal mutat rokonságot, velük egy monofiletikus csoportot képez ^{27,28}. A 4-ABS-3,4-dioxigenáz és a glutamin szintetáz nélkülözhetetlen elemei a perifériás útvonálnak. Ők felelősek a szulfanilsav szulfokatekollá történő átalakításáért. Az azonosított operon harmadik tagja, a *sadD*, egy ferredoxin-szerű fehérjét kódol, amely feltételezhetően a dioxigenáz redox partnere. Jelenléte a szulfanilsav bontásának szempontjából nem tűnt esszenciálisnak, azonban jelentősen növeli a rendszer hatékonyságát ²⁶. Az operon hiányosnak tekinthető olyan szempontból, hogy nem tartalmazza a dioxigenáz elektron transzport láncának minden elemét, mivel a ferredoxint speciálisan redukálni képes ferredoxin reduktázt kódoló gén nem része az operonak.

Az SA bontás perifériás folyamatának *in vitro* körülmények között történő vizsgálata napjainkig nem megoldott, ami megnehezíti a lejátszódó enzimatikus reakciók pontos megértését. Ilyen típusú enzimatikus folyamatot legpontosabban az *Acinetobacter sp.* YAA törzs anilin bontási útvonala vizsgálata során jellemezték ²⁸. Ez alapján feltételezhető, hogy a *sadB* által kódolt glutamin szintetáz az L-glutamátot képes a szulfanilsav amino csoportjához kapcsolni. A 4-ABS-3,4-dioxigenáz csak az így létrejövő gamma-glutamil köztitermék hidroxilálását végzi. Továbbiakban két hipotetikus ciklohexadién-diol köztiterméken keresztül 4-szulfokatekol (4-SC) képződik L-glutamát és ammónia felszabadulása mellett. Feltételezések szerint az L-glutamát a reakció során védőcsoportként funkcionál és megakadályozza a szulfanilsav amino csoportjának oxidációját ²⁹ (9. ábra).



9. ábra. A szulfanilsav bontási útvonal perifériás szakaszának lépései az anilin példája alapján

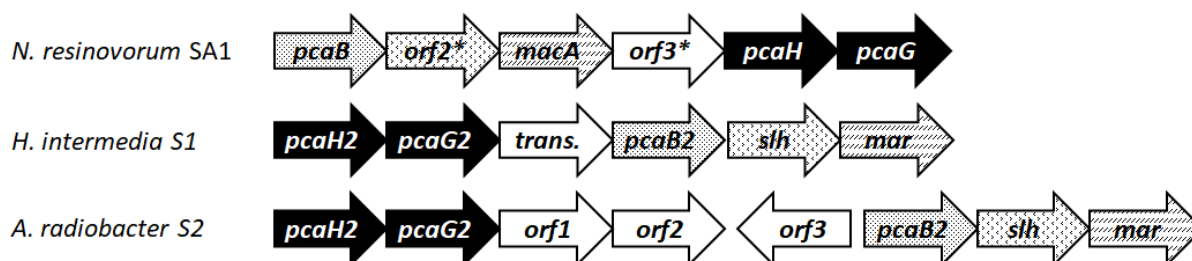
²⁸. Enzimek rövidítése: AtdA1, SadB: Glutamin szintetáz; AtdA3-5: Rieske típusú oxigenáz alfa, béta alegysége és reduktáz komponense; SadA: Rieske típusú oxigenáz alfa alegysége.

3.5.2 A szulfanilsav bontás centrális útvonala

Az SA bontásának centrális szakasza a folyamat legjobban feltérképezett része. Legkorábban a két kölcsönös kapcsolatban álló baktérium, a *H. intermedia* S1 és az *A. radiobacter* S2 fajok esetében sikerült azonosítani az útvonalnak ezt a szakaszát. Az útvonal legrészletesebb jellemzése is ezeknél a fajoknál történt meg ³⁰⁻³². Valamivel később az önálló szulfanilsav bontásra is képes *N. resinovorum* SA1 törzs esetén, proteomikai módszerekkel azonosították a korábban jellemzett enzimek ortológjait és *in vitro* módszerekkel bizonyították azok szerepét a szulfocatekol katabolizmusában ^{33,34}. Továbbá az SA bontására szintén csak kölcsönös kapcsolatban képes *H. intermedia* PBC baktérium esetén is sikerült transzpozon inszerciós mutagenézis segítségével olyan génetet találni, amelyek nélkülözhetetlenek az SA

bontásának centrális szakaszában. Ezek a gének nagyfokú hasonlóságot mutattak a korábban már azonosított génekkel ²⁵, azonban részletesebb vizsgálatuk nem történt meg.

A centrális útvonal elemeinek genomi elrendeződése a vizsgált fajok esetén nem egységes. A két legjobban vizsgált esetben az útvonal elemei két egymáshoz közel elhelyezkedő operonba rendeződnek. Mindkét esetben közös, hogy az aromás gyűrű hasításáért felelő dioxigenáz enzim két alegységét kódoló gének (*pcaHG*) az útvonal többi elemétől elkülönülő operonban találhatóak meg. Míg a *H. intermedia* S1 esetén egy transzpozáz kódozó gén választja el őket az útvonal többi elemétől, addig az *A. radiobacter* S2 esetén a TRAP transzporter rendszert elemeit kódoló gének és egy IclR típusú transzkripció faktor kódozó gén ékelődik a két operon közé ^{35,36}. Ezzel szemben a *N. resinovorum* SA1 törzsben a teljes centrális útvonal egy operonba rendeződik ³³ (10. ábra).



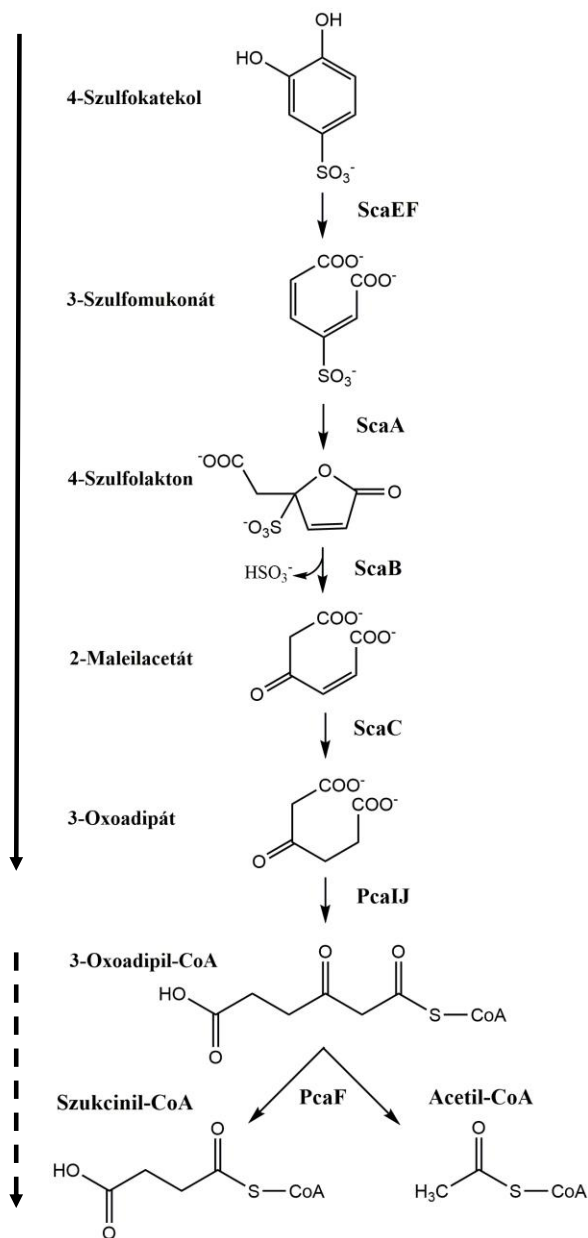
10. ábra. A szulfanilsav bontási útvonalának centrális szakaszát kódoló gének *N. resinovorum* SA1 ³³, *H. intermedia* S1 és az *A. radiobacter* S2 ^{35,36} törzsekben. A gének által kódolt enzimek: *pcaB*, *pcaB2*, (*scaA*), 3-szulfo-cis-cis-mukonát-cikloizomeráz; *orf2**, *slh*, (*scaB*), szulfolakton hidroláz; *macA*, *mar*, (*scaC*), maleilacetát reduktáz; *pcaH*, *pcaH2*, (*scaE*), szulfokatekol 3,4-dioxigenáz béta alegység; *pcaG*, *pcaG2*, (*scaF*), szulfokatekol 3,4-dioxygenáz alfa alegység; *orf3**, oxidoreduktáz; *trans.*, transzpozáz; *orf1*, *orf2*, TRAP-transzporter rendszer elemei; *orf3*, IclR típusú transzkripció faktor

A centrális útvonal első lépése során történik a perifériás útvonal alatt keletkező 4-szulfokatekol (4-SC) oxidációja, ami a vegyület orto-hasításán keresztül 3-szulfomukonát (3-SM) képződéshez vezet. Az oxidációt végző enzimek szerkezetük és fehérjeszekvenciájuk alapján nagyfokú hasonlóságot mutatnak a protokatekol-3,4-dioxigenázokkal (P34O), amelyek a protokatekol (PC) orto-hasításában vesznek részt 3-karboxi-cis-cis-mukonát (3-CM) képződése mellett. Szubsztrát specificitásuk alapján azonban jelentősen eltérnek a klasszikus P34O-októl, mivel a PC mellett a 4-szulfokatekol (4-SC) bontására is képesek – igaz csekély aktivitással. A két enzim típus fehérje szekvenciájának az összehasonlítása és célzott mutagenézis útján rámutatt arra, hogy az új szubsztrát alkalmazásához nem volt szükséges jelentős változásra az enzim szekvenciájában, már egy aminosav cseréje is elegendő ahhoz,

hogy egy klasszikus P34O képes legyen a 4-SC hasítására ^{31,32}. Azonban a hatékonysága jelentősen elmarad az ehhez specifikusan alkalmazkodott enzimekétől. A 4-SC hatékony bontáshoz evolválódott P34O filogenetikai analízisek alapján is jól elkülönülnek a csak PC bontására képes P34O-októl. Ezek alapján a P34O-ok két típusát különítik el. Amíg a klasszikus, csak PC-t hasító P34O oxigenázokat I. (Type I) típusúaknak, addig a 4-SC bontására is képes P34O-okat II. (Type II) típusúaknak tekintik ³².

A bontás következő lépése is követi a 3-oxoadipát útvonal protokatekol ágánál megismerteket. Amíg a klasszikus protokatekol útvonal esetén a 3-CM reciklizálását egy 3-karboxi-cisz-cisz-mukonát cikloizomeráz végzi 4-karboximukonolakton (4-CML) képződése mellett, addig az SA bontása során keletkező 3-SM-okat egy 3-szulfó-cisz-cisz-mukonát cikloizomeráz alakítja tovább 4-szulfolaktonná (4-SL). A két enzimtípus szubsztrát specificitásában nincs jelentős eltérés, mindegyikük képes a 3-SM reciklizálására, azonban a 3-SM-ra specializálódott enzimek hatékonysága lényegesen nagyobb ³⁶. Filogenetikai vizsgálatok alapján a két enzimtípus ebben az esetben is jól elkülöníthető egymástól. A tapasztalt különbségek miatt a klasszikus protokatekol útvonalban résztvevő enzimeket I-es típusúnak (Type I), míg a szulfanilsav bontásában résztvevő enzimet II-es típusúnak (Type II) tekintik ^{36,37}.

Míg a korábbi reakciók átfedtek a β -ketoadipát útvonal protokatekol ágával addig 4-szulfolakton deszulfonálása egy speciális enzimikus reakciót igényel. Ezt a lépést egy aminohidroláz szupercsaládba tartozó, metallo-függő hidroláz végzi, amely szulfit felszabadulása mellett 2-maleilacetátot (2-MA) hoz létre ³⁵. A keletkező maleilacetátot egy maleilacetát reduktáz alakítja tovább 3-oxoadipáttá (béta-ketoadipáttá). Ez a vegyület a klasszikus béta-ketoadipát útvonal utolsó két lépése során alakul tovább szukcinil-CoA-vá és acetyl-CoA-vá, melyek belépnek a TCA ciklusba. Az útvonal utolsó két lépését katalizáló enzimek nem szerves részei a központi útvonalnak. Mivel az általuk katalizált reakciók az összes párhuzamos β -ketoadipát útvonalat képesek kiszolgálni, ezért egy globális közös útvonal elemeinek tekinthetők (11. ábra).



11. ábra. A szulfanilsav bontás útvonalának centrális és közös szakasza.

Az enzimek rövidítése: ScaE (PcaH2), szulfokatekol 3,4-dioxygenáz béta alegység; ScaF (PcaG2), szulfokatekol 3,4-dioxygenáz alfa alegység; SacA (PcaB2), 3-szulfo-cis-cis-mukonát-cikloizomeráz; ScaB (4SLH), 4-szulfolakton hidroláz; ScaC (MAR), maleilacetát reduktáz; PcaI, 3-Oxoacid CoA transzferáz A alegység; PcaJ, 3-Oxoacid CoA transzferáz B alegység; PcaF, 3-oxoadipil-CoA tioláz. Az ábra bal oldalán futó folytonos nyíl jelzi az útvonal centrális szakaszát, a szaggatott nyíl pedig a közös szakaszát.

3.6 A szulfanilsav bontási útvonalának kiegészítő elemei

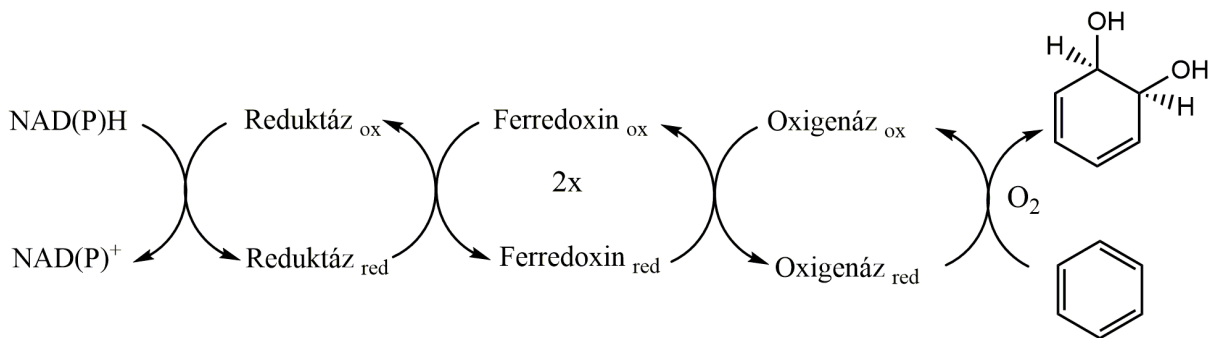
Amíg a szulfanilsav bontásának útvonalát viszonylag kevés faj esetén, de sikerült felismerni és jellemezni, addig a bontáshoz nélkülözhetetlen kiegészítő rendszerek alig kutatottak. Ennek a fő oka, hogy ezek a rendszerek szétszórtan helyezkednek el a genomban és nagy változatosságot mutatnak. Klasszikus módszerekkel nehezen azonosíthatóak.

3.6.1 SA dioxygenáz elektron transzport láncának lehetséges elemei

A Rieske típusú aromás gyűrűhidroxiláló oxigenázok, amelyek közé a szulfanilsav hidroxilálását végző oxigenáz is tartozik, rendszerint több komponensből álló rendszerek. A

hidroxiláció folyamata során az elektrontranszport lánc elemei szállítják az elektront a NAD(P)H-tól az oxigenáz katalitikus nagy alegységéhez, amely felhasználja azt a levegő oxigén molekulájának (O₂) aktiválásához. A katalitikus folyamat során az oxigenáz az aktivált oxigén molekulát használja fel az aromás vegyület régió- és sztereospecifikus hidroxilálásához³⁸. A rendszer központi eleme a terminális oxigenáz, amely lehet homo-multimer (α_n) vagy hetero-multimer (α_nβ_n) szerkezetű³⁹. Mindkét esetben a nagy α alegység felelős a katalitikus aktivitásért, míg a kis β alegység strukturális szerepet tölt be. Az α alegység két fő doménből épül fel. Az N-terminális végén található Rieske típusú [2Fe-2S] vas-kén klaszter kötése képes domén (pfam: PF00355) és egy C-terminális katalitikus domén alkotja, amely egy konzervált mononukleáris vas-kötőhellyel (pfam: PF00848) rendelkezik⁴⁰.

A központi elem működéséhez szükséges folyamatos elektron utánpótlást egy kisegítő elektrontranszport lánc biztosítja. Az elektrontranszport lánc állhat egy reduktázból vagy egy reduktázból és egy ferredoxin fehérjéből^{41,42} (12. ábra.).



12. ábra. A Rieske típusú dioxigenázok elektrontranszport folyamata

A reduktázokat domén szerkezetük alapján három csoportba lehet sorolni⁴³. Az Rd-I csoport a glutation reduktázok (GR) típusát foglalja magába. Az ilyen típusú reduktázok egy N-terminális FAD/NAD(P) kötő doménnel (pfam: PF07992) és egy C-terminális reduktáz doménnel (pfam: PF14759) rendelkeznek. A Rd-II és Rd-III csoport a ferredoxin-NADP⁺-reduktázok típusát (FNR) foglalják magukba. A két csoport tagjai ugyanolyan flavin, NAD(H) és [2Fe-2S] kofaktorokat kötő doméneket tartalmaznak. Az eltérést egyedül csak a domének sorrendjében van. Amíg az Rd-II (FNRC) csoport tagjai esetén a fehérje N-terminális végén a flavin-kötő domén (pfam: PF0097) helyezkedik el, majd azt követi a NAD-kötő domén (pfam: PF00175) és végül a növényi típusú [2Fe-2S]-kötő domén (pfam: PF00111) addig az Rd-III (FNRn) csoport tagjai esetén a fehérje N-terminális végén a [2Fe-2S]-kötő domén, majd a flavin-kötő domén végül a NAD-kötő domén helyezkedik el⁴³.

Az elektrontranszport lánc legbizonytalanabb pontja a ferredoxin alegység. Megléte nem minden esetben szükséges és nagy változatosságot mutat. Domén szerkezetük alapján alapvetően két fő csoportba lehet őket sorolni ⁴³. A ferredoxinok első csoportja, a Fd-I, egy [4Fe-4S] vas-kén kofaktor kötő doménnel (pfam: PF13459), ezzel szemben a ferredoxinok második csoportja, az Fd-II, egy Rieske típusú [2Fe-2S] vas-kén kofaktor kötő doménnel (pfam: PF00355) rendelkezik. Ezekon kívül azonban olyan ferredoxinok is előfordulnak, amelyek rubredoxin és növényi típusú [2Fe-2S] kofaktor kötő doménnel bírnak (pfam: PF00301; PF00111) ⁴⁴.

A terminális oxigenáz és az elektrontranszport lánc egy funkcionális egységet képeznek, de az őket kódoló gének nem minden esetben rendeződnek egy transzkripciós egységbe ^{45,46}. Ez igaz a szulfanilsav hidroxilálását végző oxigenázra is *H. intermedia* PBC esetén. Az elektrontranszport lánc elemei közül csak a növényi típusú (PF00111) ferredoxin fehérjét kódoló *sadD* gén található egy operonban a terminális oxigenáz génjével (*sadA*) ²⁶. Ez a moduláris felépítés nagymértékben megnehezíti a RHO teljes folyamatának a megismerését.

Számos kísérlet történt már a RHO-k csoportosítására és a csoportosításon keresztül az elektrontranszport láncuk felépítésének megértésére. Kezdeti próbálkozások a terminális oxigenázok katalitikus α alegységeinek fehérjeszekvencia hasonlóságán alapultak ⁴⁷, majd később figyelembe vették az oxigenázok α alegységeinek fehérjeszekvencia hasonlóságamellát a kis β alegységek meglétét és az elektrontranszport lánc összetételét is ^{43,44}. Azonban az RHO csoportok elektrontranszport láncai nem mutatnak egységes képet ami megnehezíti azonosításukat.

3.6.2 Az aromás savak transzporterei

Az SA egyik fontos kiegészítő eleme a katabolikus folyamatokat megelőző transzport folyamat. Az aromás savak csak nem-disszociált formájukban képesek passzív diffúzió segítségével átjutni a baktériumok membránján. Ilyen formában azonban csak savas közegben találhatóak. Normál és lúgos körülmények között a disszociált változatuk fordul elő, amely nem képes diffúzióval átjutni a baktériumok membránján, ezért hatékony felvételük aktív transzportekhez kötött. Ilyen aktív transzporterek elterjedtek a baktériumok között ⁴⁸⁻⁵⁴. Ezek a transzporterek nagyon változatosak lehetnek, azonban a legnagyobb részük egy elkülönülő aromás sav:H⁺ szimporter (Aromatic Acid:H⁺ Symporter; AAHS) fehérjecsaládot hoz létre a fő transzportsegítő supercsaládon (major facilitator superfamily; MFS) belül ⁵⁵. Ebbe a fehérjecsaládba tartozik a 4-hidroxibenzoát (4-HB) és a protokatekol transzportere (PcaK) ⁵⁶, a

3-hidroxibenzoát transzportere (MhbT) ⁵⁷, a 2,5-dihidroxibenzoát transzportere (GenK) ⁵⁸, a benzoát transzportere (BenK) ⁵⁹ és a vanillát transzportere (VanK) ⁶⁰.

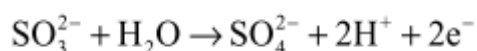
Már számos esetben megfigyelték hogy az SA diffúzióval nem képes átjutni a baktériumok membránján ⁶¹, azonban a hatékony felvételét lehetővé tevő specifikus transzportert még nem sikerült azonosítani.

3.6.3 Szulfid oxidáló enzimek

Az SA bontása során várhatóan jelentős mennyiségű szulfid (SO_3^{2-} vagy HSO_3^-) szabadul fel. A szulfid erős redukálószer, de emellett erős nukleofil tulajdonsággal is rendelkezik. Sejten belül a nagy reaktivitása miatt számos - a sejt számára káros - reakcióban képes részt venni. Ezek közé tartozik az irreverzibilis kötődési képessége a biomolekulák aldehid és keton csoportjához, emellett képes reverzibilisen kötődni az uracil és citozin csoportokhoz, ami deaminációhoz és transzaminációhoz vezet, aminek eredményeként mutációk jönnek létre. Továbbá ezek az anionok számos olyan reakciót indukálnak, amelyek reaktív oxigén gyökök (ROS) képződéséhez vezetnek.

A sejten belül keletkező szulfid káros hatásainak megelőzésére a mikroorganizmusok számos módszerrel rendelkeznek. Ezek a módszerek a szulfid exportját, redukcióját és direkt vagy indirekt oxidációját foglalják magukba.

A szulfanilsav bontása során jelentős mennyiségű szulfát megjelenése volt megfigyelhető a tápoldatban ^{17,18,20,21,34}. Ez arra utal, hogy a felszabaduló szulfid detoxifikálásában az oxidáció játssza a fő szerepet. A mikroorganizmusok szulfid oxidációjának két lehetséges útja ismert. A direkt oxidáció, amelyet molibdoprotein kofaktort (Mo-co) tartalmazó szulfid-oxidáló enzimek (SOE-k) végeznek és az indirekt, adenzin-monofoszfát (AMP) függő oxidáció, amely adenzin-5'-foszfo-szulfáton (APS') keresztül történik meg. A két út eltérő szerepet tölt be a mikroorganizmusokban. Amíg az indirekt útvonal a citoplazmában (a sejtmembrán citoplazmatikus oldalán) játszódik le és fő szerepe a redukált kénvegyületek oxidációjával történő energia előállításában van, addig a kéntartalmú szerves vegyületek bontása során felszabaduló szulfid detoxifikációja kizárólag a periplazmatikus térben történik meg a direkt útvonalon keresztül. Az összes eddig jellemzett SOE a „mononuclear molybdenum” enzimek egy fehérjecsaldájába tartozik. A molibdén-tartalmú enzimek egy különleges tulajdonsága, hogy képesek a vizet felhasználni és annak oxigénjét átvinni a szubsztrát molekulára ⁶² (13. ábra.).



13. ábra. A SOE által katalizált reakció általános egyenlete

A SOE-k széles körben elterjedtek az élővilágban. A növényekben és a gerincesekben azonosított ilyen típusú enzimek jellemzően homodimer felépítésűek. A N-terminális végükön egy molibdén kötésre alkalmas és egy dimerizációért felelős domén található, a C-terminális végén pedig a hem-kofaktort nem-kovalensen kötő citokrom *b*₅ domén.⁶³

Ezzel szemben a baktériumokban azonosított SOE-k lényegesen változatosabb szerkezettel rendelkeznek. A legjobban jellemzett bakteriális szulfid-oxidáló enzim a *Starkeya novella* törzsben (korábban: *Thiobacillus novellus*) azonosított SorAB heterodimer enzim komplex. Az enzim két alegységből épül fel. A SorA fehérje tekinthető a rendszer központi alegységének, amely tartalmazza a molibdén kötéséhez szükséges domént (pfam: Oxidored_molyb, PF00174) és egy dimerizációhoz szükséges domént (pfam: Mo-co_dimer, PF03404). Ehhez kapcsolódik a kis SorB alegység, amely a citokrom c fehérje család tagja és a hem kofaktor kötésében vesz részt a citokrom c-szerű doménjén (cytochrome c-like domain, IPR009056) keresztül.

A katalízis során két elektron jut át a szulfitról a SorA fehérje aktív helyén található, oxidált állapotban lévő Mo(VI) kofaktorára, amely ennek hatására redukálódik. A két elektron ezt követően szakaszosan a SorB fehérje által kötött hem kofaktorra kerül, majd onnan egy külső elektron akceptorra vándorol, amely azt elszállítja a légzési lánc feltételezhetően citokrom oxidáz elemére^{64,65}. A *Starkeya novella* esetén a külső elektron akceptor szerepét egy citokrom c550 fehérje tölti be.

A szulfid-oxidáló enzimek (SOE-k) másik jól jellemzett tagja a *Sinorhizobium meliloti* fajban azonosított SorT fehérje⁶⁶. A SorA(B)-val ellentétben a SorT homodimer formában fordul elő. A SorT fehérje a SorA-val megegyező domén szerkezettel rendelkezik. Ebben az esetben azonban a SorT aktív helyén lévő és a redox folyamat során redukált állapotba kerülő Mo (IV) két elektronját szakaszosan közvetlenül egy külső elektron akceptorra adja tovább, ami azt feltételezhetően szintén a légzési lánc citokrom oxidáz elemének adja tovább. A külső elektron akceptor szerepét egy citokrom c fehérje tölti be, amelyet a SorT fehérjét kodoló génnel egy operonban lévő *sorU* gén kódol^{67,68}.

3.6.4 Szulfít transzporterek

Mivel a szulfít oxidációja a periplazmatikus térben játszódik le, ezért nélkülözhetetlen olyan transzportereknek a megléte, amelyek a citoplazmában keletkező szulfítot a reakció helyére szállítják. Eddig a szulfít exportjában résztvevő transzportereknek 4 típusát azonosították. Az első ilyen típusú transzportert a *Paracoccus denitrificans* NKNIS törzsben azonosított, *tauZ* gén kódolja⁶⁹. Sejten belüli szerepe a taurin bontása során felszabaduló szulfít exportja. Eddig két ortológját sikerült azonosítani a *Paracoccus pantotrophus* NKNCYSA törzsben⁷⁰. Ebben az esetben az egyik ortológ szintén a taurin bontásában vesz részt, míg a másik *suyZ* által kódolt transzporter feltételezhetően a ciszteát (ciszteinsav sója) bontása során felszabaduló szulfít exportját végzi⁷⁰. A *tauZ* egy másik ortológja, a *Ruegeria pomeroyi* DSS-3 törzsben azonosított *cuyZ* gén, ami szintén a 3-szulfo-alanin (ciszteát) bontása során játszik szerepet⁷¹. Egy hasonló gént azonosítottak *Roseovarius nubinhibens* ISM törzsben, melynek terméke feltételezhetően a szulfolaktása során felszabaduló szulfít transzportjáért felelős⁷².

A *Cupriavidus necator* H16 törzsben azonosított, *tauE* gén, amely a negyedik típusú szulfít transzportert kódolja, nem tekinthető a *tauZ* ortológjának, azonban kifejeződését számos szulfonált vegyület bontása során kimutatták⁷³.

3.7 A *Sphingomonadaceae* család tagjainak változatos genomi összetétele

A *Sphingomonadaceae* család tagjait már régóta komoly figyelem övezi rendkívül változatos anyagcseréjük miatt. Az ide tartozó fajok számos természetesen és mesterségesen előállított vegyület bontására képesek⁷⁴. Sajátosságuk, hogy a bontást lehetővé tevő metabolikus útvonalak génjei gyakran nem alkotnak egységes operont, hanem a genomban szétszórva, kisméretű génklaszterekbe rendeződve helyezkednek el. Ezek a génklaszterek sok esetben a család tagjainál gyakran előforduló, nagyméretű plazmidokon találhatóak meg⁷⁵. A változatos génkészlet és feldarabolt operon szerkezet létrejöttének fő oka a plazmidok által közvetített génátvitel, a kromoszóma és a plazmid közötti rekombináció, továbbá a fágok és transzpozonok által közvetített genomi átrendeződések⁷⁶. A *N. resinovorum* SA1 törzs esetén megismert SA bontási útvonalra is jellemző ez a feldarabolt operon szerkezet. Az útvonal centrális szakaszának génjei egy egységes klaszterbe rendeződnek, amely fizikailag elkülönül az útvonal többi, még ismeretlen elemétől (ld később).

Az SA bontásának képessége a *Sphingomonadaceae* család tagjai között is ritka. Ez idáig csak az *N. resinovorum* SA1 törzs esetén sikerült bizonyítani ennek a tulajdonságnak a meglétét.

3.8 Genomszekvenálás

A DNS szekvenálás óriási fejlődésen ment keresztül 1977 óta, amikor Fred Sanger kifejlesztette dideoxi láncterminációs szekvenálási technikáját ⁷⁷. A baktérium genomszekvenálás forradalma az 1990-es évek elején indult meg, amikor hozzákezdtek a modellorganizmusok, mint például *Escherichia coli* és *Bacillus subtilis*, genomjának szekvenálásához ^{78,79}. A forradalom első igazi nagy eredménye 1995-ben következett be, amikor shotgun („sörétes puska”) módszer használatával sikerült meghatározni a *Haemophilus influenzae* teljes genomszekvenciáját ⁸⁰. A módszer lényeg hogy a fág genomját véletlenszerűen feldarabolták majd a darabokból létrehozott klonkönyvtárat megszekvenálták. Az első generációs szekvenálás csúcsát a humán genom megszekvenálása jelentette ^{81,82}. A szekvenálás technikájában a következő nagy előrelépés a 2000-es évek közepén következett be a nagy áteresztőképességű vagy más néven újgenerációs (második generációs) szekvenáló technikák megjelenésével ^{83,84}. Ezek a módszerek a kisebb kutatócsoportok számára is elérhetővé tették a genomszekvenálást. Azonban a technikában rejlő korlátok miatt, amely főleg a rövid leolvasási (read) hosszra vezethetők vissza, a szekvenált genomok teljes összeillesztése (complete genome) továbbra sem volt könnyen megvalósítható. Emiatt a szekvenált genomok jelentős része befejezetlen (draft genome) formában maradt. Amíg a teljesen összeillesztett genom a végleges formát képviseli, ahol a szekvencia összes bázispárjának orientációját és minőségét nagy biztonsággal meghatározták, addig a befejezetlen, draft állapotban lévő genomok különböző hosszúságú szekvenciadarabokból (kontigokból) épülnek fel. Ezeknek a kontigoknak sok esetben se a sorrendje, se az irányultsága nem ismert. A DNS-szekvenálás következő fejlődési lépése, az úgynevezett „single-molecule, long-read sequencing” főleg a rövid leolvasás hosszából adódó problémák megoldására törekszik. Az első elérhető ilyen technológiát nyújtó platform a Pacific Bioscience „single-molecule real time ” SMRT rendszere volt ⁸⁵, amelyet később követett az Oxford Nanopore MinIon rendszere ⁸⁶. Ezek a rendszerek jelentős áttörést hoztak a leolvasás hosszában, akár több tíz kilobázis hosszúságú szekvenciák leolvasására is képessé váltak, azonban még komoly problémákkal küzdenek, amelyek közül a leg súlyosabbak a pontatlanságuk, ami elterjedésüket jelentősen korlátozza.

A genomszekvenálás a technikai fejlődésének és vele párhuzamosan csökkenő árának következtében egyre elérhetőbbé vált a kisebb kutatócsoportok számára is, ami a szekvenált genomok számának robbanásszerű növekedéséhez vezetett. 2017 augusztusában 104993 prokarióta genom volt elérhető a GenBank adatbázisában. Azonban ezeknek a genomszekvenciáknak csak egy töredéke, alig 8%-a volt teljes. A többi genom valamilyen

befejezetlen formában volt csak elérhető főleg scaffold (szuperkontig) és kontig formájában. Ennek a fő oka, az hogy a jelenleg elterjedt szekvenálási technikák által létrehozott rövid hosszúságú szekvenciák (read), amelyek (függetlenül a lefedettség mértékétől) összeillesztése egy teljesen zárt genomra komoly kihívást jelent ⁸⁷, sok esetben időigényes és költséges manuális utómunkálatokat (finishing) igényelnek ⁸⁸.

A befejezetlen állapotban lévő genomok önmagukban is alkalmasak számos fontos tudományos kérdés megválaszolására, azonban jelentős korlátokkal rendelkeznek. A több darabból álló genom szekvencia nagymértékben megnehezíti és pontatlanná teszi a genom annotációját, amely a funkcionális genomika egyik nélkülözhetetlen eleme. Továbbá nem ad teljes képet a genom pontos összetételéről, mert nem különíthetőek el egyértelműen a kromoszómális elemek az esetleg előforduló extrakromoszómális elemektől. Ez a baktériumok esetén különösen fontos mivel számos fontos metabolikus útvonal csak a plazmidjaikon találhatóak meg. Ezen kívül szinte lehetetlenné teszi a genom evolúciója során bekövetkező események, mint például gén duplikációk, gén deléciók és a horizontális gén transzferek tanulmányozását.

3.9 Transzkriptom vizsgálat

A transzkriptom egy sejt vagy egy sejtpopuláció egy időpontban kifejeződő teljes RNS készletét jelenti. Ez egy rendkívül dinamikusan változó rendszer. Vizsgálatával fény deríthető arra, hogy a sejt fejlődése során, illetve a különböző környezeti körülmények hatására, milyen típusú RNS-eket fejez ki, azok milyen szerkezetűek és az adott pillanatban mekkora mennyiségben találhatóak meg. Sok esetben azonban csak a sejt mRNS populációjában bekövetkező változás vizsgálatát jelenti. Ezeknek a vizsgálatoknak a fő célja a meghatározott körülmények között aktívan átíródó génkészlet megismerése. Ez végső soron hozzájárulhat a sejt fenotipikus változásaiért felelős genetikai rendszerek megismeréséhez.

A sejt globális transzkript összetételének vizsgálatára számos technikát fejlesztettek ki. Ezek a technikák főleg hibridizáció és szekvenálás alapú megközelítéseket követnek. A hibridizáción alapuló technikák során legtöbbször fluoreszcens festékkel jelölt, mRNS-ből reverz transzkripcióval átírt komplementer DNS (cDNS) populációkat inkubálnak szilárd felszínre rögzített, meghatározott geometriai tömbbe rendezett nagyszámú, rövid DNS szakasszal. A rögzített DNS szakaszokat leggyakrabban polimeráz-láncreakció (PCR) segítségével állítják elő ismert vagy jóslott génszekvenciák alapján ⁸⁹. Ezzel az úgynevezett „microarray” technikával akár több tízezer különböző mRNS-t/cDNS-t lehet párhuzamosan detektálni és mennyiségét meghatározni ⁹⁰. A technika egy továbbfejlesztett változata az

úgynevezett „tiling microarray”. Ebben az esetben a rögzített DNS szekvenciák a genom teljes egészét lefedik beleértve a kódoló szekvenciák mellett a nem kódoló részeket is ⁹¹. A hibridizáción alapuló technikák előnyei között szerepel a nagy áteresztőképesség, költséghatékonyság. Emellett a „tiling microarray” megjelenésével már alkalmassá váltak teljes genom feltérképezésére is. Használhatóságukat azonban jelentősen korlátozza a technika alacsony dinamikatartománya. Amíg az alacsony kifejeződéssel rendelkező gének a kereszt-hibridizáció miatt fellépő nagy háttér következtében nem érzékelhetők, addig a nagyon magas kifejeződési szintek pontos meghatározása pedig a jelek telítődése miatt nem megoldható. Mi több, az ilyen típusú vizsgálatok a genomszekvencia ismerete nélkül nem megvalósíthatóak. A legnagyobb problémát a reprodukálhatóság rossz hatásfoka jelenti, ami elsősorban a technika bonyolultságából adódik ⁹².

A szekvenálás alapú technikák a hibridizációs technikákkal szemben képesek közvetlenül meghatározni a cDNS szekvenciáját. Az első ilyen típusú technikák az első generációs Sanger szekvenálásra épültek. A cDNS szekvenálásából létrehozott EST (Expressed Sequence Tags) segítségével számos gén expressziójának megismerésére vált lehetőség ⁹³. Azonban ez a módszer viszonylag alacsony kihozattalal rendelkezik, drága és nem alkalmas mennyiségi meghatározásra (nem kvantitatív). A limitációk áthidalására különböző úgynevezett tag-alapú módszereket hoztak létre. Ezek közül a legelterjedtebbek a SAGE (serial analysis of gene expression) ⁹⁴, CAGE (cap analysis of gene expression) ⁹⁵ és az MPSS (massively parallel signature sequencing) ⁹⁶. A módszereknek az a lényege, hogy a cDNS egy rövid szakaszát (tag) felszorzozzák, a szekvencia részleteket egymáshoz ligálják, végül az így létrehozott hosszú DNS szakaszt szekvenálják meg. Ez a megközelítés nagy áteresztőképességű és képes pontosan meghatározni a gén kifejeződési szintjét. Azonban komoly problémája, hogy a rövid tag szekvenciák nagy részét nem lehet egyedi módon rátérképezni a referencia genomra.

Ahogy a második generációs szekvenálás forradalmasította a bakteriális genomok szekvenálását, úgy a szekvenálás alapú transzkriptom vizsgálat jelentős fejlődését is elősegítette. A második generációs szekvenáláson alapuló új technológiát RNA-seq nek nevezik. Az RNA-seq során a teljes RNS populációból, vagy leggyakrabban annak csak mRNS-t tartalmazó részéből cDNS könyvtárat hoznak létre. Az így létrehozott könyvtár minden molekuláját, egy amplifikációs lépés követően vagy akár annak elhagyásával, nagy áteresztőképességű technikák felhasználásával szekvenálják. Az így kapott leolvasásokat a referencia genom megléte esetén közvetlenül ráillesztik a genomra és az illeszkedő leolvasások számából következtetnek a gén kifejeződés mértékére. A referencia genom hiányában a

leolvasásokat első lépésben egymáshoz illesztve (de novo assembly) megalkotják a genom transzkriptom térképét, amelyre a következő lépésben ráillesztik a szekvenálásból kapott readeket és az illeszkedő leolvasások számából következtetnek a gének kifejeződésének mértékére. Az RNA-seq előnye, hogy nem feltétlenül igényli a referencia genom meglétét, ami alkalmazhatóságát kiterjeszti az előzőleg kevésbé ismert élőlényekre is. Képes akár egy bázispár pontossággal meghatározni a transzkriptek határait. A vizsgálat nagy dinamikatarományal rendelkezik. Nagyon alacsony háttér szignál keletkezik, mivel a kapott szekvenciák nagy biztonsággal illeszthetők a templátra, emiatt nagy érzékenység érhető el. Az illesztés felső határa sem korlátozott, emiatt nem alakul ki telítődés. További előnye, hogy viszonylag pontosan képes meghatározni az expressziós szintet. Azonban ez a technológia sem mentes a problémáktól. Mivel a második generációs szekvenáló technikák csak rövid szekvenciákat tartalmazó könyvtárak esetében használhatóak, ezért az ilyen típusú könyvtárak létrehozásához vagy az RNS molekulákat vagy a cDNS molekulákat fel kell darabolni megfelelően rövid szekvencia részletekre. Számos darabolási technikát fejlesztettek ki, de mindegyik egy rá jellemző torzítást visz a cDNS könyvtárba. Erre a harmadik generációs szekvenáló technikák elterjedése nyújthatnak majd megoldást, amelyek már képesek lesznek a teljes mRNS szekvenciájának a meghatározására. További problémát jelent a szekvenálás során keletkező óriási adatmennyiség tárolása és feldolgozása, amely komoly bioinformatikai kihívást jelent ⁹⁷.

4 Célkitűzések

Az SA biológiai úton történő hatékony lebontásához ismernünk kell azokat az útvonalakat, amelyek lehetővé teszik a baktériumok számára a vegyület/vegyülettípus teljes metabolizmusát. Mivel az *N. resinovorum* SA1 törzs képes az SA-t önállóan bontani és felhasználni azt egyedüli szén-, nitrogén-, kén- és energiaforrásként nagyszerű lehetőséget biztosít ahhoz hogy segítségével megértsük ennek a rendkívül ellenálló vegyületnek a biológiai úton történő bontási folyamatát. Különösen fontos ez, mert jelenleg nem áll rendelkezésünkre olyan vizsgálat, amely ezt átfogóan bemutatná. Az elérhető példák - amellett hogy főleg a bontást közvetlenül végző enzimekre koncentrálnak - csak nagyon kisszámú baktérium vizsgálatából származnak. Amíg a bontás perifériás szakaszának elemeit csak egyetlen baktérium esetén sikerült azonosítani és jellemezni addig a bontás viszonylag jobban ismert centrális szakaszának működését is csak két baktérium példáján keresztül tárták fel. A bontási folyamatot lehetővé tevő kiegészítő rendszer komponensei szinte teljesen ismeretlenek. A perifériás útvonal oxigenázának a működéséhez szükséges ETC elemeiről csak bizonytalan elképzeléseink vannak. Az olyan fontos kiegészítő rendszerekről, amelyek lehetővé teszik az SA felvételét és az SA bontása során felszabaduló szulfid detoxifikációját, pedig semmilyen információval sem rendelkezünk. Ezen kívül fontos kérdésként merül fel az SA speciális hatása a bontására specializált rendszerek kifejeződésére illetve a globális génkifejeződésre.

A fenti hiányosságok megválaszolására a következő célokat tűztem ki:

- Az *N. resinovorum* SA1 törzs genomszekvenciájának meghatározása második generációs szekvenáló technikák használatával.
- Az *N. resinovorum* SA1 genomjában megtalálható β -ketoacid útvonalak elemeinek azonosítása.
- Az SA bontási útvonal elemeinek összehasonlítása a már jellemzett útvonalakéval.
- Az útvonal eredetének feltárása, az esetleges horizontális géntranszfer eseményének alátámasztása nukleotid statisztikai és filogenetikai módszerekkel.
- Az SA bontási útvonalak szabályozásának megismerése a promóterek aktív régióinak azonosításán keresztül.
- Mivel az SA egy olyan nem ideális szubsztrát, amely a sejt számára “nehezen fogyasztható”, sőt a jelenléte a citoplazmatikus térben kifejezetten toxikus, ezért célom volt az éhezésnek/stressznek és az SA-nak az *N. resinovorum* SA1 törzs teljes

transzkriptomjára gyakorolt hatásának a megismerése RNA-seq technika használatával, valamint az eredmények részleges validálása RT-qPCR technikával.

- Az SA bontását lehetővé tevő kisegítő rendszerek elemeinek megismerése.

5 Anyagok és módszerek

5.1 A munka során használt baktérium törzsek, plazmidok és oligonukleotidok

5.1.1 Baktérium törzsek megnevezése és leírása

Törzsek	Genotípus	Referencia
<i>Escherichia coli</i> XL1-Blue MRF'	$\Delta(mcrA)183$, $\Delta(mcrCB-hsdSMR-mrr)173$, <i>endA1</i> , <i>supE44</i> , <i>thi-1</i> , <i>recA1</i> , <i>gyrA96</i> , <i>relA1 lac</i> [F' <i>proABlacI_qZλM15 Tn10 (Tcr)</i>] _c	Stratagene
<i>Novosphingobium resinovorum</i> SA1	vad	(Perei 2001); DSMZ ID: DSM- 103995

5.1.2 Plazmidok megnevezése és leírása

Plazmidok	Genotípus	Referencia
pBluescript SK +	ColE1 replikon, Amp ^r	Stratagene
pPROBE-NT	pBBR1, gfp reporter gén, Km ^r	(Miller 2000)
PI_S4	pPROBE-NT plazmid, amely a <i>Sma</i> I hasítási helyénél a <i>sadA</i> TSS előtti -1 től -619 ig terjedő intergenikus régiót tartalmazza, OSpromF2/OSpromR1, Km ^r	Ez a munka
PI_S5	pPROBE-NT plazmid, amely a <i>Sma</i> I hasítási helyénél a <i>sadA</i> TSS előtti -1 től -551 ig terjedő intergenikus régiót tartalmazza, OSpromF8/OSpromR1, Km ^r	Ez a munka
PI_S6	pPROBE-NT plazmid, amely a <i>Sma</i> I hasítási helyénél a <i>sadA</i> TSS előtti -1 től -485 ig terjedő intergenikus régiót tartalmazza, OSpromF9/OSpromR1, Km ^r	Ez a munka
PI_S3	pPROBE-NT plazmid, amely a <i>Sma</i> I hasítási helyénél a <i>sadA</i> TSS előtti -1 től -276 ig terjedő intergenikus régiót tartalmazza, OSpromF7/OSpromR1, Km ^r	Ez a munka
PI_S2	pPROBE-NT plazmid, amely a <i>Sma</i> I hasítási helyénél a <i>sadA</i> TSS előtti -1 től -212 ig terjedő intergenikus régiót tartalmazza, OSpromF3/OSpromR1, Km ^r	Ez a munka
PI_E1	pPROBE-NT plazmid, amely a <i>Sma</i> I hasítási helyénél a <i>sadA</i> TSS előtti -192 től -829 ig terjedő intergenikus régiót tartalmazza, OSRHF/OSpromR2, Km ^r	Ez a munka
PI_E4	pPROBE-NT plazmid, amely a <i>Sma</i> I hasítási helyénél a <i>sadA</i> TSS előtti -283 től -829 ig terjedő intergenikus régiót tartalmazza, OSRHF/OSpromR3, Km ^r	Ez a munka

PII

pPROBE-NT plazmid, amely a *Sma*I hasítási helyénél a *scaA* TSS előtti -1 től -107 ig terjedő intergenikus régiót tartalmazza, OSII_F/OSII_R, Km^r

Ez a munka

5.1.3 Az intergénikus régiók PCR segítségével történő amplifikációjához használt oligonukleotidok megnevezése és szekvenciája

Név	Szekvencia (5' - 3')
OSRHF	CCTTCAACATGGCGTTGTCG
OSpromF2	AAGAAGCTGCTGGCCGAAC
OSpromF8	TTTGGCGGAAGTTCAGG
OSpromF9	CACATCGTCGTTTCTTCG
OSpromF1	GGTGGAGAGCTTCATCTTG
OSpromF7	CCGCTTGCTTTCAGAG
OSpromF3	ACGAAATGCGCCTGTTTGTC
OSII_F	CTGGAGATCGCAATGTCAG
OSpromR1	GGGCTGGATTCTCCTTC
OSpromR2	AGACAAACAGGCGCATTTTCG
OSpromR3	TTGTGCGAGCGCACGTATC
OSpromR4	TGTTGGCTTTGACCTGAG
OSII_R	GACATCTCAGCGGACGGG

5.1.4 Az RT-qPCR során használt oligonukleotidok megnevezése és szekvenciája

Gén neve	Szekvencia (5'-3')	Oligonukleotid neve	Termék hossza (bp)
<i>gyrB</i>	TCGGACGACAATGCCTAC CGCGAACTTCATCCAGTG	gyrB_F gyrB_R	129
<i>sadA</i>	CACGTTTCGTGCACCAATC ATTCACCGGCGTCGTATAG	OSRHF_RT OSRHR_RT	147
<i>scaE</i>	CGGTCGCTATATCCACAAG CGCCAGACATTGTGCAG	OSCAEF_RT OSCAER_RT	156
<i>pcaH</i>	GATCTGGCAGGCCAATG GATCGTGGTGTAGCTGTAG	pcaH_F pcaH_R	133
<i>pcaJ</i>	CAGGCAGTTACTCTGCTTAC GCACGGAAACCTGGAAAG	pcaJ_F pcaJ_R	115
<i>ferB</i>	GATCGCTCTCAAGACTTCC GAGACAGTCGGTGATACG	OFERF_RT OFERR_RT	165
<i>pcaK</i>	GACGGCACCAATGAGTC AAGCCACTGCAGAATTC	pcaK_F pcaK_R	153

<i>sorX</i>	CGATCGCTATGACCAAGG TACGCACGCCAGCATAG	sorT_F sorT_R	126
<i>sorA</i>	ATGGGCAACGCCAAATG CGACTTGCGCAAATCCG	soe2_F soe2_R	138
<i>tauZ1</i>	CGAAGGCATCGCATTTCAG AGGATTGCGCTCACGAC	tauZ1_F tauZ1_R	141
<i>tauZ2</i>	GATATGCGGTGCGTCTG AAACCCAGTGCCTGTGAC	tauZ2_F tauZ2_R	153
<i>tauE</i>	TCGGCGTCTGGCTTTATC GCCGGTGAAACCAACAATG	tauE_F tauE_R	146

5.2 Növesztési körülmények

Az *E. coli* törzseket 37°C-on, az *N. resinovorum* SA1 törzset pedig 28°C-on növesztettem. A szilárd táptalajon történő növesztéshez légkeveréses száraz inkubátort használtam, míg a folyékony tápoldatokban a törzseket nedves rázóinkubátorban növesztettem. Az utóbbi esetben a rázatás sebessége 150 - 180 rpm-re volt beállítva. Folyékony tápoldat esetén a növesztés 100 mL-es Erlenmeyer lombikban történt, amely 30 mL tápoldatot tartalmazott. Ez alól kivétel képez az RNA-seq transzkriptom vizsgálat, amely esetén a sejteket 100 ml tápoldatot tartalmazó, 250 mL-es Erlenmeyer lombikokban növesztettem.

Az *N. resinovorum* SA1 törzs tápoldata minden esetben tartalmazott 25 µg/mL sztreptomycin antibiotikumot. Olyan esetekben, amikor a sejtek pBluescript SK + vagy pPROBE-NT alapú plazmidokat tartalmaztak, a tápoldatot 100 µg/mL ampicillin illetve 25 µg/mL kanamicin antibiotikummal egészítettem ki.

5.3 Felhasznált tápanyagok

5.3.1 Luria-Bertain (LB) tápoldat

Az LB tápoldat 10 g/L NaCl-ot, 10 g/L triptont és 5 g/L élesztő kivonatot tartalmazott. A szilárd táplemezek készítéséhez az LB tápoldathoz 1,5% agarral egészítettem ki ⁹⁸.

5.3.2 Minimál tápoldat

A minimál tápoldat a következő összetevőket tartalmazta: 0,68 g/L KH₂PO₄; 0,87 g/L K₂HPO₄; 0,125 g/L MgSO₄ x 7H₂O; 0,044 g/L CaCl₂; 0,585 g/L NaCl, 14 mg/L FeSO₄; 18,6 mg/L EDTA

x 2H₂O; 0,2 mg/L ZnSO₄; 0,06 mg/L MnCl₂ x 7H₂O; 0,6 mg/L H₃BO₃; 0,4 mg/L CoCl₂ x 6 H₂O; 0,02 mg/L CuCl₂ x 6H₂O; 0,04 mg/L NiCl₂ x 6H₂O; 0,06 mg/L NaMoO₄ x 6H₂O¹⁹.

Szénforrás: a promóter vizsgálatnál a glükózt (Glc-t) és a szulfanilsavat (SA-t) is 5 vagy 50 mM koncentrációban használtam; az RNA-seq vizsgálatnál a Glc-t és az SA-t is 10 mM-os koncentrációban alkalmaztam; az RT-qPCR vizsgálatnál a tápoldatban mind a Glc, mind az SA, mind a hidroxibenzoát (HB) koncentrációját 10mM-ra állítottam be.

Nitrogénforrás: a promóter vizsgálatnál, ahol a vizsgálat indokolta, a tápoldathoz 15 mM NH₄NO₃-ot adtam; az RNS-seq vizsgálatnál csak a Glc szénforrást tartalmazó tápoldathoz adtam 10 mM NH₄Cl-ot; az RT-qPCR vizsgálat pedig a Glc és a HB-ot tartalmazó tápoldathoz adtam 10 mM NH₄Cl-ot

Az RNS-seq és az RT-qPCR vizsgálatok során a tápoldat 25 mM MOPS (pH=7,0) puffer oldatot is tartalmazott.

5.4 Molekuláris biológiai munkák

5.4.1 Genomi DNS izolálása *N. resinovorum* SA1 törzsekből

A genomi DNS izolálást 30 ml LB tápoldatban, egy éjszakán keresztül növesztett baktérium kultúrából végeztem. Az izoláláshoz a GenElute Bacterial Genomic DNA Kit-et (Sigma) használtam. Az izolálás során a kit Gram-negatív baktériumokra vonatkozó protokollját követtem. A tisztított genomi DNS-t 100 µL vízben feloldottam és a későbbi felhasználásáig - 20°C on tároltam.

5.4.2 Az *N. resinovorum* SA1 genomjának szekvenálása

Az *N. resinovorum* SA1 genomját háromszor szekvenáltattuk meg. A szekvenálások mindhárom esetben az Illumina kémiáján alapultak. Az első szekvenálást FASTER SA (Planles-Ouates, Svájc) cég végezte, Illumina Genome IIX analyzer szekvenáló készülék használatával.

A második szekvenálást kooperációs partnerünk, Han Ming Gan és munkatársai végezték a Maláj Monash egyetemen (Szalangor, Malájzia). A szekvenáláshoz használt DNS könyvtárat a Nextera XT (Illumina) kit használatával készítették a gyártó utasításait betartva. A szekvenálás MiSeq (Illumina) asztali szekvenáló készülékkel hajtották végre.

A harmadik szekvenálást a SeqOmics Biotechnology Ltd. (Mórahalom, Magyarország) végezte. A szekvenáláshoz használt könyvtárat a Nextera Mate Pair Sample Prep Kit (Illumina)

használatával készítették a gyártó utasításait betartva. A szekvenálást MiSeq (Illumina) asztali szekvenáló készülékkel hajtották végre.

5.4.3 Kémiai kompetens *E. coli* sejtek készítése és transzformálása

Az XL1-Blue *E. coli* sejteket 50 mL SOB tápoldatban $OD_{600nm} = 0,4 - 0,6$ -ig növesztettem 18°C-on, a sejkultúrát 10 percig jégen tároltam, majd 4°C-on, 2 500 x g-vel 10 percig centrifugáltam. A felülúszó eltávolítása után, az összegyűlt sejteket jégen tartva 16 mL hideg TB pufferben (10 mM PIPES, 5 mM $CaCl_2$, 250 mM KCl, 55 mM $MnCl_2$, pH = 6,70) szuszpendáltam fel, melyet 300 μ L DMSO hozzáadását követően 10 percig jégen inkubáltam. Ezután a kompetens sejtuszpenziót 100 μ L-enként folyamatosan hűtött 1,5 mL-s mikrocentrifugacsövekbe osztottam szét, végül folyékony nitrogénben fagyasztottam és a további felhasználásáig -80°C-on tároltam.

A -80°C-on tárolt kémiai kompetens sejteket használat előtt jégen felolvasztottam. A felolvasztott sejtekhez adtam hozzá 10 μ L végtérfogatban a transzformálni kívánt plazmidot, majd alapos keverés után 15 percig jégen tartottam, amit egy 45 másodpercig tartó 42°C-os hőkezelés követett. A hőkezelés után a sejtoldatot 8 μ L 50%-os glükózt tartalmazó 900 μ L SOC (1 L-re számolva: 0,5 g/L NaCl, 20 g/L tripton, 5 g/L élesztőkivonat, 10 mM $MgSO_4$) oldatban vettem fel. Az elegyet 1 órán át 37°C-on inkubáltam 500 rpm-es rázatás mellett. Végül a sejteket a megfelelő antibiotikummal kiegészített LB lemezre szélesztettem.

5.4.4 Elektrokompetens *N. resinovorum* SA1 törzsek létrehozása és transzformálása

Az *N. resinovorum* SA1 sejteket 25 μ g/mL sztreptomycin antibiotikumot tartalmazó, 100 mL LB tápoldatban növesztettem $OD_{600nm} = 0,4 - 0,6$ közötti értékig. A sejkultúrát 10 percig jégen tároltam, majd 4°C-on, 3 000 x g-vel, 10 percig centrifugáltam. A felülúszó eltávolítása után, az összegyűlt sejteket jégen tartva, 100 mL jéghideg deionizált desztillált vízben szuszpendáltam fel és 4°C-on, 3 000 x g-vel, 10 percig centrifugáltam. Ezt a lépést megismételtem 100 mL, 50 mL jéghideg ionkezelt desztillált víz használatával és végül 25 mL 10%-os glicerin oldattal is. Az így mosott sejteket 1 mL, 20%-os glicerin oldatban szuszpendáltam fel és az elektrokompetens sejtuszpenziót 40 μ L-enként szétosztottam folyamatosan hűtött 1,5 mL-s mikrocentrifugacsövekben. Végül az így szétosztott sejteket folyékony nitrogénben fagyasztottam és a további felhasználásáig -80°C-on tároltam.

Az *N. resinovorum* SA1 törzsekbe a DNS-t elektroporálás módszerével juttattam be. A módszer lényege hogy a DNS-t egy rövid ideig tartó elektromos sokk segítségével juttatjuk be

a sejtekbe. Az elektroporáláshoz a Bio-Rad MicroPulse készülékét és 0,1 cm elektródtávolságú küvetát használtam. Az elektromos sokk 2,5 msec-ig tartó, 1,7 kV feszültség használatával történt. A folyamat első lépéseként az előkészített elektrokompetens sejteket jégen kiolvasztottam majd 1 µg plazmidot kevertem hozzájuk. A sejtoldatot az előre jégen lehűtött elektroporáló küvetta alján található szűk vályúba juttattam, figyelve arra, hogy légbuborékok ne juttassak be a sejtekkel együtt. Elektroporálás után a sejteket 950 µl SOC oldatban vettem fel. Az elegyet 3 órán át 28°C-on inkubáltam 500 rpm-n történő rázatás mellett. Végül a sejteket a megfelelő antibiotikummal kiegészített LB lemezre szélesztettem.

5.5 Promóter régió azonosítása

5.5.1 Plazmid DNS izolálása *E. coli* és *N. resinovorum* SAI törzsekből

A plazmid DNS izolálást 100 ml megfelelő antibiotikumot tartalmazó LB tápoldatban, egy éjszakán keresztül növesztett baktérium kultúrából végeztem. Az izoláláshoz a QIAGEN Plasmid Midi Kit-et (QIAGEN) használtam. Az izolálás során követtem a gyártó által javasolt protokollt.

5.5.2 Restriktációs enzimekkel történő hasítás

A pPROBE-NT vektor hasítását a Fermentas által gyártott Fast Digest *Sma*I restriktációs endonukleázzal végeztem. Használata során követtem a gyártói előírásokat.

5.5.3 DNS polinukleotidkináz kezelés

A PCR termékek végeinek kináz kezelését a Fermentas által gyártott T4 PNK (T4 Polynucleotide Kinase) enzimmel végeztem. Használata során követtem a gyártói előírásokat.

5.5.4 Alkalikus foszfatáz kezelés

A hasított plazmidok végeinek defoszforizálását a Fermentas által gyártott FastAP (Fast Alkaline Phosphatase) enzimmel végeztem. Használata során követtem a gyártói előírásokat.

5.5.5 Ligálás

A beépítendő DNS fragmentumot háromszoros feleslegben adtam a meghasított vektorhoz. A ligáláshoz a Fermentas által gyártott T4 DNS ligáz enzimet alkalmaztam 5x-ös Rapid Ligáz Pufferben a gyári előírásokat betartva.

5.5.6 Agaróz gélelektroforézis

A tisztított DNS mintákat (genomi, plazmid), PCR reakció termékeit és a restrikciós emésztés után kapott fragmenteket analitikai és preparatív célokból gélelektroforézis segítségével választottam el. Az alkalmazott gél 1% agarózt és 0,1 µg/mL etidium-bromidot tartalmazott. A gél előállításához és a futtatás során is 1 x TAE puffert (4 mM Tris-acetát, 1 mM EDTA pH = 8,00) használtam. A futtatás 90 V-os feszültségen történt.

5.5.7 DNS fragment izolálás

Az agaróz gélből a megfelelő méretű DNS szakaszt tartalmazó géldarabot steril pengével vágtam ki. A kivágott géldarabban lévő DNS izolálását a QIAquick Gel Extraction Kittel (QIAGEN) végeztem. Az izolálás során követtem a gyártói előírásokat.

5.5.8 Nukleotidsorrend meghatározása

A DNS szekvencia meghatározások automata kapilláris szekvenálással készültek szekvenspecifikus oligonukleotidok használatával az MTA-SZBK szekvenáló szolgáltató egységénél.

5.5.9 A promóter régió meghatározásához szükséges promóter próba vektor (pPROBE) változatok elkészítése

A pPROBE vektor egy széles gazdaspecificitású, pBBR replikációs origóval rendelkező plazmid, amely tartalmaz egy promóter nélküli *gfp* riporter gént⁹⁹. Az *N. resinovorum* SA1 genomját templátként használva, szekvencia specifikus oligonukleotidok (primer) párok segítségével felsokszoroztam a *sadA* (perifériás útvonal) és *scaA* (centrális útvonal) gének előtti intergenikus régiót. A felsokszorozáshoz PCR technikát használtam, nagy pontosságú Phusion HF DNA polimeráz (Thermo Scientific) segítségével. A PCR termékeket kináz kezelés után, *SmaI* restrikciós endonukleázzal hasított, majd alkalikus foszfatázzal kezelt pPROBE-NT plazmidba ligáltam. A beépülés irányultságát PCR reakcióval ellenőriztem majd -végső ellenőrzésképpen - a beépült régió szekvenciájának pontosságát szekvenálással vizsgáltam meg.

5.5.10 Promóter régió azonósításának menete

A promóter régiók feltérképezése során a perifériás és a centrális útvonal operonjai előtti genomi régiókra specifikus primerpárokat terveztem, majd PCR technika használatával a primerpárok közötti régiót felszorzoztam. Az így kapott rövid DNS szakaszok végeit DNS polinukleotid kináz használatával foszforiláltam és korábban hasított majd defoszforizált promóter próba vektorba (pPROBE) illesztettem. Az így kapott plazmid konstrukciókat *E. coli* XL1 Blue törzsbe transzformáltam. A transzformált sejteket kolónia PCR segítségével szűrtem a megfelelő orientációjú inszertet tartalmazó plazmidokra, amelyeket izoláltam majd szekvenálással ellenőriztem. A megfelelő plazmidokat *N. resinovorum* SA1 törzsbe elektroporáltam. A plazmidról kifejeződő GFP mértékét fluoriméter használatával mértem. A mérés során a gerjesztő fény hullámhossza 485 nm, a mért emittált fény hullámhossza pedig 538 nm volt

5.5.11 Polimeráz-lánreakció (PCR)

A reakciókat az Eppendorf Mastercycler Eppgradient típusú készülékével végeztem. Ellenőrzés céljából végzett PCR vizsgálatok esetén DyNAzyme (Finnzymes) DNS polimerázt használtam. Olyan esetekben, amikor az átírás pontossága különösen fontos volt, Phusion (Thermo) polimerázt használtam. Az alábbi koncentrációjú reagensekből a végtérfogat 1/10 részét adtam a reakcióhoz: primerek: 10 pmol/ μ L; dNTP: 2 mM. DyNAzyme polimeráz használata esetén 1U enzimet (0,033 U/ μ L végkoncentráció) és 10 x DyNAzyme puffert, míg Phusion polimeráz használata esetén 0,6 U (0,02 U/ μ L végkoncentráció) enzimet és 5 x HF Phusion puffert adtam a reakcióhoz. A reakcióelegyet 1 μ L templát hozzáadása után a reakciót desztillált vízzel 30 μ L-es végtérfogatra egészítettem ki.

A PCR reakció menetét meghatározó program 35 ciklusból állt. A ciklusokat megelőzte egy 1 percre tartó 94°C vagy 98°C-on történő elődenaturáció és lezárta egy 3 percre tartó 72°C fokon történő termosztálás. Az elődenaturáció és a denaturáció lépéseit a használt enzim határozta meg. A DyNAzyme polimeráz esetén 94 °C-ot míg a Phusion polimeráz esetén 98°C-ot használtam. A lánreakció ciklusai az alábbi lépésekből épültek fel: denaturáció lépése 15 másodperc, 94°C vagy 98°C-on; primer-templát hibridizáció lépése 15 másodperc, 60°C-on; elongáció lépése 40 másodperc, 72°C-on.

5.6 Analitikai módszerek

5.6.1 A GFP kifejeződésének meghatározása

A növesztés során adott időpillanatokban a kultúrákból 1 ml mintát vettem, amelyet 1,5 ml-es Eppendorf csövekben 1 percen keresztül centrifugáltam, 13 000 x g-n. A tápoldatot tartalmazó felülúszót leöntöttem a leülepedett sejtekről és a sejteket 1 ml TBS (Tris-HCl puffer pH=7,4) oldatban felfuszpendáltam majd ismét 1 percen keresztül centrifugáltam, 13.000 x g-n. Ezt a mosási lépést még egyszer megismételtem és végül a sejteket 1 ml TBS-ben szuszpendáltam fel. A sejtoldat OD (optikai denzitás) értékét 600 nm-en mértem Bio-Rad SmartSec 3000 fotométer használatával. A sejtek GFP tartalmát Thermo Fluoroskan Ascent FL fotó fluoriméterrel mértem. A mérés során a gerjesztő fény hullámhossza 485 nm, a mért emittált fény hullámhossza pedig 538 nm volt.

A normalizált kifejeződést az alábbi képlet segítségével számoltam: $(GFP_{\text{minta}} / OD_{\text{minta}}) - (GFP_{\text{NT}} / OD_{\text{NT}})$. Az NT az üres pPROBE-NT plazmidot tartalmazó sejtet jelenti.

5.6.2 A szulfanilsav fogyása mérése

Az SA koncentrációváltozását a sejtmentes tápoldat (5 perc centrifugálás, 13 000 x g-n) 248 nm-en mért abszorbancia értékéből határoztam meg Bio-Rad SmartSec 3000 fotométer segítségével.

5.6.3 Glükóz koncentráció mérése

A Glc koncentráció változását a sejtmentes tápoldatból (5 perc centrifugálás, 13 000 x g-n) egy refraktív index detektorral (RI Chromaster detector 5450) felszerelt HPLC (nagy teljesítményű folyadékkromatográfia, Hitachi Chromaster) segítségével határoztam meg. Az elválasztás egy IC Sep ICE-COREGEL-64H (300 x 7,8 mm) oszlopon keresztül történt. A mérés során az oszlop és a detektor hőmérséklete 50°C és 41°C volt. Az elúcióhoz 0,01 M H₂SO₄ oldatot használtam 0,8 mL/perc áramlási sebességgel.

5.7 Transzkriptomikai munkák

5.7.1 Teljes RNS tisztítása *N. resinovorum* SAI-ből

A növesztés különböző szakaszaiban lévő (exponenciális, stationer) és különböző tápanyagforrást tartalmazó (Glc, SA, HB) sejt kultúrákból 6 mL mintát vettem ki RNS tisztítás céljából. A kivett mintákat 5 percig, 4 °C-on, 17 226 x g-n centrifugáltam. A tápanyagot tartalmazó felülúszót eltávolítottam, majd a centrifuga falához tapadt sejteket folyékony nitrogénben lefagyasztottam és a mintákat -80 °C-on tároltam. Későbbiekben a lefagyasztott sejteket 350 µL Zymo DNA/RNA Shield (Zymo Research) RNáz gátlószerben vettem fel. Az RNáz gátolt sejtoldathoz 700 µL RLT puffert (QIAGEN) adtam, amit előtte 7 µL β-merkaptoetanollal egészítettem ki. Az oldatot alapos összekeverés után átvittem egy 0,8 g, 0,5 mm átmérőjű üveggyöngyöket (Scientific Industries Inc.) tartalmazó 2 mL-es Eppendorf csőbe. A csövet egy TurboMix kiegészítővel (Scientific Industries Inc.) ellátott Vortex-Genie 2 rázóató (Scientific Industries Inc.) maximális fokozatán 5 percig rázóttam. A feltárt sejtek törmelékeitől és a feltáráshoz használt gyöngyöktől 10 másodperc, 9 600 x g-n történő centrifugálással szabadultam meg. Az RNS tisztítás további lépéseit a RNeasy Mini (QIAGEN) kittel végeztem el. Tisztítás során az oszlophoz kötött RNS mintát, a gyártó javaslata szerint, DNázal kezeltem. Az így tisztított RNS oldatot egy ismételt DNáz I (Invitrogen) kezelésnek vettem alá, hogy az esetleg még nyomokban jelen lévő genomi DNS-től is megszabaduljak. Az RNS mintákat MgCl₂ -ot tartalmazó, 10x DNáz I Reakció Pufferben (Invitrogen) és 1 U/µL DNáz I jelenlétében 30 percig, 37 °C-on inkubáltam. A reakciót 10 perces, 65°C-on történő hőkezeléssel állítottam le 1 mM EDTA (Invitrogen) jelenlétében. Az így kapott RNS-t a RNeasy MinElute Clean Up kittel (QIAGEN) tovább tisztítottam a gyártó utasításai szerint. A tisztított RNS-t 30 µL nukleáz-mentes vízben vettem fel és a későbbi felhasználásáig -80°C-on tároltam.

5.7.2 Az *N. resinovorum* SAI teljes transzkriptomjának a meghatározása RNA-seq módszerével

A tisztított teljes RNS-ből történő cDNS könyvtár készítését és a könyvtár elemeinek Illumina kémiai alapú szekvenálását a SeqOmics Biotechnology Ltd. (Mórahalom, Magyarország) végezte. A szekvenáláshoz használt cDNS könyvtár elkészítése előtt a mintákból a teljes RNS populáció 98%-át is kitevő riboszómális RNS-eket (rRNS) a baktérium specifikus Ribo-Zero rRNA removal kit (Illumina) használatával távolították el. A könyvtár készítéséhez a ScriptSeq

V2 RNA-Seq könyvtár készítő kitet (Illumina) használták. A könyvtár indexelésére a ScriptSeq index PCR primereket (Illumina) használták. Az elkészített könyvtárat Illumina MiSeq készülékével szekvenálták, MiSeq reagens kit V2 (300 ciklus) használatával. A szekvenálás nyers adatait elérhetővé tettem az NCBI GEO (Gene Expression Omnibus) adatbázisában (GSE102626).

5.7.3 Az RT-qPCR folyamata

A reverz transzkripciót követő kvantitatív PCR első szakaszában elvégeztem a DNáz I kezelt RNS minták cDNS-é történő átírását. A cDNS készítéséhez 0,25 µg random hexamert és 0,5 µg DNáz I kezelt teljes RNS-t használtam 15 µL végtérfogatban. Az oldatot 70 °C-on, 5 percig hőkezelttem, majd utána azonnal jégre tettem. A mintát jégen tartva 5 µL 5x M-MLV puffert (Promega), 1,25-1,25 µL 10-10 mM dATP-t, dCTP-t, dGTP-t, dTTP-t, 1 µL 200 U M-MLV RT enzimet (Promega) adtam, végül pedig kiegészítettem 25 µL végtérfogatra nukleázmentes vízzel. Ezután 37 percig, 60°C fokon hőkezelttem. Az így elkészített cDNS-t kétszeresére hígítottam és használtam fel templátként a kvantitatív PCR lépése során.

A qPCR-t ABI 7500 Real Time PCR System (Life Technology Co.) készülékkel végeztem el Power SYBERGreen PCR Master Mix (Life Technology Co.) használatával. A reakcióhoz tervezett génspecifikus primerpárok a gének közepére térképeződnek és megközelítőleg 150 bp hosszú régiót sokszoroznak fel. A 25 µL végtérfogatú reakcióelegy összetétele az alábbi volt: 12.5 µL 2x Power SYBERGreen PCR Master Mix, 1 µL cDNS templát, 1,25 µL forward és reverse primer, 9 µL ultratiszta víz. Az alkalmazott program: 50°C/2 perc; 95°C/10 másodperc, 95°C/15 másodperc, 60°C/1 perc. A program első két lépése egyszer, utolsó két lépése pedig 40-szer ismétlődött meg.

A reakciók ciklus küszöbértéke (C_t) az ABI sequence detection software v1.4 (Life Technology Co.) segítségével lett megállapítva. A minták normalizált ciklus küszöbértékei (ΔC_t) a referenciagénként használt *gyrB* C_t értéke segítségével lettek meghatározva. A gének \log_2 kifejeződésének változása megfelel a $-\Delta C_t$ értéknek, amelyet úgy számoltam, hogy kivontam a referenciának választott minta ΔC_t értékét a vizsgált minta ΔC_t értékéből.

5.8 Bioinformatikai vizsgálatok

5.8.1 Az *N. resinovorum* SAI genomjának összeillesztése és annotációja.

A genom összeillesztéséhez három - két párosított végű (paired-end) és egy társított párosított végű (mate paired-end) - genomszekvenálásból származó leolvasásokat használtam. A

leolvasások tisztításához és összeillesztéshez a CLC Genomics Workbench v. 7.0 (Qiagen Bioinformatics) programcsomagot és a Scaffolder programcsomagot ¹⁰⁰ használtam. Az összeillesztett genomot feltöltöttem az NCBI Genome adatbázisába (NCBI Reference Sequence azonosító: NZ_CP017075.1-9.1, Ensembl Bacteria azonosító: CP017075-9). A kódoló régiók azonosítását és annotációját az NCBI Prokaryotic Annotation Pipeline segítségével végeztem ¹⁰¹. A genom elérhetővé vált a KEGG adatbázisában is ¹⁰² (KEGG azonosító: T04490). A genom megjelent az Uniprot adatbázisában is (Uniprot azonosító: UP000094626), ahol számos referencia adatbázis használatával (Gene3D, InterPro, PANTHER, Pfam, SUPERFAM) történt a fehérjéket kódoló gének annotálása.

A fehérjék COG ¹⁰³ kategóriákba rendezését az emapper.py ¹⁰⁴ python program segítségével végeztem az eggNOG 4.5.1 adatbázisának ¹⁰⁵ használatával.

5.8.2 NCBI BLAST keresés

A homológ fehérjék keresését az NCBI RefSeq Representative Genome adatbázisában végeztem, mert ez az adatbázis tartalmazza a RefSeq Genomes adatbázisból kiválogatott legjobb minőségű referencia és reprezentatív genom szekvenciákat. A kereséshez az NCBI TBLASTN online elérhető változatát használtam. A találatokat szűrtem az $1e^{-100}$ nál kisebb "e" értékekre (evaluate < $1e^{-100}$)

5.8.3 Filogenetikai vizsgálatok

A filogenetikai vizsgálatok során a kiválasztott fehérjék szekvenciáját a mafft 7 programcsomag ¹⁰⁶, E-INS-I módszerének használatával illesztettem. A filogenetikai fa elkészítése a maximum likelihood módszerével történt, a programcsomag által ajánlott LG+FG modell használatával ¹⁰⁷. A filogenetikai vizsgálatokat a Mega 6.0.6 programcsomag használatával végeztem ¹⁰⁸.

5.8.4 Az *N. resinovorum* SA1 RNA-Seq során kapott leolvasások informatikai feldolgozása

A szekvenálás után kapott nyers leolvasásról eltávolítottam az adapter szekvenciáját és a rossz minőségű szekvencia részleteket. A megmaradt jó minőségű leolvasásokat rátérképeztem a referenciaként használt *N. resinovorum* SA1 törzs genomjának kódoló régiójára. Ezeket a lépéseket a CLC Genomic Workbench 7 (Qiagen Bioinformatics) programcsomag használatával végeztem el.

5.8.5 *Megváltozott kifejeződést mutató gének vizsgálata*

A megváltozott kifejeződést mutató gének (DEG, Differentially Expressed Gene) vizsgálatát az EdgeR¹⁰⁹ R programcsomag segítségével végeztem el. A vizsgálatba csak azok a gének kerültek be, amelyek 3 minta esetén rendelkeztek legalább 1 CPM (Count Per Million read; (génre térképezett readszám/összes readszám)*1 000 000) leolvasásszámmal. A kifejeződés változáshoz számolt statisztikai értéket (p értéket) Benjamini-Hochberg korrekciós eljárás használatával korrigáltam (FDR, False Discovery Rate). Azon gének kifejeződését tekintettem szignifikánsan változónak, amelyek kifejeződése legalább négyszeresen változott ($\text{abs}(\log_2) > 2$), a hozzá tartozó korrigált FDR érték pedig $\leq 0,001$ volt.

5.8.6 *Egyéb statisztikai vizsgálatok*

A főkomponens analízist (MDS) az R edgeR programcsomag plotMDS funkciójának segítségével végeztem a gének CPM ((génre térképezett readszám) /összes readszám))*1 000 000) értékének használatával.

A korrespondancia elemzést az R factor extra programcsomag használatával végeztem. A vizsgálatból kihagytam a kis változást mutató kodonokat (atg, taa, tag, tga, tgg).

A COG funkciós csoportokba tartozó gének Student's t-teszten alapuló túlreprezentációs analíziséhez a R stats programcsomag t.test függvényét használtam. A KEGG alapú géncsoport dúsulás vizsgálatát (Gene Set Enrichment Analysis, GSEA) a clusterProfiler programcsomag gseKEGG függvényének használatával végeztem el.

6 *Eredmények és tárgyalásuk*

6.1 *Az N. resinovorum SA1 törzs genom szekvenciájának meghatározása.*

A *Novosphingobium* nemzetség számos tagjának a genom szekvenciája elérhető, azonban ezeknek nagy része nincs befejezett állapotban. A dolgozat írása során csak három faj, a *Novosphingobium aromaticivorans* DSM, *Novosphingobium pentaromativorans* US6-1 és a *Novosphingobium sp.* PP1Y rendelkezett teljesen összeszerelt genom szekvenciával. Ezeknek a fajoknak az átlagos kromoszómamérete 3,82 Mb, átlagos GC tartalma pedig 64,14% -os. Mindhárom faj több plazmiddal is rendelkezik, amelyek száma 2 és 5 között mozog. A legnagyobb méretű plazmid a *Novosphingobium sp.* PP1Y törzsben található meg. Ez az Mpl-nek nevezett plazmid 1,16 Mb méretű.

Az *N. resinovorum* SA1 genomjának szekvenálása során a fő cél a genom teljes összeillesztése illetve annotálása volt. A szekvenálási projekt során három különálló genomi könyvtár szekvenálása történt meg. A szekvenálások mindhárom esetben, második generációs szekvenálási technikán alapuló Illumina MiSeq platform használatával történtek. Az első szekvenálás során megközelítőleg 7,5 millió, átlagban 35 bp hosszú párosított végű (paired-end) leolvasást kaptam (párok közötti átlagos távolság 256,85 bp, átlagos eltérés 10,47). A második szekvenálás során megközelítőleg 4,5 millió, átlagban 150 bp hosszú párosított végű (paired-end) read-et (átlagos távolság 352,87 bp, átlagos eltérés 129,44), míg a harmadik szekvenálás során megközelítőleg 64 ezer, átlagosan 250 bp hosszú társított párosított végű (mate paired-end) read-et kaptam (átlagos távolság 7767,16 bp, átlagos eltérés 639,05). A genom összeillesztéséhez hibrid megközelítést követtem, amelyhez felhasználtam mindhárom szekvenálásból származó read-eket. A bizonytalan vagy esetleg hiányos részeket polimeráz-lánreakció (PCR) és kapilláris szekvenálás alkalmazásával tisztáztam. Az összeillesztés eredményeképpen megkaptuk a *N. resinovorum* SA1 végleges genomját, amely a kromoszóma mellett további 4 plazmidból épül fel (1. táblázat). A kromoszóma mérete és GC összetétele hasonló a *Novosphingobium* nemzetség többi tagjánál megfigyeltékhez, azonban figyelemre méltó az ebben a törzsben jelen lévő plazmidok nagy száma illetve mérete. Az *N. resinovorum* SA1 négy plazmidjából három nagyobb, mint 100 kb. A legnagyobb mérettel a pSA1 plazmid rendelkezik, amely eléri a 1,76 Mb hosszt, amellyel megközelíti a kisebb kromoszómák mérettartományát.

Jellemzők	Kromoszóma	pSA1	pSA2	pSA3	pSA4
Nyilvántartási szám (INSDC)	CP017075	CP017076	CP017077	CP017078	CP017079
Méret (Mbp)	3,78	1,76	0,96	0,35	0,07
Szerkezet	cirkuláris	lineáris*	cirkuláris	cirkuláris	cirkuláris
GC összetétel (%)	65,6	65,3	63	61,4	62,5
Fehérjét kódoló gének száma	3456	1415	835	324	83
Pszudogének száma	34	26	55	17	3
rRNS kódoló gének száma (5S, 16S, 23S)	1, 1, 1	2, 2, 2	0	0	0
tRNS kódoló gének száma	52	7	0	0	0
Egyéb RNS kódoló gének száma (ncRNA, tmRNA)	2, 1	0	0	0	0

1. táblázat. Az *N. resinovorum* SA1 genomjának általános jellemzői. A pSA1 plazmid szerkezete esetén szerplő csillag arra utal, hogy a lineáris szerkezett nem feltétlenül tükrözi a valóságot. A szekvenálásból származó adatok alapján nem tudtam megbízhatóan elvégezni a végek zárását.

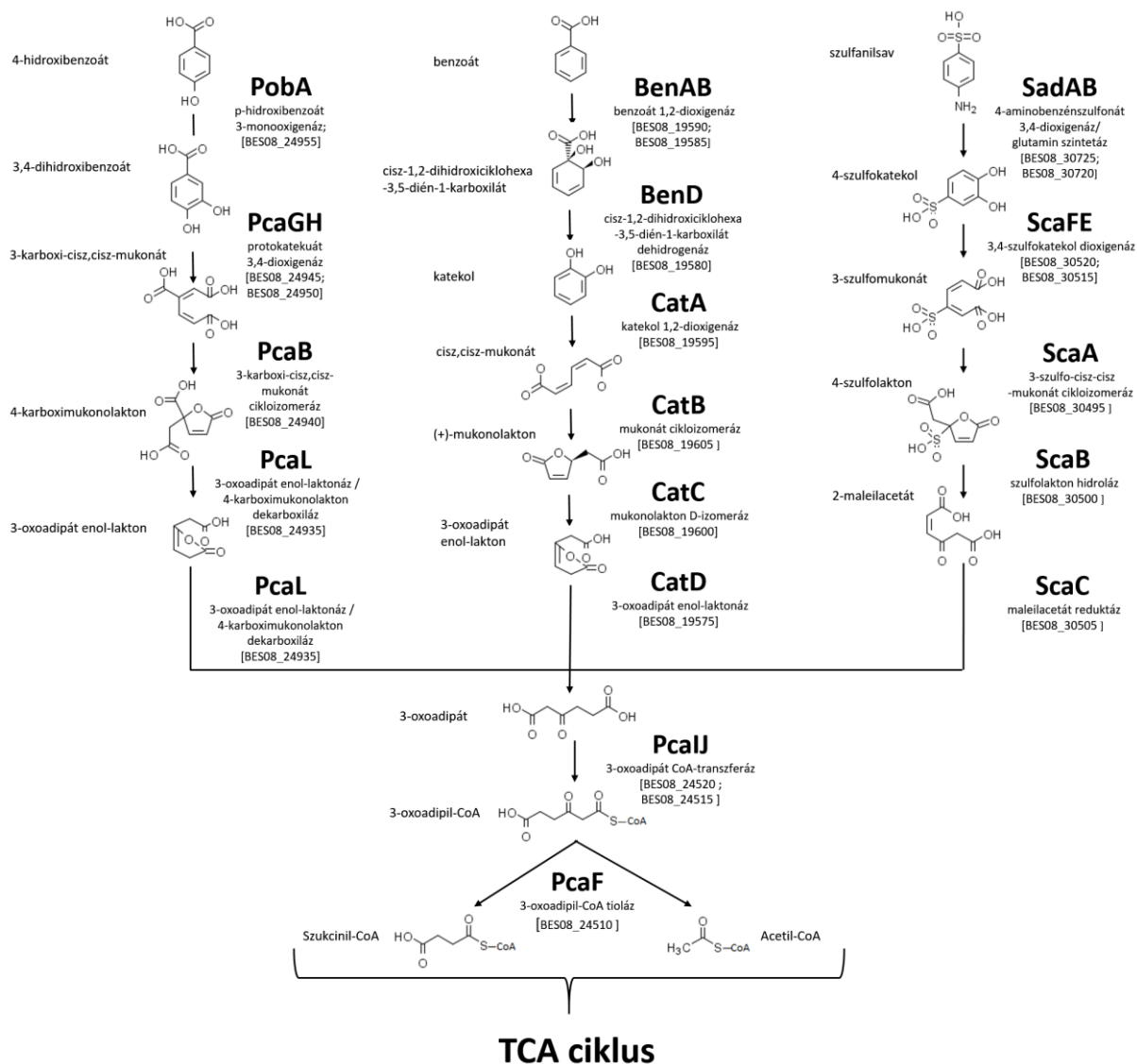
A kódoló genomi régiók jóslása és a jóslt régiók annotálása az NCBI (National Center for Biotechnology Information) PGAAP (Prokaryotic Genome Autonomic Annotation Pipeline) rendszerén keresztül történt ¹⁰¹ és elérhetővé vált az NCBI GenBank és az Ensemble Genomes adatbázisokban. A jóslt fehérjék ismételt annotálása az UniProt (Universal Protein Resource) és KEGG (Kyoto Encyclopedia of Genes and Genomes) rendszerein keresztül is megtörténtek és szintén elérhetővé váltak nyilvános adatbázisaikban. A fehérjéket COG (Clusters of Orthologous Groups of proteins) funkciós kategóriákba ¹¹⁰ sorolását az eggNOG-mapper programcsomag segítségével végeztem el az eggNOG 4.5.1 adatbázis használatával ¹⁰⁴. Ezzel a módszerrel a gének 84.8%-át sikerült valamelyik funkciós csoportba elhelyezni (2. táblázat).

COG kód	Kromoszóma		pSA1		pSA2		pSA3		pSA4		Genom		COG kategória leírása
	érték	% (3456)	érték	% (1415)	érték	% (835)	érték	% (324)	érték	% (83)	érték	% (6113)	
B	0	0.00	0	0.00	1	0.12	0	0.00	0	0.00	1	0.02	Kromoszóma szerkezet és dinamika
C	195	5.64	89	6.29	43	5.15	22	6.79	1	1.20	350	5.73	Energia termelés és átalakítás
D	30	0.87	2	0.14	2	0.24	5	1.54	3	3.61	42	0.69	Sejt ciklus szabályozás, sejt osztódás, kromoszóma szétválás
E	186	5.38	60	4.24	18	2.16	5	1.54	0	0.00	269	4.40	Aminosav transzport és metabolizmus
F	80	2.31	13	0.92	3	0.36	5	1.54	0	0.00	101	1.65	Nukleinsav transzport és metabolizmus
G	128	3.70	105	7.42	26	3.11	3	0.93	1	1.20	263	4.30	Szénhidrát szállítás és metabolizmus
H	100	2.89	29	2.05	7	0.84	4	1.23	0	0.00	140	2.29	Koenzim transzport és metabolizmus
I	131	3.79	58	4.10	49	5.87	7	2.16	1	1.20	246	4.02	Lipid transzport és metabolizmus
J	151	4.37	12	0.85	2	0.24	2	0.62	0	0.00	167	2.73	Transzláció, riboszóma szerkezet és biogenezis
K	168	4.86	108	7.63	62	7.43	26	8.02	6	7.23	370	6.05	Transzkripció
L	161	4.66	57	4.03	70	8.38	56	17.28	16	19.28	360	5.89	Replikáció, rekombináció és hibajavítás
M	189	5.47	65	4.59	21	2.51	5	1.54	3	3.61	283	4.63	Sejtfa/membrán/hártya biogenezis
N	36	1.04	1	0.07	4	0.48	0	0.00	0	0.00	41	0.67	Sejt mobilitás
O	123	3.56	22	1.55	13	1.56	5	1.54	0	0.00	163	2.67	Poszt transzlációs módosítás, fehérjeforgalom, chaperonok
P	170	4.92	150	10.60	40	4.79	25	7.72	1	1.20	386	6.31	Szervetlen ion transzport és metabolizmus
Q	65	1.88	73	5.16	23	2.75	11	3.40	1	1.20	173	2.83	Másodlagos metabolitok bioszintézise, szállítása és katabolizmusa
S	854	24.71	349	24.66	232	27.78	70	21.60	12	14.46	1517	24.82	Ismeretlen funkció
T	110	3.18	65	4.59	45	5.39	10	3.09	0	0.00	230	3.76	Jelátviteli mechanizmusok
U	69	2.00	25	1.77	8	0.96	14	4.32	3	3.61	119	1.95	Sejten belüli szállítás, szekréció, és vezikuláris transzport
V	30	0.87	12	0.85	4	0.48	3	0.93	1	1.20	50	0.82	Védekező mechanizmusok
-	525	15.19	144	10.18	176	21.08	48	14.81	36	43.37	929	15.20	Nincs a COG-ok között

2. táblázat. Az általános COG funkciós kategóriákba sorolható fehérjét kodoló gének száma.

6.2 A szulfanilsav bontási útvonalának komparatív vizsgálata

Az irodalmi és annotációs adatok felhasználásával rekonstruáltam a *N. resinovorum* SA1 genomjában található három párhuzamos β -ketoadipát útvonalat (14. ábra.). Ezek az útvonalak biztosítják az *N. resinovorum* SA1 törzs korábban megfigyelt hidroxibenzoát (HB), benzoát (B) és szulfanilsav (SA) bontó képességét¹⁹. Az útvonalért felelős gének több plazmidon szétszóródva helyezkednek el. A HB bontási útvonalának elemeit kodoló gének (*pobR-pobA-pcaH-pcaG-pcaB-pcaL*), a B bontási útvonal elemeit kodoló gének (*benM-catB-catC-catA-benA-benB-benD-catD*) és az összes párhuzamos β -ketoadipát útvonal esetén közös szakasz elemeit kodoló gének (*pcaF-pcaJ-pcaI*) a legnagyobb méretű, pSA1 plazmidon találhatóak meg három elkülönült klaszterbe rendeződve. Ezzel szemben az SA bontásáért felelős módosult β -ketoadipát útvonal perifériás és centrális elemei két géncsoportba rendeződve helyezkednek el a pSA3 plazmidon. A bontási útvonal perifériás szakaszának elemeit kodoló géncsoport (*sadA-sadB*) és az útvonal centrális szakaszának elemeit kodoló géncsoport (*scaA-scaB-scaC-scaD-scaE-scaF*) között viszonylag nagy, 40 kbp távolság figyelhető meg.



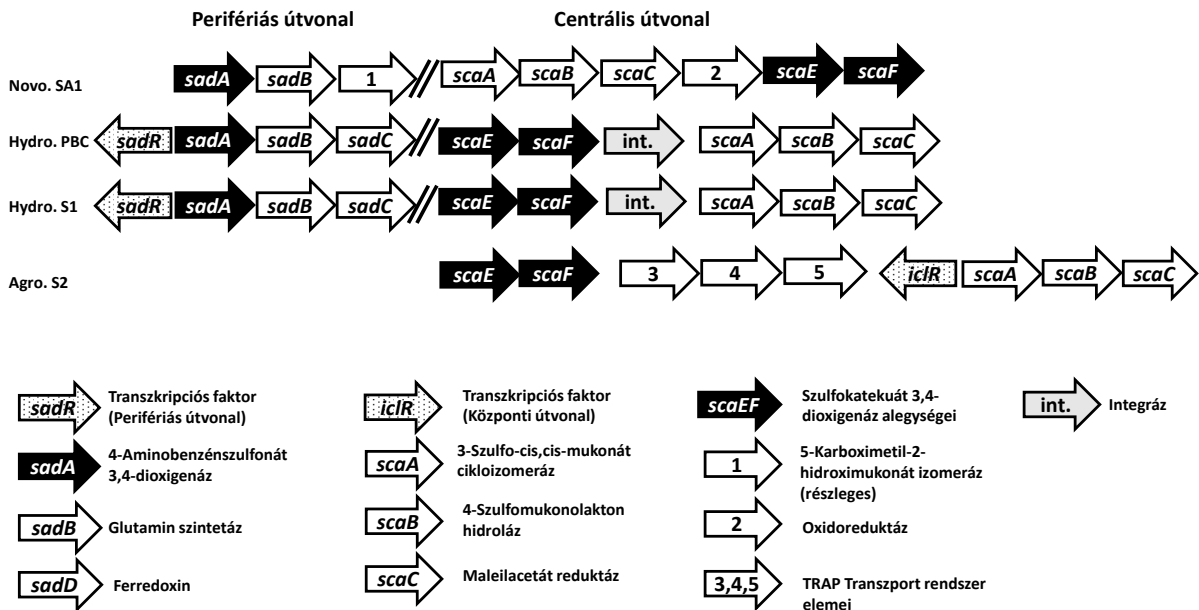
14. ábra. A három azonosított béta-ketoadipát útvonal elemei és az általuk katalizált enzimatikus lépések *N. resinovorum* SA1-ban.

Összehasonlítottam az *N. resinovorum* SA1 esetén rekonstruált, az SA bontásért felelős útvonalat a már korábban jellemzett SA-bontó törzsekben megtalálható útvonalakkal (15. ábra.). Az összehasonlítást megnehezítette, hogy a dolgozat írása során a SA bontására képes törzsek közül csak a két nagyon közeli rokon *Hydrogenophaga intermedia* S1 és a PBC rendelkezett elérhető genomszekvenciával (NCBI Assembly: GCF_000723405.1, GCF_000263795.1)¹¹¹, illetve a SC bontására képes *Agrobacterium radiobacter* S2 genomja volt ismert (NCBI Assembly: GCF_000723345.1). A dolgozat írásáig ezek közül is csak *H. intermedia* PBC genomja vált teljessé (NCBI Assembly: GCA_000263795.2)¹¹².

A különböző fajokban megtalálható útvonalak nagyfokú hasonlóságot mutatnak, azonban számos eltérés is megfigyelhető. A *H. intermedia* PBC esetén jellemzett perifériás útvonal dioxigenáz (*sadA*) és glutamin szintetáz (*sadB*) génei mindhárom SA-bontó baktériumban megtalálhatóak, egy génklaszterbe rendeződve. A *Hydrogenophaga* fajok esetén a génklaszter kiegészül egy ferredoxin fehérjét kódoló génnel (*sadD*), amely közvetlenül a *sadB* gén után helyezkedik el, illetve egy *GntR* típusú transzkripciós faktort kódoló génnel (*sadR*), amely a *sadA* géntől 5' irányban található meg, azzal ellentétes orientációban. Ennek a génnek a terméke feltételezhetően a perifériás útvonal szabályozásában játszhat szerepet. Az *N. resinovorum* SA1 esetében se a ferredoxint, se a transzkripciós faktort kódoló gén nem része a klaszternek. A ferredoxint kódoló gén helyett egy 3-karboximetil-2-hidroximukonát izomeráz enzimet kódoló gén töredéke található meg, amely feltételezhetően nem vesz részt a SA bontásának folyamatában.

A centrális útvonal elemei a *H. intermedia* S1, az *A. radiobacter* S2 és az *N. resinovorum* SA1 törzsek esetén jól ismertek. Az útvonalat alkotó gének egyedül csak a *N. resinovorum* SA1 esetén rendeződnek egy génklaszterbe. Az *A. radiobacter* S2 esetén az aromás gyűrű hasítását végző oxigenáz génei (*scaEF*) és az útvonal többi elemét kódoló gének (*scaABC*) közé egy TRAP (tripartite ATP-independent periplasmic) transzportrendszert kódoló génklaszter ékelődött be. További érdekesség, hogy csak ebben az esetben volt megfigyelhető egy IclR típusú transzkripciós faktort kódoló gén a *scaA* génjétől 5' irányban, azzal ellentétes orientációban. A transzkripciós faktor feltételezhetően ennek az útvonalnak a szabályozásában játszhat szerepet.

A *H. intermedia* fajok esetén is megfigyelhető a centrális útvonal feldaraboltsága. A *H. intermedia* PBC és a *H. intermedia* S1 esetén egy IS6 típusú mozgó genetikai elem ékelődött a gyűrű hasítását végző enzimek génei (*scaEF*) és az útvonal többi elemét kódoló gének (*scaABC*) közé.



15. ábra. A különböző baktériumokból származó SA bontási útvonalak génszerveződésének összehasonlítása

Az SA teljes bontás folyamatában résztvevő fehérjék szekvenciájának páronkénti összehasonlításából látható, hogy a *Hydrogenophaga* fajokból származó enzimek szekvenciái egymással szinte 100%-os azonosságot mutatnak (Függelék 1.). Az *N. resinovorum* SA1 törzsben található perifériás útvonal két fehérjéje, a SadA és a SadB, valamint *Hydrogenophaga*-ban található ortológjaik között csak 54% és 42%-os azonosság figyelhető meg. A centrális útvonal elemei (ScaA-F) esetén ennél kicsivel nagyobb, 56-68%-os az azonosság mértéke. A *Novosphingobium* fehérjék a legnagyobb fehérjeszintű azonosságot (61-82%) az *A. resinovorum* S2 törzsben található centrális útvonal elemeivel mutatnak.

6.3 A szulfanilsav bontási útvonal evolúciós eredetének tanulmányozása

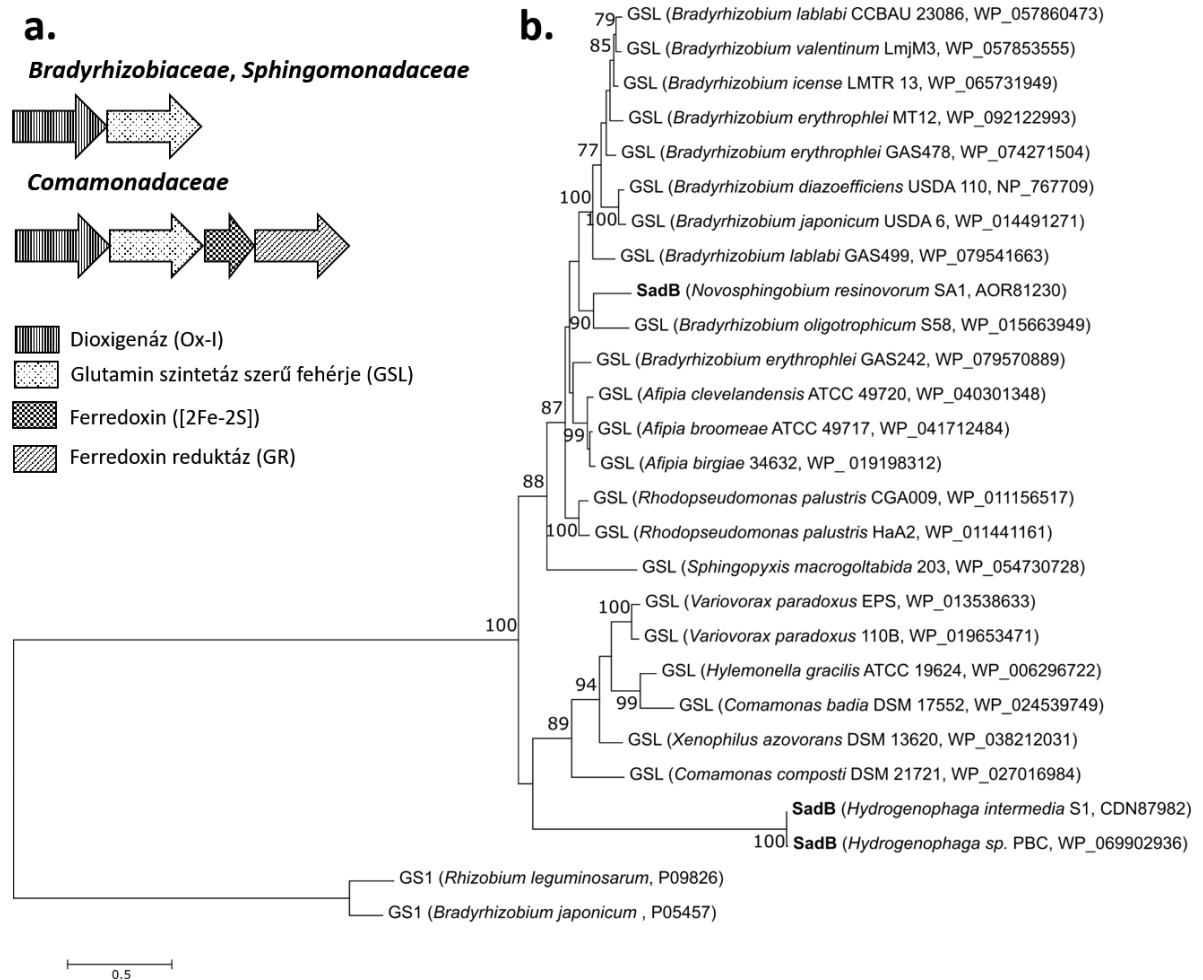
A β -keto adipát útvonalak különböző változatai elég elterjedtek a *Sphingomonadaceae* család tagja között ⁷⁶. A B és a HB útvonal elemeit az *N. resinovorum* SA1 törzs két legközelebbi rokon fajában, a *Novosphingobium resinovorum* KF1 (Taxonomy ID: 158500) és a *Sphingomonas sp.* LH128 (Taxonomy ID: 473781) esetén is azonosítottam (Függelék 2.). Az útvonal elemei az *N. resinovorum* SA1 esetén megismert módon rendeződnek génklaszterekbe és az ortológ fehérjék aminosavsorrendjei 100%-os azonosságot mutatnak. Az SA bontás útvonalának elemei azonban nem voltak megtalálhatóak ezekben a közeli rokon fajokban, ami alapján feltételezhető, hogy azok esetleg valamilyen genomátrendezés eredményeképpen

elvesztek ezekből a fajokból vagy valamilyen horizontális géntranszfer útján jutottak az *N. resinovorum* SA1 genomjába.

A lehetséges horizontális géntranszfer eseményét az NCBI „RefSeq Representative genomes” adatbázisában található baktériumok (taxid:2) genomján végzett TBLASTN ($E < 1e^{-100}$) keresés is megerősíti. A perifériás útvonal fehérjei (SadAB) legnagyobb hasonlóságot nem a *Sphingomonadaceae* családba, hanem a szintén az alfa-proteobaktériumok osztályába tartozó *Bradyrhizobiaceae* családba tartozó baktériumok fehérjéivel mutatnak.

A *Comamonadaceae* családba tartozó *Hydrogenophaga* nemzetség fajai esetén megfigyelték hogy a *sadAB*-szerű gén klaszter előfordulása általánosnak tekinthető¹¹². Megvizsgáltam, hogy ugyanez elmondható-e az *N. resinovorum* SA1 törzset is tartalmazó *Sphingomonadaceae* család illetve a *Bradyrhizobiaceae* család tagjaira is. A keresést az *N. resinovorum* SA1 SadAB fehérjéivel végeztem el az NCBI RefSeq Representative genomes adatbázisban található *Comamonadaceae*, *Sphingomonadaceae* és *Bradyrhizobiaceae* családba tartozó fajok genomján az NCBI TBLASTN keresőjének használatával. Olyan *Comamonadaceae*, *Sphingomonadaceae* és *Bradyrhizobiaceae* családba tartozó fajokat kerestem, ahol a *sadAB* génklaszter megtalálható. A keresés alapján kiderült, hogy nem csak a *Comamonadaceae* család tagjai között, hanem a *Bradyrhizobiaceae* család tagjai között is általános a *sadAB* gén klaszter előfordulása. Ezzel ellentétben a *Sphingomonadaceae* család tagjai között csak nagyon ritkán található meg. Az *N. resinovorum* SA1 mellett csak a *Sphingopyxis macrogoltabida* 203 törzsben volt azonosítható az általunk használt szigorú keresési paraméterek mellett (kicsi de jó minőségű adatbázis használata, a két gén egy klaszter részét kellett hogy képezze, e-value $< 1e^{-100}$). Jóval megengedőbb paraméterek mellett ez a szám kis mértékben növekedett, de a jelentősen megnövekedő redundáns találatok miatt a munkánk során végül a szigorú paraméterek alkalmazásával kapott találatok feldolgozása mellett döntöttem.

Mind a három családból a 10 legnagyobb hasonlóságot mutató fehérjével elvégeztem a SadA (Függelék 3.) és a SadB (16. ábra.) fehérjék filogenetikai vizsgálatát.



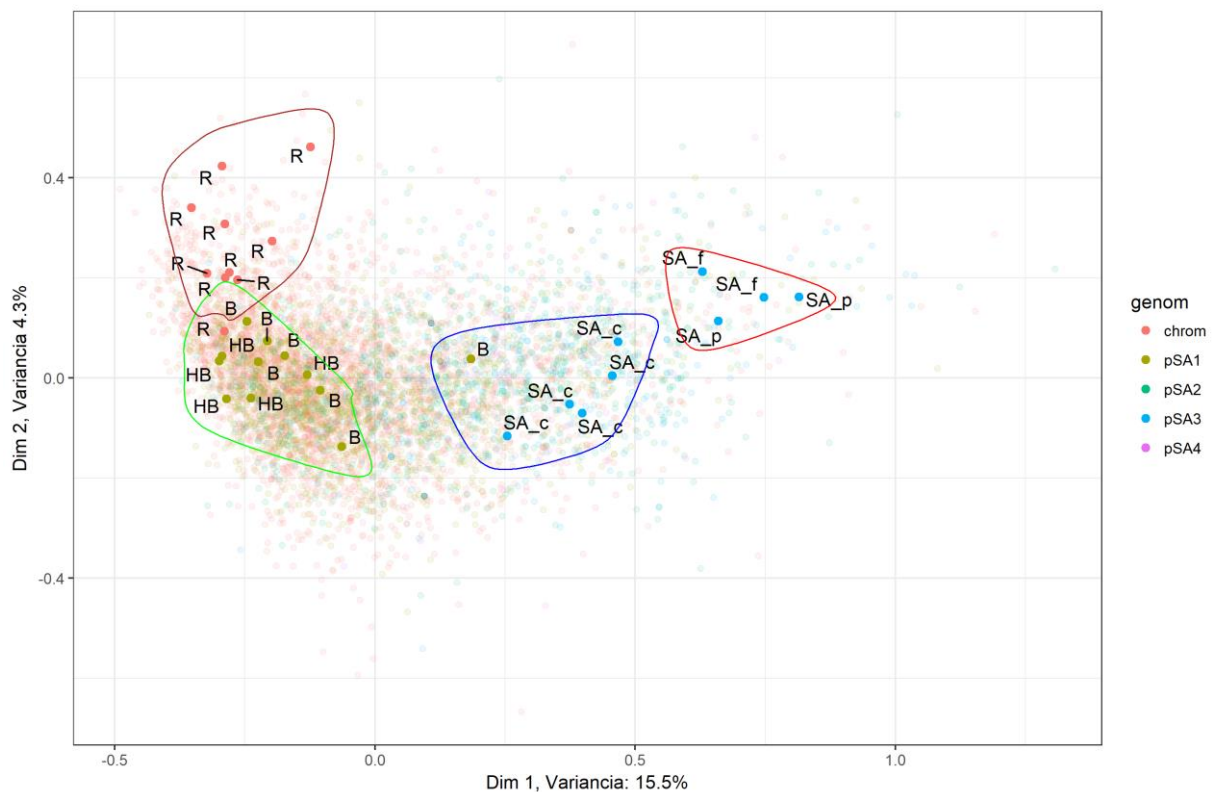
16. ábra. A) A *Bradyrhizobiaceae*, a *Sphingomonadaceae* és a *Comamonadaceae* család tagjaiban azonosítható *sadAB* génklaszterek sematikus felépítése. B) A génklaszterek által kódolt SadB fehérjék filogenetikai törzsfája. A fehérjék közötti evolúciós kapcsolat maximum likelihood módszerrel volt meghatározva, melynek során a GS1 fehérjék szolgálták kulcsoportként. Az egység az aminosavanként történő szubsztitúciók számát mutatja.

A filogenetikai vizsgálatok azt mutatták, hogy az egy családba tartozó fajokból származó fehérjék jól elkülönülő csoportokat alkotnak. Ez utalhat arra, hogy ezek a fehérjék génjei konzerváltan megtalálhatóak ezeknek a családoknak a tagjaiban és azokkal közös evolúciós utat járnak be. Egyedül csak az *N. resinovorum* SA1 SadAB fehérjéi mutatnak ez alól kivételt, mivel azok valamilyen *Bradyrhizobium* fajból származó fehérjével mutattak nagyobb hasonlóságot. A fehérjeszekvencia hasonlóság alapján azonosított génklaszterek felépítése bizonyos család specifikus sajátosságokat mutatnak. A *Comamonadaceae* család tagjai esetén a *sadAB* gének mellett nagy gyakorisággal előfordul egy 2Fe-2S fer2 (pfam: PF00111) típusú ferredoxint kódoló gén és bizonyos esetekben a ferredoxint kódoló gén után egy glutation reduktázt (GR) típusú ferredoxin reduktázt kódoló gén (pfam: PF07992, PF14759) is megtalálható. Ezzel

szemben a *Bradyrhizobiaceae* és *Sphingomonadaceae* család tagjainál azonosított *sadAB* gén klaszterek nem tartalmazták az ilyen típusú fehérjék génjeit.

A korábban használt paraméterek mellett megismételtem a keresést a centrális útvonal elemeivel is (*ScaABCEF*). A már ismert szulfokatekol bontására képes törzseken kívül azonban nem sikerült újabb olyan törzseket azonosítani, amelyek tartalmazták volna a génklaszter összes elemét magas hasonlósági kritériumok alkalmazásával.

Az *N. resinovorum* SA1 törzs SA bontási útvonalának perifériás szakaszát számos mobilis genetikai elem határolja, ami horizontális géntranszferre utal. Azért, hogy további bizonyítékot találjunk az SA bontási útvonal eredetére, korrespondencia analízis módszerével (Correspondence analysis – CA) ¹¹³ összehasonlítottam az *N. resinovorum* SA1 törzs összes, fehérjét kódoló génjének relatív szinonim kodon használatát (Relative Synonymous Codon Usage - RSCU) ¹¹⁴ (17. ábra.).

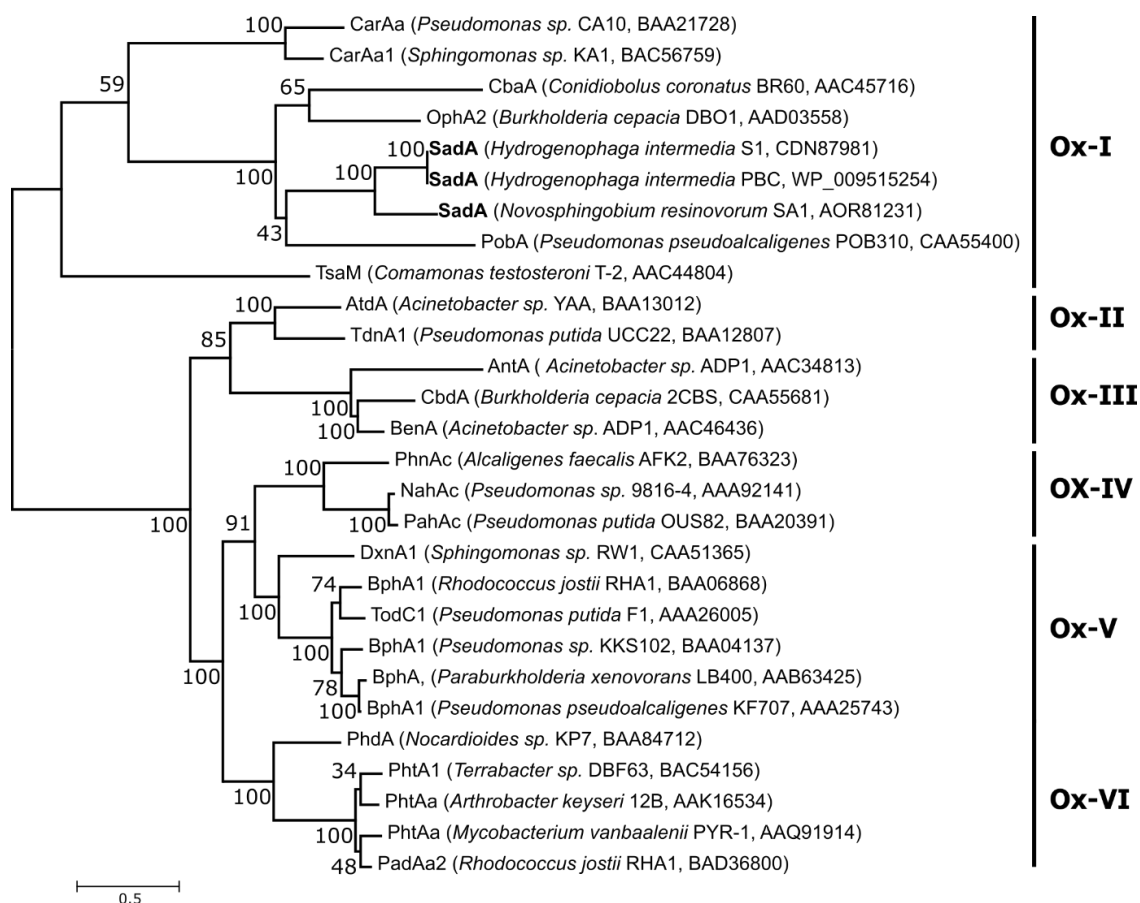


17. ábra. Az *N. resinovorum* SA1 génjeik relatív szinonim kodon használatát bemutató korrespondencia analízis eredménye. Kiemelve jelöltem az R – a főbb riboszómális fehérjék génjeit; B – Benzoát bontási útvonal génjeit; HB – Hidroxi benzoát bontási útvonalának génjeit; SA_p – Szulfanilsav bontási útvonal perifériás szakaszának a génjeit; SA_c – Szulfanilsav bontási útvonal centrális szakaszának a génjeit; SA_f – a Szulfanilsav bontási útvonal perifériás szakaszához szükséges elektrontranszport lánc elemeinek a génjeit. A körülöttük látható vonal jelzi a klaszter analízis segítségével elkülöníthető csoportokat.

A korrespondencia analízisből az látszik, hogy az SA bontás útvonalát felépítő gének kodon használata jelentősen eltér a sejt génjeinek kodon preferenciájától. Ez különösen igaz a perifériális útvonal génjeire.

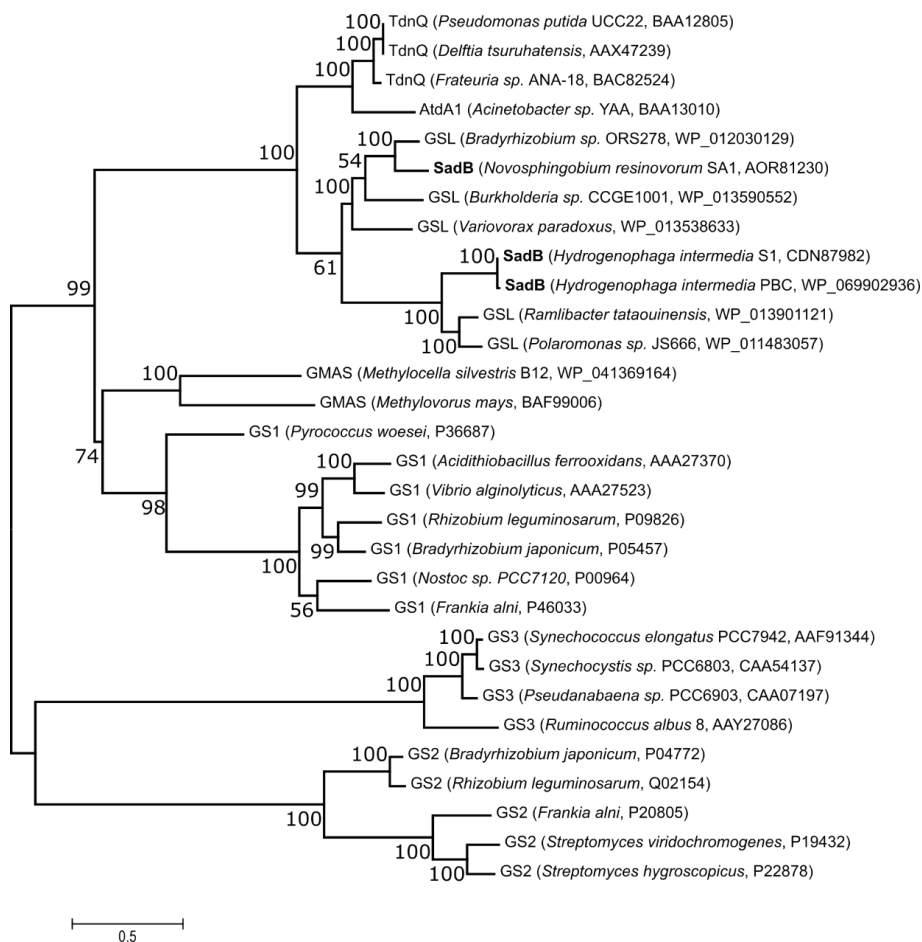
Számos jel utal arra, hogy az SA bontás perifériás szakaszát alkotó génklaszter a baktériumok több családjában is elterjedt, azonban a *N. resinovorum* SA1 törzsbe feltételezhetően horizontális géntranszfer útján került. A horizontális géntranszfer eseménye nem történhetett nagyon régen, mivel a fehérjéket kódoló gének kodon preferenciája még nem tudott alkalmazkodni az új gazdabaktériumra jellemző kodon preferenciájához. A friss eseményt tovább erősíti az is, hogy a legközelebbi rokon *Novosphingobium* törzsek még nem rendelkeznek ezen gének ortológjaival és a *Sphingomonadaceae* család többi tagja között is csak nagyon ritkán fordulnak elő. A perifériás szakaszt kódoló gének eredeti funkciója más baktériumokban nem ismert, feltehetően valamilyen szubsztituált, feltehetően amin-csoportot tartalmazó aromás vegyületek konverzióját végzik. Ugyanez elmondható az útvonal centrális részéről is, amely alapvetően egy béta-ketoadipát útvonalnak tekinthető, amely a szulfonált aromás vegyületek orto-hasításán keresztül történő bontásához alkalmazkodott.

A Nam és mtsai. által létrehozott ⁴⁷, a Rieske oxigenázok (RO) osztályozására alkalmas rendszert követve Gan és mtsai. ²⁶ a *H. intermedia* PBC-ben azonosított, az SA hidroxilálását végző SadA enzimet az oxigenázok I-es csoportjába sorolta be. A Kweon és mtsai féle ⁴³ kicsivel fejlettebb rendszert követve megismételtem a SadA fehérje csoportba sorolását. A filogenetikai vizsgálat során felhasználtam a *H. intermedia* S1 ben és az *N. resinovorum* SA1-ben megtalálható SadA ortológjait is (18A. ábra.). Ez alapján az *N. resinovorum* SA1 törzsben megtalálható SadA fehérje - a két szinte teljesen azonos *H. intermedia* törzsekből származó SadA fehérjékkel közösen - egy jól elkülönült csoportot alkotnak az I es típusú oxigenázok (Ox-I) csoportján belül. Az ilyen típusú oxigenázok elektrontranszport lánc nem tartalmaz ferredoxin komponenst, a teljes funkciót egy FNRC típusú reduktáz látja el. Ez azonban nem feltétlenül vonatkozik az SA hidroxilálásában szerepet játszó dioxigenázra. Számos példa mutatja, hogy ezek az oxigenázok gyakran alkotnak közös génklasztert növényi típusú ferredoxint és GR típusú reduktázt kódoló génekkel.



18A. ábra. A *H. intermedia* és az *N. resinovorum* SA1 fajokból származó SadA ortológok és a Kweon és munkatársainak által standardként használt jól jellemzett Rieske típusú oxigenázok közötti filogenetikai rokonsági kapcsolat. Az oxigenázok csoportosítása Kweon és munkatársainak csoportosítását követte. A fehérjék közötti evolúciós kapcsolat maximum likelihood módszerrel volt meghatározva. Az egység az aminosavanként történő szubsztitúciók számát mutatja.

A Gan és munkatársai által elvégzett filogenetikai vizsgálat megmutatta, hogy a *H. intermedia* PBC-ben található SadB fehérje nem része az általános glutamin szintetáz családnak amely a nitrogén metabolizmus egyik kulcsenzimét foglalja magába ²⁶. A vizsgálatból az is kiderült, hogy a SadB fehérje az anilin bontási útvonalban szerepet játszó glutamin szintetáz-szerű fehérjékkel (TdnQ, AtdA1) rendeződik egy csoportba, illetve olyan glutamin szintetáz-szerű fehérjékkel, amelyek génjei egy oxigenáz fehérjét kódoló génekkel alkotnak *sadAB*-ez hasonló génklasztert. Ezeknek a glutamin szintetáz-szerű fehérjéknek a funkciója nem tisztázott.



18B ábra. A *Hydrogenophaga* fajokból és az *N. resinovorum* SA1 -ből származó SadB fehérjék és jól jellemzett glutamin szintetáz fehérjék közötti filogenetikai kapcsolat. A fehérjék közötti evolúciós kapcsolat maximum likelihood módszerrel lett meghatározva. Az egység az aminosavanként történő szubsztitúciók számát mutatja.

A *H. intermedia* S1 és az *N. resinovorum* SA1 törzsekből származó SadB fehérjével megismételve a filogenetikai vizsgálatot hasonló megfigyelést lehet tenni (18B. ábra). Azonban érdekes módon az *N. resinovorum* SA1 sejtéből származó SadB fehérje nem a *H. intermedia* törzsekből származó SadB fehérjékkel alkot egy közös csoportot, ahogy a SadA fehérje esetén láttuk, hanem a *Bradyrhizobium* törzsből származó glutamin szintetáz-szerű fehérjével, amelynek a funkciója nem vizsgált.

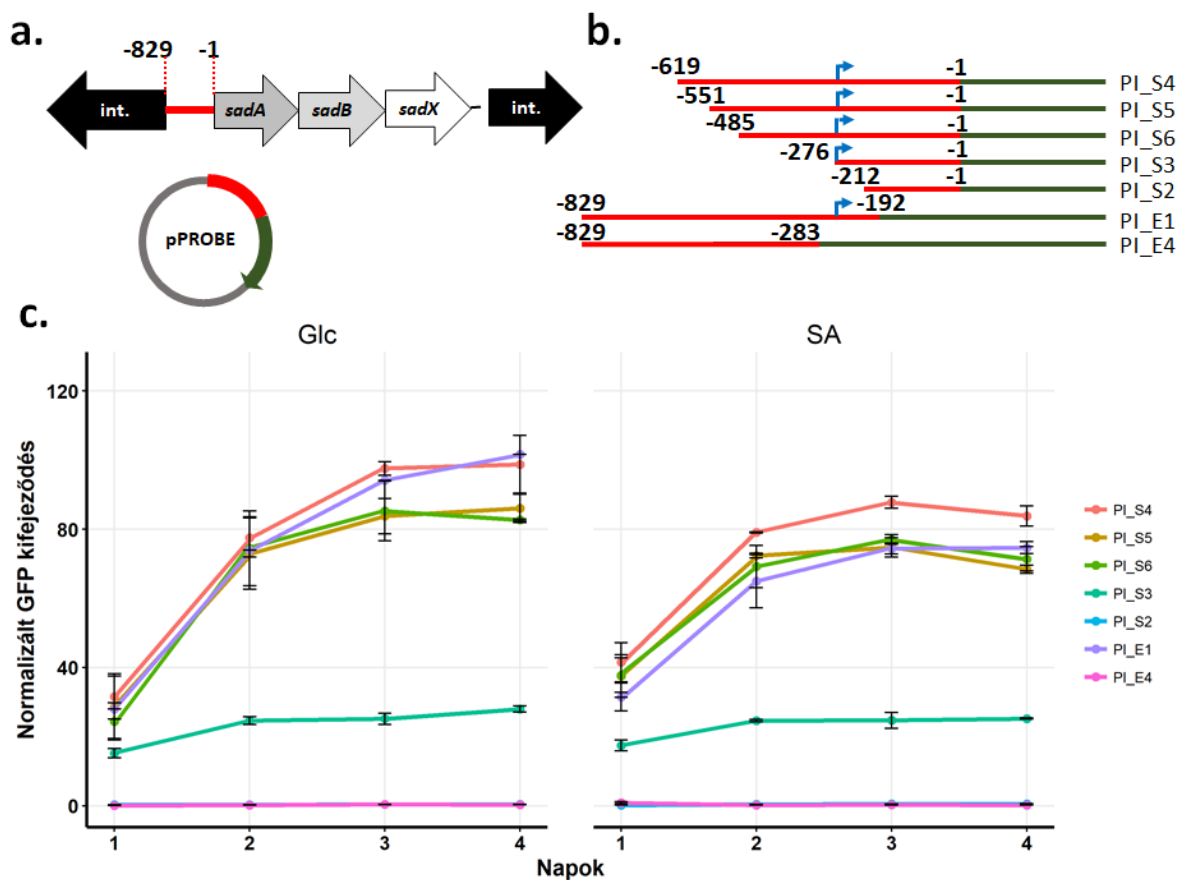
6.4 A perifériás és centrális útvonal génjeinek transzkripciós szabályozó régiói

Az *N. resinovorum* SA1 törzs SA bontását lehetővé tevő katabolikus útvonal elemei nem olyan régen válhattak a törzs változatos katabolikus rendszerének részévé. Azért, hogy megtudjuk, ilyen rövid idő alatt hogyan voltak képesek beépülni a baktérium szabályozó rendszerébe

promóter próba vektor ⁹⁹ segítségével feltérképeztem a perifériás és a centrális útvonalak promóter/szabályozó cisz elemeit illetve megvizsgáltam azok különböző szénforrásokra (Glc, SA) adott válaszát.

A perifériás útvonal génjeit tartalmazó gén klaszter (*sadABX*) két mobilis genetikai elem (BES08_30730, BES08_30710) közé ékelődve található meg az *N. resinovorum* SA1 genomjában. A *sadA* gén és az előtte elhelyezkedő, vele ellentétes orientációban átíródó integráz génje (BES08_30730) között 845 bp hosszú nem kódoló, intergénikus régió található. Feltételezhetően ez a viszonylag hosszú DNS szakasz tartalmazza a *sadABX* átírásáért felelős promóter szekvenciát és az azt esetleg szabályozó transzkripció faktor kötődési helyét. Azért, hogy feltérképezzük ezeknek a potenciális cisz elemeknek az elhelyezkedését, polimeráz-lánreakció (PCR) használatával a két gén közötti intergénikus régió különböző hosszúságú szakaszait hoztam létre (19. ábra.). Az így létrehozott rövid szekvenciákat egy saját promóterrel nem rendelkező, zöld fluoreszkáló riporter fehérjét (GFP) kódoló gént tartalmazó promóter próba vektorba (pPROBE) illesztettem. A különböző hosszúságú szekvenciákat tartalmazó pPROBE vektort az *N. resinovorum* SA1 törzsbe transzformáltam és fluoriméter segítségével (Thermo Fluoroskan Ascent FL) több napon keresztül követtem a Glc és SA szénforráson növesztett sejtek GFP kifejeződési szintjét (19. ábra.). A tápoldatok, az alkalmazott szénforrásokat viszonylag alacsony, 5 mM koncentrációban tartalmzták, amit a sejtek a negyedik napra már teljesen el is fogyasztottak. A vizsgálat alapján úgy tűnik, hogy a perifériás útvonal génjeit tartalmazó gén klaszter kifejeződése egyetlen promóterhez köthető. Ez a promóter a PI-S3-as szekvencia 5'-végének közelébe térképezhető, a translációs kezdési helyétől (TSS) kevesebb, mint 276 bázispárral. Azt, hogy ez a szekvencia részlet felelős egyedül a perifériás útvonal kifejezéséért megerősíti az, hogy a PI-S3 5'- vége előtti (PI_E4) és kicsivel utáni (PI_S2) szekvencia részletek nem voltak képesek a GFP kifejezésére (19. ábra.). A legkisebb, kifejeződést előidézni képes szekvenciaryészlet (PI-S3) esetén konstitutív kifejeződés figyelhető meg. Ennél kicsivel hosszabb szekvencia részlet (PI-S6) esetén azonban már valamilyen szintű szabályozás is megfigyelhető, ami a szekvencia hosszának tovább növelésével már nem változik lényegesen (PI-S5, PI-S4). A szabályozás – ha van egyáltalán - valószínűleg aktiváción keresztül történik meg. Érdekes módon az aktiváció alig függ a használt szénforrástól, annál inkább a növesztés idejétől. Ez alapján feltételeztem, hogy a sejtek növekedése során elfogyasztott tápanyagok miatt kialakuló éhezés lehet az aktiváció fő oka. Az aktivációért felelős transzkripció faktor kötési helye feltételezhetően a PI-S3 szekvencia fölött helyezkedik el valahol a TSS helytől számított -485 és -276 pozíciók között.

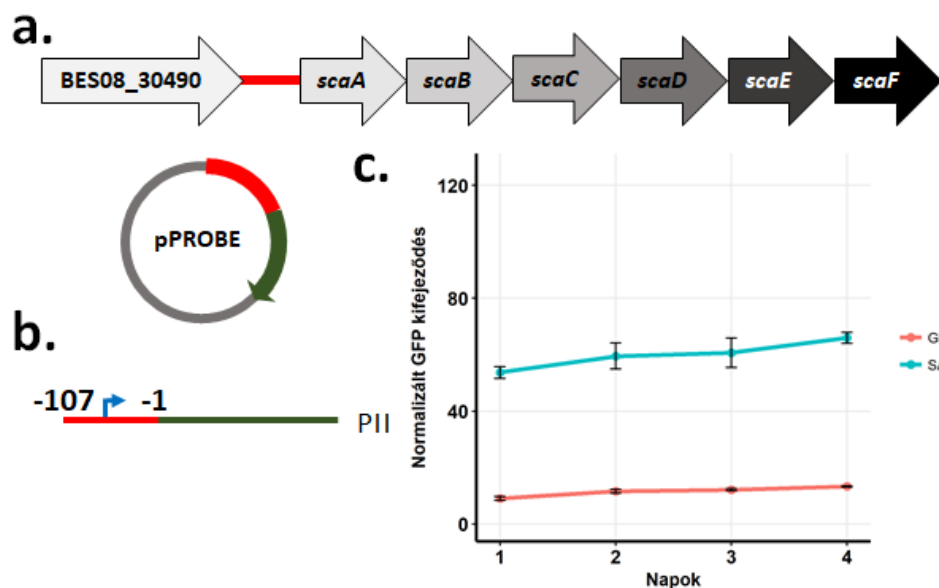
A bakteriális promóterek és transzkripciós faktor kötőhelyeinek azonosítására alkalmas BPROM¹¹⁵ program segítségével megvizsgáltam a -276 pozíció fölött és mögött 100-100 bázispárnyira elnyúló szekvenciát (-375 és -176 közötti szekvenciát). A program által azonosítottam egy bakteriális promóter szekvenciát (-35 box: TTGCTT, -10 box: AGGCATTTT, konzervált szekvenciák közötti rész „inter-box gap”: 14 bp) amely a PI-S3 fragmenten található meg, annak 5’ végétől 4 bázispárnyira. A programnak a promóter előtt 5 bázispárnyira egy feltételezett IHF (integrációs host faktor) fehérje kötőhelyet (TTTTATTT) is sikerült azonosítania.



19. ábra. A) A SA bontási útvonal perifériás szakasza előtt elhelyezkedő intergénikus régió (piros). **B)** A promóter régió feltérképezésére létrehozott pPROBE vektor változatokban megtalálható intergénikus régió darabok (piros), feltételezett promóter helye (kék nyíl) és a GFP kódoló gén (zöld) sematikus képe. **C)** A különböző szénforráson növesztett, eltérő fragmentet hordozó pPROBE vektort tartalmazó *N. resinovorum* SA1 törzsek által kifejezett GFP fehérjék szintjének változása az idő függvényében. A hibásávok jelzik a három biológiai párhuzamos esetén tapasztalt szórást.

Hasonló módszerekkel elvégeztem a centrális útvonal génjeit tartalmazó *scaABCDEF* operon promóter/szabályozó régiójának az azonosítását is. Ebben az esetben az operon első génje (*scaA*) és a tőle 5’ irányban elhelyezkedő feltételezett szulfid exportert TauE/SafE kódoló gén

(BES08_30490) között lényegesen kisebb, 107 bp hosszúságú intergénikus régió helyezkedik el (21. ábra.). A rövid hossz miatt a vizsgálathoz a teljes 107 bp hosszúságú szekvenciát felhasználtam. A GFP kifejeződés mértéke alapján az látszik, hogy a centrális útvonal kifejeződésért felelős promóter szénforrá- függő szabályozással rendelkezik. SA szénforrás használata esetén az *N. resinovorum* SA1 lényegesen nagyobb mértékben fejezte ki a GFP fehérjét, mint glükóz használata esetén. A növesztés idejének és az esetleges éhezés kialakulásának nincs befolyásoló hatása a gének kifejeződésére. A GFP kezdeti kifejeződési szintje a 4 napi növesztés után is szinte változatlan maradt (20. ábra.).

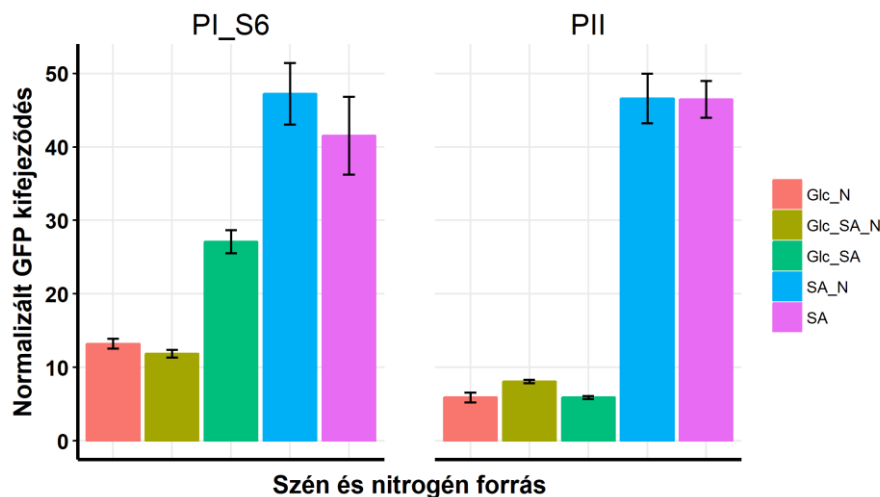


20. ábra. a.) Az SA bontási útvonal centrális szakasza előtt elhelyezkedő intergénikus régió (piros) kiterjedése. b.) A promóter régió feltérképezésére létrehozott pPROBE vektorban megtalálható intergénikus régió darab (piros), feltételezett promóter helye (kék nyíl) és a GFP kódoló gén (zöld) sematikus képe. c.) A különböző szénforráson növesztett, pPROBE vektort tartalmazó *N. resinovorum* SA1 törzsek által kifejezett GFP fehérjék szintjének változása az idő függvényében. A hibasávok jelzik a három biológiai párhuzamos esetén tapasztalt szórást.

Ebben az esetben is a BPROM program használatával sikerült azonosítani egy lehetséges promóter szekvenciát a TSS-től megközelítőleg -50 bázispárra (-10 box: ACATAGAAT, -35 box: ATCAAT, konzervált szekvenciák közötti rés „inter-box gap”: 15 bp). Potenciális transzkripciós faktor kötőhelyet ebben az esetben azonban nem találtam.

Amíg az *N. resinovorum* SA1 törzs az SA-at egyedüli szén, nitrogén, kén és energia forrásként is képes felhasználni, addig a glükózt, annak kémiai összetétele miatt, csak szén és energiaforrásként képes hasznosítani. A glükóz szénforráson növesztett sejtek tápoldatát emiatt

könnyen felvehető nitrogén forrással egészítettem ki. A kénforrás külön biztosítása nem szükséges mivel a tápoldat só komponensei megfelelő mennyiségben tartalmazták azt. Ez ahhoz vezethet, hogy a glükózon növesztett sejtek nemcsak könnyen hasznosítható szénforrással, hanem nitrogén forrással is rendelkeznek. Ezzel szemben az SA-n növesztett sejtek számára nem csak a szén-, hanem a nitrogénforrás is nehezebben hozzáférhető. Azért, hogy megvizsgáljam a könnyen hozzáférhető szén- és nitrogénforrás hatását az SA bontási útvonalának szabályozására a pPROBE/PI-S6 és pPROBE/PII plazmidokat tartalmazó *N. resinovorum* SA1 törzset olyan tápanyagon növesztettem, amely a szén- (SA, Glc) és nitrogénforrásokat (SA, NH₄NO₃) különböző kombinációban tartalmazta (21. ábra.). Ebben az esetben a szénforrást a tápoldat viszonylag magas koncentrációban, 50 mM, tartalmazta, hogy elkerüljem a korábban tapasztalt, feltételezhetően az éhezés által kiváltott jelenséget.



21. ábra. Az SA bontás perifériás és centrális útvonalainak promóter és szabályozó régióit is tartalmazó pPROBE változatok (perifériás: PI_S6, centrális: PII) *N. resinovorum* SA1-ben megfigyelhető GFP kifejeződési szintjének változása különböző szén és nitrogénforrások hatására. A hibasávok jelzik a két biológiai párhuzamos esetén tapasztalt szórást

A GFP kifejeződési szintje ebben az esetben is azt mutatja, hogy az SA bontási útvonalának perifériás szakasza más jellegű szabályozással rendelkezik, mint a centrális szakasza. Abban az esetben, amikor megfelelő minőségű szén- és nitrogénforrás áll a sejt rendelkezésére (Glc_N, Glc_SA_N) a perifériás útvonal promótere által kifejezett GFP szintje alacsony marad. Az SA jelenléte nem befolyásolta a GFP kifejeződési szintjét (Glc_SA_N). Hogyha a nitrogén vagy a szén szükséglet csak nehezebben elérhető forrásból elégíthető ki (Glc_SA, SA_N, SA) a GFP kifejeződése lényegesen megnő. A Glc_SA esetén tapasztalt köztes érték magyarázható azzal,

hogy a baktérium az SA-t csak nitrogénforrásként hasznosította, amit viszonylag kevés SA bontásával már képes volt kielégíteni.

A centrális útvonal promótere csak abban az esetben eredményez megnövekedett GFP kifejeződési szintet, hogyha a tápoldat nem tartalmaz könnyen felvehető szénforrást, csak szulfanilsavat (SA_N, SA). Ebben az esetben a nitrogénforrásnak nem volt látható hatása a GFP kifejeződésére. Glc_SA esetében meglepő az alacsony GFP kifejeződési szint, hiszen a szabályozás két legvalószínűbb indukálószerre, az SA és a bontás első jelentősebb köztterméke, az SC, is jelen van a sejtben. Mivel a sejt nitrogén igényét a perifériás útvonal során felszabaduló ammónia képes kielégíteni, ezért könnyen felvehető szénforrás jelenlétében a centrális útvonal szakasz kifejeződése nem létfontosságú. A tapasztalt alacsony kifejeződési szint pedig magyarázható a mikroorganizmusok esetén gyakran megfigyelt katabolit represszió (CCR: carbon catabolic repression) jelenségével. Ez a jelenség felelős azért, hogy a nehezebben bontható szénforrások hasznosításáért felelős gének kifejeződése mindaddig alacsony szinten maradjanak, amíg könnyebben felhasználható szénforrás is elérhető a baktérium számára^{116,117}.

Az SA bontási útvonal szabályozása más fajok esetén alig ismert. A *H. intermedia* PBC az egyetlen olyan törzs, amely esetén nem csak az útvonal enzimeinek azonosítására és jellemzésére koncentráltak, hanem vizsgálták azok kifejeződésének szabályozását is. Azonban a vizsgálatok ebben az esetben is csak a bontási útvonal perifériás szakaszára korlátozódtak²⁶. Viszont az jól látszik, hogy a *N. resinovorum* SA1 törzssel ellentétben a *H. intermedia* PBC törzsben a *sadABD* operon kifejeződése nem konstitutív. A bontási útvonal szubsztrátjával (SA) és az útvonal első fő közttermékével (4-SC) is indukálható.

A *H. intermedia* PBC törzsben megtalálható *sadABD* operon előtt, azzal ellentétes orientációban, egy GntR típusú transzkripciós faktor génje helyezkedik el. A *sadABD* operonhoz kötődő GntR típusú transzkripciós faktor általánosnak tekinthető a *Hidrogenophaga* fajok esetén. Sok esetben leírták aromás bontási útvonalak szabályozásában betöltött szerepét¹¹⁸. Ezek alapján feltételezhető, hogy ebben az esetben az *sadABD* operon esetén is hasonló szerepet tölt be.

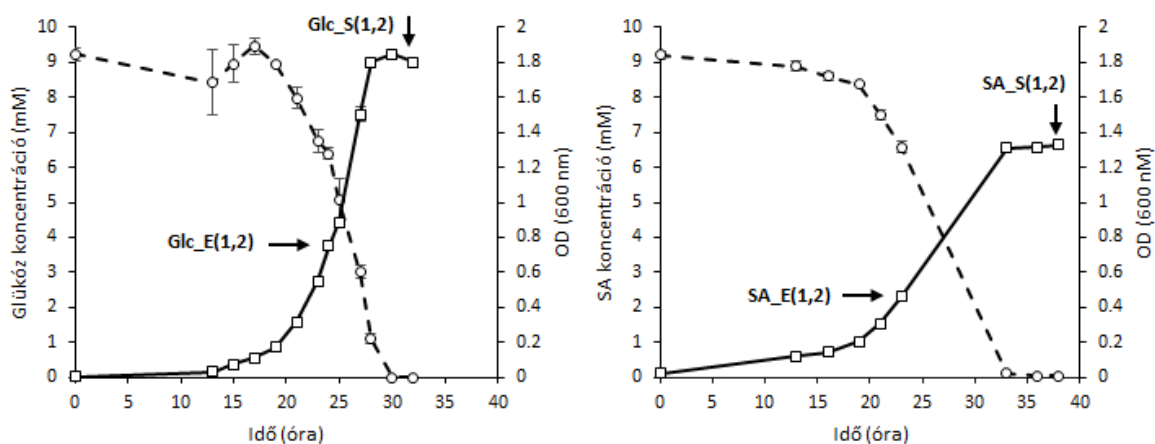
Az *N. resinovorum* SA1 törzs *sadABX* operonja előtt egy mobilis genetikai elem (integráz) génje található meg. Ettől 5' irányban a *sadABX* operonnal azonos orientációban elhelyezkedő maleilacetát redukáz (AOR81346) kódoló gén és vele ellentétes orientációban pedig egy IclR típusú transzkripciós faktort (AOR81347) kódoló gén található. Ezek a fehérjék szoros kapcsolatban állnak az aromás bontási útvonalakkal. Amíg a maleilacetát redukáz számos esetben a β -ketoacid út fontos építőeleme^{119–121}, addig az IclR típusú transzkripciós

faktor az aromás bontási útvonalak egyik fontos transzkripciós faktor típusa ^{122,123}. Feltételezhető, hogy az integráz beépülése előtt ezek a gének a *sadABX* génekkel közös operont alkottak és egy mára ismeretlen bontási útvonal elemét képezték, azonban a ma megfigyelhető SA bontási útvonalában se építő, se szabályozó szerepet nem játszanak.

6.5 Teljes RNA-Seq alapú transzkriptom vizsgálat

Az *N. resinovorum* SA1 klasszikus módszerekkel nehezen tanulmányozható baktérium törzs, aminek a fő oka, hogy napjainkig nem sikerült kialakítani megbízhatóan működő technikákat genetikai módosítására. A vizsgálatot tovább nehezíti a törzs komplex genom összetétele és az SA bontás útvonalát felépítő gének genomon belüli szétszórt elhelyezkedése. Amíg a bontási útvonal jól ismert elemeinek szekvencia homológia alapján történő azonosítására ¹²⁴ az irodalomból ismert adatok használatával megoldható, addig a kevésbé ismert elemek feltárására nagyon kevés lehetőség áll rendelkezésünkre. Az RNA-Seq alapú transzkriptom vizsgálat, amely egyre elterjedtebb technika a baktérium teljes transzkripciós mintázatának feltérképezésére, az ilyen problémákhoz is megoldást nyújthat. Segítségével lehetőség nyílik olyan nehezen azonosítható rendszerek, mint a bontási útvonal perifériás szakaszában megtalálható oxigenáz elektrontranszport láncát felépítő gének, az SA felvételét végző transzporter génje illetve a bontás során nagy mennyiségben felszabaduló szulfid detoxifikálását végző rendszer génjeinek azonosítására. Továbbá pontosabb képet alkothatunk a bontási folyamat többértű szabályozásáról, amelyben az SA indukáló hatása mellett más fiziológias hatásokra, pl. az éhezésre és esetleg az SA toxikus jellegére adott válaszok is megmutatkozhatnak.

Azért, hogy minél átfogóbb képet kapjak az *N. resinovorum* SA1 transzkriptomjában bekövetkező változásokról az RNA-seq vizsgálatához szükséges teljes RNS mintát a 10 mM glükózt vagy 10 mM SA-at tartalmazó minimál tápoldaton növesztett, exponenciális (Glc_E/SA_E) és stationer (Glc_S/SA_S) növekedési fázisban lévő sejtekből izoláltam (22. ábra.). A glükóz minimál tápoldat esetén a nitrogén forrást 10 mM NH₄Cl hozzáadásával biztosítottam, ezzel szemben a SA minimál tápoldat esetén a nitrogén forrást az SA biztosította. Hasonlóan a korábbi promóter vizsgálatokhoz, a növesztés során itt sem tapasztaltam nitrogén limitációra utaló jelet. Minden vizsgálatot két biológiai párhuzamos használatával végeztem el.



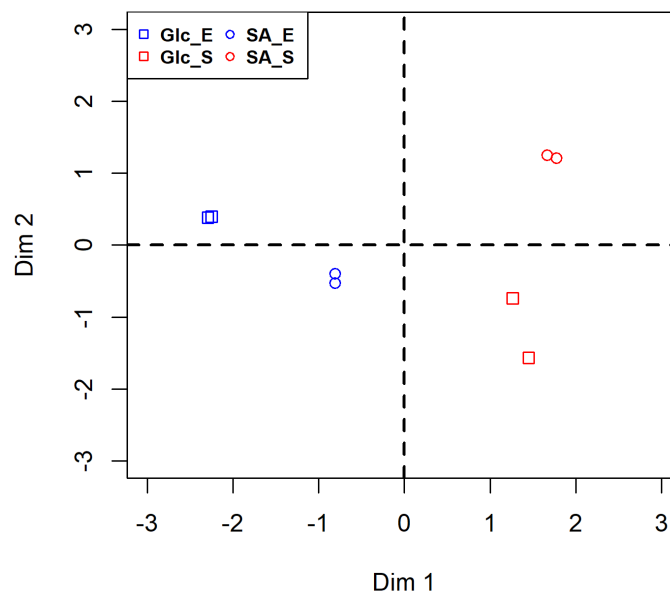
22. ábra. A glükózon (Glc) és szulfanilsvon (SA) növesztett *N. resinovorum* SA1 törzs sejtszámának (négyzet) és a tápanyagforrás koncentrációjának (kör) változása az idő függvényében. A hibaszávok jelzik a két biológiai párhuzamos esetén tapasztalt szórást. A nyilak jelzik azokat a pontokat ahonnan az RNA-seq vizsgálathoz felhasznált minták származnak.

A sejtek exponenciális növekedési szakaszából származó minták szekvenálásából kapott tisztított (trimmed) readok közel 50% -ka térképezhető az *N. resinovorum* SA1 genomjának fehérjét kódoló régiójára. A stacioner növekedési szakaszból származó minták esetén ez az érték valamivel alacsonyabb (2. táblázat). Sajnálatos módon, feltételezhetően a szekvenálás során bekövetkezett hiba miatt, a glükózon növesztett, stacioner növekedési szakaszban lévő sejtekből származó minták egyik biológiai párhuzamosa (Glc_S2) nagyon alacsony kódoló régióra térképezhető reads számot eredményezett. Mivel a Glc_S1 és a Glc_S2 minták RNA-Seq profilja között erős korreláció volt megfigyelhető a read-szám különbség ellenére, a Glc_S2 mintát nem távolítottam el a vizsgálatból.

Minta ID.	Teljes readszám	Párban lévő, tisztított (trimmed) readok száma	Géekre térképeződött, párban lévő, tisztított (trimmed) readok száma	%
Glc_E1	2 989 506	1 388 729	664 559	47.9
Glc_E2	3 151 936	1 467 686	752 965	51.3
Glc_S1	1 437 618	699 741	368 393	52.6
Glc_S2	3 650 414	1 321 804	90 268	6.8
SA_E1	3 319 482	1 558 226	834 220	53.5
SA_E2	2 694 052	1 319 173	736 529	55.8
SA_S1	1 822 426	784 532	268 548	34.2
SA_S2	1 736 428	802 557	292 106	36.4

2. táblázat. Az RNA-seq vizsgálathoz használt MiSeq szekvenálás adatainak összefoglalása

A minták génjeinek kifejeződési mintázata közötti eltérések ábrázolására MDS szóródási diagrammot (Multidimensional scaling plot) használtam (23. ábra.). Az ábrán megfigyelhető, hogy a biológiai párhuzamosok közötti eltérések nem jelentős. A Glc_S minták közötti eltérés valamilyen mértékben nagyobb, mint a többi mintapár esetén azonban ez nem mondható igazán jelentősnek. Ezzel szemben a különböző körülmények közül származó minták között jelentős eltérés mutatkozik. A legnagyobb hasonlóság a különböző tápanyagon növesztett stacioner növesztési fázisban lévő sejtek génkifejeződési mintázatai között figyelhető meg (Glc_S és SA_S). Érdekes módon azonban a SA_E minták közel azonos mértékben térnek el a Glc_E mintáktól, mint a Glc_S mintától. Ez alapján akár az is feltételezhető lehet, hogy a SA_E esetén tapasztalható kifejeződési mintázatot a Glc_S mintázatát is nagyrészt meghatározó, éhezés/stressz által kiváltott szabályozó folyamatok is befolyásolhatják.



23. ábra. A minták génkifejeződési mintázata közötti eltérés ábrázolása MDS plot (Multidimensional scaling plot) használatával. A glükózon növesztett mintákat (Glc) négyzet, a szulfanilsavon növesztett mintákat (SA) pedig kör jelöli. Az exponenciális (E) növekedési fázisban lévő mintákat kék, a stacioner (S) növekedési fázisban lévő sejteket pedig piros szín jelöli.

A különböző körülmények között megfigyelhető génkifejeződés változás vizsgálatához EdgeR programcsomagot használtam ¹⁰⁹. Csak azoknak a géneknek a kifejeződését vizsgáltam, amelyek a 8 mintából legalább 3 esetben rendelkeztek legalább 1 CPM (Count Per Million) értékkel. A vizsgált gének közül azoknak a változását tekintettem szignifikánsnak, amelyek az

összehasonlított mintapárok esetén legalább négyszeres ($\text{abs}(\log_2) > 2$) kifejeződési szint (FC – Fold Change) változást mutattak, FDR (False Discovery Rate) javított p érték ≤ 0.001 mellett. A három mintapár összehasonlításából kapott megváltozott kifejeződést mutató gének (DEG, Differentially Expressed Gene) összegzését a 3. táblázat tartalmazza. Az azonos stacioner növekedési fázisban lévő, de eltérő szénforráson növesztett sejtek génjeinek kifejeződését számos tényező befolyásolhatja. Összehasonlításuk során kapott génkifejeződés változások csak nagyon nehezen értelmezhetőek, ezért munkám során eltekintettem tőle.

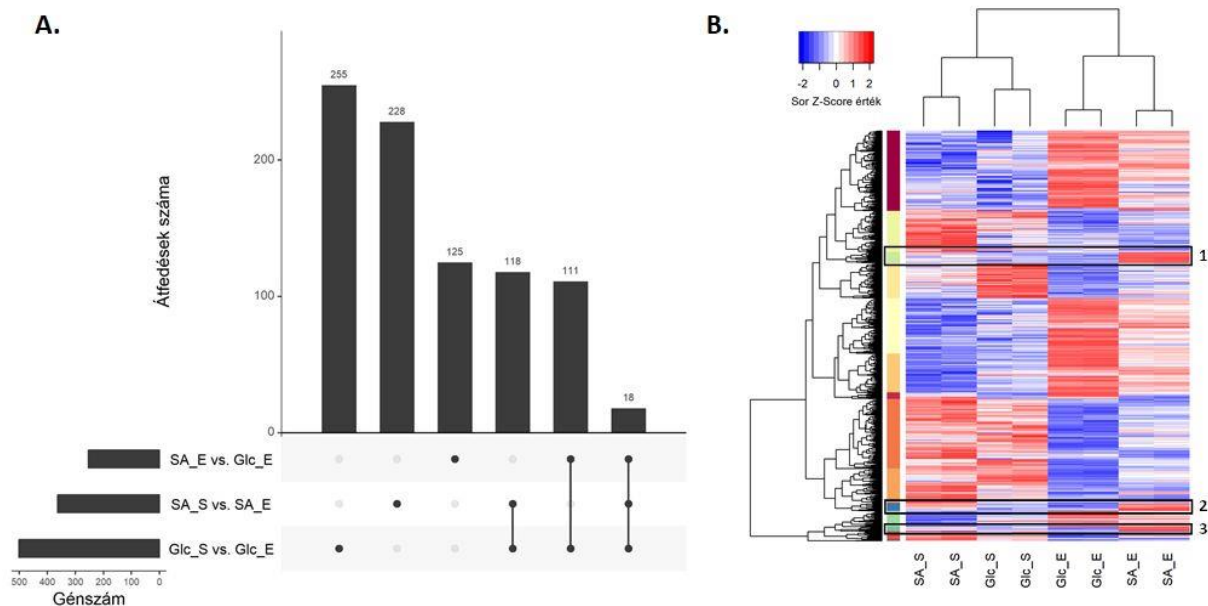
	Glc_S vs. Glc_E	SA_E vs. Glc_E	SA_S vs. SA_E
Összes DEG-ek száma (%)	1056 (17,6%)	432 (7,2%)	792 (13,2%)
Növekedett kifejeződésű DEG-ek száma (%)	502 (8,4%)	254 (4,2%)	364(6,1%)
Csökkenett kifejeződésű DEG-ek száma (%)	554 (9,2%)	178 (3%)	428(7,1%)

3. táblázat. A csoportok összehasonlítása során szignifikánsan megváltozott kifejeződést mutató gének száma. Az *N. resinovorum* SA1 törzs 6113 fehérjét kodoló génje közül 6003 rendelkezett legalább 3 minta esetén 1 vagy ennél nagyobb CPM értékkel.

A DEG-ek számában a legnagyobb változást a sejtek növekedési fázisai közötti váltás idézte elő. A Glc-n és az SA-n növesztett sejtek génjeinek a 17.6% és 13.2% -ának a kifejeződése változott meg az exponenciális növekedési fázisból a stacioner növekedési fázisba történő átmenet során. Mivel a stacioner növesztési fázisban lévő sejtek, a kezdetben rendelkezésre álló tápanyagforrást teljesen felhasználták (22. ábra.), feltételezésem szerint az itt tapasztalható változás elsősorban a kialakult éhezésnek tulajdonítható. Az SA, az éhezés hatásától kisebb, de így is jelentős kifejeződési szint változást idézett elő a sejtekben. Az eltérő tápanyagon növesztett exponenciális növekedési fázisban lévő sejtek (SA_E vs. Glc_E) génjeinek 7.2%-a mutatott valamilyen változást.

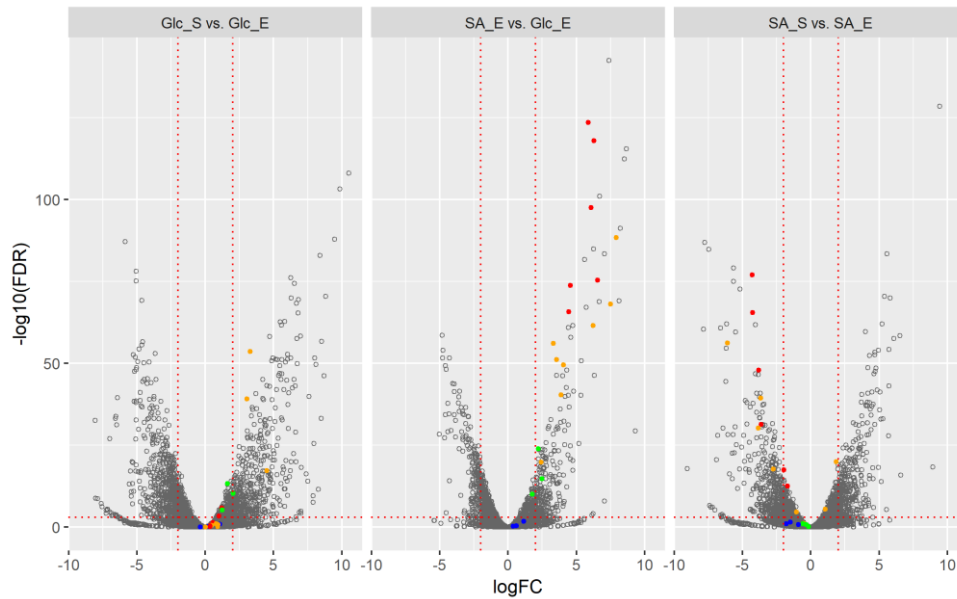
A különböző mintapárok összehasonlításából kapott, megnövekedett génkifejeződést mutató génkészletek közötti átfedés mértékét a 24A ábrán látható UpSet plot szemlélteti ¹²⁵. Az ábra alapján az látszik, hogy számos olyan gén kifejeződése nő meg szignifikánsan, amely csak a Glc –on vagy a SA –on növesztett sejtek éhezési folyamatára jellemző (255, 228) illetve olyan géneké, amely mindkét éhezési folyamatban szerepet játszanak (118). 125 olyan gén van azonban, amely csak a SA hatására, a SA_E vs. Glc_E összehasonlítása során mutatott megnövekedett kifejeződést. Meglepő módon sok olyan gén is előfordul, amely SA hatására és Glc-on történő növesztés során kialakuló éhezés hatására is megnövekedett kifejeződést mutat (111). Elhanyagolhatóak azonban azoknak a géneknek a száma (18), amelyek mind a három esetben mutatnának megnövekedett kifejeződést.

A DEG gének különböző körülmények között mutatott kifejeződési mintázatát az 24B. ábrán látható hőtérkép szemlélteti (23. ábra.). Az MDS plot hoz hasonlóan ezen az ábrán is megfigyelhető, hogy a biológiai párhuzamosok jól elkülönülő klasztereket képeznek. Emellett az is látható hogy a hasonló növekedési fázisból származó, de eltérő tápanyagforráson növesztett sejtek is különálló klaszterekbe rendeződnek. Ez utalhat arra, hogy a sejtek génjeinek kifejeződését jelentős mértékben a sejt növekedési fázisa határozza meg. Az eltérő tápanyagforrás szintén számottevő változásokat eredményez. A hőtérképen három olyan klaszter figyelhető meg, amelyben az SA függő kifejeződést mutató gének legnagyobb részét ($n = 83$; 66,4%) magába foglalja. A hármas számú klaszter 17 gént foglal magába. Ezek közé tartalmazzik a jól szabályozott centrális útvonal génjei, az azok közvetlen közelében elhelyezkedő IclR típusú transzkripciós faktor (BES08_30535) és vele feltételezhetően egy operonba tartozó Fad-kötő dehidrogenáz (BES08_30540) génje. Továbbá egy 4-hidroxibenzoát transzporter (*pcaK*) (BES08_30350), szulfid exporter (*tauE*) (BES08_30490), glutamináz (BES08_30685) génje. Ezek a gének feltételezhetően valamilyen kisegítő szerepet játszanak a SA bontási folyamatában. Az első és a második klaszter (42, 32) is számos további gént tartalmaz, amelyek elősegíthetik a bontás folyamatát. Ilyenek gének a feltételezhetően szulfid transzportert kodoló gének (*tauZ1*, *tauZ2*) (BES08_29790, BES08_30485), a szulfid oxidációjában szerepet játszó gének (*sorA*, *sorX*) (BES08_29830, BES08_29795).



24. ábra. A.) A három mintapár összehasonlításából kapott megnövekedett kifejeződést mutató gének mintapárok közötti átfedésének ábrázolása UpSet plot segítségével. B.) A DEG gének kifejeződésének ábrázolása hő térkép segítségével. A kifejeződési értéke a CPM kettes alapú logaritmusának felelnek meg. Az oszlopok/sorok hierarchikus klaszterezéséhez euklideszi távolságszámítást és teljes láncmódszer szerinti klaszterezést alkalmaztam. A számozott kiemelés azokat a klasztereket jelöli, amelyek olyan géneket foglalnak magukba, amelyek kifejeződése a SA_E körülmények között a legnagyobb és a többi körülmény között jelentősen visszaesik.

Az *N. resinovorum* SA1 génkifejeződésének változását (FC) és a változás szignifikancia szintjét (FDR p-érték) ábrázoló vulkán-diagram (25. ábra.) tovább erősíti a korábban látottakat. A biológiai párhuzamosok közötti eltérés csekély. Az összehasonlított körülmények során a növekvő FC növekvő szignifikancia szinttel párosul. A perifériás, centrális, közös és a kiegészítő útvonalak (lásd később) elemeinek kiemelése esetén láthatjuk hogy a jól szabályozott centrális és kiegészítő útvonalak elemeinek kifejeződése glükózon növesztett sejtek esetén (Glc_S vs. Glc_E) alig változik, szulfanilsav hatására azonban (SA_E vs. Glc_E) szignifikáns növekedést mutatnak, amely jelentősen visszaesik a SA elfogyasztásának következtében (SA_S vs. SA_E).



25. ábra. Az RNA-Seq adatok ábrázolása vulkán-diagram segítségével. Az x tengely ábrázolja a gének kifejeződésében megfigyelhető változás kettes alapú logaritmusát (logFC). Az y tengely ábrázolja a statisztikai teszt p-értékének korrigált, tízes alapú negatív logaritmusát. A függőleges, piros színű szaggatott vonal jelzi a 4 szeres kifejeződésváltozás (logFC = +/- 2) határait. A vízszintes pedig a $-\log_{10}(0,001 \text{ FDR})$ szignifikanciaszint határát jelöli ki. A zöld színű pont jelöli a perifériás, a piros a centrális, a kék a közös, a narancssárga pedig a kisegítő útvonal elemeit (lásd később).

Azért, hogy képet kapjunk arról, hogy a változások milyen funkciós csoportba tartozó fehérjék génjeit érintik, Fisher's egzakt teszten alapuló gén szett túlréprezentáltsági analízist végeztem (26. ábra) (GSOA: Gene set overrepresentation analysis) a gének COG funkcionális annotációjának használatával (2. táblázat).

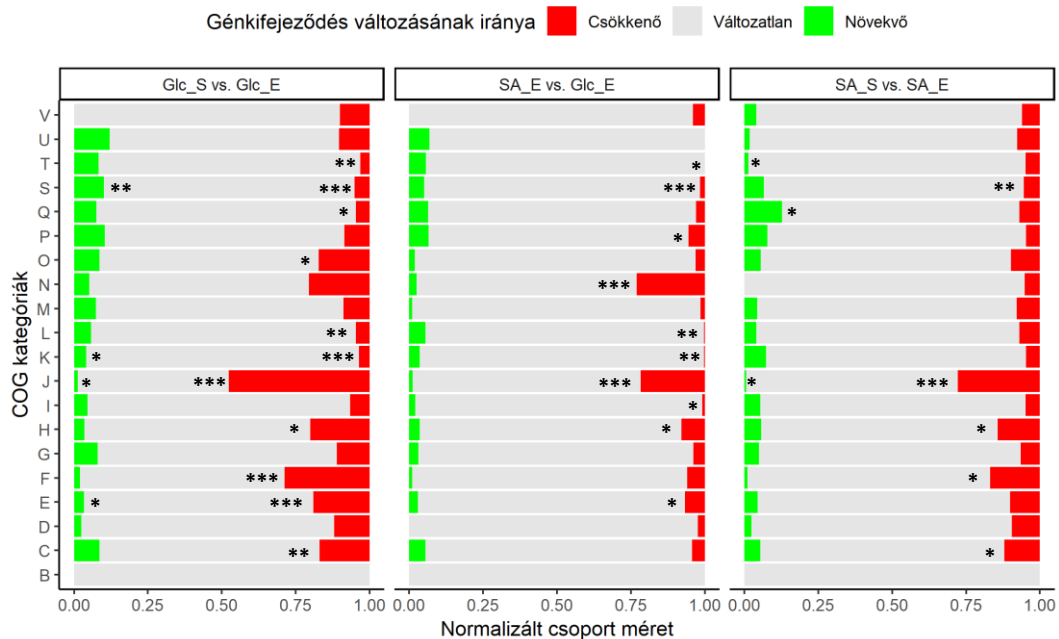
A Glc-on történő növesztés során kialakuló éhezés hatása (Glc_S vs. Glc_E) eredményezte a legtöbb funkciós kategórián belüli változást. A kategóriák többségében a csökkenő kifejeződést mutató gének szignifikáns mértékű felhalmozódása figyelhető meg. Az SA hatása (SA_E vs. Glc_E) szintén viszonylag sok és kivétel nélkül a csökkenő kifejeződést mutató gének felhalmozódását eredményezte. Az SA-on történő növesztés során kialakuló éhezés (SA_S vs. SA_E) lényegesen kevesebb funkciós csoport összetételében okozott változást. Azonban ahol igen ott abban az esetben is többnyire a csökkenő kifejeződésű gének felhalmozódását eredményezte.

Az SA drámai, a Glc-on kialakuló éhezést tükröző hatása markánsan megfigyelhető. Mindkét esetben közös, hogy a sejt hatékony replikációjához szükséges, alapvető funkciós csoportokba, mint a transzláció, transzkripció, replikáció (J, K, L) tartozó gének szignifikáns számban mutattak csökkenő kifejeződést. A replikáció sebességének csökkenését jól tükrözi az SA-on növesztett sejtek növekedési görbéje (22. ábra.). Azonban olyan létfontosságú funkciós

kategóriák, mint az energia termelés, aminosav metabolizmus, nukleinsav metabolizmus (C, E, F) tartozó gének kifejeződésének drámai csökkenése előszörban a Glc-on és az SA-on kialakuló éhezésre jellemző. Az SA esetén a translációban szerepet játszó (J) gének kifejeződésének további csökkenése is szembevető. Ez jól egybecseng azzal, hogy a sejtek többsége ilyen körülmények között már egyáltalán nem osztódódik, ami a növekedési görbén is megfigyelhető (22. ábra.).

Érdekes módon az SA hatására egyetlen COG kategóriában sem halmozódtak fel szignifikáns számban megnövekedett kifejeződésű gének. Ennek az lehet az oka, hogy a SA bontást viszonylag kis számú és változatos kategóriákba tartozó gének végzik, amelyek feldúsulása nem kimutatható. Esetleg a kifejeződésük csak kis mértékben változik, mint ahogy a SA bontás perifériás szakaszánál tapasztaltuk.

A GSOA alapján az tételezhető fel, hogy az SA a sejtekben erős éhezésre jellemző választ vált ki, amely az SA elfogyasztásával tovább erősödik és lényegesen nem tér el a Glc esetén kialakuló éhezéstől.



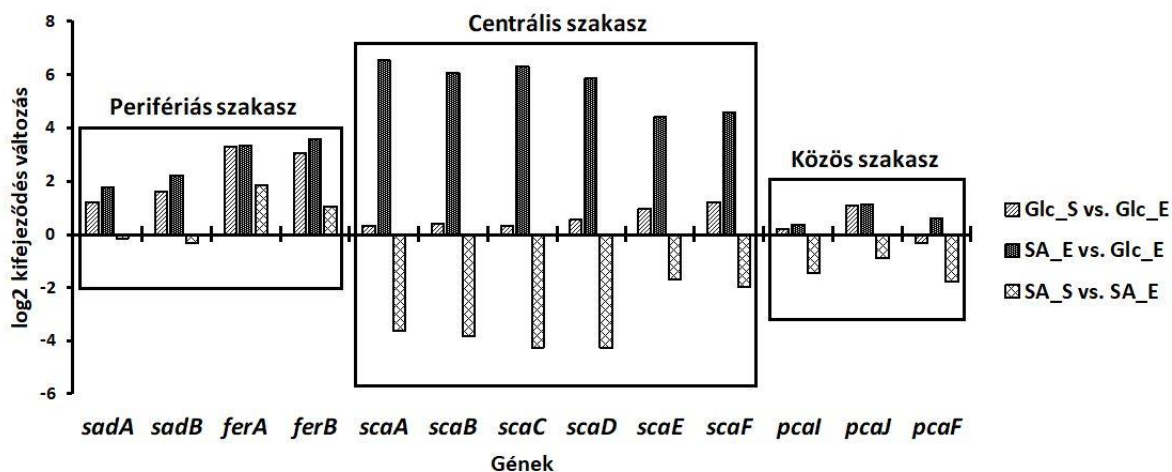
26. ábra. Az *N. resinovorum* SA1 törzs COG kategóriákba tartozó, a mintapárok összehasonlítása során változó kifejeződést mutató géneinek túlréprezentáltációs analízise. Az x tengely mutatja a COG kategóriák normalizált méretét. A jobboldali piros sávok az adott kategóriába tartozó csökkent kifejeződéssel rendelkező gének, míg a baloldali zöld sávok a megnövekedett kifejeződéssel rendelkező gének arányát mutatják. A több kategóriába sorolható géneket többször vettem figyelembe. A betűk a különböző COG kategóriákat jelölik. A felülreprezentált csoportok Fisher's egzakt teszt használatával számolt FDR javított szignifikancia értékeit csillaggal jelöltem (* $p \leq 0.05$, ** $p \leq 0.001$, *** $p \leq 0.0001$).

Hasonló tendencia volt megfigyelhető a sejt útvonalait funkciójuk alapján csoportokba rendező KEGG adatbázissal ¹⁰² elvégzett géncsoport dúsulási vizsgálat (Gene Set Enrichment Analysis, GSEA) ¹²⁶ során is. Szinte az összes feldúsult géncsoport csökkenő kifejeződést mutató géneket tartalmazott. Egyedül csak az aromás vegyületek bontásának (nre01220) kategóriájába tartozó gének mutattak emelkedett kifejeződési szintet SA jelenlétében (Függelék 5).

6.6 A szulfanilsav bontásában résztvevő gének kifejeződése

Az SA perifériás és centrális szakaszát kódoló operonok promóter régiójának a vizsgálatával már korábban képet kaphattunk arról, hogy milyen szabályozó folyamatok hatnak az útvonal elemeinek kifejeződésére. Ezt a képet a transzkriptom vizsgálat tovább erősíti.

A bontási útvonal perifériás szakaszának génjei már a glükózon növesztett, exponenciális növekedési fázisban lévő sejtek esetén is figyelemreméltó kifejeződési szinttel rendelkeznek (Glc_E). A *sadA* (gyűrű hidroxiláló dioxigenáz) és a *sadB* (glutamin szintetáz) gének a baktérium 20 legerősebben kifejeződő génjei közé tartoznak, olyan génekkel együtt, mint a tipikusan magas kifejeződési szinttel rendelkező riboszómális fehérjék, chaperonok és translációs iniciációs faktorok génjei. A glükózon növesztett sejtek *sadAB* génjeinek alap kifejeződési szintje majdnem négyszeresére növekedett az exponenciális fázisból stacioner fázisba történő váltás során (Glc_S vs. Glc_E). A különböző szénforráson növesztett, exponenciális növekedési fázisban lévő sejtekből származó minták között ennél csak minimálisan nagyobb növekedés figyelhető meg (SA_E vs. Glc_E). Ahogy az SA_S és az SA_E minták összehasonlítása mutatja (SA_S vs. SA_E), a megnövekedett kifejeződési szint az SA-n növesztett sejtek stacioner fázisában is megmaradt, értéke csak kismértékben csökkent. Ez azért érdekes, mert az SA_S minták esetén, az SA már elfogyott a tápoldatból, ezért ennek indukáló hatása már nem érvényesülhet, emiatt a *sadAB* gének kifejeződés szintjének jelentős csökkenését várhatnánk (27. ábra).



27. ábra. Az SA bontási útvonal perifériás szakaszát, annak kiegészítő elektrontranszport láncát, továbbá az útvonal centrális és közös szakaszát felépítő géneknek a kifejeződési szintjében megfigyelhető változás látható. Az ábrán szereplő gének a következő enzimeket kódolják: *sadA*: 4-aminobenzolszulfonát 3,4-dioxigenáz; *sadB*: glutamin szintetáz; *ferA*: ferredoxin; *ferB*: ferredoxin reduktáz; *scaA*: 3-szulfo-cisz,cisz-mukonát cikloizomeráz; *scaB*: szulfolakton hidroláz; *scaC*: maleilacetát reduktáz; *scaD*: lehetséges oxidoreduktáz; *scaE*: szulfokatekol 3,4-dioxigenáz béta alegysége; *scaF*: szulfokatekol-3,4-dioxigenáz alfa alegysége; *pcaI*: 3-oxoadipát CoA transzferáz A alegység; *pcaJ*: 3-oxoadipát CoA transzferáz B alegység; *pca*: 3-oxoadipil CoA tioláz

A perifériás útvonal génjeivel szemben a centrális útvonal génjei (*scaABCDEF*) alacsony alap kifejeződési szinttel rendelkeznek a Glc_E minták esetén, ami nem növekedett jelentősen a stacioner növekedési fázisba történő váltás után sem (Glc_S vs. Glc_E). Viszont az SA-n növesztett exponenciális növekedési fázisban lévő sejtek esetén közel 64-szeres kifejeződési szint emelkedés volt megfigyelhető a szintén exponenciális fázisban lévő, de Glc tápanyagon növesztett sejtekhez képest (SA_E vs. Glc_E). Az SA elfogyasztása a stacioner fázisára a korábban észlelt megnövekedett génkifejeződés csökkenését eredményezte (SA_S vs. SA_E), megközelítőleg a Glc_E és Glc_S mintáknál tapasztalt szintig (27. ábra).

Az összes β -ketoadipát útvonalon keresztül történő bontási folyamat közös jellemzője a 3-oxiadipát köztitermék megjelenése¹²⁷. Amíg a baktériumok számos párhuzamos β -ketoadipát útvonallal rendelkezhetnek, amelyek képesek a legváltozatosabb aromás vegyületeket 3-oxoadipáttá (β -ketoadipáttá) alakítani, addig a 3-oxoadipát továbbalakítására csak egyetlen közismert útvonallal áll rendelkezésükre. Ez az úgynevezett közös útvonal felelős a 3-oxoadipát szukcinil-CoA és acetyl-CoA vegyületekké alakításáért, amelyek végső soron képesek belépni a TCA ciklusba. Ebben a folyamatban három enzim, a PcaIJ és PcaF vesz részt. Ezeket az enzimeket kódoló gének az *N. resinovorum* SA1 törzs pSA1-es plazmidján találhatóak meg minden bizonnyal egy operonba rendeződve. Habár ennek az útvonalszakasznak a szerepe ismereteink alapján nélkülözhetetlen lenne az SA bontása során, mégis azt tapasztaltam, hogy

a felépítésében részt vevő gének kifejeződése rendkívül alacsony volt az összes vizsgált körülmény között. Érdekes módon az SA jelenlétének alig volt hatása ezeknek a géneknek a kifejeződésére (27. ábra.).

6.7 A szulfanilsav bontást kisegítő elemek azonosítása

6.7.1 A szulfanilsav gyűrűhidroxiláló dioxigenáz elektrontranszport láncának elemei

Az SA bontás perifériás szakaszának a legfontosabb lépése az aromás gyűrű dihidroxilálása, amelyet a NAD(P)H oxidálása és az SA aminocsoportjának eltávolítása kísér. Ezt a reakciót a *sadA* gén által kódolt Rieske típusú oxigenáz (RO) enzim katalizálja. Az RO működése folyamatos elektronellátáshoz kötött. A szükséges elektronokat egy többé-kevésbé oxigenáz specifikus elektrontranszport lánc (ETC) szállítja a NAD(P)H redox kofaktorokról. Az oxigenázokat kiszolgáló ETC állhat csak egy flavoprotein reduktázból illetve kiegészülhet egy elektron szállításért felelős ferredoxin fehérjével ⁴⁷.

Az *N. resinovorum* SA1 pfam annotációjának felhasználásával hat olyan lehetséges flavoprotein reduktázt kódoló gént sikerült azonosítani, amelyek szerepet játszhatnak az RO elektron igényének kielégítésében. Kweon osztályozását követve a géntermékek közül ötöt lehetett a reduktázok egyes (Rd-I, GR típusú), míg egyet a hármas (Rd-III, FNR_N típusú) csoportba besorolni. A transzkriptom vizsgálat alapján az azonosított lehetséges flavoprotein reduktázok közül csak egy Rd-I csoportba tartozó reduktáz génje, a *ferB* (BES08_30695), rendelkezik jelentős, SA-függő kifejeződési szinttel a sejtek exponenciális növekedési fázisában (SA_E vs. Glc_E). Érdekes módon az SA bontási útvonal perifériás szakaszának többi eleméhez hasonlóan, a *ferB* gén már a Glc-on növesztett, exponenciális szakaszában lévő sejtek esetén is magas kifejeződési szinttel rendelkezik, ami tovább növekszik a sejtek stacioner fázisba történő átmenete során (Glc_S vs. Glc_E). A többi lehetséges flavoprotein reduktáz génje viszonylag alacsony és változatlan kifejeződési szinttel rendelkezett az összes vizsgált mintában.

A *ferB* gén az *N. resinovorum* SA1 törzs pSA3-as plazmidján található, megközelítőleg 3 kbp-nyi távolságra a *sadAB* génektől 3' irányban. A pSA3 plazmidon nem önállóan helyezkedik el, hanem közös génklasztert alkot egy Fer2 (PF00111) pfam doménnel rendelkező növényi típusú ferredoxin fehérjét kódoló génnel, a *ferA*-val (BES08_30690). Hasonló pfam doménnel rendelkező ferredoxin fehérjét kódoló gén található a *H. intermedia* PBC törzs, *sad* gén klaszterében, azonban ebben az esetben a ferredoxin reduktáz génje nem képezi a klaszter részét ²⁶. Más esetekben a filogenetikai vizsgálatok azt mutatták (16. ábra.), hogy a *Comamonadaceae*

család egyes tagjainál a *sadAB* és *ferAB* szerű gének egy génklaszterbe is rendeződhetnek, ami a funkciójuk közötti kapcsoltságot tovább erősíti.

6.7.2 Szulfanilsav transzporter

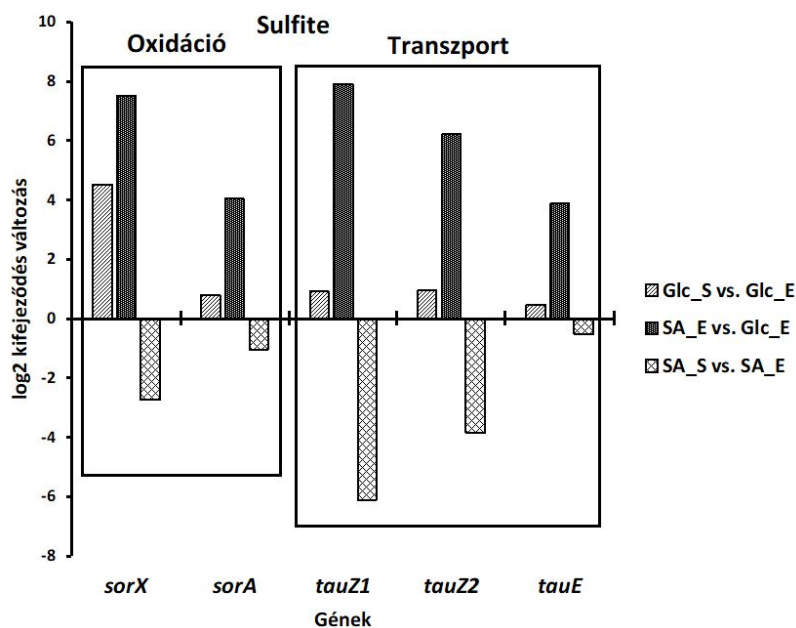
Az SA egy poláris, ikerionos molekula. Sok aromás vegyülettel ellentétben a sejtek passzív diffúzió útján nem képesek felvenni ⁶¹, csak valamilyen aktív transzport rendszer használatával. Az SA felvételét lehetővé tevő transzport rendszert még nem sikerült azonosítani. A KEGG adatbázis BRITE funkcionális osztályozása alapján az *N. resinovorum* SA1 törzs 360 olyan fehérje génjét tartalmazza, amely feltételezhetően valamilyen transzport funkcióban játszik szerepet („Transporters”, nre02000). Ezek közül csak 13 gén mutatott megnövekedett kifejeződést SA hatására. Közülük 12-en valamilyen ion szállításában vesz részt, elsősorban a vas szállításában (Függelék 6). Az egyik ilyen lehetséges vas transzporter génje (BES08_23675) határozottan erős SA függő kifejeződési szinttel rendelkezik, ami nem meglepő mivel a SA bontási útvonal számos vas-tartalmú fehérjét tartalmaz, amelyek emelkedett szintje alapján megnövekedett vas igényre számíthatunk. Azonban az azonosított 13 gén között egy olyan (BES08_30350) is található, amely terméke az MFS (major facilitator superfamily) szupercsalád része, azon belül is AAHS (aromatic acid/H⁺ symporter) család tagja. Ez a transzporter család azért különösen érdekes, mert számos eleméről bizonyították már, hogy szerepet játszanak az aromás savak sejtmembránon keresztül történő átjutásában. Az AAHS család tagjaira jellemző az MFS_1 (PF07690) pfam domén megléte. Az *N. resinovorum* SA1 67 olyan gént tartalmaz, amely terméke rendelkezik ilyen doménnel. Ezek közül csak a BES08_30350 mutatott SA függő kifejeződést.

6.7.3 A szulfít sorsa a sejtben

Az SA tápanyagforrásként történő hasznosítása során elkerülhetetlen, hogy nagy mennyiségű szulfít képződjön a sejtben, amely toxikus a sejt számára. Emiatt a keletkező szulfítot a sejtnek át kell alakítani kevésbé toxikus kénvegyületekké. Az átalakítás történhet a szulfít redukcióján keresztül, amely kén esetleg szulfid képződéséhez vezet, vagy oxidációján keresztül, amely szulfát képződéséhez vezet. Az *N. resinovorum* SA1 az SA-ból felszabaduló szulfítot csak kis részét használja fel anyagcsere folyamataihoz, a fel nem használt szulfítot szulfát formájában jelenik meg a tápoldatban ³⁴. Ez alapján arra lehet következtetni, hogy a detoxifikáció fő módja a szulfít oxidációja szulfáttá, ami egyébként a legelterjedtebb módszer a baktériumok között ¹²⁸.

A baktériumokban a szulfid oxidációja történhet a citoplazmában, indirekt módon, APS képződésén keresztül (indirekt útvonal), vagy a periplazmatikus térben a szulfid oxidáló enzimek (SOE) által (direkt útvonal). A két lehetséges útvonal közül a direkt útvonal a legelterjedtebb és legismertebb a baktériumok között. Munkahipotézisünk szerint az SA bontása során felszabaduló szulfid detoxifikációjában is ez az útvonal játssza a fő szerepet, ezért azonosítottam az útvonal elemeit és meghatároztam a génjeinek kifejeződésében tapasztalható változásokat a vizsgált körülmények között.

A SOE enzimek egy jellegzetes Oxidored_molyb (PF00174) pfam doménnel rendelkeznek. Két olyan gén (BES08_29795, BES08_29830) található az *N. resinovorum* SA1 genomjában, amely terméke rendelkezett ilyen típusú doménnel. Az összehasonlító transzkriptom vizsgálat alapján mindkét gén esetében SA-függő kifejeződés mutatható ki (SA_E vs. Glc_E). Továbbá az SA-n növesztett, stacioner növekedési fázisában lévő sejtek esetén a csökkenő szulfid képződés miatt a gének kifejeződésének csökkenése is megfigyelhető (SA_S vs. SA_E) (28. ábra.).



28. ábra. A lehetséges szulfid oxidáló enzimek és szulfid transzporterek génjeinek kifejeződési szintjeinek változása.

Genetikai elhelyezkedése alapján mindkét gén az *N. resinovorum* SA1 törzs pSA3 plazmidján található meg, viszonylag kis távolságra egymástól. *In silico* vizsgálatok kimutatták, hogy mindkét gén által kódolt fehérje rendelkezik a SO enzimekre jellemző Oxidored_molyb (PF00174) domén mellett egy Mo-co_dimer (PF03404) doménnel is¹²⁸, amely alapján az SOE II-es csoportjába, azon belül is a 2C csoportjába sorolhatóak be. A BES08_29830 által kódolt

fehérjeszekvenciája viszonylag nagy hasonlóságot (megközelítőleg 60%) mutat a SorA fehérjével, amely a 2C csoport egyik legjobban jellemzett tagja ⁶⁴. A hasonlóságot növeli, hogy az *N. resinovorum* SA1 törzsben található SorA-szerű fehérjéje is rendelkezik egy TAT szignállal (Twin-arginine translocation pathway signal, IPR006311), amely lehetővé teszi a fehérje számára a TAT útvonalon keresztül történő kijutást a periplazmatikus térbe. Ezenfelül a SorA fehérje génje egy klaszterbe rendeződik egy IPR009056 InterPro doménnel rendelkező citokróm c-szerű fehérjét kódoló génnel (BES08_29825), amely doménje alapján akár a SorA SOE hem-tartalmazó alegységének (SorB) felelhet meg.

A másik lehetséges SOE kódoló gén (BES08_29795) termékének a besorolása a többi SOE-hoz mutatott alacsony szekvencia hasonlósága miatt nem volt lehetséges. Emiatt egyszerűen SorX-nek neveztem el.

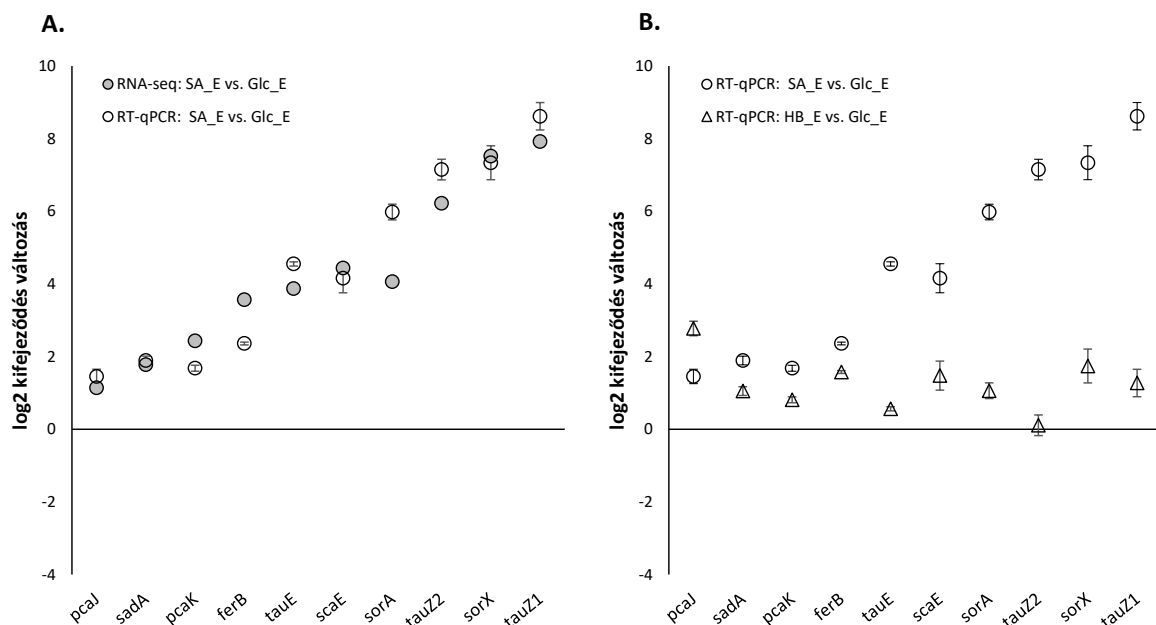
Mivel a SOE enzimek a periplazmatikus térben helyezkednek el, ezért a detoxifikáció első lépésében a szulfidot a sejtnek speciális transzportereken keresztül ide kell juttatnia. A szulfit transzportereket pfam doménjük alapján két csoportba lehet sorolni. Az első csoportra, amelybe a TauZ/CuyZ/SuyZ transzporterek is tartoznak ⁶⁹⁻⁷¹, a Cons_hypoth698 (PF03601) pfam domén megléte a jellemző, míg a második csoportra, ahova a TauE fehérje tartozik, a TauE (PF01925) pfam domén a jellemző ⁷³.

Az *N. resinovorum* SA1 törzsben 13 lehetséges szulfit transzportert kódoló gént sikerült azonosítani. Közülük 4 gén terméke rendelkezik a PF03601, míg 9 a PF01925 doménnel. Ezek közül három lehetséges transzportert kódoló gén az SA_E mintában a Glc_E mintához képest megnövekedett kifejeződési szinttel rendelkezik (26. ábra.). Mindegyik esetben a megnövekedett kifejeződési szint jelentősen visszaesett a stationer fázisban (SA_S vs. SA_E). A három, azonosított gén közül kettő, a BES08_29790 és a BES08_30485 által kódolt fehérje rendelkezik PF03601 doménnel és mutat viszonylag nagy (46.4 és 79.5%) hasonlóságot a TauZ fehérjékkel. A harmadik gén (BES08_30490) által kódolt transzporter PF01925 doménnel rendelkezik, ami alapján feltételezhetően egy TauE típusú fehérje.

Mind a három SA által indukált transzporter génje az *N. resinovorum* SA1 törzs pSA3 plazmidján található meg, akárcsak a SOE enzimek génjei. Továbbá az egyik lehetséges transzportert kódoló gén BES08_29790 (TauZ1) az egyik SOE génjének BES08_29795 (SorX) szomszédságában található, míg a másik két lehetséges transzporter génje, a BES08_30485 (TauZ2) és a BES08_30490 (TauE) közvetlenül a centrális útvonal génjeit kódoló *sca* operon előtt helyezkednek el.

6.8 Az RNA-seq eredményeinek alátámasztása RT-qPCR használatával

A korábban bemutatott RNA-seq vizsgálatból származó eredmények megerősítésének céljából RT-qPCR vizsgálatot végeztem. A vizsgálatához exponenciális növekedési fázisban lévő, Glc-n, SA-n és HB-n növesztett sejteket használtam (27. ábra.). A vizsgálat során a következő gének kifejeződését néztem: *sadA*, az SA bontás perifériás szakaszában szerepet játszó oxigenáz génje; *scaE* és az SA centrális szakaszában részt vevő gyűrű-hasító dioxigenáz alegységének a génjei; *pcaJ*, a bontás közös szakasza során, a 3-oxoadipát 3-oxoadipát-CoA átalakításában szerepet játszó enzim génje; *ferB* ferredoxin reduktázt kódoló gén (BES08_30695); *pcaK*, az SA által indukált transzporter génje (BES08_30350); továbbá a szulfid oxidálásában, szállításában és szerepet játszó enzimek génjei (*sorA*, *sorX*, *tauZ1*, *tauZ2*, *tauE*).



29. ábra. A.) Az RT-qPCR és RNA-seq módszerével kapott SA hatására (SA_E vs. Glc_E) bekövetkező génkifejeződés változás összehasonlítása. B.) A kiválasztott gének génkifejeződésében SA (SA_E vs. Glc_E) és HB (HB_E vs. Glc_E) hatására kialakuló változások összehasonlítása RT-qPCR segítségével. Az ábrán az RT-qPCR vizsgálatokból származó kifejeződési szint változását világos kör/háromszög, az RNA-seq vizsgálatból származó kifejeződési szint változását sötét kör jelöli. A hibaszávok a három biológiai párhuzamos esetén tapasztalt szórást jelölik.

A vizsgált gének SA_E vs. Glc_E esetén mutatott kifejeződés változását a 29A. ábra szemlélteti. Az RT-qPCR és az RNA-seq által kapott értékek nem térnek el lényegesen

egymástól közöttük magas korreláció mutatható ki ($r = 0.94$, Pearson). Ez alapján az látszik, hogy az RT-qPCR megerősíti a teljes transzkriptom vizsgálat során megfigyelteket.

Hogy elkülönítsük a SA specifikus génkifejeződést az általános aromás savakra adott választól kontrollkén elvégeztük a HB_E vs. Glc_E összehasonlítást is (29B. ábra). Ebből az látszik, hogy a SA hatására magas kifejeződést mutató gének (centrális útvonal elemei, kisegítő útvoanal elemei) HB esetén is megőrzik SA függő kifejeződésüket. Az SA hatására kevésbé magas kifejeződést mutató gének is nagyobb mértékben fejeződnek ki SA hatására. Ez alól kivételt képez a közös útvonal génje (*pcaJ*), amely HB-on növesztett sejtek esetén az SA-hoz képest kicsivel nagyobb kifejeződést képes elérni.

7 Összefoglalás

A mesterségesen előállított szubsztituált aromás szerves molekulák egyik széles körben használt vegyülete a szulfanilsav (SA). Olyan kémiai módosításokat tartalmaz, amelyek alkalmassá teszik arra, hogy az ipar számos területe felhasználhassa fontos termékek alapanyagaként, mint például a szulfonamid gyógyszerek és különböző festékek. A szulfanilsavat kémiai természete miatt a sejtek nehezen veszik fel, de ha bejutott, akkor toxikus hatású: a folsav bioszintézis kompetitív inhibitora.

Bizonyos esetekben azonban kialakulhatnak olyan speciálisan alkalmazkodott mikroorganizmus törzsek, amelyek még ezeket a természetidegen vegyületeket is képesek bontani és felhasználni anyagcsere folyamataikhoz. Egy ilyen mikroorganizmus törzs az *N. resinovorum* SA1 is, amely az első olyan izolátum, mely képes más baktériumok segítségével az SA-t egyedüli szén-, nitrogén-, kén- és energiaforrásként felhasználni.

Munkám során céлом volt minél részletesebben feltárni azokat a metabolikus/biokémiai rendszereket, amelyek lehetővé teszik az N. resinovorum SA1 törzs számára hogy képes legyen az SA-t, mint egyedüli szén-, nitrogén-, kén- és energiaforrást felhasználni

Eredményeim az alábbi pontokban foglalhatóak össze.

- Illumina MiSeq második generációs szekvenáló rendszer használatával meghatároztam az *N. resinovorum* SA1 genomjának szekvenciáját.
- A nyers szekvenciákból a genom teljes összeillesztése egy kromoszómát és négy plazmidot (pSA1-4) eredményezett. A plazmidok közül három is jelentős, 100 kbp is nagyobb mérettel rendelkezik.
- A teljes genomot annotáltam és a géneket több adatbázis (Pfam, KEGG és COG) segítségével funkcionális csoportokba osztottam.
- Az *N. resinovorum* SA1 genomja három párhuzamos β -ketoadipát útvonalat tartalmaz, amelyek szerepet játszanak a törzsre jellemző szulfanilsav (SA), benzoát (B) és hidroxibenzoát (HB) bontó képesség kialakításában.
- A B, HB és a β -ketoadipát útvonalak közös szakasza a törzs legnagyobb, a pSA1-es plazmidján található meg, míg az SA katabolikus útvonal elemei a pSA3-as plazmidon szóródnak szét.
- A korábban azonosított SA-bontó útvonalakat hasonló gének kódolnak, amelyek három, különálló génklaszterbe rendeződnek. Az útvonal perifériás szakasza egy 4-aminobenzol-szulfonát-3,4-dioxigenáz (SadA) és egy glutamin-szintetáz (SadB)

enzimekből épül fel. Ezek génjeitől fajonként eltérő távolságban helyezkedik el az útvonal centrális génjeinek genomi régiója, amely 3-szulfo-cisz-cisz-mukonát cikloizomerázt (ScaA), szulfolakton hidrolázt (ScaB), maleilacetát reduktázt (ScaC) és a 3,4-szulfokatekol dioxigenáz két alegységét (ScaFE) kódoló génekből épül fel. A teljes katabolizmus közös szakaszának génjei az 3-oxoadipát CoA-transzferáz (PcaIJ) két alegységét és a 3-oxoadipil CoA-tioláz (PcaF) enzimek génjeit foglalja magába.

- Amíg az *N. resinovorum* SA1 törzs SadA és SadB fehérjéi 54% és 42%-os azonosságot mutatnak a *Comamonadaceae* családba tartozó *Hydrogenophaga* fajokban található ortológ párjaikkal, addig az ScaA-ScaF fehérjék legnagyobb mértékben az *Bradyrhizobiaceae* családba tartozó *A. resinovorum* S2-ben található ortológjaikhoz hasonlítanak, 56% és 68%, közötti azonossági szinttel.
- Az NCBI adatbázisán végzett TBLASTN keresés alapján az *N. resinovorum* SA1 törzs SadA és SadB fehérjéi legnagyobb hasonlóságot a *Bradyrhizobiaceae* család tagjaiban előforduló homológjaikkal mutatnak.
- A SadA és SadB fehérjék ismételt TBLASTN keresése *Comamonadaceae*, *Sphingonadaceae* és *Bradyrhizobiaceae* család tagjaiban azt mutatta, hogy az SA bontás kulcsenzimeikhez hasonló fehérjék széles körben előfordulnak a *Comamonadaceae* és a *Bradyrhizobiaceae* család törzseiben, viszont a *Sphingomonadaceae* család tagjai esetén nem általánosak. Ez természetesen nem jelenti azt, hogy számos *Comamonadaceae* és a *Bradyrhizobiaceae* törzs képes lenne az SA bontására, de azt igen, hogy hordoznak ilyen típusú gyűrűhidroxiláló dioxigenázt illetve ehhez kapcsolódóan aminotranszferázt/glutamin-szintetázt.
- A centrális útvonal elemei lényegesen ritkábban fordulnak elő a baktériumok között, csak azokban a baktérium fajokban voltak megtalálhatóak, amelyek képesek a szulfokatekol (SC) bontására.
- Az *N. resinovorum* SA1 törzs génjein végzett korrespondencia analízis azt mutatta, hogy az SA bontás útvonal génjeinek kodonpreferenciája lényegesen eltér a sejt többi génjeitől. Ez különösen igaz a perifériás útvonal génjeire.
- A kodonpreferencia valamint a gének szerveződésének, genomi kontextusának analízise alapján arra a következtetésre jutottam, hogy az SA bontás útvonala egy nem olyan régen bekövetkezett horizontális géntranszfer eredményeképpen alakulhatott ki az *N. resinovorum* SA1 törzs esetén.

- Az SA bontás perifériás útvonalának eredeti funkciója feltételezhetően valamilyen természetben előforduló amino-csoportot tartalmazó aromás vegyület bontásában lehetett. Ezt erősíti meg a SadB fehérje filogenetikai vizsgálata is, amely alapján egy csoportba sorolható olyan glutamin-szintetáz-szerű fehérjékkel, amelyek az anilin bontásában játszanak szerepet.
- Promóter próba vektor (pPROBE) segítségével sikerült feltérképezni a perifériás és a centrális útvonal kifejeződésért felelős genomi régió cisz elemeit. Ez a régió a perifériás útvonal esetén megközelítőleg 276 bp-nyira található az operon első elemének translációs start pontjától. A centrális útvonal esetén ez a távolság lényegesen rövidebb, megközelítőleg 50 bp.
- A perifériás és a centrális útvonal promótereinek vizsgálata alapján az látszik, hogy a két útvonalszakasz eltérő szabályozási logikát követ. Amíg a perifériás útvonal viszonylag magas alap kifejeződési szinttel rendelkezik, amely éhezés hatására tovább növekszik, addig a centrális útvonal minimális alap kifejeződési szintje az SA hatására jelentősen megugrik.
- A centrális útvonal szabályozását tovább színesíti a karbon katabolit represszió hatása, amely megakadályozza az operon SA-függő kifejeződését könnyen hasznosítható szénforrás jelenlétében.
- Hogy pontosabb képet kapjunk az SA bontási útvonalának szabályozásáról, illetve megismerjük a bontást lehetővé tevő kisegítő rendszerek elemeit RNA-seq technikán alapuló teljes transzkriptom vizsgálatot végeztünk exponenciális és stacioner növekedési fázisban lévő glükózon (Glc) és szulfanilsavon (SA) növesztett sejtek esetén. A stacioner fázisban mind a Glc mind az SA szubsztrát elfogyott így ezek a sejtek tápanyaghiányos/éheztető környezetben voltak.
- Legtöbb gén kifejeződése az exponenciális növekedési fázisból a stacioner fázisba történő átmenet során változott. A glükózon (Glc) növesztett sejtek génjeinek 17,7%-ának, míg a szulfanilsavon (SA) növesztett sejtek génjeinek 13,5%-ának az expressziója változott valamilyen irányba. Az eltérő szénforráson növesztett, exponenciális fázisban lévő sejtek esetén kisebb, 7,3%-os változás volt megfigyelhető
- Az elvégzett géncsoport túlreprezentációs elemzés (GSOA) és a géncsoport dúsulási vizsgálat (GSEA) alapján az exponenciális fázisból a stacioner fázisba történő váltás során szénforrástól függetlenül, nagyrészt megegyező funkciójú csoportokban volt megfigyelhető szignifikáns változás. A változás minden esetben csökkenő volt és

tükrözte az éhezés sejtre gyakorolt hatását. Az eltérő szénforráson növesztett, exponenciális fázisban lévő sejtek esetén is inkább az éhezésre utaló jelek voltak megfigyelhetők. Érdekes módon az éhezésen kívül egyéb stresszválasz nem volt megfigyelhető.

- Az SA bontásban résztvevő gének kifejeződése a promóter vizsgálatok esetében tapasztaltakat tükrözte. A perifériás útvonal génjei magas alap kifejeződéssel rendelkeznek, amely az éhezés hatására tovább növekszik. Ezzel szemben a centrális útvonal génjeinek szabályozása SA-függő. A közös útvonal szakasz pedig érdekes módon nem reagált se a szulfanilsav (SA), se az esetleges éhezés hatására.
- Sikerült azonosítani a centrális útvonal lehetséges elektrontranszport láncát, amely egy glutation-reduktáz (GR) és egy növényi típusú ferredoxin (Fer2) fehérjéből épül fel.
- A transzkriptom analízis alapján az SA felvételéért egy, az MFS szupercsaládba tartozó, transzporter felelős.
- Az SA-t tápanyagforrásként használó sejt számára fontos a bontás során felszabaduló szulfid elleni hatékony védekezés. Az *N. resinovorum* SA1 törzs esetében több olyan enzim génjét sikerült azonosítani, amelyek ebben játszhatnak szerepet. Ezek közül három szulfid transzporter, amelyek feladata a keletkező szulfidnak a sejt perifériás terébe történő juttatása, kettő pedig szulfid oxidáló enzim amelyek a periplazmatikus térben jutott szulfidot szulfidáttá történő átalakítását végzik.
- Az RNA-seq vizsgálat során azonosított fontosabb gének kifejeződését RT-qPCR használatával is igazoltam.
- Átfogó képet mutattam be arról hogy egy széles körben használt xenobiotikum az *N. resinovorum* SA1 törzsben milyen metabolikus változásokat okoz.

8 Summary

Sulfanilic acid (SA) is a widely used synthetic sulfonated aromatic organic compound. It contains specific chemical modifications allowing its wide industrial applications as starting material for synthesis of important chemicals, such as sulfa drugs and various dyes. Owing to the chemical structure of SA, its cellular uptake is difficult, but once it gets into the cells, it has toxic effects: competitively inhibits the folic acid biosynthesis.

In certain cases, organisms capable of degradation of such xenobiotics can be evolved and they can use these toxic compounds for their metabolic processes. A bacterium, the *N. resinovorum* SA1 strain, was the first isolate, which could use the sulfanilic acid (SA) as only carbon, nitrogen, sulfur and energy source without the assistance of any other organism.

During my work, my main purpose was to disclose those complex metabolic/biochemical systems, which endow N. resinovorum SA1 strain to use SA as only carbon, nitrogen, sulfur and energy source.

My results can be summarized in the following points.

- The *N. resinovorum* SA1 strain genome was sequenced by means of Illumina MiSeq second-generation sequencing platform.
- From the raw sequences, the genome assembly resulted in a chromosome and four plasmids. Out of these, three plasmids have remarkable, more than 100 kbp size.
- The full genome was annotated using three databases (Pfam, KEGG és COG) and the annotated genes were divided into various functional groups.
- The *N. resinovorum* SA1 genome contains the genes of three parallel β -ketoacid pathways, which play roles in degradation of sulfanilic acid (SA), benzoate (B) and hydroxy-benzoate (HB).
- The genes of the “B”, “HB” and the common pathway elements of all three β -ketoacid pathways are located on the largest pSA1 plasmid while the genes of the specific SA catabolite pathway are scattered on the pSA3 plasmid.
- So far, the genes encoding similar proteins involved in SA degradation are located in three, separate gene clusters. The peripheral part of the pathway consists of the genes coding for a 4-aminobenzenesulfonate-3,4-dioxygenase (SadA) and a glutamine-synthetase (SadB). The central part of the pathway encoded by the genes

of 3-sulfo-cis-cis-muconate cycloisomerase (ScaA), sulfolactone hydrolase (ScaB), maleylacetate reductase (ScaC) and the two subunit of the 3,4-sulfocatechol dioxygenase (ScaFE). The genomic distances of the genes of the peripheral and central pathways change by species. Moreover, the common part of the pathways consists of two subunits of the 3-oxoadipate CoA-transferase (PcaIJ) and the 3-oxoadipyl CoA thiolase (PcaF).

- While, the SadA and SadB proteins of *N. resinovorum* SA1 strains are more similar to their orthologues in *Hydrogenaphaga* strains belonging to *Comamonadaceae* family (a 54% and 42% similarity level), the ScaABCEF enzymes are more similar to their orthologues of *A. resinovorum* S2 strain belonging to *Bradyrhizobiaceae* family (similarity levels are between 56% and 68%).
- According to the TBLASTN search in the NCBI databases, the SadA and SadB proteins of the *N. resinovorum* SA1 strain show the highest similarity to their homologs in the members of the *Bradyrhizobiaceae* family.
- Repeated TBLASTN search using SadA and SadB proteins as query sequences in the sequences databases of *Comamonadaceae*, *Bradyrhizobiaceae* and *Sphingomonadaceae* families showed that appearance of proteins similar to the key peripheral enzymes of the SA degradation is common in the *Comamonadaceae* and *Bradyrhizobiaceae* strains and rarely occurring in the *Sphingomonadaceae* family. Of course, this does not mean that many strains of the *Comamonadaceae* and *Bradyrhizobiaceae* families are able to degrade SA, but it indicates that they possess such a ring hydroxylating dioxygenases and aminotransferases/glutamine synthetases.
- The components of the central part of the SA degradation pathway are less common among bacteria. They were identified exclusively in those bacteria, which were capable of SC degradation.
- The correspondence analysis performed on the *N. resinovorum* SA1 genes showed that the codon preference genes involved in SA degradation substantially differs from that of the other – e.g. housekeeping - genes. This is especially true for the genes of the peripheral pathway.
- Based on the codon preference analysis, gene organization/genomic context, I concluded that emerge of the SA degradation pathway is a result of a recent horizontal gene transfer event.

- Presumably, the peripheral part of the SA pathway originally played a role in the degradation of a naturally occurring aminated aromatic compound. It is strengthened by the phylogenetic analysis of the SadB protein, which shows that it clusters together with other glutamine synthetase like proteins, which presumably take part in the aniline degradation.
- By means of modified promoter probe vectors (pPROBE), it was possible to map the genomic regions responsible for the expression of the peripheral and central part of the pathway. In the case of the peripheral pathway, this region is located 276 bp upstream of the translation start codon of the first gene of the operon. In the case of the central pathway, this distance was significantly shorter, approximately 50 bp.
- According to the analysis of the promoter regions, it seems that, the regulation of peripheral and central part of the pathway substantially differs. While the peripheral genes have high basal expression, which further increase during starvation, the genes of the central part have a basal expression levels which are really SA inducible.
- The carbon catabolite repression influences the regulation of the central part of the pathway by blocking the SA induced expression in the presence of an easily consumable carbon source.
- In order to gain insight into the regulation of the SA degradation pathway and to identify accessory elements assisting for the degradation processes, we performed RNA-seq based whole-transcriptome analyses of cells grown on glucose (Glc) and sulfanilic acid (SA) sampled in the exponential and stationary phase. In the stationary phase, both glucose (Glc) and sulfanilic acid (SA) were completely consumed from the media thus, cells were in carbon source limited/starving conditions.
- The largest changes in gene expression could be observed during the transition from the exponential to stationary growth phase. While the expression of 17,6% of the genes of the cells grown on glucose (Glc) changed, in the case of cells grown on SA, this percentage is 13,2%. The gene expression levels in cells grown on various carbon sources in the exponential phase changed in 7,2% of the genes, only.
- Gene set overrepresentation (GSOA) and gene set enrichment analyses (GSEA) showed that during the transition from the exponential to stationary growth phase the differentially expressed genes have enriched mostly in similar functional groups

independently of the carbon source. In all cases, just downregulation could be observed which reflected the starvation effect on the cell. Similar starvation-like effect could be seen in cells being in similar exponential phase but grown on different carbon source. Interestingly, beside the starvation, other stress response was not detectable.

- The transcript profiles of the genes taking part in the SA degradation supported the conclusions of the promoter mapping experiments. The genes of the peripheral part had a high basal expressions, which further increased as a result of starvation. In contrast, the regulation of the genes of the central part was clearly SA dependent. Interestingly, the gene expression level of the common part of the pathway did not respond to the presence of SA or to starvation.
- It was possible to identify the elements of the electron transport chain linked to the peripheral pathway. This was composed of a glutathione reductase (GR) and plant type ferredoxin (Fer2) coding gene.
- Based on the transcriptome analysis, I identified a gene which encodes a transporter belonging to the MFS superfamily and likely responsible for the SA uptake.
- For cells, using SA as sole nutrients, an effective protection against sulfite released during the degradation is important. In the case of the *N. resinovorum* SA1 strain, multiple genes were identified with encode proteins likely involved in this function. Three of them are sulfite transporters participating in sulfite transport into the periplasmic space and two of them are sulfite oxidizing enzymes which catalyze the oxidation of sulfite to sulfate basically in the periplasmic space.
- The expression levels of the most important genes identified with RNA-seq analyses were further confirmed by means of RT-qPCR analyses.
- We presented a comprehensive picture about the metabolic changes induced by a widely used xenobiotic compound in *N. resinovorum* SA1.

9 Hivatkozások jegyzéke

1. Bugg TDH, Ahmad M, Hardiman EM, Rahmanpour R. Pathways for degradation of lignin in bacteria and fungi. *Nat Prod Rep.* 2011;28(12):1883.
2. Mitra S, Roy P. BTEX : A Serious Ground-water Contaminant. *Res J Environ Sci.* 2011;5(5):394-398.
3. Othmer K. BTX Processing. *Kirk-Othmer Encycl Chem Technol.* 2007;(X):1-17.
4. Morrison RT, Boyd RN. *Organic Chemistry.* Prentice Hall; 1992.
5. Lindner O, Rodefeld L. *Benzenesulfonic Acide and Their Derivatives.* 5th edn.; 1985.
6. Ben Mansour H, Corroler D, Barillier D, Ghedira K, Chekir L, Mosrati R. Evaluation of genotoxicity and pro-oxidant effect of the azo dyes: acids yellow 17, violet 7 and orange 52, and of their degradation products by *Pseudomonas putida* mt-2. *Food Chem Toxicol.* 2007;45(9):1670-1677.
7. Stolz A. Basic and applied aspects in the microbial degradation of azo dyes. *Appl Microbiol Biotechnol.* 2001;56(1-2):69-80.
8. Yemashova N, Kalyuzhnyi S. Microbial conversion of selected azo dyes and their breakdown products. *Water Sci Technol.* 2006;53(11):163-171.
9. Tan NCG, van Leeuwen A, van Voorthuizen EM, et al. Fate and biodegradability of sulfonated aromatic amines. *Biodegradation.* 2005;16(6):527-537.
10. Khan R, Bhawana P, Fulekar MH. Microbial decolorization and degradation of synthetic dyes: A review. *Rev Environ Sci Biotechnol.* 2013;12(1):75-97.
11. Hoogkamp-Korstanje JAA. In-vitro activities of ciprofloxacin, levofloxacin, lomefloxacin, ofloxacin, pefloxacin, sparfloxacin and trovafloxacin against Gram-positive and Gram-negative pathogens from respiratory tract infections. *J Antimicrob Chemother.* 1997;40(3):427-431.
12. Hruska K, Franek M. Sulfonamides in the environment: A review and a case report. *Vet Med (Praha).* 2012;57(1):1-35.
13. Boreen AL, Arnold WA, McNeill K. Photochemical fate of sulfa drugs in then aquatic environment: Sulfa drugs containing five-membered heterocyclic groups. *Environ Sci Technol.* 2004;38(14):3933-3940.
14. Heider J, Fuchs G. Anaerobic metabolism of aromatic compounds. *Eur J Biochem.* 1997;243(3):577-596.
15. Thurnheer T, Kohler T, Cook AM, Leisinger T. Orthoanilic acid and analogues as carbon

- sources for bacteria: growth physiology and enzymic desulphonation. *Microbiology*. 1986;132(5):1215.
16. Pitter P. Determination of biological degradability of organic substances. *Water Res.* 1976;10(3):231-235.
 17. Feigel BJ, Knackmuss H-J. Bacterial catabolism of sulfanilic acid via catechol-4-sulfonic acid. *FEMS Microbiol Lett.* 1988;55(1):113-117.
 18. Gan HM, Shahir S, Ibrahim Z, Yahya A. Biodegradation of 4-aminobenzenesulfonate by *Ralstonia* sp. PBA and *Hydrogenophaga* sp. PBC isolated from textile wastewater treatment plant. *Chemosphere*. 2011;82(4):507-513.
 19. Perei K, Rákhely G, Kiss I, Polyák B, Kovács KL. Biodegradation of sulfanilic acid by *Pseudomonas paucimobilis*. *Appl Microbiol Biotechnol.* 2001;55(1):101-107.
 20. Singh P, Birkeland NK, Iyengar L, Gurunath R. Mineralization of 4-aminobenzenesulfonate (4-ABS) by *Agrobacterium* sp. strain PNS-1. *Biodegradation*. 2006;17(6):495-502.
 21. Wang YQ, Zhang JS, Zhou JT, Zhang ZP. Biodegradation of 4-aminobenzenesulfonate by a novel *Pannonibacter* sp. W1 isolated from activated sludge. *J Hazard Mater.* 2009;169(1-3):1163-1167.
 22. Dangmann E, Stolz A, Kuhm AE, et al. Degradation of 4-aminobenzene sulfonate by a two-species bacterial coculture: Physiological interactions between *Hydrogenophaga palleronii* S1 and *Agrobacterium radiobacter* S2. *Biodegradation*. 1996:223-229.
 23. Kim K, Gan HM. A glimpse into the genetic basis of symbiosis between *Hydrogenophaga* and their helper strains in the biodegradation of 4-aminobenzen. *J Genomics*. 2017;5(July):77-82.
 24. Feigel BJ, Knackmuss HJ. Syntrophic interactions during degradation of 4-aminobenzenesulfonic acid by a 2 species bacterial culture. *Arch Microbiol*. 1993;159:124-130.
 25. Gan HM, Ibrahim Z, Shahir S, Yahya A. Identification of genes involved in the 4-aminobenzenesulfonate degradation pathway of *Hydrogenophaga* sp. PBC via transposon mutagenesis. *FEMS Microbiol Lett.* 2011;318(2):108-114.
 26. Gan HM, Shahir S, Yahya A. Cloning and functional analysis of the genes coding for 4-aminobenzenesulfonate 3,4-dioxygenase from *Hydrogenophaga* sp. PBC. *Microbiology*. 2012;158(Pt 8):1933-1941.
 27. Fukumori F, Saint CP. Nucleotide sequences and regulational analysis of genes involved

- in conversion of aniline to catechol in *Pseudomonas putida* UCC22(pTDN1). *J Bacteriol.* 1997;179(2):399-408.
28. Takeo M, Ohara A, Sakae S, et al. Mechanism of Bacterial Aniline Oxidation via γ -Glutamylanilide: Function of a Glutamine Synthetase-Like Protein. *J Bacteriol.* July 2013.
 29. Walker MC, van der Donk WA. The many roles of glutamate in metabolism. *J Ind Microbiol Biotechnol.* 2016;43(2-3):419-430.
 30. Hammer A, Stolz A, Knackmuss HJ. Purification and characterization of a novel type of protocatechuate 3,4-dioxygenase with the ability to oxidize 4-sulfocatechol. *Arch Microbiol.* 1996;166(2):92-100.
 31. Contzen M, Stolz A. Characterization of the Genes for Two Protocatechuate 3,4-Dioxygenases from the 4-Sulfocatechol-Degrading Bacterium *Agrobacterium radiobacter* Strain S2. *J Bacteriol.* 2000;182(21):6123-6129.
 32. Contzen M, Bürger S, Stolz A. Cloning of the genes for a 4-sulphocatechol-oxidizing protocatechuate 3,4-dioxygenase from *Hydrogenophaga intermedia* S1 and identification of the amino acid residues responsible for the ability to convert 4-sulphocatechol. *Mol Microbiol.* 2001;41(1):199-205.
 33. Magony M, Perei K, Medzihradzky K, Kovács K, Rákhely G. Szulfonált aromás vegyületek lebontásának molekuláris mechanizmusa. *Biokémia.* 2004;(28):26–36.
 34. Magony M, Kákonyi I, Gara A, et al. Overlaps between the various biodegradation pathways in *Sphingomonas subarctica* SA1. *Acta Biol Hung.* 2007;58 Suppl:37-49.
 35. Halak S, Basta T, Bürger S, et al. 4-Sulfomuconolactone Hydrolases from *Hydrogenophaga intermedia* S1 and *Agrobacterium radiobacter* S2. *J Bacteriol.* 2007;189(19):6998-7006.
 36. Halak S, Basta T, Bürger S, Contzen M, Stolz A. Characterization of the genes encoding the 3-carboxy-cis,cis-muconate-lactonizing enzymes from the 4-sulfocatechol degradative pathways of *Hydrogenophaga intermedia* S1 and *Agrobacterium radiobacter* S2. *Microbiology.* 2006;152(11):3207-3216.
 37. Halak S, Lehtiö L, Basta T, et al. Structure and function of the 3-carboxy-cis,cis-muconate lactonizing enzyme from the protocatechuate degradative pathway of *Agrobacterium radiobacter* S2. *FEBS J.* 2006;273(22):5169-5182.
 38. Bugg TDH, Ramaswamy S. Non-heme iron-dependent dioxygenases: unravelling catalytic mechanisms for complex enzymatic oxidations. *Curr Opin Chem Biol.*

- 2008;12(2):134-140.
39. Butler CS, Mason JR. *Structure-Function Analysis of the Bacterial Aromatic Ring-Hydroxylating Dioxygenases*. Vol 38. Academic Press; 1997.
 40. Jiang H, Parales RE, Lynch NA, Gibson DT. Site-directed mutagenesis of conserved amino acids in the alpha subunit of toluene dioxygenase: Potential mononuclear non-heme iron coordination sites. *J Bacteriol*. 1996;178(11):3133-3139.
 41. Senda M, Kishigami S, Kimura S, Fukuda M, Ishida T, Senda T. Molecular mechanism of the redox-dependent interaction between NADH-dependent ferredoxin reductase and Rieske-type [2Fe-2S] ferredoxin. *J Mol Biol*. 2007;373(2):382-400.
 42. Ashikawa Y, Fujimoto Z, Noguchi H, et al. Electron transfer complex formation between oxygenase and ferredoxin components in Rieske nonheme iron oxygenase system. *Structure*. 2006;14(12):1779-1789.
 43. Kweon O, Kim S-J, Baek S, et al. A new classification system for bacterial Rieske non-heme iron aromatic ring-hydroxylating oxygenases. *BMC Biochem*. 2008;9:11.
 44. Chakraborty J, Ghosal D, Dutta A, Dutta TK. An insight into the origin and functional evolution of bacterial aromatic ring-hydroxylating oxygenases. *J Biomol Struct Dyn*. 2012;30(4):419-436.
 45. Chang HK, Zylstra GJ. Novel organization of the genes for phthalate degradation from *Burkholderia cepacia* DBO1. *J Bacteriol*. 1998;180(24):6529-6537.
 46. Armengaud J, Happe B, Timmis KN. Genetic analysis of dioxin dioxygenase of *Sphingomonas* sp. strain RW1: Catabolic genes dispersed on the genome. *J Bacteriol*. 1998;180(15):3954-3966.
 47. Nam J-W, Nojiri H, Yoshida T, Habe H, Yamane H, Omori T. New Classification System for Oxygenase Components Involved in Ring-Hydroxylating Oxygenations. *Biosci Biotechnol Biochem*. 2001;65(2):254-263.
 48. Allende JL, Gibello A, Martin M, Garrido-Pertierra A. Transport of 4-hydroxyphenylacetic acid in *Klebsiella pneumoniae*. *Arch Biochem Biophys*. 1992;292(2):583-588.
 49. Allende JL, Suarez M, Gallego M, Garrido-Pertierra A. *4-Hydroxybenzoate Uptake in Klebsiella Pneumoniae Is Driven by Electrical Potential*. Vol 300. Academic Press; 1993:142-147.
 50. Groenewegen PEJ, Driessen AJM, Konings WNILN, Bont JANAMDE, De Bont JAM. Energy-dependent uptake of 4-chlorobenzoate in the coryneform bacterium NTB-1. *J*

- Bacteriol.* 1990;172(1):419-423.
51. Harwood CS, Nichols NN, Kim M, Diitty JL, Parales RE. Identification of the *pcaRKF* Gene Cluster from *Pseudomonas*: involvement in chemotaxis, biodegradation, and transport of 4-hydroxybenzoate. *Biotechniques*. 1994;176(21):6479-6488.
 52. Locher HH, Poolman B, Cook AM, Konings WN. Uptake of 4-toluene sulfonate by *Comamonas testosteroni* T-2. *J Bacteriol.* 1993;175(4):1075-1080.
 53. Miguez CB, Greer CW, Ingram JM, MacLeod RA. Uptake of benzoic acid and chloro-substituted benzoic acids by *Alcaligenes denitrificans* BRI 3010 and BRI 6011. *Appl Environ Microbiol.* 1995;61(12):4152-4159.
 54. Saint CP, Romas P. 4-Methylphthalate catabolism in *Burkholderia* (*Pseudomonas*) *cepacia* Pc701: a gene encoding a phthalate-specific permease forms part of a novel gene cluster. *Microbiology*. 1996;142:2407-2418.
 55. Pao SS, Paulsen IANT, Saier MH. Major facilitator superfamily. *Microbiol Mol Biol Rev.* 1998;62(1):1-34.
 56. Nichols NN, Harwood CS. *PcaK*, a high-affinity permease for the aromatic compounds 4-hydroxybenzoate and protocatechuate from *Pseudomonas putida*, *PcaK*, a High-Affinity Permease for the Aromatic Compounds 4-Hydroxybenzoate and Protocatechuate from *Pseudomonas putida*. *J Bacteriol.* 1997;179(16):5056-5061.
 57. Xu Y, Gao X, Wang SH, Liu H, Williams PA, Zhou NY. *MhbT* Is a specific transporter for 3-hydroxybenzoate uptake by gram-negative bacteria. *Appl Environ Microbiol.* 2012;78(17):6113-6120.
 58. Xu Y, Wang SH, Chao HJ, Liu SJ, Zhou NY. Biochemical and molecular characterization of the gentisate transporter *genK* in *Corynebacterium glutamicum*. *PLoS One.* 2012;7(7):1-8.
 59. Collier LS, Nichols NN, Neidle EL. *benK* encodes a hydrophobic permease-like protein involved in benzoate degradation by *Acinetobacter* sp. strain ADP1. *J Bacteriol.* 1997;179(18):5943-5946.
 60. Chaudhry MT, Huang Y, Shen XH, Poetsch A, Jiang CY, Liu SJ. Genome-wide investigation of aromatic acid transporters in *Corynebacterium glutamicum*. *Microbiology*. 2007;153(3):857-865.
 61. Hwang SY, Berges DA, Taggart JJ, Gilvarg C. Portage transport of sulfanilamide and sulfanilic acid. *J Med Chem.* 1989;32(3):694-698.
 62. Hille R. Structure and function of mononuclear molybdenum enzymes. *JBIC J Biol Inorg*

- Chem.* 1996;1(5):397-404.
63. Schrader N, Fischer K, Theis K, Mendel RR, Schwarz G, Kisker C. The crystal structure of plant sulfite oxidase provides insights into sulfite oxidation in plants and animals. *Structure*. 2003;11(10):1251-1263.
 64. Kappler U, Bennett B, Deutzmann R, Mcewan AG, Dahl C. Sulfite : Cytochrome c Oxidoreductase from *Thiobacillus novellus*. *J Biol Chem*. 2000;275(18):13202-13212.
 65. Kappler U, Bailey S. Molecular basis of intramolecular electron transfer in sulfite-oxidizing enzymes is revealed by high resolution structure of a heterodimeric complex of the catalytic molybdopterin subunit and a c-type cytochrome subunit. *J Biol Chem*. 2005;280(26):24999-25007.
 66. Wilson JJ, Kappler U. Sulfite oxidation in *Sinorhizobium meliloti*. *Biochim Biophys Acta - Bioenerg*. 2009;1787(12):1516-1525.
 67. Low L, Kilmartin JR, Bernhardt P V, Kappler U, Dahl C. How are “ atypical ” sulfite dehydrogenases linked to cell metabolism? Interactions between the SorT sulfite dehydrogenase and small redox proteins. 2011;2(March):1-10.
 68. McGrath AP, Laming EL, Casas Garcia GP, et al. Structural basis of interprotein electron transfer in bacterial sulfite oxidation. *Elife*. 2015;4(December2015):1-26.
 69. Brüggemann C, Denger K, Cook AM, Ruff J. Enzymes and genes of taurine and isethionate dissimilation in *Paracoccus denitrificans*. *Microbiology*. 2004;150(4):805-816.
 70. Rein U, Gueta R, Denger K, Ruff J, Hollemeyer K, Cook AM. Dissimilation of cysteate via 3-sulfolactate sulfo-lyase and a sulfate exporter in *Paracoccus pantotrophus* NKNCYSA. *Microbiology*. 2005;151(3):737-747.
 71. Denger K, Smits THM, Cook AM. L-cysteate sulpho-lyase, a widespread pyridoxal 5'-phosphate-coupled desulphonative enzyme purified from *Silicibacter pomeroyi* DSS-3(T). *Biochem J*. 2006;394(Pt 3):657-664.
 72. Denger K, Mayer J, Buhmann M, Weinitschke S, Smits THM, Cook AM. Bifurcated degradative pathway of 3-sulfolactate in *Roseovarius nubinhibens* ISM via sulfoacetaldehyde acetyltransferase and (S)-cysteate sulfolylase. *J Bacteriol*. 2009;191(18):5648-5656.
 73. Weinitschke S, Denger K, Cook AM, Smits THM. The DUF81 protein TauE in *Cupriavidus necator* H16, a sulfite exporter in the metabolism of C2 sulfonates. *Microbiology*. 2007;153(9):3055-3060.

74. Stolz A. Molecular characteristics of xenobiotic-degrading sphingomonads. *Appl Microbiol Biotechnol.* 2009;81(5):793-811.
75. Stolz A. Degradative plasmids from sphingomonads. *FEMS Microbiol Lett.* October 2013:1-10.
76. Aylward FO, McDonald BR, Adams SM, et al. Comparison of 26 Sphingomonad Genomes Reveals Diverse Environmental Adaptations and Biodegradative Capabilities. *Appl Environ Microbiol.* 2013;79(12):3724-3733.
77. Sanger F, Nicklen S, Coulson R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* 1977;74(12):5463-5467.
78. Burland V, Plunkett G, Daniels DL, Blattner FR. DNA sequence and analysis of 136 kilobases of the Escherichia coli genome: organizational symmetry around the origin of replication. *Genomics.* 1993;16(3):551-561.
79. Glaser P, Kunst F, Arnaud M, et al. Bacillus subtilis genome project: cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees; *Mol Microbiol.* 1993;10(2):371-384.
80. Fleischmann RD, Adams MD, White O, et al. Whole-Genome Random Sequencing and Assembly of Haemophilus-Influenzae Rd. *Science (80-).* 1995;269(5223):496-512.
81. Venter JC, Adams MDD, Myers EWW, et al. The sequence of the human genome. *Science (80-).* 2001;291(5507):1304-1351.
82. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822):860-921.
83. Metzker ML. Emerging technologies in DNA sequencing. *Genome Res.* 2005;15(12):1767-1776.
84. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010;11(1):31-46.
85. Eid J, Fehr A, Gray J, et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science (80-).* 2009;323(January):133-138.
86. Mikheyev AS, Tin MMY. A first look at the Oxford Nanopore MinION sequencer. *Mol Ecol Resour.* 2014;14(6):1097-1102.
87. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly. *Nat Methods.* 2011;8(1):61-65.
88. Mavromatis K, Land ML, Brettin TS, et al. The Fast Changing Landscape of Sequencing

- Technologies and Their Impact on Microbial Genome Assemblies and Annotation. *PLoS One*. 2012;7(12):1-6.
89. Schulze A, Downward J. Navigating gene expression using microarrays — a technology review. *Nat Cell Biol*. 2001;3(8):E190-E195.
 90. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science (80-)*. 1995;270:467-470.
 91. Selinger DW, Cheung KJ, Mei R, et al. RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. *Nat Biotechnol*. 2000;18(12):1262-1268.
 92. Wilkes T, Laux H, Foy CA. Microarray Data Quality—Review of Current Developments. *Omi A J Integr Biol*. 2007;11(1):1-13.
 93. Adams MD, Kelly JM, Gocayne JD, et al. Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science (80-)*. 1991;252(1990):1651-1656.
 94. Velculescu VE, Zhang L, Vogelstein B, Kinzler KE. Serial Analysis of Gene Expression. *Science (80-)*. 1995;270(5235):484-487.
 95. Shiraki T, Kondo S, Katayama S, et al. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc Natl Acad Sci*. 2003;100(26):15776-15781.
 96. Brenner S, Johnson M, Bridgham J, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol*. 2000;18(6):630-634.
 97. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009;10(1):57-63.
 98. Ausubel FM, Brent R, Kingston RE, et al., eds. *Current Protocols in Molecular Biology*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2001.
 99. Miller WG, Leveau JH, Lindow SE. Improved gfp and inaZ broad-host-range promoter-probe vectors. *Mol Plant Microbe Interact*. 2000;13(11):1243-1250.
 100. Barton MD, Barton HA. Scaffolder - software for manual genome scaffolding. *Source Code Biol Med*. 2012;7(1):4.
 101. Tatusova T, DiCuccio M, Badretdin A, et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res*. 2016:1-11.
 102. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44(D1):D457-D462.
 103. Tatusov RL. The COG database: a tool for genome-scale analysis of protein functions

- and evolution. *Nucleic Acids Res.* 2000;28(1):33-36.
104. Huerta-Cepas J, Forslund K, Szklarczyk D, Jensen LJ, von Mering C, Bork P. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *bioRxiv.* 2016;(Fitch 1970).
 105. Huerta-Cepas J, Szklarczyk D, Forslund K, et al. EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 2016;44(D1):D286-D293.
 106. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772-780.
 107. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol.* 2008;25(7):1307-1320.
 108. Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725-2729.
 109. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2009;26(1):139-140.
 110. Galperin MY, Makarova KS, Wolf YI, Koonin E V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 2015;43(D1):D261-D269.
 111. Gan HM, Chew TH, Tay YL, Lye SF, Yahya A. Genome sequence of *Hydrogenophaga* sp. strain PBC, a 4-aminobenzenesulfonate-degrading bacterium. *J Bacteriol.* 2012;194(17):4759-4760.
 112. Gan HM, Lee YP, Austin CM. Nanopore Long-Read Guided Complete Genome Assembly of *Hydrogenophaga intermedia*, and Genomic Insights into 4-Aminobenzenesulfonate, p-Aminobenzoic Acid and Hydrogen Metabolism in the Genus *Hydrogenophaga*. *Front Microbiol.* 2017;8(October):1-9.
 113. Bolla Marianna., Krámlí András. *Statisztikai Következtetések Elmélete.* Typotex; 2005. Accessed November 2, 2017.
 114. Sharp PM, Tuohy TMF, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 1986;14(13):5125-5143.
 115. Solovyev V, Salamov A. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In: *Metagenomics and Its Applications in Agriculture,*

Biomedicine and Environmental Studies. ; 2011.

116. Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol.* 2008;6(8):613-624.
117. Rojo F. Carbon catabolite repression in *Pseudomonas*: Optimizing metabolic versatility and interactions with the environment. *FEMS Microbiol Rev.* 2010;34(5):658-684.
118. Tropel D, Meer JR Van Der. Bacterial Transcriptional Regulators for Degradation Pathways of Aromatic Compounds. *Society.* 2004;68(3):474-500.
119. Kolvenbach BA, Dobrowinski H, Fousek J, et al. An unexpected gene cluster for downstream degradation of alkylphenols in *Sphingomonas* sp. strain TTNP3. *Appl Microbiol Biotechnol.* July 2011.
120. Kaschabek SR, Reineke W. Maleylacetate reductase of *Pseudomonas* sp. strain B13: dechlorination of chloromaleylacetates, metabolites in the degradation of chloroaromatic compounds. *Arch Microbiol.* 1992;158(6):412-417.
121. Mu TA, Byrde SM, Werlen C, Meer JR Van Der, Kohler HE. Genetic Analysis of Phenoxyalkanoic Acid Degradation in *Sphingomonas herbicidovorans* MH. *Society.* 2004;70(10):6066-6075.
122. Ferrández A, Garcíá JL, Díaz E. Genetic characterization and expression in heterologous hosts of the 3-(3-hydroxyphenyl)propionate catabolic pathway of *Escherichia coli* K-12. *J Bacteriol.* 1997;179(8):2573-2581.
123. Molina-Henares AJ, Krell T, Eugenia Guazzaroni M, Segura A, Ramos JL. Members of the IclR family of bacterial transcriptional regulators function as activators and/or repressors. *FEMS Microbiol Rev.* 2006;30(2):157-186.
124. Hegedüs B, Kós PB, Bálint B, et al. Complete genome sequence of *Novosphingobium resinovorum* SA1, a versatile xenobiotic-degrading bacterium capable of utilizing sulfanilic acid. *J Biotechnol.* 2017;241:76-80.
125. Conway JR, Lex A, Gehlenborg N. UpSetR: An R package for the visualization of intersecting sets and their properties. *Bioinformatics.* 2017;33(18):2938-2940.
126. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-15550.
127. Harwood CS, Parales RE. The beta-ketoadipate pathway and the biology of self-identity. *Annu Rev Microbiol.* 1996;50:553-590.

128. Kappler U. Bacterial sulfite-oxidizing enzymes. *Biochim Biophys Acta - Bioenerg.* 2011;1807(1):1-10.

10 Köszönetnyilvánítás

Szeretnék köszönetet mondani mindenkinek, aki segített abban, hogy ez a munka elkészülhessen.

Különösen hálás vagyok témavezetőmnek **Dr. Rákhely Gábornak**, hogy vállalta a szakmai irányításomat, ezáltal elindított a kutatói pályán. Köszönöm az éveken át tartó türelmét, iránymutatását és önzetlen segítségét.

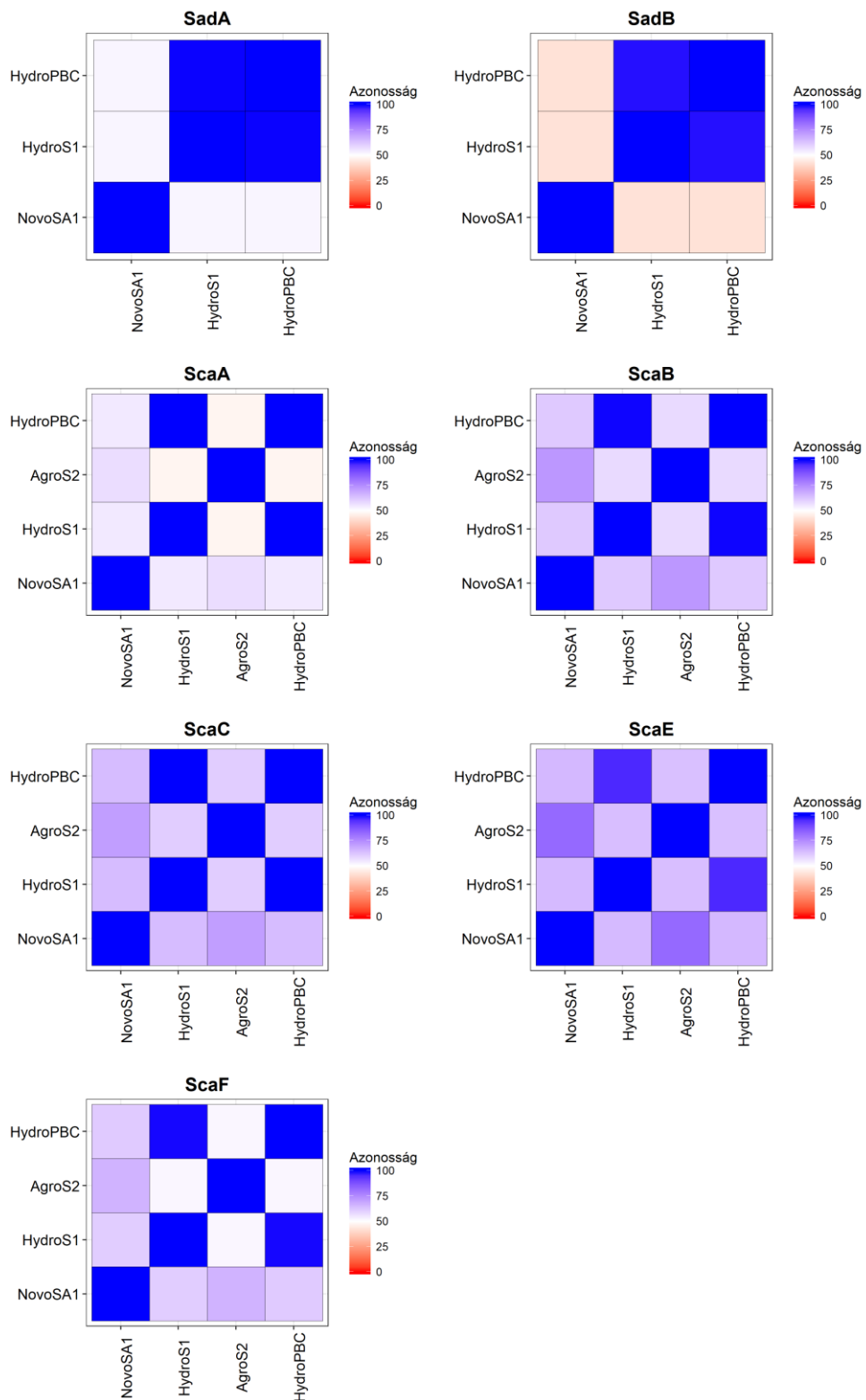
Köszönöm Professzor Dr. Kovács Kornélnak hogy a doktori éveim kezdetén lehetőséget teremtett a munkám elvégzéséhez.

Köszönöm szakdolgozóink szorgalmas munkáját.

Továbbá köszönettel tartozom az SZTE Biotechnológiai Tanszék és az MTA Szegedi Biológiai Kutatóközpont Biofizikai Intézet minden dolgozójának, a munkám sikeréhez való hozzájárulásukért. Különösen Dúzs Ágnesnek és Klem Józsefnek, akinek a tanácsaira, segítségére minden körülményke között számíthattam.

Végül, de nem utolsó sorban mérhetetlen hálát érzek a **Családom** és **Barátaim** iránt, akik támogatása nélkül nem tarthatnék itt.

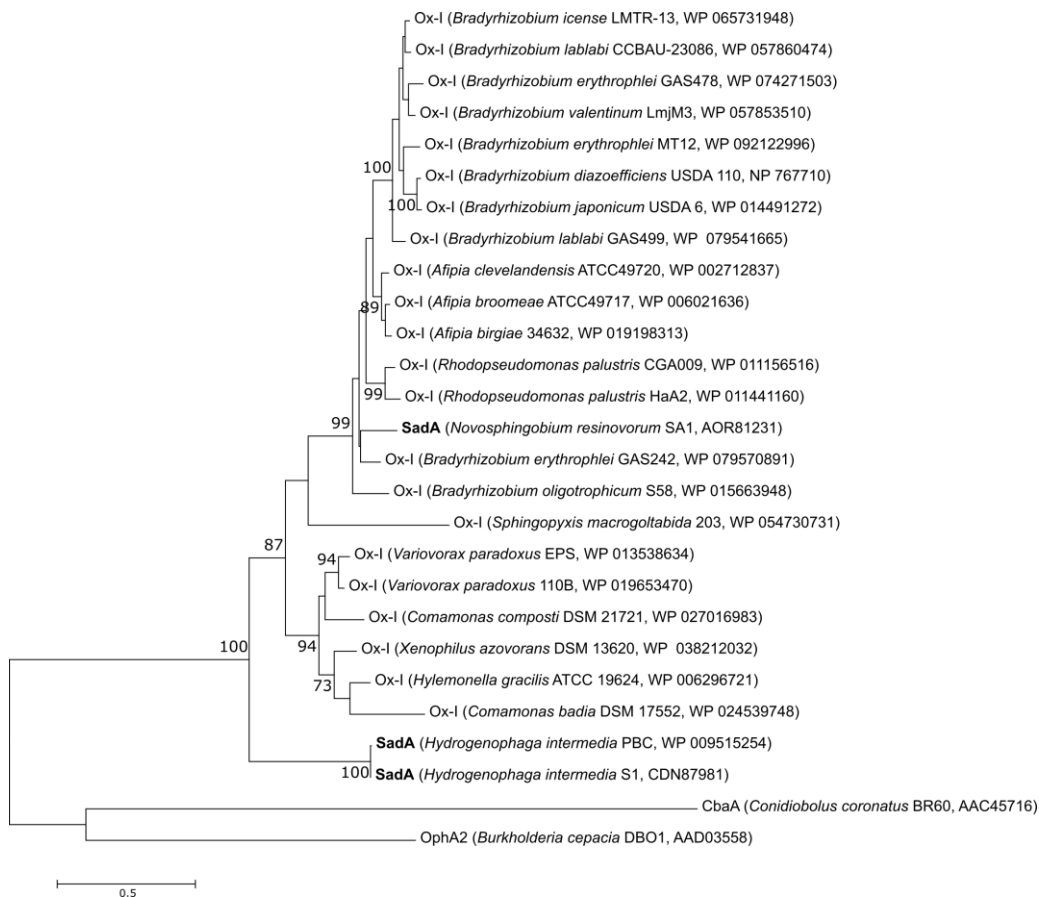
11 Függelék



1. függelék. A különböző baktérium törzsekben megtalálható, SA bontásában szerepet játszó fehérjék páronkénti összehasonlítása.

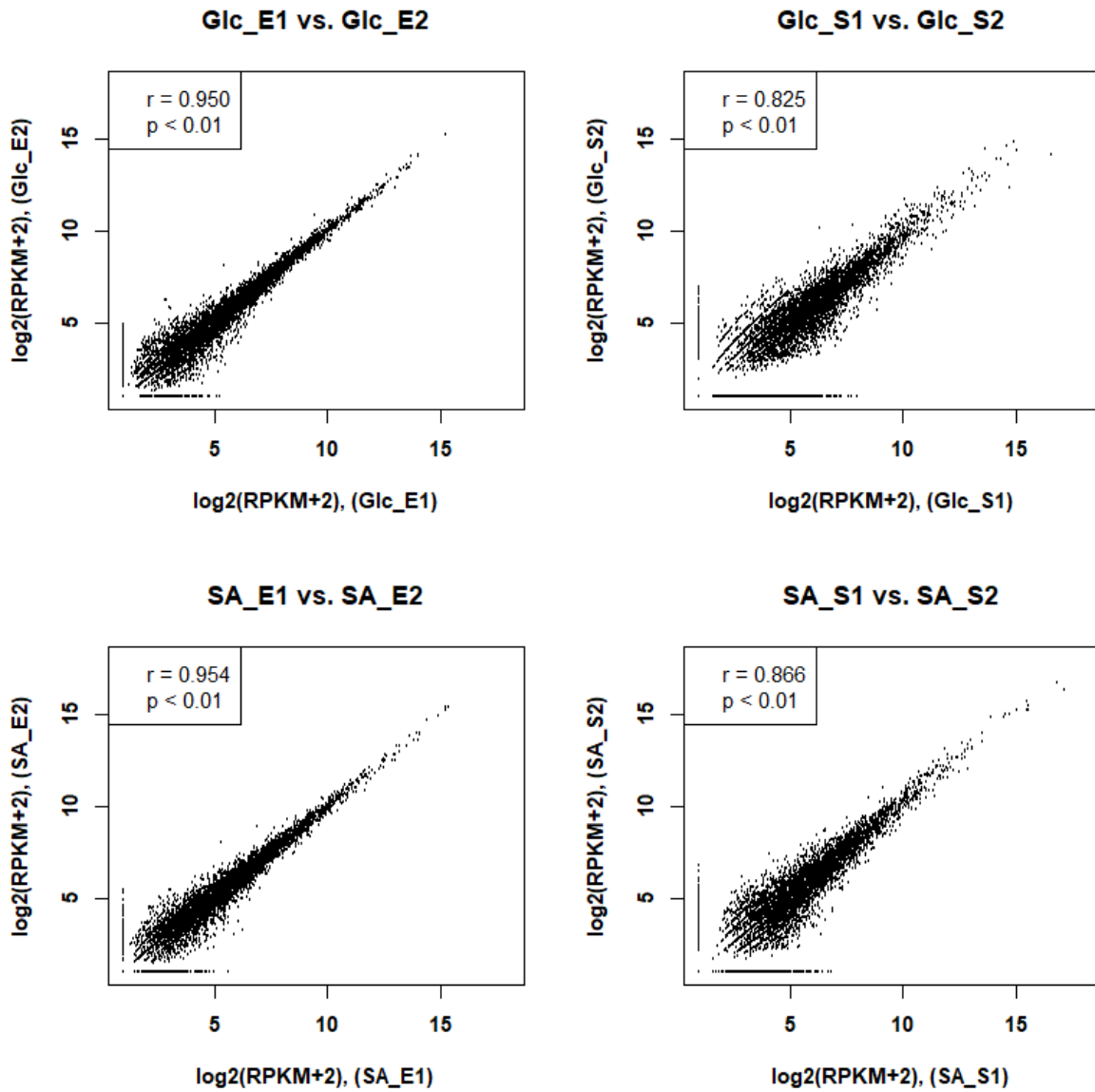
Enzimek	<i>N. resinovorum</i> SA1	<i>S. sp.</i> LH128	<i>N. resinovorum</i> KF1	Bontási útvonalak
PobA	AOR79986.1	EJU09115.1	EZP81133.1	HB
PcaG	AOR80188.1	EJU09117.1	EZP81135.1	HB
PcaH	AOR79985.1	EJU09116.1	EZP81134.1	HB
PcaB	AOR79984.1	EJU09118.1	EZP81136.1	HB
PcaL	AOR79983.1	EJU09119.1	EZP81137.1	HB
BenA	AOR79096.1	EJU09764.1	EZP82700.1	B
BenB	AOR79095.1	EJU09763.1	EZP82701.1	B
BenD	AOR79094.1	EJU09762.1	EZP82702.1	B
CatA	AOR79097.1	EJU09765.1	EZP82699.1	B
CatB	AOR79099.1	EJU09767.1	EZP82697.1	B
CatC	AOR79098.1	EJU09766.1	EZP82698.1	B
CatD	AOR79093.1	EJU09761.1	EZP82704.1	B

2. függelék. A *N. resinovorum* SA1 két legközelebbi rokonfajában megtalálható HB és B bontási útvonalak elemei

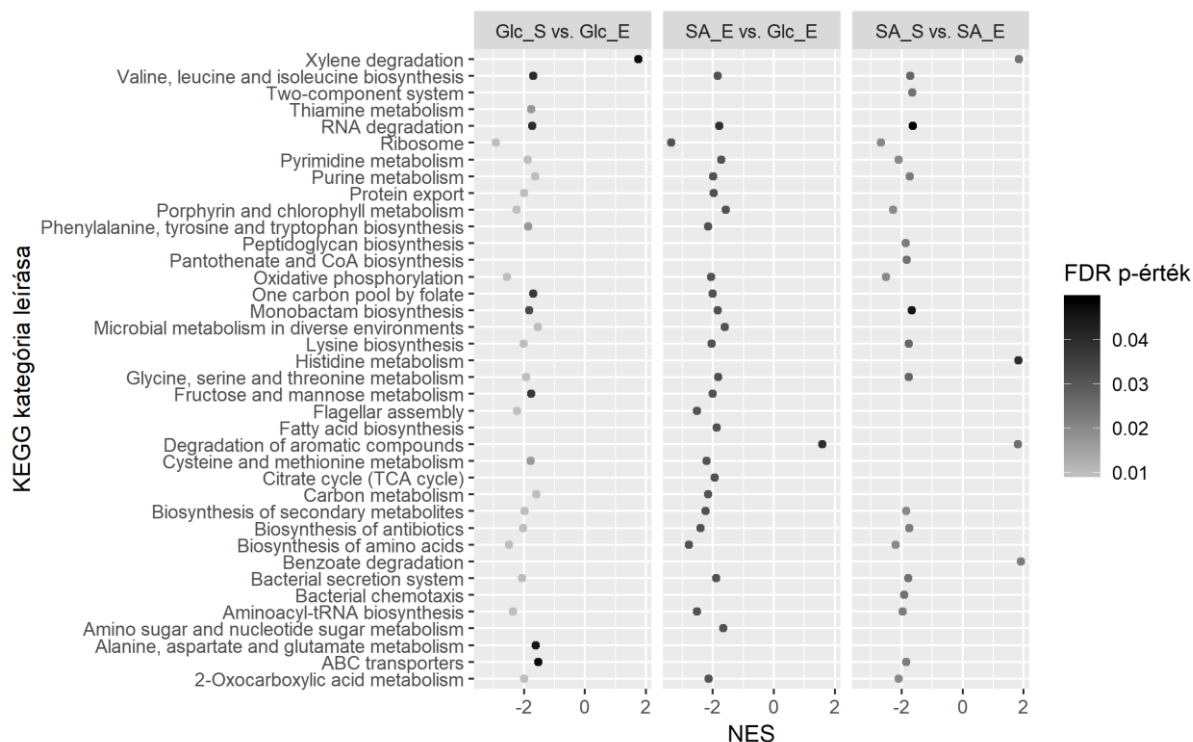


3. függelék. A *Bradyrhizobiaceae*, a *Sphingomonadaceae* és a *Comamonadaceae* család tagjaiban azonosítható *sadAB* gén klaszterek SadB fehérjéinek filogenetikai vizsgálata. A fehérjék közötti evolúciós kapcsolat maximum likelihood módszerrel volt meghatározva. A

CbaA és OphA2 fehérjék szolgáltak kulcsoportként. Az egység az aminosavanként történő szubsztitúciók számát mutatja.



4. függelék. Az RNA-seq vizsgálat biológia párhuzamosai esetén tapasztalt génkifejeződés Pearson korrelációs analízise.



5. függelék. A KEGG adatbázis alapján funkciós csoportokba rendezett fehérjék génjeinek géncsoport dúsulási vizsgálata (Gene Set Enrichment Analysis, GSEA). A színek az FDR–el javított p értékeket fejezik ki. Az x tengely a NES (normalizált dúsulási értéket, normalized enrichment score) értéket, az y tengely pedig a KEGG csoportok elnevezését mutatja.

NCBI locus tag	Kategória	Funkció
BES08_09020	ABC Transporters, Prokaryotic Type	iron complex transport system ATP-binding protein [EC:3.6.3.34]
BES08_30350	Major Facilitator Superfamily (MFS)	MFS transporter, AAHS family, 4-hydroxybenzoate transporter
BES08_08830	Other Transporters	iron complex outermembrane receptor protein
BES08_14745	Other Transporters	outer membrane immunogenic protein
BES08_18055	Other Transporters	iron complex outermembrane receptor protein
BES08_18430	Other Transporters	iron complex outermembrane receptor protein
BES08_18580	Other Transporters	iron complex outermembrane receptor protein
BES08_22625	Other Transporters	iron complex outermembrane receptor protein
BES08_23675	Other Transporters	iron complex outermembrane receptor protein
BES08_24205	Other Transporters	outer membrane immunogenic protein
BES08_26825	Other Transporters	iron complex outermembrane receptor protein
BES08_28345	Other Transporters	magnesium transporter
BES08_28525	Other Transporters	Na ⁺ :H ⁺ antiporter, NhaA family

6. függelék. SA hatására (SA_E vs. Glc_E) megváltozott kifejeződést mutató lehetséges transzporter fehérjét kódoló gének