

**Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta**

**Porovnání genomických evolučních změn a metabolické
kapacity symbiotických bakterií rodu *Legionella* u vší**

Bakalářská práce

Natálie Furiková

Školitel: prof. RNDr. Václav Hypša, CSc.

České Budějovice 2022

Furiková, N., 2022: Porovnání genomických evolučních změn a metabolické kapacity symbiotických bakterií rodu *Legionella* u vši. [Comparison of genomic evolutionary changes and metabolic capacities in symbiotic bacteria of the genus *Legionella* in lice. Bc. Thesis, in Czech.] -41 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

The genomes of 29 different lines of endosymbiotic bacteria *Legionella polyplacis* were annotated and compared. Genes have been identified and assigned to the corresponding annotated proteins of the online BLAST database. Annotated proteins were assigned to metabolic pathways and the functionality of the metabolic pathways and the level of degradation were compared to a non-symbiotic species *Legionella pneumophila subsp. pneumophila*.

Prohlašuji, že jsem autorem této kvalifikační práce a že jsem ji vypracoval(a) pouze s použitím pramenů a literatury uvedených v seznamu použitých zdrojů.

V Českých Budějovicích,

dne

Podpis studenta

Touto cestou bych chtěla poděkovat především prof. RNDr. Václavu Hypšovi, CSc. za vstřícnost, trpělivost, konzultace a připomínky při vedení mé bakalářské práce. Dále bych ráda poděkovala mé rodině a příteli, kteří mi byli podporou během psaní této bakalářské práce.

1 Obsah

1	Úvod.....	5
1.1	<i>Bakteriální symbionti hmyzu</i>	5
1.1.1	<i>Obecná charakteristika</i>	5
1.1.2	<i>Evoluce genomu symbiotických bakterií.....</i>	6
1.2	<i>Symbiotické bakterie vši</i>	8
1.2.1	<i>Legionella polyplacis.....</i>	9
2	Cíle práce.....	10
3	Materiály a metody	11
4	Výsledky	13
4.1	<i>Velikost genomu</i>	13
4.2	<i>Počet společných genů</i>	14
4.3	<i>Identifikace hypotetických genů a jejich úrovně degradace</i>	15
4.4	<i>Identifikace a porovnání hypotetických genů algoritmy tBLASTn, BLASTp a Conserved domains</i>	17
4.5	<i>Rekonstrukce proteinů metabolických drah</i>	19
4.5.1	<i>Rekonstrukce proteinů metabolických drah OG0000474.....</i>	19
4.5.2	<i>Rekonstrukce proteinů metabolických drah OG0000475.....</i>	21
4.5.3	<i>Rekonstrukce proteinů metabolických drah OG0000476.....</i>	26
4.5.4	<i>Rekonstrukce proteinů metabolických drah OG0000478.....</i>	26
4.5.5	<i>Rekonstrukce proteinů metabolických drah OG0000480.....</i>	27
4.5.6	<i>Rekonstrukce proteinů metabolických drah OG0000481</i>	29
5	Diskuse.....	30
5.1	<i>Porovnání hlavních charakteristik genomů.....</i>	30
5.2	<i>Sdílené geny.....</i>	32
5.3	<i>Hypotetické geny</i>	33
5.4	<i>Metabolické dráhy.....</i>	34
6	Závěr.....	37
7	Seznam literatury	38

1 Úvod

1.1 Bakteriální symbionti hmyzu

1.1.1 Obecná charakteristika

Hmyz je charakteristický svou schopností vytvářet široké spektrum symbiotických vztahů s bakteriemi od parazitismu až po obligátní mutualismus (Kikuchi, 2009). Tyto mutualistické vztahy jsou typické zejména pro ekologické skupiny hmyzu živící se rostlinnými šťávami nebo sajícími krev. Vzhledem k těmto druhům potravy hrají symbionti u fytofágního a hematofágního hmyzu důležitou roli (McFall-Ngai et al., 2013; Wierz et al., 2021). Symbiotické bakterie jsou schopné svému hostiteli poskytovat živiny, především esenciální aminokyseliny a vitaminy, které v potravě chybí (Brownlie and Johnson, 2009; Baumann, 2005; Zhang et al., 2015). Jako příklad je možné uvést endosymbionta mšic proteobakterii *Buchnera aphidicola*. *Buchnera* je schopna dodávat tryptofan a riboflavin, které jsou důležité pro vývoj a reprodukci mšic (Zhang et al., 2015). Rovněž jsou symbiotické bakterie schopné pomáhat hostiteli v reprodukci. Například bakterie tvořící biofilmy na vajíčkách hmyzu dokážou zvýšit rychlost líhnutí a také mají schopnost produkovat antifungální látky, které vajíčka chrání před choroboplodnými zárodky (Duperron, 2017). Symbionti v neposlední řadě hrají důležitou roli při ochraně hostitele před přirozenými nepřáteli, kterými mohou být například parazitoidi, hlístice, nebo patogenní houby. Kromě výše zmíněných dvou ekologických skupin hmyzu však symbionti mohou hrát důležitou roli také u ostatních ekologických skupin (Douglas, 1998; Werren, 1997).

Bakteriální symbionty lze rozdělit na dvě skupiny: obligátní tzv. primární symbionty a na fakultativní tzv. sekundární symbionty. Primární symbionti zastávají nutriční funkci a jsou tak schopni hostiteli syntetizovat potřebné živiny (Baumann, 2005). Sekundární symbionti nejsou esenciální pro svého hostitele, ale dokážou být prospěšní mutualisté nebo komenzálové. Někteří sekundární symbionti mohou být parazité a mít negativní vliv na fitness svého hostitele. Dokáží například manipulovat s reprodukčním potenciálem hostitele jako je tomu v případě některých kmenů gramnegativní bakterie rodu *Wolbachia*, která dokáže například u cizopasných vosiček vyvolat partenogenezi (Kaiwa et al., 2010; Moran et al., 2008). U některých skupin hmyzu je možné pozorovat přítomnost jak primárních symbiontů, tak i sekundárních symbiontů (Moran et al., 2008). Jako příklad je možné uvést mouchu tse-tse. Mouchy tse-tse jsou spojeny s konstantní a omezenou střevní mikroflórou (Rio et al., 2016). Přičemž střevní mikroflóra mouchám tse-tse dominují dvě gammaproteobakterie a to primární

symbiont *Wigglesworthia glossinidia* (Aksoy, 1995) a získaný sekundární symbiont *Sodalis glossinidius* (Dale and Maudlin, 1999).

Obligátní mutualističtí symbionti jsou především vnitrobuněční a vyskytují se ve specializovaných hmyzích buňkách zvaných bakteriocyty (Baumann, 2005). Dle druhu hostitele bakteriocyty mohou být součástí střevní stěny, tukového tělesa, nebo mohou být oddělené od tkání, a také mohou tvořit samostatné specializované orgány tzv. bakteriomy, které bakteriím zajišťují potřebné živiny a vhodné podmínky k životu (Braendle et al., 2003; Moran et al., 2008). Nedávné studie uvádí, že bakteriocyty jsou schopné také bakterie chránit před imunitním systémem hostitele pomocí exprese určitých imunitních genů, které vedou k lokální toleranci a jsou rovněž schopné kontrolovat hustotu a umístění bakterií (Masson et al., 2016; Masson et al., 2015).

Primární symbionti vytvářejí se svými hostiteli mutualistické vztahy, které mohou mít dlouhou koevoluční historii. Naopak sekundární symbionti vytvářejí různé symbiotické vztahy, které tak dlouhou koevoluční historii jako primární symbionti nemají (Moran et al., 1993). Aby byla zajištěna dlouhotrvající koevoluce endosymbionta s hostitelem jsou primární i sekundární symbionti přenášeni vertikálně z matky na potomstvo (Longdon and Jiggins, 2012; Baumann, 2005).

1.1.2 *Evoluce genomu symbiotických bakterií*

Genomy symbiotických bakterií jsou odlišné od genomů volně žijících bakterií a vyznačují se některými jedinečnými vlastnostmi, které získaly během svého soužití s hostitelem (Hildebrand et al., 2010). Mezi tyto vlastnosti patří převážně degenerativní změny, které jsou důsledkem života bakterie v predikovatelném prostředí, které jim zajišťuje řadu funkcí včetně ochrany a metabolismu. Rovněž jsou tyto změny důsledkem sníženého selekčního tlaku prostředí a životem bakterií v malých populacích podléhajících bottleneck efektu (Moran, 2002; McCutcheon and Moran, 2012).

Degenerativní změny mají za následek degradaci genomu, kdy dochází ke změně jak jeho velikosti, tak i nukleotidového složení. Genomy obligátních symbiontů mohou být oproti volně žijícím bakteriím velmi redukovány, v extrémních případech například u bakterie *Nasuiana* došlo k redukci genomu až na 112 kb a u bakterie *Candidatus Carsonella ruddii* na velikost 160 kb (Wernegreen, 2015; McCutcheon and Moran, 2012). Vlivem tohoto zkracování genomu dochází ke ztrátě genů, které nejsou esenciální pro život bakterie v hostiteli. Změny genomu tedy zahrnují ztrátu metabolických funkcí, které doplňuje hostitel, genů pro ochranné

mechanismy, regulaci genové exprese, virulenci a genů účastnících se oprav DNA. Dochází tedy k zachování pouze genů nezbytných pro fungování vlastní buňky a genů nezbytných pro vývoj a/nebo reprodukci hostitele (Martínez-Cano et al., 2015; Gosalbes et al., 2010; McCutcheon and Moran, 2012). Symbiotické bakterie se vlivem degradace genomu tak stávají zcela závislé na svém hostiteli. Postupnou degradací genomu může dojít i ke ztrátě genů nahrazujících metabolické nedostatky hostitele. V takové situaci může dojít ke ztrátě původního symbionta a dojde k získání symbionta jiného, který dokáže nahradit nebo doplnit biosyntetické vlastnosti symbionta původního (McCutcheon and Moran, 2010).

S postupnou degenerací a ztrátou genů u symbiontů se pojí i pseudogenizace. Při tomto jevu dochází ke ztrátě funkce genu vlivem narušení jeho kódovací sekvence nejčastěji v důsledku akumulace škodlivých mutací v sekvenci (Grus and Zhang, 2008; McCutcheon and Moran, 2012). V bakteriálním genomu jsou pseudogeny zastoupeny v relativně malém množství (1 až 5 % z celého genomu). Přičemž nízká hladina těchto pseudogenů ve většině bakterií implikuje to, že vlivem silné selekce jsou tyto pseudogeny velmi rychle z genomů odstraňovány (Liu et al., 2004). Mezi intracelulárními endosymbionty ovšem existují případy, kdy množství přítomných pseudogenů v genomu dosáhlo velmi vysoké úrovně jako je tomu u symbionta *Sodalis glossinidius*. Důsledkem tohoto velkého množství pseudogenů došlo u bakterie ke snížení kódovací kapacity o 50 % (Toh et al., 2006).

Aby bylo možné geny dále zkoumat je potřeba je nějakým způsobem identifikovat a přiřadit jim funkci. Tento proces je nazýván jako anotace, přičemž jedním z nejnámějších a nejpoužívanějších programů pro anotaci je program RAST (Abril and Castellano, 2019; Aziz et al., 2008). Při anotaci genomu se může stát, že se v genomu nacházejí geny, které jsou tak degradované, nebo dosud jsou jejich funkce neznámé, že je program nedokáže rozpoznat. V tomto případě jsou pak tyto sekvence programem označeny jako tzv. hypotetické proteiny, nebo ponechány zcela bez anotace. Metody genové anotace se opírají o vyhledávací nástroje založené na homologii jako je například algoritmus BLAST. Tento algoritmus využívá specifické databáze k vyhledání homologních genů a výsledné informace pak využije k anotaci genů (Pevsner, 2009). Pro další biologické výzkumy je také důležité stanovit fylogenetické vztahy mezi jednotlivými sekvencemi genů. Pro zjištění těchto vztahů byla vyvinuta řada programů, které na základně vstupních sekvencí genů dokážou fylogenetické vztahy mezi genovými sekvencemi identifikovat (Altenhoff et al., 2016; Nichio et al., 2017). Programy při této identifikaci využívají data podobnosti sekvencí (nebo očekávaných hodnot) získaných z algoritmu BLAST nebo jeho alternativ (Camacho et al., 2009). Mezi nejpoužívanější

programy patří InParanoid, OrthoMCL, OMA a OrthoFinder (Ostlund et al., 2009; Li et al., 2003; Altenhoff et al., 2010; Emms and Kelly, 2015). Všechny z těchto zmíněných programů vytvářejí různá výstupní data. Některé z těchto softwarů identifikují orthoskupiny, některé ortology a paralogy a některé dokážou identifikovat obojí (Altenhoff et al., 2016; Trachana et al., 2011).

Dalším typickým znakem obligátních symbiontů je také velmi nízké procentuální zastoupení nukleotidových bází guaninu a cytosinu (GC) oproti volně žijícím bakteriím (Moran, 2002; McCutcheon and Moran, 2012). V GC páru jsou mezi guaninem a cytosinem tři vodíkové můstky, což je energeticky náročnější, a to je také důvodem, proč u symbiotických bakterií probíhá mutace genetické informace ve prospěch nukleotidových bází adeninu a thyminu. Výsledkem redukce genomu je tedy snižování energetických nákladů symbiotických bakterií. Nejznámějším příkladem symbiotické bakterie s vysoce redukovaným obsahem GC je bakterie rodu *Buchnera* a její průměrný obsah GC je pod 30 %. V dalším extrémním případě u betaproteobakterie *Candidatus Zinderia insecticola* došlo k redukci obsahu GC až na 13,5 % (McCutcheon and Moran, 2012)

1.2 Symbiotické bakterie vší

Vši (Anoplura) se řadí mezi hostitelsky specifické ektoparazity savců a jsou taktéž typickým příkladem obligátních hematofágů. Vzhledem k typu potravy, kterou se vší živí a která postrádá esenciální látky, které vší potřebují, staly se též hostiteli symbiotických bakterií (Baumann, 2005; Boyd et al., 2017; Moran and Telang, 1998). Příkladem významného symbionta u vší druhu *Pediculus humanus* je gammaproteobakterie *Candidatus Riesia pediculicola* (Allen et al., 2007). Dalším významným symbiontem u vší druhu *Pediculus schaeffi* je také *Candidatus Riesia pediculischaeffi* a u vší rodu *Polyplax* hraje důležitou roli symbiont *Legionella polyplacis* (Hammoud et al., 2022; Říhová et al., 2017). Fylogenetické studie ukázaly, že různé skupiny vší sdílejí endosymbiotické vztahy s různými druhy symbiontů (Boyd and Reed, 2012). Tyto symbiotické vztahy u různých druhů vší se pravděpodobněji vyvinuly u různých druhů několikrát nezávisle na sobě nežli ze společného předka (Buchner, 1965; Hypša and Křížek, 2007).

1.2.1 *Legionella polyplacis*

L. polyplacis z čeledi *Legionellaceae* patří mezi obligátní mutualisty s velmi malým genomem, který je následkem dlouhého symbiotického vztahu bakterie se svým hostitelem. V evoluci genomu této bakterie sehrál důležitou roli horizontální přenos DNA, který se týká pouze přenosu kompletního biotinového operonu, který bakterii pravděpodobně pomohl s přechodem z intracelulárního symbionta améb do role vnitrobuněčného symbionta vší *Polyplax serrata* a *Polyplax spinulosa* žijících na hlodavcích. V prostředí jsou legionely vázané na biofilmy, kde tvoří symbiotické intracelulární vztahy s amébami nebo jinými protisty. Vzhledem k této životní strategii jsou všechny známé druhy legionel popsány buď z nemocnic anebo z vodních zdrojů. Jedinou výjimku v tomto směru tvoří tedy zmíněná bakterie *L. polyplacis* (Fields, 1996; Říhová et al., 2017).

Hostiteli vší rodu *Polyplax* jsou nejčastěji myši rodu *Apodemus*, přičemž většina doposud popsaných druhů těchto myši se vyskytuje po celé Evropě a Asii (Štefka and Hypša, 2008). Z předchozích výzkumů myši vyplývá, že některé fylogeneticky a ekologicky blízké druhy jako *Apodemus sylvaticus* a *Apodemus flavicollis* se mezi sebou mohou značně lišit zejména jejich evoluční historií nebo umístěním glaciálních refugií (Michaux et al., 2005). Důsledkem této odlišné historie došlo k ovlivnění populační historie vší (Štefka and Hypša, 2008). Navazující studii bylo prokázáno, že veš druhu *P. serrata* se skládá ze dvou genetických linií s odlišnou hostitelskou specifitou a geografickým rozložením. Tyto linie jsou rozdělené na tzv. specifické (S) linie a nespecifické (N) linie. Ačkoliv jsou tyto linie úzce příbuzné, liší se zmíněnou hostitelskou specifitou. Vší z S linie dokážou parazitovat pouze na druhu myši *A. flavicollis*, avšak vší z N linie dokážou parazitovat jak na druhu *A. flavicollis*, tak i na druhu *A. sylvaticus*. *A. flavicollis* důsledkem rekolonizace ze dvou refugií utvořil dva geneticky vzdálené klastry, které se rozšířily po celé Evropě. Vší ovšem tento proces rozšíření po Evropě nenásledovaly. Podlinie vší rozdělené na specifické východní (SE) a specifické západní (SW) se u S linie přestaly šířit ve středu Evropy. Poněvadž se vší dostaly do sekundární kontaktní zóny, která zastavila jejich šíření a vytvořila se tak hybridní zóna mezi SE a SW liniemi (Martinů et al., 2020; Martinů et al., 2018; Hvala and Wood, 2012).

2 Cíle práce

- Porovnání a anotace 29 dostupných genomů odlišných linií symbionta vší *Legionella polyplacis*
- Identifikace chybějících genů
- Zrekonstruování metabolických drah anotovaných genů a porovnání změn v metabolické kapacitě
- Porovnání genů s divokými a nesymbiotickými druhy *Legionella* s pomocí vytvořené porovnávací databáze v Geneious prime

3 Materiály a metody

Výchozí genomy 29 linií endosymbiotické bakterie *Legionella polyplacis* (Říhová et al., 2017) byly dodány ve formátu fasta. Následně jsem provedla anotaci všech 29 genomů s pomocí internetového softwaru RAST (Aziz et al., 2008) metodou RASTtk (Brettin et al., 2015), včetně funkce automatické opravy chyb. Kompletní anotované genomy byly následně přeloženy programem Geneious prime 2021.1.1. do sekvencí aminokyselin s parametrem odstranění finálního stop kodonu. Za účelem zjištění syntenie jsem provedla porovnání celých 29 genomů programem MAUVE implementovaném v programu Geneious prime.

Všechny přeložené sekvence aminokyselin jsem vyexportovala jako tabulky a následně je upravila v programu Excel 2202, kde jsem přidala číslování jednotlivých genů (odstranění duplikovaných názvů) a symbol <. Dále jsem tyto tabulky pomocí textového editoru upravila na textové soubory ve formátu fasta.

Tyto fasta soubory jsem analyzovala softwarem Orthofinder 2.4.0. (Kelly and Emms, 2019). Výsledné orthoskupiny byly programem vyexportovány jako textový soubor, který jsem následně upravila do nul a jedniček v programu Excel (tj. přítomnost a nepřítomnost genu formálně kódována pro analýzu v programu R). Tabulku jsem dále analyzovala programem Rstudio 2021.09.1 (RStudio Team, 2021) a s pomocí příkazů a balíčku UpSetR 1.4.0. (Gehlenborg, 2019) jsem sestrojila graf, který je přiložen níže v textu (Graf 1).

Na základě takto připraveného přehledu sdílených genů jsem určila několik skupin genů k podrobnější analýze: 5 genů přítomných u specifických východních linií 39RAD_SE, 51RAD_SE, SE_99ZLb_16S, SE_98Zlc_16S, SE_133Zlb, 44RADb_SE, SE_17Zlb, SE_164Zlb, 2RAD_SE, SE_98Zle, 1MAN_SE, 6RAD_SE, 106MAN_SE, 29DR_SE (dále jen jako SE) a specifických západních linií (2HBc_SW, 98Zld_SW, 47MAN_SW, Doup5c_SW, 1STRf_SW, 30VOJ_SW, 29DZ_SW, NE125b_16S_SW (dále jen jako SW), ale zároveň chybějících u linií 4ROK_N, 7ROK_N, PNG_RR_Lspin, k124_9727, L26_Afr_Arvican a N17_Afr_Steno1. Dále jsem vyhledala sekvence 2 genů přítomných pouze u SE linie a zároveň chybějících u linií 4ROK_N, 7ROK_N, PNG_RR_Lspin, k124_9727, L26_Afr_Arvican, N17_Afr_Steno1 a SW linie. Těchto 7 sekvencí jsem u SE linie vyhledala v MAUVE porovnání celých 29 genomů v programu Geneious prime. Přítomné sekvence u SE linie byly programem RAST ve všech sedmi případech anotovány jako hypotetický protein (Tab. 3).

Sekvence anotované jako hypotetický protein jsem u všech linií extrahovala z MAUVE porovnání všech 29 celých genomů a následně jsem je porovnávala v programu MAFFT

implementovaném v programu Geneious prime. Sekvence hypotetických proteinů označené jako Orthogroups 00000474-00000481 jsem následně porovnála s mnou vytvořenou databází genomů nesymbiotických bakterií rodu *Legionella* taktéž v programu Geneious prime a vytvořila jsem dle výsledků procentuální tabulku shody (Tab. 4)

Pro bližší identifikaci genů, jež program RAST zhodnotil jako hypotetický protein u nespécifické linie 4ROK_N a 7ROK_N (dále jen jako N), linie PNG_RR_Lspin, k124_9727, L26_Afr_Arvican, N17_Afr_Steno1 a SW linie jsem využila algoritmy NCBI protein BLAST (Camacho et al, 2009, Altschul et al, 1997, Altschul et al, 1990) a NCBI Conserved Domains (Bryant and Marchler-Bauer, 2004). Arbitrálně jsem vybrala sekvence hypotetických proteinů u linie 106MAN_SE a nahrála je do algoritmů s upřesňujícím parametrem, kdy hodnotu čísla Expect threshold jsem změnila na hodnotu 10 s cílem zachytit i nižší podobnosti. Výsledky algoritmů následně byly přiřazeny k odpovídajícím proteinům databáze. Dále byly tyto stejné sekvence přiřazeny k podobným sekvencím programem tBLASTn (Camacho et al, 2009, Altschul et al, 1997, Altschul et al, 1990). Výsledky jsem zanesla do tabulek (Tab. 4 a Tab. 5).

Za účelem rekonstrukce metabolických drah z příslušných anotovaných proteinů linií bakterie *L. polyplacis* L26_Afr_Arvican, N17_Afr_Steno1, 4ROK_N, 7ROK_N, k127_9727, PNG_RR_Lspin a SW nacházející se v místě sekvence anotované u SE linie jako hypotetický protein jsem fasta soubory linií importovala do softwaru BlastKoala (Kanehisa et al., 2016). Následně jsem dle KEGG čísel přiřazených k proteinům ve výstupním textovém souboru ze softwaru BlastKoala proteiny vyhledala a zmapovala softwarem KEGG mapper (Kanehisa et al., 2020) jejich metabolické dráhy.

Na závěr jsem tyto metabolické dráhy porovnála s nesymbiotickým druhem bakterie *Legionella pneumophila subsp. pneumophila* a určila jejich úroveň degradace.

4 Výsledky

4.1 Velikost genomu

Velikosti genomu u jednotlivých dostupných linií bakterie druhu *L. polyplacis* jsou odlišné. U všech 29 linií se pohybují v rozmezí od 529 749 bp do 551 286 bp, přesné velikosti se nacházejí v tabulce (Tab. 1). Podíl guanino-cytosinového komplementárního páru u jednotlivých linií bakterie není moc odlišný. Podíl GC je v rozmezí 21,7 - 23,6 %, podrobněji jsou hodnoty popsány v tabulce (Tab. 1).

Tab. 1: Přehled velikosti genomů, počtu genů a obsahu GC u 29 linií bakterie druhu *L. polyplacis*.

Název linie	Velikost genomu (bp)	Počet genů	GC (%)
39RAD_SE	529 749	525	23
51RAD_SE	529 749	526	23
SE_99ZLb_16S	529 749	526	23
SE_98Zlc_16S	529 751	525	23
SE_133Zlb	529 752	525	23
44RADb_SE	529 754	527	23
SE_17Zlb	529 756	526	23
SE_164Zlb	529 756	525	23
2RAD_SE	529 757	524	23
SE_98Zle	529 757	525	23
1MAN_SE	529 758	526	23
6RAD_SE	529 758	525	23
106MAN_SE	529 758	526	23
29DR_SE	529 760	525	23
SE_182Zlb	529 774	525	23
SW_98Zld	529 775	521	23
SW_NE125b_16S	529 775	524	23
47MAN_SW	529 779	522	23
SW_2HBc	529 781	522	23
29DZ_SW	529 782	523	23
SW_1STRf	529 782	523	23
30VOJ_SW	529 783	523	23
SW_Doup5c	529 785	522	23
4ROK_N	530 992	529	23,6
7ROK_N	530 992	529	23,6
PNG_RR_Lspin	532 177	527	23,1
K127_9727	532 296	526	23,1
L26_Afr_Arvican	546 533	527	23,1
N17_Afr_Steno1	551 289	547	21,7

4.3 Identifikace hypotetických genů a jejich úrovně degradace

Po analýze přítomných hypotetických proteinů u SE linie z celkového porovnání 29 linií *L. polyplacis* programem MAFFT v Geneious prime bylo prokázáno, že zdánlivě chybějící geny (tj. chybějící anotace) ve skutečnosti u příslušných linií PNG_RR_Lspin, k124_9727, L26_Afr_Arvican, N17_Afr_Steno1, SW linie a N linie nechybí. V místech sekvencí u linií PNG_RR_Lspin, k124_9727, L26_Afr_Arvican, N17_Afr_Steno1, SW a N linií se nacházejí identické sekvence, které jsou do určité míry rozpadlé, nebo v nich došlo k nahrazení několika nukleotidů, nachází se tam anotovaný gen či jsou sekvence zkrácené anebo se tam nachází celé, jen je program RASTk nebyl schopný anotovat.

Tyto zdánlivě chybějící hypotetické geny byly označeny dle výsledků programu Orthofinder jako Orthogroups 00000474-00000481, dále v textu pod zkratkou OG (Tab. 2).

Tab. 2: Přehled chybějících orthogroups u jednotlivých linií bakterie *L. polyplacis*.

Orthogroups	Chybějící u linie	
00000474		
00000475	4ROK_N	7ROK_N
00000476	L26_Afr_Arvican	N17_Afr_Steno1
00000477	PNG_RR_Lspin	K127_9727
00000478		
00000480 00000481	29DZ_SW	30VOJ_SW
	47MAN_SW	SW_1STRf
	SW_2HBc	SW_98Zld
	SW_Doup5c	SW_NE125b_16S
	N17_Afr_Steno	L26_Afr_Arvican
	PNG_RR_Lspin	k127_9727
	4ROK_N	7ROK_N

Všechny orthogroups očíslované 00000474-00000481 byly u SE linie anotovány jako hypotetické proteiny. Jednotlivé orthogroups a jejich anotace v genomu u jednotlivých linií jsou více specifikovány v tabulce (Tab. 3).

Tab. 3: Přehled anotovaných orthogroups u jednotlivých linií bakterie *L. polyplacis*.

Linie	Anotace						
	OG00000474	OG00000475	OG00000476	OG00000477	OG00000478	OG00000480	OG00000481
SE	Hypotetický p.	Hypotetický p.	Hypotetický p.	Hypotetický p.	Hypotetický p.	Hypotetický p.	Hypotetický p.
4ROK_N a 7ROK_N	Hypotetický p.	Bez anotace	Hypotetický p.	Hypotetický p.	Phosphatidyl glycerophosphatase A	N-acetylmuramoyl-L-alanine amidase	Bez anotace
k127_9727	Bez anotace	Heme O synthase	Hypotetický p.	Bez anotace	Phosphatidyl glycerophosphatase A	N-acetylmuramoyl-L-alanine amidase	Iron binding protein SufA for iron-sulfur cluster assembly
PNG_RR_Lspin	Bez anotace	Heme A synthase	Bez anotace	Bez anotace	Phosphatidyl glycerophosphatase A	N-acetylmuramoyl-L-alanine amidase	Iron binding protein IscA for iron sulfur cluster assembly.
N17_Afr_Steno1	Ribonuclease III	Bez anotace	Hypotetický p.	Bez anotace	Hypotetický p.	N-acetylmuramoyl-L-alanine amidase	Bez anotace
L26_Afr_Arvican	DedA protein	Bez anotace	Pap2 superfamily p.	Bez anotace	Phosphatidyl glycerophosphatase A	N-acetylmuramoyl-L-alanine amidase TsaE protein	Bez anotace
SW	Hypotetický p.	Hypotetický p.	Hypotetický p.	Hypotetický p.	Hypotetický p.	N-acetylmuramoyl-L-alanine amidase	Bez anotace

4.4 Identifikace a porovnání hypotetických genů algoritmy tBLASTn, BLASTp a Conserved domains

Hypotetické geny byly algoritmem tBLASTn a mnou vytvořenou databází z nesymbiotických druhů legionel v programu Geneious prime na základě procentuální podobnosti přiřazeny ke konkrétním genům nesymbiotických druhů legionel (Tab. 4).

Dále byla většina hypotetických genů algoritmy BLASTp a Conserved domains přiřazena ke konkrétním genům. Tyto geny jsou více specifikovány v tabulce (Tab. 5).

Tab. 4: Přehled procentuálních podobností hypotetických genů u bakterie *L. polyplacis* s konkrétními geny nesymbiotických druhů bakterie rodu *Legionella*.

Orthogroup (106MAN_SE)	tBLASTn (<i>Legionellaceae</i>)	Geneious Prime databáze
00000474	Inner membrane protein YabI (61 % <i>L. clemsonensis</i>) Hypotetický protein (38 % <i>L. geestiana</i>)	Prolyl aminopeptidase (75 % <i>L. birminghamensis</i>) Bez anotace (61 % <i>L. jamestowniensis</i>)
00000475	Heme O synthase (57 % <i>L. pneumophila</i> , <i>L. oakridgensis</i> , 55 % <i>L. jordanis</i> , 52 % <i>L. sainthelensi</i> , <i>L. geestiana</i>)	Heme O synthase (57 % <i>L. pneumophila</i> , <i>L. londiniensis</i> , 55 % <i>L. jordanis</i> , <i>L. busanensis</i> , 52 % <i>L. gratiana</i>)
00000476	Nepřiřazen	Nepřiřazen
00000477	Nepřiřazen	Nepřiřazen
00000478	Phosphatidylglycerophosphatase A (64 % <i>L. clemsonensis</i> , 63 % <i>L. fallonii</i> , 62 % <i>L. antarctica</i> , 55 % <i>L. hackeliae</i>)	Serine hydrolase (83 % <i>L. gratiana</i> , 75 % <i>L. pneumophila</i> , 66 % <i>L. wadsworthii</i>) Phosphatidylglycerophosphatase A (66 % <i>L. jamestowniensis</i> , 62 % <i>L. moravika</i>)
00000480	Nepřiřazen	Nepřiřazen
00000481	Nepřiřazen	Nepřiřazen

Tab. 5: Přehled identifikace hypotetických genů algoritmy BLASTp a Conserved Domains u jednotlivých orthogroups u bakterie *L. polyplacis*.

Orthogroup (106MAN_SE)	BLASTp (<i>Legionellaceae</i>)	Conserved Domain
00000474	Nepřiřazen	Pyridoxamine 5'-phosphate oxidase (<i>Crocospaera subtropica</i>) Ribonucleotide-diphosphate reductase subunit beta (<i>Shewanella sp.</i>)
00000475	Heme O synthase	Heme O synthase (<i>Caulobacter sp.</i>)
00000476	Nepřiřazen	Mitochondrial-associated sphingomyelin phosphodiesterase (Euteleostomi – člověk) Sensor protein RstB (<i>Chromobacterium violaceum</i>)
00000477	Nepřiřazen	Acetyl-CoA decarboxylase/synthase complex subunit beta (<i>Archaeoglobus fulgidus</i>) Cytochrome P450 family 1 (Gnathostomata – člověk)
00000478	Serine hydrolase, Beta-lactamase family p.	Nepřiřazen
00000480	Nepřiřazen	Nepřiřazen
00000481	Nepřiřazen	Second immunoglobulin (Ig)-like domain of interleukin-1 receptor (Euteleostomi- člověk) translation initiation factor IF-2 (<i>Methanocaldococcus jannaschii</i>)

4.5 Rekonstrukce proteinů metabolických drah

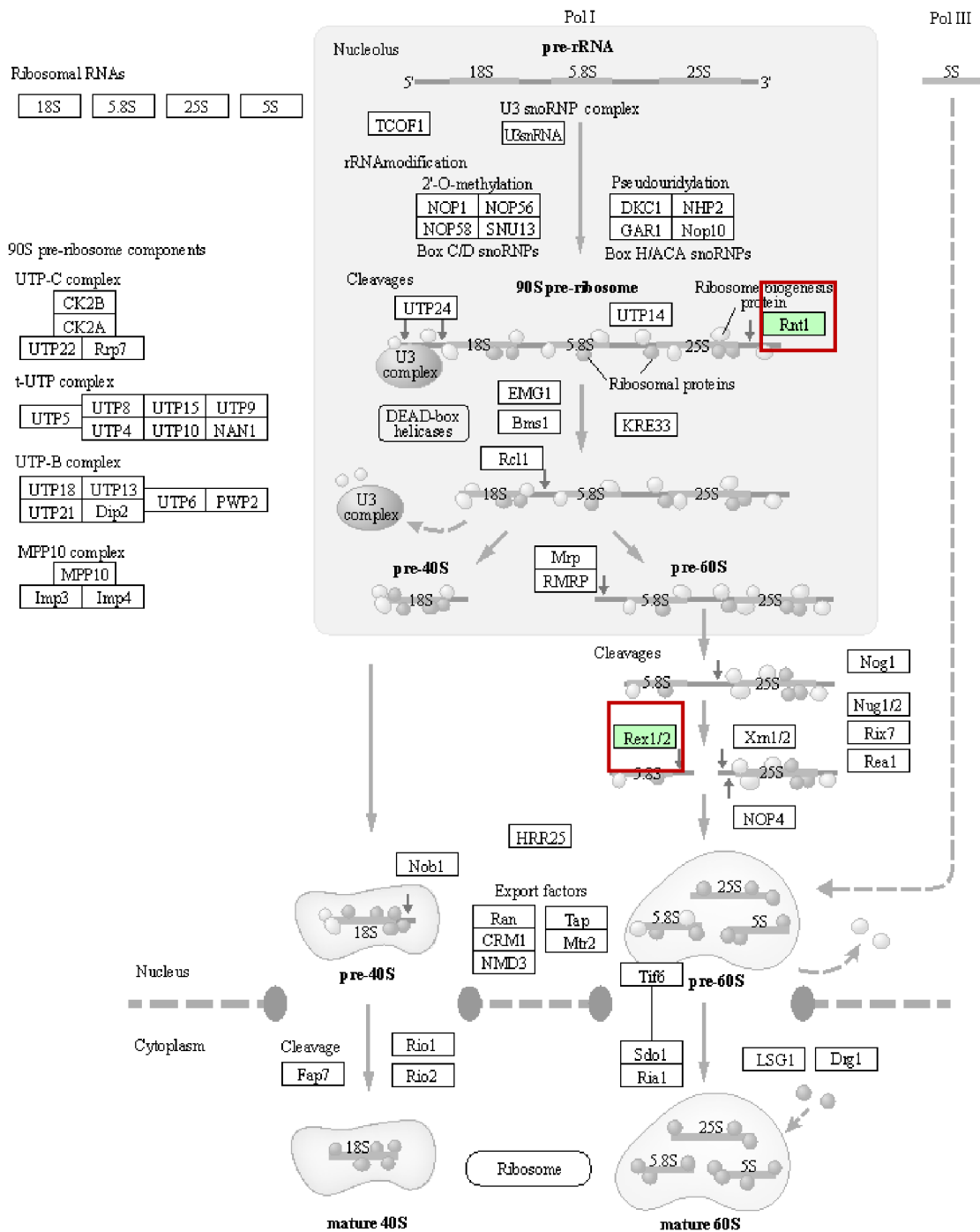
Příložené obrázky metabolických drah mají přiřazené KEGG identifikátory programem KEGG mapper (Obr. 1-7). Zelená barva KEGG identifikátorů symbolizuje, že proteiny jsou přítomné v příslušné metabolické dráze. V případě, že proteiny chybí, jsou KEGG identifikátory nevybarvené. Na obrázcích jsou pro přehlednost tyto KEGG identifikátory u analyzovaných proteinů navíc zvýrazněny červeným rámečkem.

4.5.1 Rekonstrukce proteinů metabolických drah OG0000474

V místě sekvence u linie L26_Afr_Arvican byl, jak již výše bylo zmíněno, namísto genu označeného jako hypotetický protein u původní sekvence u SE linie nalezen protein anotovaný jako DedA protein. U tohoto proteinu se nepodařilo najít žádnou metabolickou dráhu, které by byl tento protein součástí.

V případě linie N17_Afr_Steno1 se v místě sekvence nacházel enzym anotovaný jako Ribonuclease III. Enzym byl přiřazen pouze k jediné metabolické dráze, a to dráze ribozomální biogeneze u eukaryot. U bakterií rodu *Legionella* podobná dráha, ke které by mohl být protein programem KEGG mapper přiřazen, nebyla nalezena (Obr. 1).

RIBOSEME BIOGENESIS IN EUKARYOTES



03008 5/24/21
(c) Kanehisa Laboratories

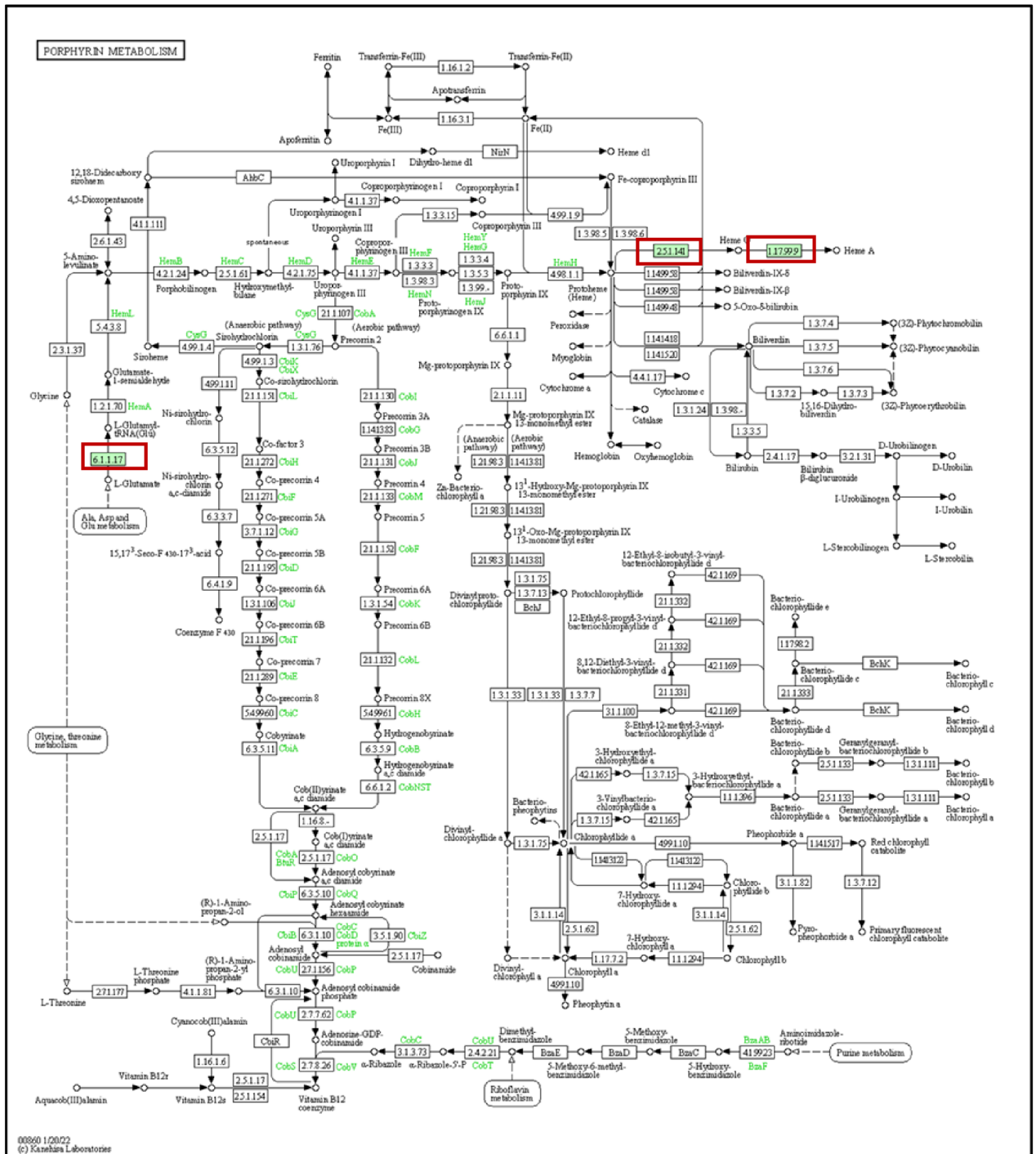
Obr. 1: Enzym Ribonuclease III (červeně zvýrazněný) v metabolické dráze ribozomální biogeneze u eukaryot.

4.5.2 *Rekonstrukce proteinů metabolických drah OG0000475*

V místě sekvence namísto genu označeného u SE linie jako hypotetický protein byl u linie k127_9727 nalezen enzym anotovaný jako Heme O synthase. Tento protein byl primárně přiřazen metabolické dráze oxidativní fosforylace.

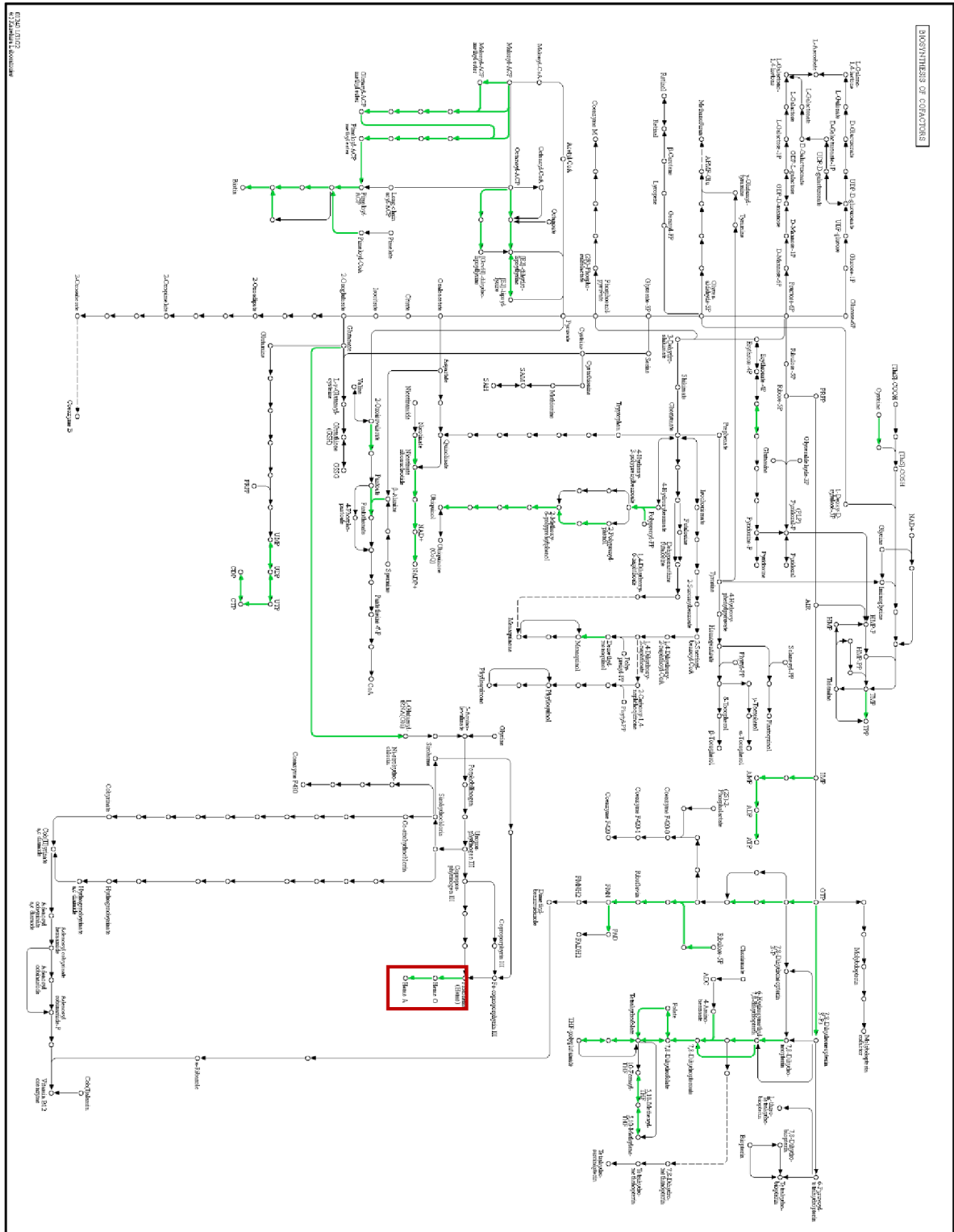
V metabolické dráze oxidativní fosforylace komplex I, tedy NADH dehydrogenáza obsahuje všechny příslušné enzymové podjednotky. Stejně tak je tomu i u komplexu II sukcinátdehydrogenázy. Naopak komplex III ubichinol-cytochrom-c-oxidoreduktáza postrádá všechny funkční enzymové podjednotky a komplexu IV cytochrom c oxidáze chybí pouze jediný funkční enzym, a to Cytochrome c oxidase subunit IV. Finálně i bakteriální ATP syntáza typu F obsahuje všechny funkční enzymové podjednotky (Obr. 2).

Další dráhou, kterou je enzym Heme O synthase součástí, je dráha metabolismu porfyrinů. Tato metabolická dráha je sdílena taktéž enzymem Heme A synthase. Metabolická dráha je zcela nekompletní z důvodu absence velkého množství funkčních enzymů. Přítomny jsou pouze tři enzymy Glutamyl-tRNA synthetase, Heme A synthase a Heme O synthase (Obr. 3).



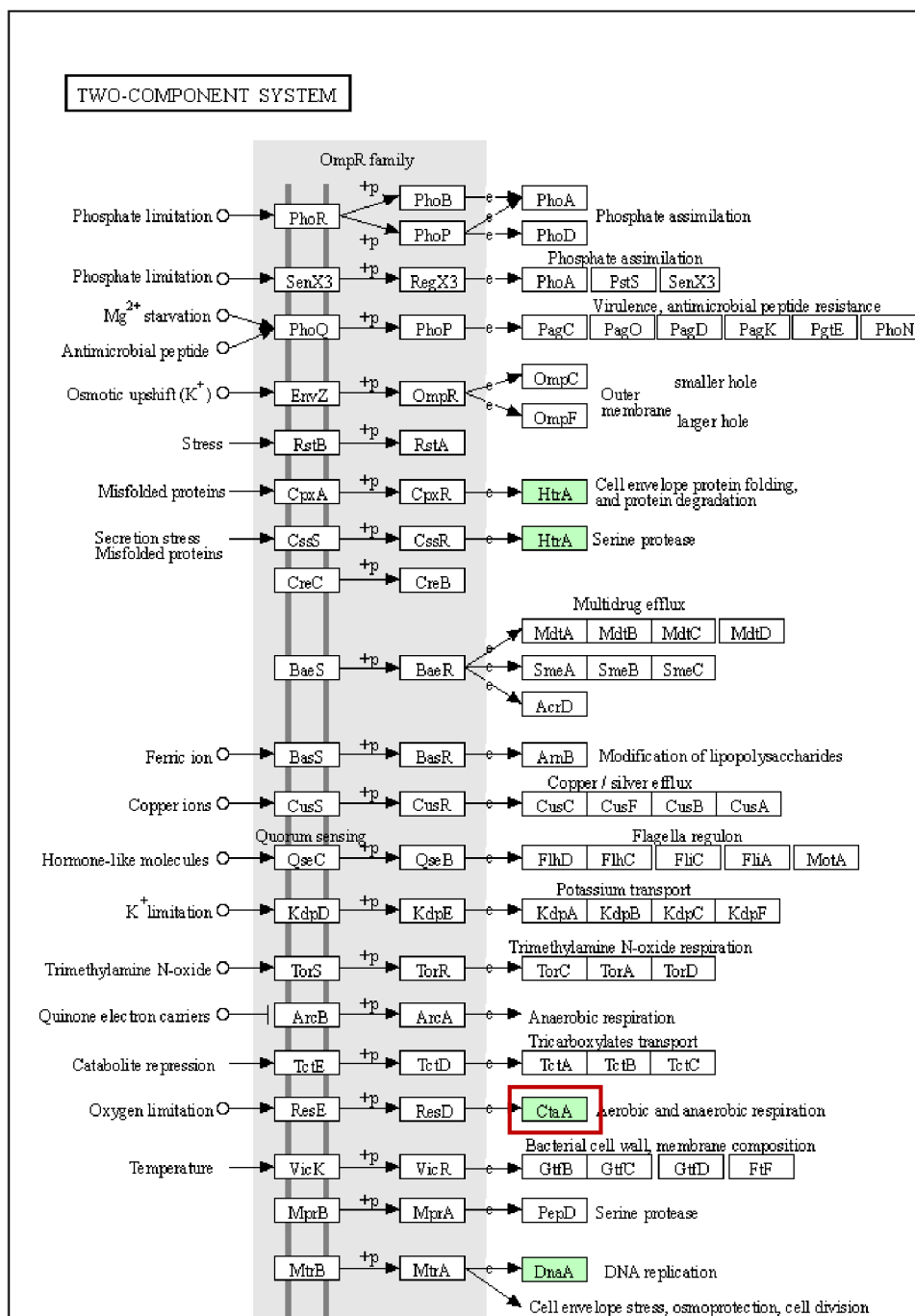
Obr. 3: Enzymy Glutamyl-tRNA synthetase, Heme A synthase a Heme O synthase (červeně zvýrazněné) v metabolismu porfyrinu u bakterie *L. polyplacis*.

Třetí metabolickou dráhou, kterou je enzym Heme O synthase součástí, a kterou zároveň sdílí s enzymem Heme A synthase, je dráha biosyntézy kofaktorů. Tato metabolická dráha postrádá velké množství funkčních enzymů (Obr. 4).



Obr. 4: Enzymy Heme A synthase a Heme O synthase (červeně zvýrazněné) v metabolismu kofaktorů u bakterie *L. polyplacis*.

Enzym Heme A synthase je také součástí dvousložkového systému. V celé metabolické dráze bylo zachováno jen 5 proteinů: Heme A synthase, Chromosomal replication initiator protein, Malate dehydrogenase, Carbon storage regulator a Glutamine synthetase. Zbytek metabolické dráhy je silně degradován (Obr. 5).



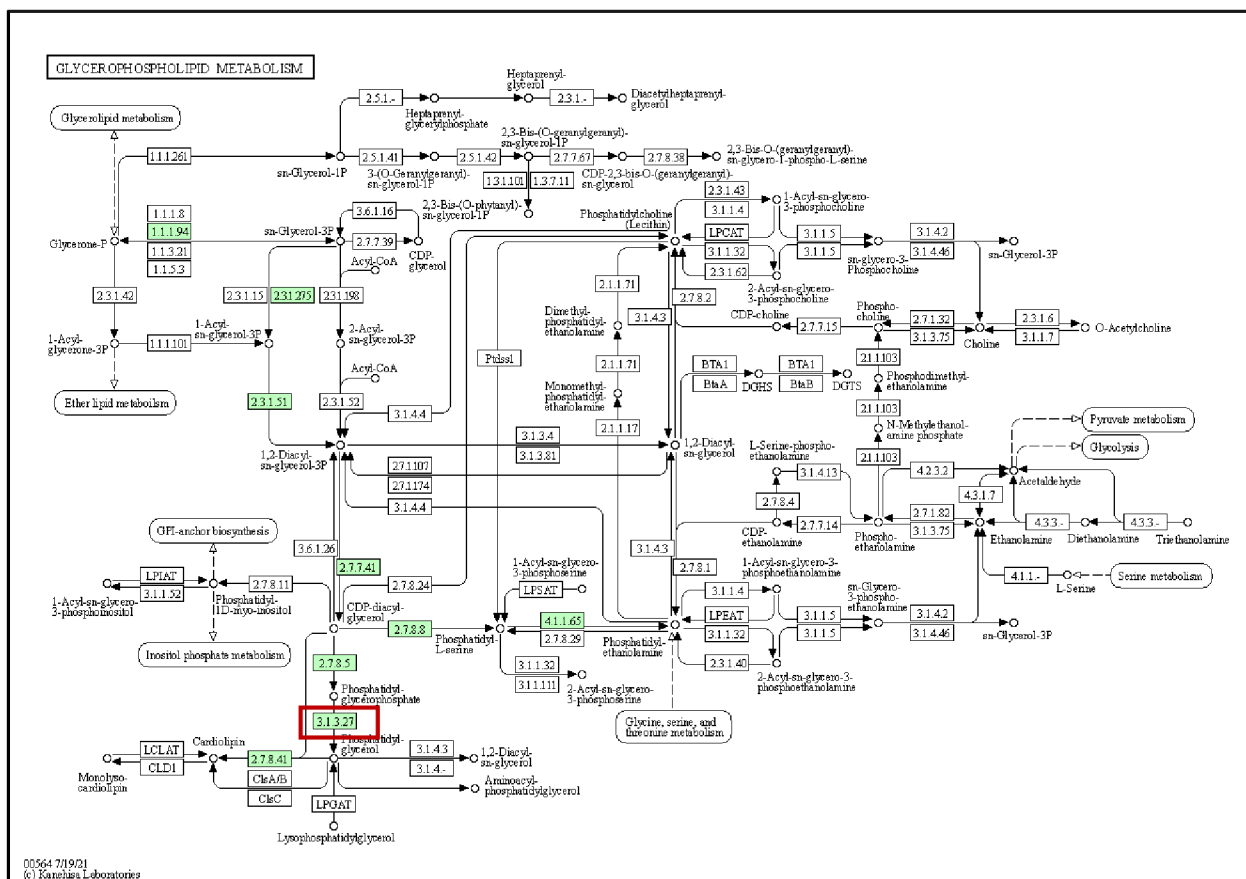
Obr. 5: Část dvousložkového systému s enzymem Heme A synthase (červeně zvýrazněný) u bakterie *L. polyplacis*.

4.5.3 *Rekonstrukce proteinů metabolických drah OG0000476*

V místě sekvence u linie L26_Afr_Arvican jak již bylo zmíněno výše, namísto genu označeného u SE linie jako hypotetický protein byl nalezen protein anotovaný jako Pap2 superfamily protein. U tohoto proteinu se nepodařilo najít žádnou metabolickou dráhu, které by byl součástí.

4.5.4 *Rekonstrukce proteinů metabolických drah OG0000478*

Sekvence byla nalezena u linií 4ROK_N, 7ROK_N, k127_9727, PNG_RR_Lspin a L26_Afr_Arvican jako anotovaný protein Phosphatidylglycerophosphatase A namísto genu označeného jako hypotetický protein u SE linie. Tento enzym je součástí dráhy metabolismu glycerolfosfolipidů a přímo se účastní fosfatidylglycerol biosyntézy (Obr. 6). V celé metabolické dráze jsou přítomny proteiny: Glycerol-3-phosphate dehydrogenase, Acyl phosphate:glycerol-3-phosphate acyltransferase, Lysophosphatidate acyltransferase, Lysophosphatidic acid acyltransferase / lysophosphatidylinositol acyltransferase, Lysophosphatidate acyltransferase, Lysocardiolipin and lysophospholipid acyltransferase, Lysophospholipid acyltransferase ½, Lysophospholipid acyltransferase, TAG lipase / steryl ester hydrolase / phospholipase A2 / LPA acyltransferase, 1-acylglycerol-3-phosphate O-acyltransferase, 1-acyl-sn-glycerol-3-phosphate acyltransferase, Phosphatidate cytidyltransferase, CDP-diacylglycerol---serine O-phosphatidyltransferase, CDP-diacylglycerol---serine O-phosphatidyltransferase, CDP-diacylglycerol---glycerol-3-phosphate 3-phosphatidyltransferase, Phosphatidylglycerophosphatase GEP4, Phosphatidylglycerophosphatase A, Phosphatidylglycerophosphatase B, Phosphatidylglycerophosphatase C a Cardiolipin synthase. Proteiny ve zbytku metabolické dráhy chybí.

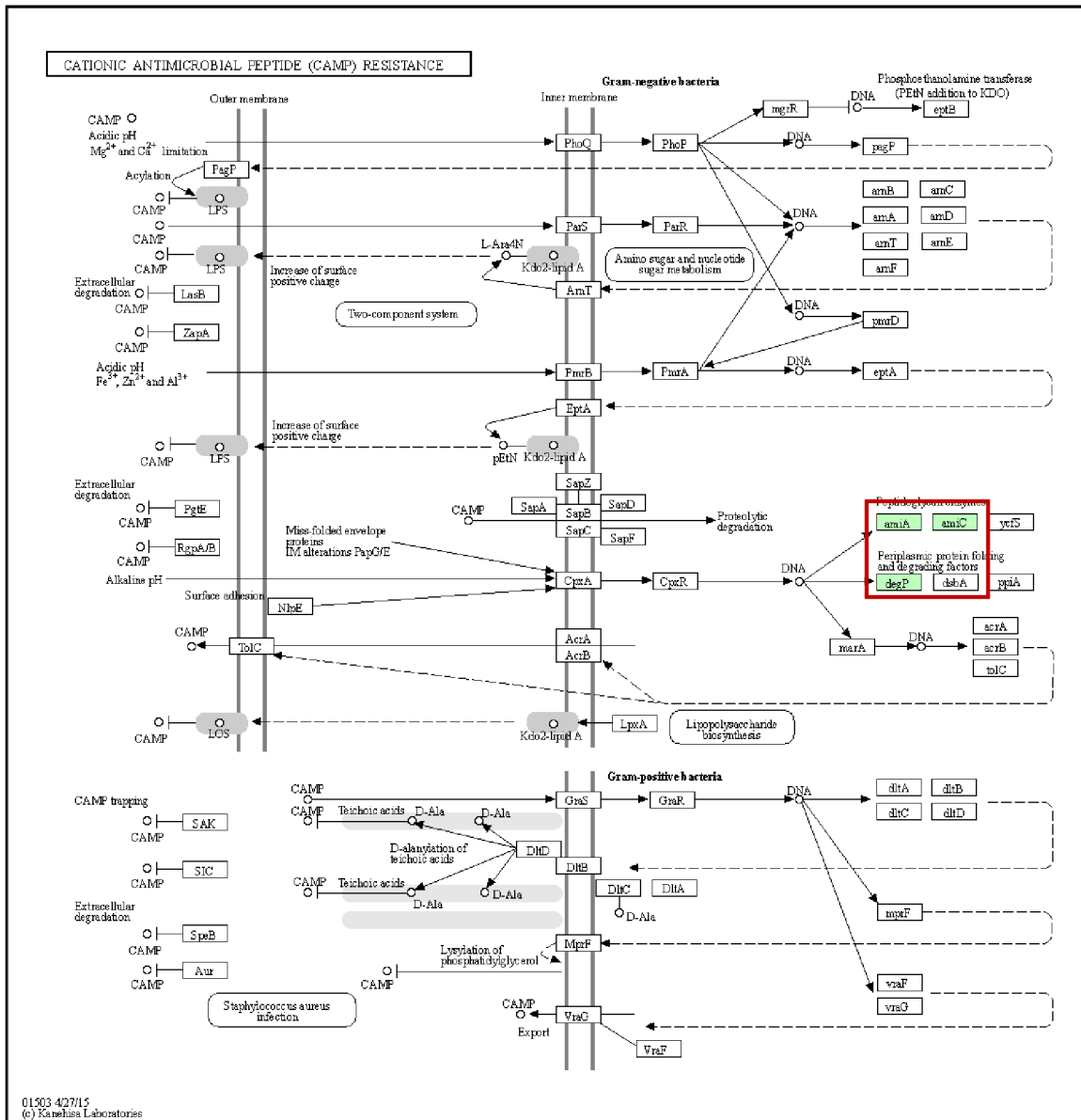


Obr. 6: Metabolismus glycerofosfolipidů s enzymem Phosphatidylglycerophosphatase A (červeně zvýrazněný) u bakterie *L. polyplacis*.

4.5.5 Rekonstrukce proteinů metabolických drah OG0000480

V místě sekvence byl u linie 4ROK_N, 7ROK_N, dále u SW linie a u linie k127_9727 přítomný namísto hypotetického genu enzym N-acetylmuramoyl-L-alanine amidase. Tento enzym je součástí pouze jedné metabolické dráhy, a to dráhy pro rezistenci kationtových antimikrobiální peptidů. Kromě zmíněného enzymu N-acetylmuramoyl-L-alanine amidase se v metabolické dráze nachází také enzym Serine protease Do (Obr. 7).

Linie L26_Afr_Arvican kromě enzymu N-acetylmuramoyl-L-alanine amidase ve druhé polovině sekvence obsahuje také TsaE protein, required for threonylcarbamoyladenosine t (6)A37 formation in tRNA. U tohoto proteinu se nepodařilo najít žádnou metabolickou dráhu, které by byl protein součástí.



Obr. 7: Enzymy N-acetylmuramoyl-L-alanine amidase a Serine protease Do (červeně zvýrazněné) v schématu rezistence kationtových antimikrobiální peptidů u bakterie *L. polyplacis*.

4.5.6 *Rekonstrukce proteinů metabolických drah OG0000481*

V první polovině sekvence u linie k127_9727 se nenacházel žádný anotovaný protein. Ve druhé polovině této sekvence byl ovšem přítomen anotovaný protein Iron binding protein SufA for iron sulfur cluster assembly.

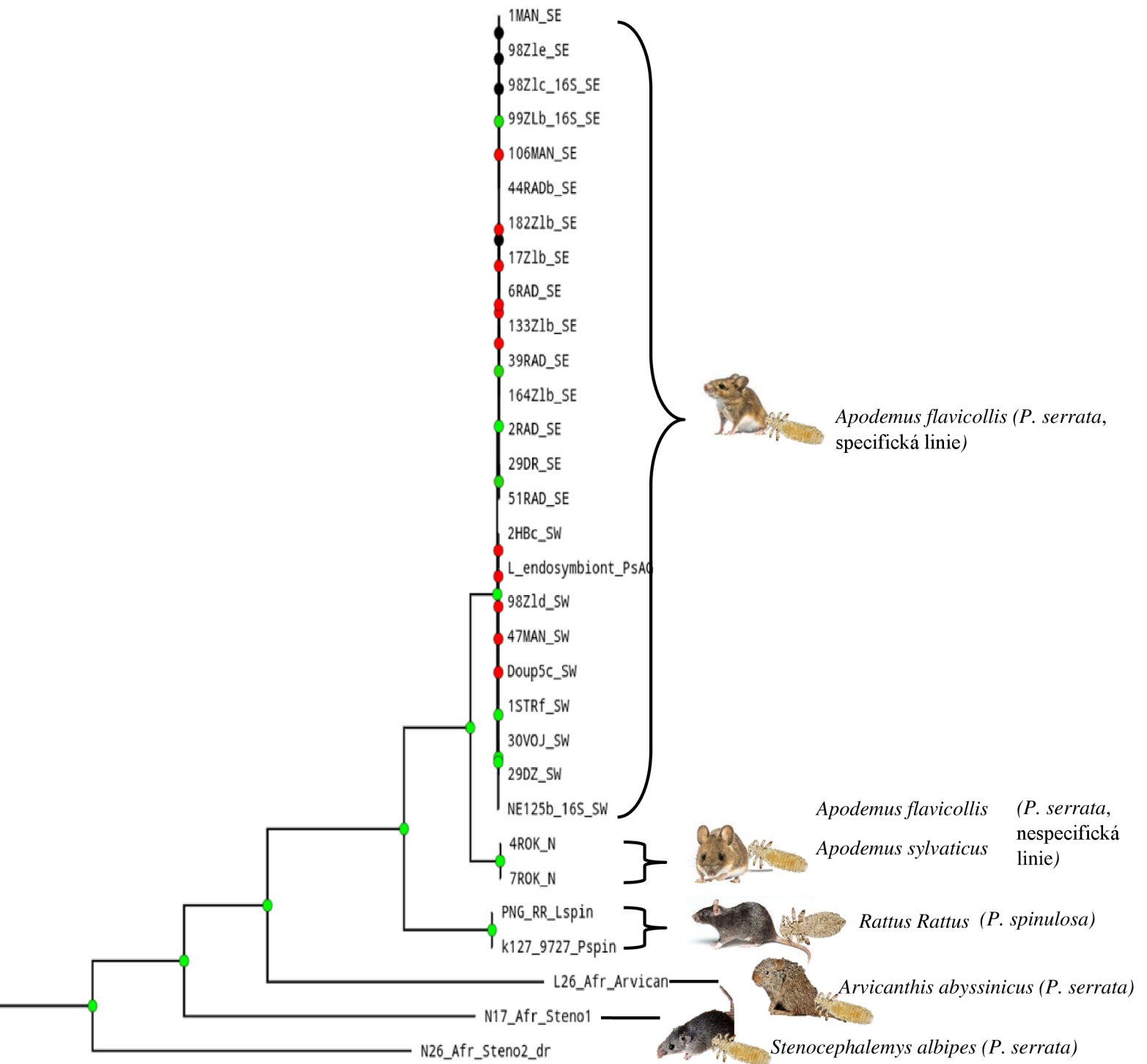
U linie PNG_RR_Lspin taktéž první polovina neobsahovala žádný anotovaný protein. Ve druhé polovině byl přítomen anotovaný protein Iron binding protein IscA for iron-sulfur cluster assembly.

U obou těchto proteinů se nepodařilo najít žádné metabolické dráhy, kterých by byly tyto proteiny součástí.

5 Diskuse

5.1 Porovnání hlavních charakteristik genomů

Porovnáním 29 genomů odlišných linií bakterie *L. polyplacis* vyšlo najevo, že všechny genomy jsou si velmi podobné. Mají dokonalou synteni, velmi podobné velikosti genomů a obsah GC. I přes tyto podobnosti je však možné říct, že se genomy mezi sebou mírně liší (Tab. 1). Zároveň došlo u všech genomů k určité úrovni degradace, což potvrzuje studii, která tvrdí, že bakterie *L. polyplacis* se vlivem dlouhého symbiotického vztahu se svým hostitelem stala jeho obligátním symbiontem (Říhová et al., 2017). Obsah GC je mezi jednotlivými liniemi velmi podobný. Jedinou linií, která se svou hodnotou GC od ostatních výrazněji liší, je fylogeneticky vzdálenější linie N17_Afr_Steno1 (Tab. 1). Menší rozdíly v obsahu GC jsou ale také mezi ostatními liniemi. Zároveň tyto linie pochází ze stejné monofyletické linie symbiontů, a jsou tedy stejně dlouhou dobu v symbiotickém vztahu se stejným hostitelem (tj. vši rodu *Polyplax*), který má pravděpodobně stejné nároky na potravu a potřeby na svého symbionta. Rozdíly v genomech by mohly být vysvětleny adaptivně v důsledku odlišných nároků na symbionta, vyplývajících z toho, že různé vsi žijí na odlišných hostitelích. Například vsi druhu *P. serrata* ze specifických linií parazitují na hostiteli *A. flavicollis* a vsi druhu *P. spinulosa* (linie k127_9727 a PNG_RR_Lspin) parazitují na hostiteli *Rattus Rattus* (Obr. 8). Změny v genomu by tedy mohly teoreticky souviset s ekologickou charakteristikou hlodavčích hostitelů. Druhou možností vysvětlující změny v genomu je striktně neutrální evoluce. Genomy symbiontů se nacházejí v pokročilejší fázi evoluce a dochází k jejich postupné degradaci, přičemž jejich velikost je v dnešní době silně redukována. U symbiotických bakterií však byly zaznamenány i případy daleko menších genomů, například již výše zmíněný redukovaný genom bakterie *Nasuiana*. V případě jednotlivých linií by degradace genomů mohla dál pokračovat. Vzhledem k tomu, že všechny linie jsou nezávislé a degradace genomu probíhá stochasticky se u různých linií bakterie *L. polyplacis* degradace genomů projevují odlišně. Je možné předpokládat, že ve změnách v genomu by se mohlo podařit nalézt i nějaké funkční sekvence genů. Za předpokladu, že změny odpovídají druhé možnosti, je velice pravděpodobné, že změny budou probíhat náhodně a budou zasahovat pouze neesenciální geny.



Obr. 8: Fylogenetický strom jednotlivých linií bakterie *L. polyplacis* s jejich hostiteli (obrázek převzat z dosud nepublikované analýzy spolupracovníka Juana Pablo Cardenase).

5.2 Sdílené geny

Prvním faktorem v posouzení, proč se počet společných genů liší může být způsob jejich sdílení a jejich rozdílů. Vzhledem k našim analýzám vypadají nejzajímavěji následující 4 kategorie sdílených genů (Graf 1). První kategorii charakterizuje největší množství sdílených genů, které mají všechny linie dohromady oproti linii N17_Afr_Steno1. Rozdíly poukazují na to, že degenerace genomu a jeho změny neprobíhají v korelaci s fylogenezí. Druhou kategorií je počet 15 genů sdílených pouze v rámci N linie, který ji odlišuje od všech ostatních linií. N linie je zastoupena výhradně dvěma vzorky pocházející z nespécifické linie vší *P. serrata*, které žijí na obou druzích myších *A. flavicollis* a *A. sylvaticus*. Třetí kategorii tvoří celkový počet genů sdílených liniemi SE a SW (tedy specifickými liniemi). Poslední kategorii charakterizují geny nacházející se pouze u SE linie. Z ekologického hlediska nemáme žádné informace o rozdílech mezi SE a SW linií. Jediná známá informace, která tyto linie rozděluje, je jejich genetická struktura v důsledku hybridní zóny, které linie tvoří (Martinů et al., 2020).

Rozdíly, které je možné vidět v grafu mohou být dány dvěma faktory. Prvním faktorem může být opravdu chybějící ortolog, který linie ztratily. Druhým faktorem může být nerozeznání ortolog programem Orthofinder (například proto, že v genomech linií došlo k určité úrovni degradace) Proto není možné usuzovat s jistotou, že nenalezené geny doopravdy v genomu chybí. V následném dohledání sedmi genů označených jako chybějících (tj. nenalezených pomocí Orthofinder) vyšlo najevo, že se v genomu nachází. Velmi vysoká podoba a syntenie těchto genů navíc poukazuje na to, že všechny linie (i těch ze vzdálených hostitelů jako je například N17_Afr_Steno1) jsou součástí jednoho fylogenetického klastru neboli jsou potomky jednoho symbiotického předka.

V této bakalářské práci jsem se zaměřila především na další zhodnocení genů ze symbiontů vší druhu *P. serrata*, neboť bylo k dispozici více vzorků a sloužily jako věrohodný podklad. Genom linie N26_Afr_Steno2_dr není součástí této bakalářské práce, protože v době zpracování nebyl k dispozici.

5.3 Hypotetické geny

Hypotetické geny se v genomech linií nacházely ve dvou podobách. První podoba, která se nacházela u dvou ortoskupin OG00000477 a OG000000481 byla taková, že se v místě sekvence nacházel gen anotovaný jako hypotetický gen a v místě sekvence u jiných linií byla přítomna pouze neoznačená sekvence. U zbylých ortoskupin se v místě sekvence nacházel anotovaný hypotetický gen a u jiných linií byl anotovaný konkrétní gen (Tab. 3). V důsledku zařazení těchto genů výhradně do kategorie hypotetických genů není možné zhodnotit jejich potenciální význam. Vzhledem k tomu, že se jedná o vysoce degenerované genomy, je možné, že anotační programy geny nemusely rozeznat. Kvůli odlišným algoritmům programů Orthofinder a RAST došlo také k rozdílnému rozeznávání sekvencí. Sekvence tedy nebyly anotovány stejně. Z důvodu rozdílných výsledků obou programů jsem se u hypotetických genů snažila vyhledat podobnost s konkrétními geny z čeledi *Legionellaceae* s pomocí algoritmu BLASTp a dále s pomocí algoritmu Conserved domains v rámci všech známých sekvencí (Tab. 5). Ve výsledcích těchto algoritmů byl jediný případ, kdy metody identifikovaly konkrétní gen stejně. Jednalo se o gen Heme O synthase (OG00000475). Hem O i hem A jsou deriváty protohemu IX a oba tyto kofaktory jsou potřeba pro správnou funkci základního procesu buněčného dýchání a ke dráze oxidativní fosforylace byl i gen Heme O synthase přiřazen. Tyto dva hemy se uplatňují při katalyzaci reakce, při níž dochází k redukci molekulárního kyslíku na vodu v komplexu cytochrom c oxidázy (Mogi et al., 1994; Brown et al., 2004). Můj výsledek ukazuje, že nekompletnost této metabolické dráhy v některých genomech může být dána špatnou anotací anotačním programem, neboť došlo k takové degeneraci genu, že jej nebylo možné programem rozeznat. V případě proteinů Serine hydrolase a Beta-lactamase family protein (OG00000478) došlo k identifikaci genů pouze algoritmem BLASTp. Serinové hydrolázy jsou jednou z největších doposud popsáných tříd enzymů. Třída je charakterizována především přítomností nukleofilního serinu, který je využíván pro hydrolýzu substrátu. Do této skupiny patří například amidázy, cholinesterázy nebo fosfolipázy (Simon and Cravatt, 2010). Enzymy ze skupiny beta-laktamáz patří do třídy serin hydroláz a poskytují bakteriím rezistenci vůči určité skupině antibiotik označených jako beta-laktamy, které bakteriím zabraňují v tvorbě buněčné stěny. Enzymy dokážou otvírat beta-laktam, tedy čtyř atomový kruh nacházející se v molekulární struktuře antibiotik a deaktivovat tak antibakteriální vlastnosti molekuly (Abraham and Chain, 1940; Neu, 1969)

Výše zmíněný gen Serine hydrolase a Beta-lactamase family protein u OG00000478 se podařilo najít algoritmem tBLASTn a ve vytvořené Geneious Prime BLAST databázi

nesymbiotických druhů legionel (Tab. 4). Ovšem výsledek obou algoritmů se jen částečně shoduje. Algoritmus tBLASTn identifikoval gen jako Phosphatidylglycerophosphatase A a Geneious prime BLAST identifikoval gen jako Phosphatidylglycerophosphatase A a také jako Serine hydrolase. U OG00000474 je naopak výsledek obou algoritmů úplně odlišný. Patrně se tedy jedná o gen, který ztratil svou charakteristiku a degeneruje u všech linií. Zajímavým případem je gen N-acetylmuramoyl-L-alanine amidase u OG00000480, který byl anotován u všech linií kromě SE linie (Tab. 3), ale zároveň se jej ani nepodařilo dohledat algoritmem BLASTp proti legionelám a ani algoritmem Conserved domains bez omezení (Tab. 5). Podobným případem je gen Iron binding protein u OG00000481, kdy byl gen u dvou linií shodně anotován (Tab. 3), ale u ostatních linií chybí a nepodařilo se jej dohledat. Příčinou u SE linie může být obrovská degradace genu a ztráta jeho funkčnosti. V závěru je tedy možné říct, že jsou pouze dva geny Heme O synthase a Phosphatidylglycer-ophosphatase A u symbiotických legionel potvrzeny.

Celkově je u všech linií velké množství hypotetických genů a je tedy možné, že zdánlivě chybějící metabolické kapacity v nich mohou být schované. Pro potvrzení této hypotézy by byla potřebná další analýza.

5.4 *Metabolické dráhy*

Anotovaný gen Ribonuclease III u OG00000474 byl přiřazen metabolické dráze ribozomální biogeneze u eukaryot (Obr. 1). Tato metabolická dráha je zásadní pro většinu buněčných procesů jako například růst a buněčné dělení (Kumar, 2021). Enzym Ribonuclease III hraje také důležitou roli v regulaci exprese genů a zpracování ribozomální RNA (Lejars et al., 2021). Vzhledem k tomu, že program KEGG mapper enzym přiřadil k eukaryotické metabolické dráze namísto k dráze prokaryotické není možné usoudit, zda je dráha u bakterie *L. polyplacis* a u nesymbiotických druhů legionel funkční.

Enzym Heme O synthase u OG00000475 byl primárně přiřazen k metabolické dráze oxidativní fosforylace (Obr. 2). Není zdaleka jisté, zda tato dráha je funkční, neboť mitochondriální komplex III tedy ubichinol-cytochrom-c-oxidoreduktáza postrádá všechny funkční enzymové podjednotky. Ke ztrátě enzymových podjednotek v komplexu III většinou dochází následkem mutací v několika genech (Garone et al., 2018; Fernández-Vizarra et al., 2009; Bénit et al., 2009). Ovšem v některých případech jako je tomu v případě gramnegativní bakterie *Flavobacterium johnsoniae* jsou bakterie schopné si vytvořit alternativní mechanismy, kterými mohou ubichinol oxidovat (Lorencik et al., 2021; Lennarz a Lane,

2013). Z toho je možné usoudit že *L. polyplacis* by mohla některý z alternativních mechanismů využívat. Co se týče komplexu IV, tomu chyběl pouze jeden funkční enzym, tedy Cytochrome c oxidase subunit IV. Práce Zuzany Zadinové ale ukazuje, že enzym Heme O synthase by měl v komplexu IV chybět, což ovšem výsledky mé bakalářské práce vyvracejí, protože sekvence tohoto genu byla u jednotlivých linií přítomna, ale bez anotace nebo pod anotací hypotetický gen. Následně byla sekvence dohledána a přiřazena k enzymu Heme O synthase algoritmy Conserved domains a BLASTp. Je tedy možné, že tento komplex v případě dodání potřebného enzymu od hostitele může být funkční anebo může být enzym obsažen v nějakém hypotetickém genu. Další dráhou, kterou je tento enzym součástí společně s enzymem Heme A synthase, je dráha metabolismu porfyrinů (Obr. 3). Porfyriny jsou meziprodukty v biosyntéze esenciálních molekul, včetně hemu nebo kobalaminu (Barnard et al., 2020). V případě degradace této dráhy je u bakterií negativně ovlivněna funkce oxidativní fosforylace a syntézy hemu. Tato metabolická dráha je u *L. polyplacis* pravděpodobně nefunkční, protože přítomny jsou pouze tři enzymy Glutamyl-tRNA synthetase, Heme A synthase a Heme O synthase. Heme O synthase a Heme A synthase spolu také sdílejí metabolickou dráhu pro biosyntézu kofaktorů (Obr. 4). Ta je vzhledem k absenci velkého množství enzymů jen částečně funkční. Vzhledem k této úrovni degradace bakterie nedokáže syntetizovat velké množství látek například aminokyseliny cystein a glycin nebo vitaminy pyridoxin a tiamin. Je možné, že některé dráhy v biosyntéze kofaktorů by mohly být funkční pouze v případě dodání potřebných látek či enzymů hostitelem. Heme A synthase se finálně účastní dvousložkového systému (Obr. 5). Tento systém díky přenosu signálu umožňuje bakterii vnímat, reagovat a přizpůsobovat se změnám ve vnějším prostředí nebo naopak v intracelulárním stavu. Je složený z histidin kinázy a regulátoru cytoplazmatické odpovědi. Při detekci specifického signálu dochází k autofosforylaci a přenosu fosfátové skupiny na regulátor. V důsledku přenosu regulátor působí jako transkripční faktor, který způsobí změnu genové exprese (Mitrophanov and Groisman, 2008). Vzhledem k životní strategii bakterie *L. polyplacis* a životem v předvídatelném prostředí je možné, že je dráha z tohoto důvodu výrazně degradovaná.

K metabolismu glycerolfosfolipidů byl u OG00000478 přiřazen enzym Phosphatidylglycerophosphatase A, který se přímo účastní fosfatidylglycerol biosyntézy (Obr. 6). Fosfatidylglyceroly se nachází téměř ve všech typech bakterií, kde se uplatňují jako hlavní složky buněčných membrán. Vzhledem k přítomnosti enzymů je možné, že je tato biosyntéza funkční nebo je funkční v případě dodání potřebných enzymů hostitelem.

Enzym N-acetylmuramoyl-L-alanine amidase u OG0000480 byl přiřazen k metabolické dráze rezistence kationtových antimikrobiální peptidů (CAMP) (Obr. 7). Tato metabolická dráha bakteriím pomáhá při obraně proti CAMP, neboť tyto peptidy narušují celistvost vnitřní a vnější membrány bakterií a následně zabíjejí bakteriální buňky (Dorin et al., 2015). Vzhledem k přítomnosti pouze tří enzymů v celé metabolické dráze je možné považovat tuto dráhu za nefunkční. Stejně jako v případě nekompletnosti dvousložkového systému je možné vzhledem k životní strategii bakterie tuto dráhu považovat za velmi degradovanou a tedy nefunkční.

6 Závěr

Porovnáním a anotací 29 genomů linií bakterie *L. polyplacis* se ukázalo, že všechny linie se přes značnou shodu ve velikosti genomů a obsahu GC od sebe částečně liší.

S pomocí algoritmů BLASTp a Conserved domains bylo možné většinu hypotetických genů identifikovat a přiřadit ke konkrétním genům u nichž BLAST rozpoznal podobnost s konkrétními geny z čeledi *Legionellaceae* a s geny u odlišných organismů. Tato analýza nám poskytla případné informace o původních genech, které se v genomu vyskytovaly, než došlo k jejich degradaci.

Díky rekonstrukci metabolických drah anotovaných proteinů programem KEGG mapper a porovnáním s nesymbiotickým druhem bakterie *L. pneumophila subsp. pneumophila* vyšlo najevo, že většina těchto drah je silně degradovaná a pravděpodobně nefunkční vzhledem k absenci celé řady enzymů. Dráha, kterou považují za teoreticky funkční, je dráha pro biosyntézu fosfatidylglycerolu v metabolismu glycerolfosfolipidů. K potvrzení této skutečnosti by se tato problematika mohla stát cílem při dalším zkoumání. Stejně je to i v případě metabolické dráhy pro oxidativní fosforylaci.

Menšina hypotetických genů byla vytvořenou databází z nesymbiotických druhů bakterie rodu *Legionella* přiřazena ke konkrétním genům, přičemž se některé výsledky shodovaly nebo částečně shodovaly s algoritmy BLASTp a Conserved domains.

7 Seznam literatury

- Abraham, E. P., Chain, E. (1940). An Enzyme from Bacteria able to Destroy Penicillin. *Nature*, 146(3713), 837-837.
- Abril, J. F., Castellano, S. (2019). Genome Annotation. In *Encyclopedia of Bioinformatics and Computational Biology*. Elsevier.
- Aksoy, S. (1995). *Wigglesworthia gen. nov.* and *Wigglesworthia glossinidia sp. nov.*, Taxa Consisting of the Mycetocyte-Associated, Primary Endosymbionts of Tsetse Flies. *International Journal of Systematic Bacteriology*, 45(4), 848-851.
- Allen, J. M., Reed, D. L., Perotti, M. A., Braig, H. R. (2007). Evolutionary Relationships of “*Candidatus Riesia spp.*” Endosymbiotic Enterobacteriaceae Living within Hematophagous Primate Lice. *Applied and Environmental Microbiology*, 73(5), 1659-1664.
- Altenhoff, A. M., Boeckmann, B., Capella-Gutierrez, S., Dalquen, D. A., DeLuca, T., Forslund, K., Huerta-Cepas, J., Linard, B., Pereira, C., Prysycz, L. P., Schreiber, F., da Silva, A. S., Szklarczyk, D., Train, C. -M., Bork, P., Lecompte, O., von Mering, C., Xenarios, I., Sjölander, K., et al. (2016). Standardized benchmarking in the quest for orthologs. *Nature Methods*, 13(5), 425-430.
- Altenhoff, A. M., Schneider, A., Gonnet, G. H., Dessimoz, C. (2010). OMA 2011: orthology inference among 1000 complete genomes. *Nucleic Acids Research*, 39
- Altschul F.S., Gish W., Lipman D.J., Miller W., Myers E.W. (1990). Basic Local Alignment Search Tool. *Journal Of Molecular Biology*. 215(3), 403-410.
- Altschul, F.S., Lipman, D.J., Madden, L.T., Miller, W., Schäffer, A. A., Zhang, J., Zhang, Z. (1997). Gapped Blast And Psi-Blast: A New Generation Of Protein Database Search Programs. *Nucleic Acids Research*. 25 (17), 3389-3402.
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., et al. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9(1).

Barnard, E., Johnson, T., Ngo, T., Arora, U., Leuterio, G., McDowell, A., Li, H., D’Orazio, S. E. F. (2020). Porphyrin Production and Regulation in Cutaneous Propionibacteria. *MSphere*, 5(1).

Baumann, P. (2005). Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking Insects. *Annual Review of Microbiology*, 59(1), 155-189.

Bénit, P., Lebon, S., Rustin, P. (2009). Respiratory-chain diseases related to complex III deficiency. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1793(1), 181-185.

Boyd, B. M., Allen, J. M., Nguyen, N. -P., Vachaspati, P., Quicksall, Z. S., Warnow, T., Mugisha, L., Johnson, K. P., Reed, D. L. (2017). Primates, Lice and Bacteria: Speciation and Genome Evolution in the Symbionts of Hominid Lice. *Molecular Biology and Evolution*, 34(7), 1743-1757.

Boyd, B. M., Reed, D. L. (2012). Taxonomy of lice and their endosymbiotic bacteria in the post-genomic era. *Clinical Microbiology and Infection*, 18(4), 324-331.

Braendle, C., Miura, T., Bickel, R., Shingleton, A. W., Kambhampati, S., Stern, D. L., Benjamin Normark. (2003). Developmental Origin and Evolution of Bacteriocytes in the Aphid–Buchnera Symbiosis. *PLoS Biology*, 1(1).

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Thomason, J. A., Stevens, R., Vonstein, V., Wattam, A. R., Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, 5(1).

Brown, B. M., Wang, Z., Brown, K. R., Cricco, J. A., Hegg, E. L. (2004). Heme O synthase and Heme A synthase from *Bacillus subtilis* and *Rhodobacter sphaeroides* Interact in *Escherichia coli*. *Biochemistry*, 43(42), 13541-13548.

Brownlie, J. C., Johnson, K. N. (2009). Symbiont-mediated protection in insect hosts. *Trends in Microbiology*, 17(8), 348-354.

Bryant H.S., Marchler-Bauer A. (2004). D-Search: Protein Domain Annotations On The Fly. *Nucleic Acids Research*. 32, 327-331.

- Buchner, P. (1965). *Endosymbiosis of Animals with Plant Microorganisms*. Interscience Publishers. New York.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10(1).
- Dale, C., Maudlin, I. (1999). *Sodalis gen. nov.* and *Sodalis glossinidius sp. nov.*, a Microaerophilic Secondary Endosymbiont of the Tsetse Fly *Glossina morsitans morsitans*. *International Journal of Systematic and Evolutionary Microbiology*, 49(1), 267-275.
- Dorin, J. R., McHugh, B. J., Cox, S. L., and Davidson, D. J. (2015). Mammalian Antimicrobial Peptides; Defensins and Cathelicidins. In *Molecular Medical Microbiology*. Elsevier.
- Douglas, A. E. (1998). Nutritional Interactions in Insect-Microbial Symbioses: Aphids and Their Symbiotic Bacteria Buchnera. *Annual Review of Entomology*, 43(1), 17-37.
- Duperron, S. (2017). Symbiosis and Other Functions. In *Microbial Symbioses*, Elsevier.
- Emms, D. M., Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, 16(1).
- Fernández-Vizarra, E., Tiranti, V., Zeviani, M. (2009). Assembly of the oxidative phosphorylation system in humans: What we have learned by studying its defects. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1793(1), 200-211.
- Fields, B. S. (1996). The molecular ecology of legionellae. *Trends in Microbiology*, 4(7), 286-290.
- Garone, C., Minczuk, M., Ghezzi, D., Zeviani, M. (2018). Human diseases associated with defects in assembly of OXPHOS complexes. *Essays in Biochemistry*, 62(3), 271-286.
- Gehlenborg, N. (2019). UpSetR: A More Scalable Alternative to Venn and Euler Diagrams for Visualizing Intersecting Sets. *R package version 1.4.0*.
- Gosalbes, M. J., Latorre, A., Lamelas, A., Moya, A. (2010). Genomics of intracellular symbionts in insects. *International Journal of Medical Microbiology*, 300(5), 271-278.

Grus, W. E., Zhang, J. (2008). Human Lineage-specific Gene Inactivation. In *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons.

Hammoud, A., Louni, M., Missé, D., Cortaredona, S., Fenollar, F., Mediannikov, O. (2022). Phylogenetic relationship between the endosymbiont “Candidatus *Riesia pediculicola*” and its human louse host. *Parasites & Vectors*, 15(1).

Hildebrand, F., Meyer, A., Eyre-Walker, A., Nachman, M. W. (2010). Evidence of Selection upon Genomic GC-Content in Bacteria. *PLoS Genetics*, 6(9).

Hypša, V., Křížek, J. (2007). Molecular Evidence for Polyphyletic Origin of the Primary Symbionts of Sucking Lice (Phthiraptera, Anoplura). *Microbial Ecology*, 54(2), 242-251.

John A. Hvala and Troy E. Wood (2012). "Speciation: Introduction". In *Encyclopedia of Life Sciences (ELS)*. John Wiley & Sons.

Kaiwa, N., Hosokawa, T., Kikuchi, Y., Nikoh, N., Meng, X. Y., Kimura, N., Ito, M., Fukatsu, T. (2010). Primary Gut Symbiont and Secondary, Sodalis – Allied Symbiont of the Scutellerid Stinkbug *Cantao ocellatus*. *Applied and Environmental Microbiology*, 76(11), 3486-3494.

Kanehisa, M. Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28-35.

Kanehisa, M., Sato, Y., Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, 428(4), 726-731.

Kikuchi, Y. (2009). Endosymbiotic Bacteria in Insects: Their Diversity and Culturability. *Microbes and Environments*, 24(3), 195-204.

Kumar, V. (2021). Ribosomal biogenesis in eukaryotes. In *Emerging Concepts in Ribosome Structure, Biogenesis, and Function*.

Lejars, M., Kobayashi, A., Hajnsdorf, E. (2021). RNase III, Ribosome Biogenesis and Beyond. *Microorganisms*, 9(12).

Lennarz, W. J., Lane, M. D. (2013). *Encyclopedia of biological chemistry*. Academic Press.

Li, L., Stoeckert, C. J., Roos, D. S. (2003). OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Research*, 13(9), 2178-2189.

Liu, Y., Harrison, P. M., Kunin, V., Gerstein, M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biology*, 5(9).

Longdon, B., Jiggins, F. M. (2012). Vertically transmitted viral endosymbionts of insects: do sigma viruses walk alone? *Proceedings of the Royal Society B: Biological Sciences*, 279(1744), 3889-3898.

Lorencik, K., Ekiert, R., Zhu, Y., McBride, M. J., Gennis, R. B., Sarewicz, M., Osyczka, A., Gralnick, J. A. (2021). The Monoheme c Subunit of Respiratory Alternative Complex III Is Not Essential for Electron Transfer to Cytochrome aa 3 in *Flavobacterium johnsoniae*. *Microbiology Spectrum*, 9(1).

Martínez-Cano, D. J., Reyes-Prieto, M., Martínez-Romero, E., Partida-Martínez, L. P., Latorre, A., Moya, A. A. Dela, L. (2015). Evolution of small prokaryotic genomes. *Frontiers in Microbiology*, 5.

Martinů, J., Hypša, V., Štefka, J. (2018). Host specificity driving genetic structure and diversity in ectoparasite populations: Coevolutionary patterns in *Apodemus* mice and their lice. *Ecology and Evolution*, 8(20), 10008-10022.

Martinů, J., Štefka, J., Poosakkannu, A., Hypša, V. (2020). "Parasite turnover zone" at secondary contact: A new pattern in host-parasite population genetics. *Molecular Ecology*, 29(23), 4653-4664.

Masson, F., Balmand, S., Heddi, A., Vallier, A., Vigneron, A., Vincent-Monégat, C., Zaidman-Rémy, A. (2015). Systemic Infection Generates a Local-Like Immune Response of the Bacteriome Organ in Insect Symbiosis. *Journal of Innate Immunity*, 7 (3), 290-301.

Masson, F., Heddi, A., Zaidman-Rémy, A. (2016). Antimicrobial Peptides and Cell Processes Tracking Endosymbiont Dynamics. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371 (1695).

McCutcheon, J. P., Moran, N. A. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biology and Evolution*, 2, 708-718.

McCutcheon, J. P., Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, 10(1), 13-26.

McFall-Ngai, M., Hadfield, M. G., Bosch, T. C. G., Carey, H. V., Domazet-Lošo, T., Douglas, A. E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S. F., Hentschel, U., King, N., Kjelleberg, S., Knoll, A. H., Kremer, N., Mazmanian, S. K., Metcalf, J. L., Nealson, K., Pierce, N. E., et al. (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences*, 110(9), 3229-3236.

Michaux, J. R., Libois, R., Filippucci, M. -G. (2005). So close and so different: comparative phylogeography of two small mammal species, the Yellow-necked fieldmouse (*Apodemus flavicollis*) and the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. *Heredity*, 94(1), 52-63.

Mitrophanov, A. Y., Groisman, E. A. (2008). Signal integration in bacterial two-component regulatory systems. *Genes Development*, 22(19), 2601-2611.

Mogi, T., Saiki, K., Anraku, Y. (1994). Biosynthesis and functional role of haem O and haem A. *Molecular Microbiology*, 14(3), 391-398.

Moran, N. A. (2002). Microbial Minimalism. *Cell*, 108(5), 583-586.

Moran, N. A., McCutcheon, J. P., Nakabachi, A. (2008). Genomics and Evolution of Heritable Bacterial Symbionts. *Annual Review of Genetics*, 42(1), 165-190.

Moran, N. A., Munson, M. A., Baumann, P., Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 253(1337), 167-171.

Moran, N. A., Telang, A. (1998). Bacteriocyte-Associated Symbionts of Insects. *BioScience*, 48(4), 295-304.

Neu, H. C. (1969). Effect of β -Lactamase Location in *Escherichia coli* on Penicillin Synergy. *Applied Microbiology*, 17(6), 783-786.

Nichio, B. T. L., Marchaukoski, J. N., Raittz, R. T. (2017). New Tools in Orthology Analysis: A Brief Review of Promising Perspectives. *Frontiers in Genetics*, 8.

Ostlund, G., Schmitt, T., Forslund, K., Kostler, T., Messina, D. N., Roopra, S., Frings, O., Sonnhammer, E. L. L. (2009). InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Research*, 38

- Pevsner, J. (2009). *Bioinformatics and functional genomics*. Wiley-Blackwell.
- Rio, R.V. M., Attardo, G. M., Weiss, B. L. (2016). Grandeur Alliances: Symbiont Metabolic Integration and Obligate Arthropod Hematophagy. *Trends in Parasitology*, 32(9), 739-749.
- RStudio Team (2021). RStudio: Integrated Development for R. RStudio.
- Řihová, J., Nováková, E., Husník, F., Hypša, V. (2017). Legionella Becoming a Mutualist: Adaptive Processes Shaping the Genome of Symbiont in the Louse *Polyplax serrata*. *Genome Biology and Evolution*, 9(11), 2946-2957.
- Simon, G. M., Cravatt, B. F. (2010). Activity-based Proteomics of Enzyme Superfamilies: Serine Hydrolases as a Case Study. *Journal of Biological Chemistry*, 285(15), 11051-11055.
- Štefka, J., Hypša, V. (2008). Host specificity and genealogy of the louse *Polyplax serrata* on field mice, *Apodemus* species: A case of parasite duplication or colonisation?. *International Journal for Parasitology*, 38(6), 731-741.
- Toh, H., Weiss, B. L., Perkin, S. A. H., Yamashita, A., Oshima, K., Hattori, M., Aksoy, S. (2006). Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. *Genome Research*, 16(2), 149-156.
- Trachana, K., Larsson, T. A., Powell, S., Chen, W. -H., Doerks, T., Muller, J., Bork, P. (2011). Orthology prediction methods: A quality assessment using curated protein families. *BioEssays*, 33(10), 769-780.
- Wernegreen, J. J. (2015). Endosymbiont evolution: predictions from theory and surprises from genomes. *Annals of the New York Academy of Sciences*, 1360(1), 16-35.
- Werren, J. H. (1997). Biology of wolbachia. *Annual Review of Entomology*, 42(1), 587-609.
- Wierz, J. C., Gaube, P., Klebsch, D., Kaltenpoth, M., Flórez, L.V. (2021). Transmission of Bacterial Symbionts With and Without Genome Erosion Between a Beetle Host and the Plant Environment. *Frontiers in Microbiology*, 12.

Zhang, F., Li, X., Zhang, Y., Coates, B., Zhou, X., Cheng, D. (2015). Bacterial Symbionts, Buchnera, and Starvation on Wing Dimorphism in English Grain Aphid, *Sitobion avenae* (F.) (Homoptera: Aphididae). *Frontiers in Physiology*, 6.