

Jihočeská univerzita v Českých Budějovicích
Přírodovědecká fakulta

**Symbiotické bakterie vší rodu *Polyplax*:
základní fylogenetická a genomická charakterizace**

Bakalářská práce

Jana Říhová

Školitel: Prof. RNDr. Václav Hypša, CSc.

České Budějovice 2015

Říhová, J., 2015: Symbiotické bakterie vši rodu *Polyplax*: základní fylogenetická a genomická charakterizace. [The symbiotic bacteria of lice of genus *Polyplax*: general phylogenetic and genomic characterisation. Bc.Thesis, in Czech] – 66 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Anotace:

Blood-sucking lice of the genus *Polyplax* harbour two bacterial endosymbionts, one from family Legionellales and the other from family Neisseriales. Phylogenetic and genomic analyses reveal typical features of endosymbiotic bacteria and coevolution history with the host. In the genome of the endosymbiont from family Legionellales I found operon, which encodes biosynthesis of biotin, essential vitamin, especially for blood-sucking insects living on low-nutrient diet. This operon was probably horizontally transfer from *Wolbachia*, nutrient endosymbiont of bedbug *Cimex lectularius*.

Prohlašuji, že svoji bakalářskou práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své bakalářské práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

V Českých Budějovicích, 24. dubna

Jana Říhová

Poděkování

Ráda bych v první řadě poděkovala svému školiteli Václavu Hypšovi, za odborné vedení mé bakalářské práce, také za častou morální podporu a vstřícnost. Dále bych chtěla poděkovat Evě Novákové za externí konzultace a Filipovi Husníkovi za poskytnutá genomická data a pomoc při bioinformatických analýzách. Velký dík patří zároveň Evě Šochové za její cenné rady a pomoc při práci v naší laboratoři. V neposlední řadě bych ráda poděkovala Míše Matějkové, Janě Martinů a všem ostatním členům laboratoře za poskytnutí vzorků, pomoc a morální podporu při téměř ročních nezdarech se screeningem bakterií.

OBSAH

1	ÚVOD	- 1 -
1.1	Symbióza mezi bakteriemi a krevsajícím hmyzem v genomické éře	- 1 -
1.2	Historické pozadí studia symbiózy	- 1 -
1.3	Původ a význam symbiózy u hmyzu	- 2 -
1.3.1	Symbióza u hmyzu	- 2 -
1.3.2	Bakteriomy	- 3 -
1.4	Genomický vhled do symbiózy	- 3 -
1.4.1	Redukce genomu: selektivní ztráty genů a charakteristické modifikace	- 3 -
1.4.2	Výměny endosymbiontů a horizontální přenos	- 5 -
1.5	Krevsající hmyz a symbiotické bakterie	- 6 -
1.5.1	Vši – skupina s vysokou diverzitou bakteriálních endosymbiontů	- 7 -
1.5.2	Současný stav výzkumu v naší laboratoři	- 9 -
2	CÍLE	- 12 -
3	MATERIÁL A METODIKA	- 13 -
3.1	Screening bakterií	- 13 -
3.1.1	Sběr vzorků a izolace DNA	- 13 -
3.1.2	Polymerázová řetězová reakce	- 14 -
3.1.3	Gelová elektroforéza	- 16 -
3.1.4	Klonování a transformace	- 16 -
3.1.5	Amplifikace	- 17 -
3.1.6	Práce se sekvencemi a fylogenetické analýzy	- 17 -
3.2	Genomické analýzy	- 18 -
3.3	FISH	- 21 -
4	VÝSLEDKY	- 22 -
4.1	Screening bakterií	- 22 -

4.2	Fylogenetické analýzy.....	- 22 -
4.2.1	Legionellales	- 22 -
4.2.2	Neisseriales.....	- 22 -
4.3	Genomické analýzy.....	- 23 -
4.4	FISH.....	- 24 -
5	DISKUZE	- 36 -
5.1	Fylogeneze.....	- 36 -
5.2	L-symb	- 36 -
5.3	N-symb.....	- 37 -
5.4	Genomika	- 38 -
5.5	FISH.....	- 41 -
6	ZÁVĚR	- 41 -
7	POUŽITÁ LITERATURA	- 42 -
8	PŘÍLOHY	- 51 -

1 ÚVOD

1.1 Symbióza mezi bakteriemi a krevsajícím hmyzem v genomické éře

Nástup moderních genomických metod ovlivnil velké množství biologických oborů a témat. Patří mezi ně i výzkum interakcí mezi hmyzem a symbiotickými bakteriemi. Toto téma prošlo během více jak stovky let (od prvních prací českého biologa Šulce – Šulc, 1910) několika epochami a metodickými zlomy: pozorování bakterií pomocí světelného mikroskopu (Ries, 1931; Buchner, 1965 - dokumentace častého výskytu a velké variability mikroorganismů v tkáních hmyzu), nástup elektronové mikroskopie (ujasnění bakteriální povahy symbiontů), vynález PCR a následné využití molekulárních metod pro fylogenetické analýzy (taxonomické zařazení symbiontů a potvrzení dlouhodobé koevoluce) a v současné době možnost porovnání genomických charakteristik a funkčního významu symbiontů. Důsledkem recentního metodického pokroku je mimo jiné i poznání široké škály typů a funkcí bakterií v těle hmyzu, obecně zahrnovaných pod široký pojem *symbiont*.

1.2 Historické pozadí studia symbiózy

Termín *symbióza* (symbiotismus) poprvé použil v roce 1877 německý biolog Albert Bernhard Frank. Definoval ho jako koexistenci dvou rozdílných druhů bez ohledu na povahu jejich vztahu (Frank, 1877). V roce 1879 Anton de Bary použil toto slovo ve smyslu: „společné soužití jinak se nazývajících organismů“ a rozdělil symbiózu na několik typů podle povahy vztahu interagujících organismů na parazitismus, komenzálismus a mutualismus (de Bary, 1879). Na počátku 20. století začali někteří vědci používat slovo *symbióza* jako synonymum pro popis mutualistických nebo neparazitických vztahů, které jsou výhodné pro oba kooperující druhy (Cleveland, 1926, p. 51, Boucher, 1985; Saffo, 1992). Nejasnosti ohledně používání slova *symbióza* přetrvávají dodnes, ačkoli většina současných učebnic uvádí slovo *symbióza* v širším slova smyslu jako společné soužití dvou různých organismů. Nicméně v důsledku přetrvávajících debat a kontroverzí ohledně definice slova *symbióza* se dokonce objevily názory, které nesouhlasí s používáním tohoto termínu pro vysvětlení vztahů mezi organismy (Martin & Schwab, 2012). V této práci bude samotný pojem *symbióza* používán v jeho širokém smyslu, tedy pro soužití hostitelů a symbiontů bez ohledu na jeho důsledky na fitness obou organismů.

1.3 Původ a význam symbiózy u hmyzu

1.3.1 Symbiόza u hmyzu

Symbiόza s bakteriemi je u bezobratlých častým jevem studovaným zejména v rámci třídy Insecta (Costopoulos et al., 2014). Mnoho zástupců hmyzu udržuje mutualistické vztahy s jednou nebo více symbiotickými bakteriemi (Buchner, P. 1965). Bakterie, které se účastní těchto asociací, jsou označovány buď jako obligátní (P, primární) endosymbionti, kteří jsou mutualisty nezbytnými pro normální vývoj a reprodukci hostitele nebo fakultativní (S, sekundární) endosymbionti, kteří svým příležitostným nositelům mohou poskytovat určité výhody ve fitness (Moran et al., 2008). Primární i sekundární endosymbionti často koexistují v jednom hmyzím hostiteli (Dale et al., 2006).

Všeobecně se uvádí, že potrava některých ekologických skupin hmyzu, např. fytofágního a hematofágního hmyzu, je chudá na některé důležité složky (Douglas, 1989). Z toho důvodu ve svém těle tento hmyz hostí bakteriální P-ednosymbionty a často pro ně vytváří speciální útvary zvané bakteriomy. Tito endosymbionti jsou schopni využívat vlastní metabolické dráhy k syntéze esenciálních látek, které pak poskytují svému hostiteli. (Moran et al., 2008). Obligátní symbiόzy s bakteriemi vznikaly v průběhu evoluce opakovaně a jsou klíčem k ekologické diverzifikaci mnoha skupin bezobratlých.

Symbionti hmyzu vznikali z různých skupin bakterií, nejčastěji je však přechod k endosymbiόze pozorován v rámci gammaproteobakterií (McCutcheon & Moran, 2007) a to především v čeledi Enterobacteriaceae (Moran et al., 2008). Patří sem většina známých S-endosymbiontů (Moran et al., 2008) a někteří P-symbionti, jako je například *Buchnera aphidicola*, která je intenzivně studovaná jako model symbionta, jehož metabolické funkce jsou poskytovány hostiteli (Shigenobu & Wilson, 2011).

S-endosymbionti se obecně rozdělují do dvou typů podle interakce s hostitelem na fakultativní mutualisty a reprodukční manipulátory (Moran, 2008). Fakultativní mutualisti poskytují hostitelům výhody ve fitness, umožňují jim delší život například ochranou před parazitoidními vosami (Olivier et al., 2003), nebo rezistencí proti patogenním houbám (Scarborough, 2005). Vyšší reprodukce hostitelů tak zajišťuje zvyšování frekvence infikovaných jedinců a tím i jejich symbiontů. Druhým typem jsou parazitičtí reprodukční manipulátoři, jejichž přenos je zajištěn zvyšující se hostitelskou reprodukcí dcer na úkor reprodukce synů (Moran et al., 2008). Tito paraziti jsou typicky přenášeni po mateřské linii a využívají různé strategie manipulace hostitelem, jako je reprodukční inkompatibilita, zabíjení synů, feminizace samců a partenogeneze. Nejvíce studovaným reprodukčním

manipulátorem je rod *Wolbachia*, který je velmi rozšířen u hmyzu a některých dalších bezobratlých, kteří vykazují některé z uvedených fenotypů (Stouthamer et al, 1999).

1.3.2 Bakteriomy

Znalosti o morfologii symbiotických bakterií a jejich umístění v těle hostitele byly dlouhou dobu závislé jen na klasických mikroskopických technikách a elektronové mikroskopii. Například P-endosymbionti jsou typicky odděleni ve specializovaném orgánu, tzv. bakteriomu, který je složen z jednotlivých buněk zvaných bakteriocyty. P-endosymbionti žijí v cytosolu těchto specializovaných buněk. (Baumann, 2005). První bakteriom objevil Robert Hooke v roce 1664 při studiu fyziologie lidské vši *Pediculus humanus* (Hooke, 1665). V závislosti na hostitelské skupině se mohou bakteriomy skládat z tukových tělních buněk, buněk střevní stěny nebo z vysoce specializovaných buněk, které jsou vývojově determinovány v embryu (Moran et al., 2008). Narozdíl od obligátních P-symbiontů, fakultativní S-endosymbionti jsou náhodně distribuováni v různých tkáních hostitele. U skupin hmyzu vlastních bakteriomy, S-endosymbionti mohou napadat bakteriocyty, kde následně přežívají s P-endosymbionty, nebo je dokonce nahrazují. (Buchner, 1965). Výrazným posunem ve výzkumu bakteriomů se v nedávné době staly metody umožňující sledování určitého úseku DNA *in vivo*. Jednou z takových metod je FISH (=fluorescent in-situ hybridization), která má mnoho různých aplikací. Například povoluje detekci zatím nekultivovaných organismů a umožňuje sledování konkrétního druhu bakterie v komplexních mikrobiálních komunitách (Moter & Göbel, 2000). Tato metoda může být využita ke studiu umístění bakteriálních endosymbiontů v těle hostitele.

1.4 Genomický vhled do symbiózy

1.4.1 Redukce genomu: selektivní ztráty genů a charakteristické modifikace

Nejvýraznějším znakem typickým pro evoluci bakteriálních endosymbiontů je extrémní redukce jejich genomů oproti jejich volně žijícím příbuzným. Velikost jejich genomu může být snížena až na genomovou velikost některých organel, jako je mitochondrie nebo chloroplast. Množství protein-kódujících genů je ale podstatně vyšší v genomech endosymbiontů ve srovnání s podobně velkými genomy organel (McCutcheon & Moran, 2012). Ačkoli se původně předpokládalo, že bakterie s takto malým genomem představují předky ostatních organismů (Wallace, 1973), v dnešní době je prokázáno, že jsou odvozeny od bakterií s velkým genomem. Dá se tedy předpokládat, že obligátní endosymbionti vznikli

z původně fakultativních endosymbiontů (Nikoh et al., 2014). Celý proces evoluce od volně žijící bakterie přes sekundárního symbionta až k silně redukovanému P-endosymbiontovi vykazuje řadu typických charakteristik opakujících se u různých fylogenetických skupin bakterií. Genomy nedávno vzniklých endosymbiontů, které jsou na počátku degenerace, obsahují velké množství mobilních elementů a pseudogenů, dále jsou pozorována genová přeskupení a delece chromozomových fragmentů (Toh et al., 2006; Ochman & Davalos, 2006, Burke & Moran, 2011). V důsledku to znamená zvyšování počtu delečních mutací a jejich fixaci. Naopak v genomech dávných endosymbiontů, kteří po miliony let koexistují s hostitelem, byly mobilní elementy i většina pseudogenů eliminovány (McCutcheon & Moran, 2012).

S redukcí genomu souvisí další změny v genomu, jako je extrémně rychlá sekvenční evoluce, změna v přiřazování kodonů a změna v nukleotidovém složení. Předpokládá se, že hlavními příčinami těchto změn je malá populační velikost a asexualita těchto organismů (Moran, 1996). Populační struktura endosymbiontů je ovlivněna jejich těsnou vazbou na hostitele a nedostatkem, či úplnou absencí rekombinací mezi jednotlivými kmeny v různých hostitelích, což zvyšuje vliv genetického driftu, který následně odstraňuje některé geny (McCutcheon & Moran, 2012). Zmenšováním genomu přichází endosymbiont o důležité geny, které kódují především DNA reparace a rekombinace, ale vždy si ponechávají určitý set genů kódující alespoň nejzákladnější procesy těchto funkcí (Moran, 2008). Z redukovaných genů se vytváří nefunkční pseudogeny, které v důsledku končí úplnou delecí z genomu (Goodhead & Darby, 2015). Na druhou stranu, geny, které souvisí s mechanismy průniku a přetrvávání v hostitelských buňkách nebo přispívají k hostitelské fitness, jsou v případě obligátních endosymbiontů dlouhodobě udržovány.

Změny v nukleotidovém složení se u takto malých genomů, jako např. u bakteriálního endosymbionta krevsající mouchy tsetse *Wigglesworthia glossinidia*, projevují nízkým obsahem nukleotidů G+C ve prospěch A+T oproti volně žijícím příbuzným. Předpokládá se, že tyto změny mohou být výsledkem ztráty některých genů ovlivňujících reparaci DNA, v případě *W. glossinidia* je to absence SOS reparace, odstraňování chybných bazí či nukleotidů a následným vznikem mutací, které způsobují přeměnu nukleotidů G nebo C na A nebo T (Akman et al., 2002). Tento výrazný posun ve složení DNA endosymbiotických bakterií má zásadní dopad na přesnost fylogenetických analýz. Dalším z běžných rysů endosymbiotického genomu je změna translatovaných kodonů a to nejčastěji změna UGA z původního stop kodonu na kodon kódující aminokyselinu tryptofan (Trp)

(Knight, 2001). Tento jev může být způsoben mutacemi v tRNA^{Trp}, které pak umožňují kódování stop kodonu (UGA) i kodonu pro tryptofan (UGG). (McCutcheon & Moran, 2012).

Dále jsou často ztráceny geny kódující produkci mastných kyselin, fosfolipidů a peptidoglykanů jako v případě obligátního endosymbionta *Buchnera aphidicola*. Tyto molekuly jsou pak dále využívány k produkci buněčných membrán. Bakterie, kterým chybí uvedené aminokyseliny, jsou pak plně závislé na buněčné stěně vytvářené hostitelem nebo jiným endosymbiontem uvnitř jednoho bakteriocyty (Shigenobu, 2000).

Navzdory všem genovým ztrátám, některé geny mají klíčovou roli v biologii těchto bakterií a jsou zpravidla uchovávány v genomu. Patří mezi ně geny GroES-GroEL chaperonin komplex, DnaK komplex, Hsp70 (heat shock protein) chaperone komplex a geny spojené s biosyntézou aminokyselin esenciálních pro jejich hostitele (McCutcheon & Moran, 2012).

1.4.2 Výměny endosymbiontů a horizontální přenos

Z důvodu ztráty různých genů při rychlé redukci genomu endosymbiontů může docházet k narušení metabolických funkcí. Takoví endosymbionti ztrácejí své funkce a jsou často nahrazováni, nebo doplňováni jinými efektivnějšími endosymbionty. Výsledkem může být složitý komplex zahrnující několik endosymbiotických bakterií v jednom hostiteli, jako bylo pozorováno např. u některých linií podřádu Auchenorrhyncha (Koga et al., 2013). V případě úplné výměny symbiontů pak dochází ke ztrátě kongruence mezi fylogenezemi symbiontů a hostitelů.

Alternativním procesem, který může řešit ztrátu důležitých genů endosymbionta, je jejich opětovné získání prostřednictvím horizontálního genomového přenosu (HGT), který je často pokládán za hlavní hnací sílu evoluce bakterií (Heuer & Smalla, 2007). V poslední době se ukazuje, že k HGT dochází i mezi hostitelským a symbiotickým genomem. Například v genomu mšice *Acyrtosiphon pisum* byly objeveny dva nezávislé HGT z genomu jejího primárního endosymbionta *Buchnera aphidicola*, jejichž produktem jsou však nefunkční pseudogeny (Nikoh et al., 2010). Podobně genom vlnatky *Planococcus citri* obsahuje minimálně 22 horizontálně přenesených funkčních genů z různých bakterií. Ačkoli vlnatka hostí dva bakteriální endosymbionty, přenos jejich genů do genomu hostitele nebyl zaznamenán. Zajímavostí této endosymbiózy je, že jeden z endosymbiontů, *Tremblaya princeps*, vlastní nejmenší zaznamenaný symbiotický genom, si udržuje vlastního

endosymbionta *Moranella endobia*. Zdá se, že obrovská redukce genomu u *Tremblaya* je doplněna sdílením metabolitů pro syntézu esenciálních aminokyselin mezi oběma endosymbionty (Husník et al., 2013).

V některých případech může docházet pomocí HGT i k přenosu celých operonů, které kódují určité metabolické dráhy. V případě bakteriálního endosymbionta rodu *Wolbachia* izolovaného z krevsající štěnice *Cimex lectularius* došlo k přenosu operonu kódujícího biosyntézu biotinu (vitamín B7). Je známo, že krev obratlovců jako zdroj potravy je chudá zejména na B-vitamíny a kofaktory (Douglas, 1989). Obecně je bakterie rodu *Wolbachia* u hmyzu známá jako parazitický reprodukční manipulátor, který má negativní dopad na fitness hostitele (Werren et al. 2008; viz. kapitola: Symbióza u hmyzu). Nicméně u výše zmíněné štěnice funguje tato bakterie jako primární endosymbiont nezbytný pro normální růst a reprodukci hostitele. Tento přechod z fakultativní na obligátní endosymbiózu je patrně způsoben přenosem biotinového klastru genů, s jehož pomocí *Wolbachia* poskytuje hostiteli esenciální vitamín B7. Předpokládá se, že *Wolbachia* získala tento operon díky HGT pravděpodobně koinfekcí od endosymbiontů *Cardinium* nebo *Rickettsia* (shrunutí článku Nikoh et al., 2014).

1.5 Krevsající hmyz a symbiotické bakterie

Hmyz a jiní živočichové získávají vitamíny především v potravě. Krev obratlovců, jakožto zdroj obživy neposkytuje dostatečný zdroj B-vitaminů (Baumann, 2005). Z toho důvodu v sobě hematofágní hmyz hostí bakteriální endosymbionty, kteří syntetizují tyto esenciální látky (Baumann, 2006). Závislost krevsajícího hmyzu na vitamínech skupiny B je dobře známá již z experimentálních prací z doby „předmolekulární“. Eliminace endosymbiontů ze vši nebo much tsetse vedla k poškození jejich životnosti nebo reprodukční kapacity, pokud nebyla jejich potrava doplněna o vitamíny B (Puchta, 1954; Puchta, 1955; Nogge, 1976).

Z výsledků rozborů bakteriálních genomů se dá usoudit, že geny biosyntetických drah pro vitamíny skupiny B se pomocí HGT dynamicky pohybují mezi různými bakteriálními liniemi (Tanaka et al., 2005). Tyto funkční „sobecké“ operony se předávají podobně jako geny zajišťující rezistenci proti antibiotikům pomocí HGT a poskytují okamžité výhody bakteriálnímu příjemci (Lawrence & Roth, 1997; Campillos et al., 2006). Mezi často studované metabolické dráhy endosymbiontů patří syntéza biotinu (vitamín B7). Biotin (C₁₀H₁₆N₂O₃S₁) je jeden z nejpozoruhodnějších kofaktorů karboxylázových enzymů,

kteřé se podílí na metabolismu aminokyselin a mastných kyselin v prokaryotických i eukaryotických buňkách (Lin & Cronan, 2011). Tento kofaktor byl objeven na začátku 20. století jako růstový faktor kvasinek (Wildiers, 1901). Je charakterizován jako bezbarvá sloučenina vyskytující se v 8 různých stereoizomerech, nicméně pouze jeden z nich, d-biotin, má vitamínovou aktivitu. Skládá se ze dvou spojených heterocyklických kruhů s navázaným postranním řetězcem tvořeným kyselinou valerovou (Lin et al., 2010). Nenahraditelná enzymatická role biotinu byla prokázána v syntéze mastných kyselin, metabolismu aminokyselin a glukoneogeneze (Linen, 1967; Knowles, 1989; Jitrapakdee & Wallace, 2003). Studie bakteriálních genomů *Escherichia coli* a *Bacillus subtilis* potvrdily, že začátek biotinové biosyntetické dráhy může probíhat několika různými cestami. Tyto cesty produkují pimelát-thioester prekurzor, který je esenciální pro iniciaci spojení heterocyklických kruhů biotinu (Lin & Cronan, 2011).

Při studiu uspořádání genů zajišťujících biosyntézu biotinu v genomu betaproteobakterie *Methylobacillus flagellatus* se ukázalo, že vlastní 6 genů organizovaných do dvou samostatně transkribovaných jednotek. BioB, BioF, BioH, BioC, BioD tvoří jeden operon a BioA byl nalezen v různých místech bakteriálního genomu (Serebriiskii et al. 1996). V případě obligátního endosymbionta *Wolbachia* u štěnice *Cimex lectularius* byl v genomu udržen kompaktní operon složený z šesti genů (BioC, BioH, BioF, BioA, BioD, BioB) kódující biosyntézu biotinu (Nikoh et al., 2014).

1.5.1 Vši – skupina s vysokou diverzitou bakteriálních endosymbiontů

- Základní charakteristika vší

Vši (Phthiraptera, Anoplura) představují potenciálně zajímavý model pro studium symbiomy. Jsou to obligátní ektoparaziti placentálů, kteří po celý vývojový cyklus žijí na hostiteli a živí se výhradně krví. Jejich celková diverzita je v důsledku přísné hostitelské specializace vysoká (Lehane, 2005). V rámci řádu Phthiraptera tvoří vši jeden ze 4 podřádů (Amblycera, Ischnocera, Rhynchophtirina a Anoplura) a předpokládá se, že se vyvinuly z volně žijícího předka, který původně obýval hnízda a nory obratlovců (Kim, 1985). V průběhu evoluce se u nich vyvinuly adaptace typické pro obligátně parazitický způsob života, jako je zploštění, redukce křídel a tibio-tarzální drápky, které slouží k přichycení na hostiteli (Light et al., 2010). Tento podřád vší také disponuje vysoce účinným sacím aparátem, který umožňuje

saní krve přímo z cév hostitele (Kim, 2006). Nymfy procházejí během dvou týdnů třemi larválními stádii, než se vyvinou v dospělé (Lehane, 2005). Vši mohou přenášet infekční bakterie, jako bylo prokázáno u lidské vši *Pediculus humanus*, který je přenašečem bakterie *Rickettsia prowazekii* způsobující onemocnění tyfus (Lehane, 2005).

Rod *Polyplax* se vyznačuje poměrně širokým spektrem hostitelů z řádů Rodentia a Soricomorpha. (Haitlinger, 2005; Cais, 1980). Druh *Polyplax serrata* je nejčastěji vázán na hostitele rodu *Apodemus* (Štefka & Hypša, 2008). U domestikovaných kryš a potkanů (*Rattus rattus* a *Rattus norvegicus*) se běžně vykytují vši příbuzného druhu *Polyplax spinulosa* (Pratt & Karp, 1953).

- Symbiotické bakterie vší

Podobně jako další skupiny hmyzu živící se výhradně krví obratlovců hostí vši ve svém těle symbiotické bakterie. Bakteriální P-endosymbionti krevsajících vší jsou umístěni ve striktně definovaných bakteriomech, které se mezi různými skupinami liší ve struktuře a umístění. Očekává se, že vývoj bakteriomů u podřádu Anoplura je spojen s výživou na krvi permanentního hostitele (Lehane, 2005). Obvykle jsou to nepárové, oválné orgány umístěné na ventrální straně žaludku bez přímého spojení s žaludkem (např. u rodů *Polyplax* a *Linognathus*) nebo jsou s ním blízce spojeny (*Pediculus*). U druhů *Polyplax spinulosa* a *Linognathus tenuirostris* byly nalezeny orgány skládající se z jasně definovatelné agregace bakteriomů přilepených na sebe bez společného obalu. Extrémně jednoduchý bezbarvý bakteriom u vší *Polyplax spinulosa* je vyplněn malými nepravidelnými granulovanými vesikuly (Buchner, 1965). Pomocí světelné mikroskopie bylo dokázáno, že vývoj bakteriomů u vší *Polyplax spinulosa* je závislý na pohlaví jedince (Volf, 1991). Již Buchner (1965) při výzkumu bakteriomů vší zaznamenal jejich velkou morfologickou variabilitu a také rozmanitost bakteriálních endosymbiontů a předpokládal proto, že tyto mikroorganismy tvoří polyfyletickou skupinu. Po příchodu molekulárních technik se potvrdilo, že bakteriální endosymbionti vší pocházejí z různých bakteriálních skupin. Přestože dosud byla molekulárními metodami studována fylogeneze endosymbiontů jen z několika málo skupin vší, je polyfyletický charakter jejich endosymbiontů zjevný (Hypša & Křížek, 2007). Například u vší rodu *Haematopinus* parazitujících na prasatech byly identifikovány bakterie rodu *Wolbachia* (Kyei-Poku et al., 2005) a nový druh bakterie z čeledi Enterobacteriaceae (Hypša & Křížek, 2007). Další linii

symbiotických bakterií z čeledi Enterobacteriaceae můžeme najít u vši *Solenopotes capillatus* parazitující na dobytku (Perotti et al., 2009). U vši rodu *Linognathus*, které jsou známými parazity psů, kojetů, koz, ovcí a dobytka byli detekovány bakterie rodu *Wolbachia* ze superskupiny A a B (Kyei-Poku et al., 2005). Primárním endosymbiontem lidských vší *Pediculus humanus* je bakterie známá jako *Riesia pediculicola*, která je blízká příbuzná bakteriím rodu *Arsenophonus* (podle řady analýz je přímo součástí rodu *Arsenophonus* (Perotti et al., 2009). U vši rodu *Polyplax* byl identifikován bakteriální endosymbiont, který se podle fylogenetických analýz klastruje mezi bakterie rodu *Legionella*. (Hypša & Křížek, 2007). Tato velká diverzita bakteriálních endosymbiontů vší nabízí příležitost k porovnání vývoje genomu těchto symbiotických bakterií z různých skupin, které může ukázat míru konvergence nebo naopak nezávislosti symbiontů v souvislosti s omezeními danými hostitelem.

1.5.2 Současný stav výzkumu v naší laboratoři

V rámci širšího zaměření laboratoře na výzkum hmyzích symbiontů byla získána genomická data ze vší rodu *Polyplax* (viz. Materiál a metodika). Předběžná kontrola těchto dat ukázala na dominantní zastoupení dvou bakterií. Jedna potvrzuje výskyt již předpokládaného endosymbionta z příbuznosti Legionellaceae (dále už jen zkratka L-symb), druhá je nově nalezenou bakterií z řádu (nebo jeho blízkosti) Neisseriales (dále už jen zkratka N-symb). Obě bakterie tak z taxonomického hlediska představují skupiny bakterií, které nepatří k typickým zdrojům endosymbiontů hmyzu.

- Legionellales

Řád Legionellales se řadí do skupiny gramnegativních proteobakterií zvaných Gammaproteobacteria (Obr. 1). Je tvořen ze dvou čeledí Legionellaceae a Coxiellaceae (Garrity, 2005). Do tohoto řádu patří intracelulární parazité protozoí, bezobratlých a obratlovců včetně člověka. Bakterie rodu *Legionella* (Legionellaceae) se mohou vyskytovat v odlišných formách, například byly popsány jako volně žijící planktonní buňky, nebo intracelulární paraziti a mutualisti protozoí či zástupci bakteriálních komunit biofilmů (Mampel et al., 2006). Některé druhy rodu *Legionella* jsou mezofylní bakterie vyskytující se v chladných vodách (např. antarktická jezera), které se dokázali adaptovat na extrémní podmínky (Carvalho et al., 2008). Nejznámějším zástupcem tohoto rodu je *Legionella pneumophila*, která u člověka způsobuje onemocnění legionelóza (legionářská nemoc) a

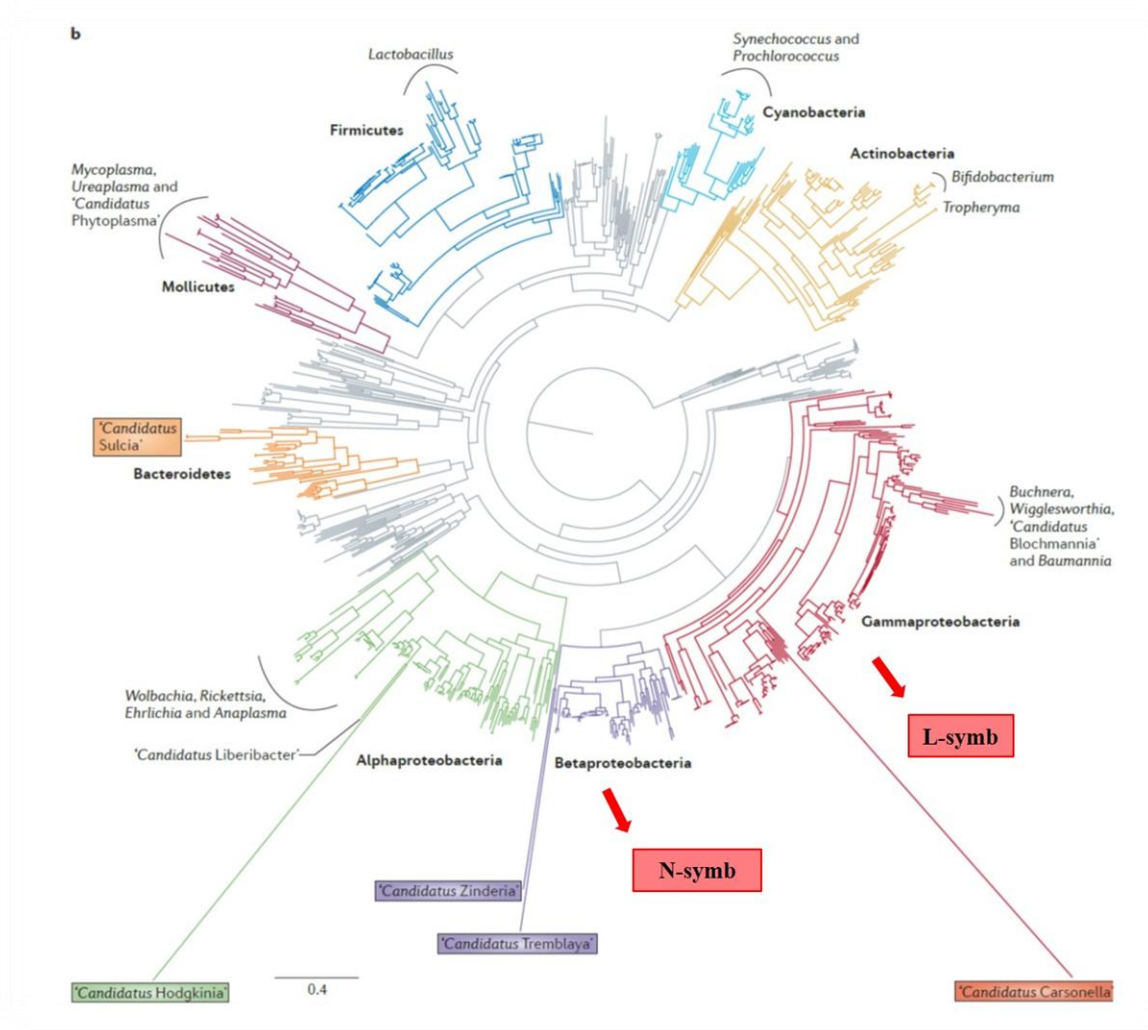
zároveň je patogenem volně žijících sladkovodních a půdních améb rodu *Acanthamoeba* a *Naegleria* (Rowbotham, 1980). Bakterie z čeledi Coxiellaceae jsou intracelulárními parazity bezobratlých i obratlovců. Druh *Coxiella burnetii* je známým endosymbiontem klíšťatovců (Metastigmata, Arthropoda) a vodních bezobratlých. Tato bakterie zároveň způsobuje onemocnění zvané Q-horečka, které je nejrozšířenější u ovcí a koz, ale přenáší se i na člověka (Garrity, 2007).

U améby *Amoeba proteus* byl objeven obligátní endosymbiotický druh *Legionella jeonii* (Park et al., 2004). Podle jednoduché fylogenetické analýzy představené v článku Hypša & Křížek (2007) byl objeven další endosymbiotický druh rodu *Legionella* u krevsajících vší *Polyplax serrata* a *Polyplax spinulosa*.

- Neisseriales

Řád Neisseriales je součástí velké skupiny proteobakterií nazývané Betaproteobacteria (Obr. 1). Z hlediska základních charakteristik jsou Neisseriales popisovány jako skupina gramnegativních nesporulujících bakterií. Vyskytují se jako patogenní, aerobní či mezofylní bakterie (Tønjum, 2005). Pro tento řád však není žádná z uvedených vlastností unikátní. V současné době bylo popsáno 32 rodů Neisseriales se širokou škálou morfologií, biotopů a nároků na růst. Zahrnují významné lidské patogeny jako je *Neisseria gonorrhoeae* a *Neisseria meningitidis* (Jamet et al., 2015; Adeolu & Gupta, 2013). *Neisseria gonorrhoeae* (Neisseriaceae) způsobuje onemocnění gonorrhoea (kapavka), které se přenáší pohlavním stykem. *Neisseria meningitidis* (Neisseriaceae) je přenášena infikovanými kapénkami ve vzduchu a způsobuje meningokoková onemocnění (Pizza & Rappuoli, 2015).

Další zajímavou bakterií z toho řádu je *Snodgrassella alvi*, která se vyskytuje jako jeden z dominantních zástupců střevní mikroflóry u včel (*Apis* spp.) a čmeláků (*Bombus* spp.) (Moran et al. 2011). Bylo zjištěno, že fylogeneze tohoto střevního mutualistického symbionta koreluje s fylogenezí hostitelských druhů více než s jejich geografickým původem, což naznačuje dlouhou koevoluční historii těchto bakterií s jejich eusociálním hostitelem (Engel & Moran, 2013). Podle současných prací se ukazuje, že tyto symbionti by mohli hrát roli v obraně proti patogenům (Koch & Schmid-Hempel, 2011) a uplatňovat se v metabolismu karboxylových kyselin a B vitamínů (Kwong et al., 2014).



Obr. 1: Rozložení bakteriálních skupin v rámci domény Bacteria. Šipkami jsou označeny skupiny bakterií, do kterých se řadí analyzovaní bakteriální endosymbionti. Převzato z článku McCutcheon & Moran, 2012.

2 CÍLE

Práce se věnuje základní fylogenetické a genomické charakterizaci dvou dominantních bakterií (Legionellales a Neisseriales) identifikovaných v genomických datech ze vší rodu *Polyplax*. Jejimi cíli je:

- Ujasnit fylogenetické postavení obou těchto bakterií

V současné době je dostupná pouze jedna sekvence bakterií N-symb a 2 sekvence L-symb 16S tRNA z izolátů vší rodu *Polyplax*. Předpokládá se, že tyto bakterie patří do příbuznosti rodů *Neisseria* a *Legionella*. Smyslem této práce je ověřit fylogenetického postavení podrobnou studií (viz. Materiál a metodika) a z jejího výsledku určit, zda tyto bakterie vytváří monofyletickou skupinu.

- Sestavit základní genomickou charakteristiku těchto bakterií

Pro potvrzení endosymbiotické povahy obou bakterií je důležité zjistit jejich základní genomické rysy (především s ohledem na charakteristiky tradičně spojované s evolucí endosymbiotických bakterií) a provést porovnání redukce genomů těchto bakterií oproti volně žijícím příbuzným.

- Vyhledat a namapovat metabolické dráhy typické pro endosymbionty

Obecná genomická charakterizace bude doplněna o konkrétní příklady metabolických drah spojených s endosymbiózou.

- Metodická příprava vizualizace endosymbiontů

Jako přípravný metodický krok pro další studium této symbiozy bude vyzkoušena cytogenetická metoda vizualizace bakterií *in vivo*, umožňující sledovat jejich lokalizaci a průběh ontogeneze ve spojení s hostitelským životním cyklem.

3 MATERIÁL A METODIKA

3.1 Screening bakterií

3.1.1 Sběr vzorků a izolace DNA

Pro screening symbiotických bakterií byla použita DNA z 24 jedinců rodu *Polyplax*. Jedinci byli sesbíráni z myšic rodu *Apodemus* ze tří oblastí – Slovensko (8), Srbsko (10 - poskytnuty Janou Martinů) a Hlinsko (6 - vlastní sběr), uchováni v 96% etanolu a následně skladováni v -20°C. K izolaci celkové DNA (tj. izolát obsahoval jak eukaryotní, tak prokaryotní DNA) byl použit izolační kit (QIAamp DNA Micro Kit, QIAGEN), postup byl dodržen podle návodu od výrobce. Získaná DNA (v některých případech předem namnožena amplifikačním kitem) byla použita jako templát pro amplifikaci požadovaného úseku pomocí metody PCR (Polymerase Chain Reaction). Koncentrace jednotlivých vzorků DNA byla zjištěna na nanofotometru (NanoPhotometer P-Class, IMPLEN). Následně byl PCR produkt podroben gelové elektroforéze.

Tab. I: Tabulka analyzovaných vzorků vši rodu *Polyplax*

Pracovní označení	Lokalita
S1	Slovensko
S2	Slovensko
S3	Slovensko
S4	Slovensko
S5	Slovensko
S6	Slovensko
S7	Slovensko
S8	Slovensko
1A*	Srbsko
1B*	Srbsko
1C*	Srbsko
1D*	Srbsko
1F*	Srbsko
1G*	Srbsko
1H*	Srbsko
1CH*	Srbsko
3	Srbsko
24	Srbsko
V1	Česká republika (Hlinsko)
V2 (1) *	Česká republika (Hlinsko)

Pracovní označení	Lokalita
V2 (2) *	Česká republika (Hlinsko)
V3 (1) *	Česká republika (Hlinsko)
V3 (2) *	Česká republika (Hlinsko)
V3 (3) *	Česká republika (Hlinsko)

*Písmeny (číslly) značení jedinci rodu *Polyplax* sesbírání z jednoho hostitele *Apodemus* sp.

3.1.2 Polymerázová řetězová reakce

Pro vyzkoušení funkčnosti reakce byly použity 4 druhy polymeráz (Taq DNA polymerase - Roche Diagnostic, FastStart High Fidelity DNA polymerase – Roche Diagnostics, Taq DNA polymerase – Top-Bio, Combi Taq DNA polymerase – Top-Bio) a to nejprve na 2 vzorky klošů, kteří byly pozitivní na bakteriální DNA (materiál poskytl Eva Nováková). U klošů se na identifikaci bakteriálních endosymbiontů nejlépe osvědčila FastStart DNA polymeráza (PCR reakce míchána do 20 μ l, počáteční denaturace: 3 min.). Při identifikaci L-symb u vší rodu *Polyplax* byla vyzkoušena ještě HotStar DNA polymeráza, která se ukázala být nejlepší volbou pro amplifikaci genů bakteriálních endosymbiontů u vší rodu *Polyplax*. Pro screening požadovaných bakterií byly vytvořeny nové primery na gen pro chaperonin GroEL a gen pro 16S malou ribozomální podjednotku obou bakterií – pro optimalizaci reakce byla provedena gradientová PCR na zkušebních izolátech rodu *Polyplax*. Všechny nově navržené primery se však ukázaly jako nevhodné pro screening obou bakterií. Pro detekci L-symb byly nakonec použity již vyzkoušené primery z článku Hypša & Křížek (2007). PCR reakce proběhla v přístroji Mastercycler (Eppendorf), ve kterém byly nastaveny vhodné programy pro PCR. (viz. Tab. III).

Tab. II: Složení PCR reakce a množství přidávaných složek.

	Popis, výrobce	Amplifikovaný vzorek	Neamplifikovaný vzorek
H ₂ O (μ l)		18,2	17,2
Pufř (μ l)	PCR Buffer, 10x, 15mM MgCl ₂ , Quiagen	2,5	2,5
Deoxynukleotidová směs (μ l)	Top-Bio	1	1
Polymeráza (μ l)	HotStar DNA polymerase, Quiagen	0,3	0,3
Forward primer (μ l)	generi biotech	1	1
Reverse primer (μ l)	generi biotech	1	1

	Popis, výrobce	Amplifikovaný vzorek	Neamplifikovaný vzorek
DNA (μl)		1	1
Celkový objem (μl)		25	25

Tab. III: Tabulka programů PCR reakcí

	Program L-symb		Program N-symb	
Počáteční denaturace	95°C	15 min	95°C	15 min
Denaturace	94°C	1,5 min	94°C	1,5 min
Nasedání primerů	54°C	1,5 min	50,5°C	1,5 min
Syntéza DNA	72°C	1,5 min	72°C	1,5 min
Konečná syntéza DNA	72°C	10 min	72°C	10 min
Počet cyklů	30		30	

Tab. IV: Tabulka primerů používaných pro PCR reakce

Označení	Sekvence (5'-3')	Gen (Zdroj)
Neis 16S F2	CGAGTAATGGGGGATAAC	Neisseriales 16S rDNA (nově navržený)
Neis 16S R2	CGGTGCTTATTCTTTAGGT	
Neis 16S F3	GAACGTGCCGAGTAATGG	Neisseriales 16S rDNA (nově navržený)
Neis 16S R3	AGCCGGTGCTTATTCTTT	
Neis GroEL F2	AGCTGAAAACATCGCCAA	Neisserie GroEL (nově navržený)
Neis GroEL R1	TTCGGAAATCACTACGCC	
Legio 16S F3	ACTATYKACTTCTGGTGC	Legionellales 16S rDNA (nově navržený)
Legio 16S R3	TGCGTCYGATTAGCTAGTT	
Legio GroEL F1	TTGAGGGACAYAAAGCAG	Legionella GroEL (nově navržený)
Legio GroEL R1	TACGATCACCAAACWCCAG	
F40	GCGGCAAGCCTAACACAT	16S rDNA (Hypša & Křížek)
R1060	CTTAACCCAACATTTCTCAACACGAG	
EUB 16S F	GCTTAACACATGCAAG	16S rDNA (O'Neill et al., 1992)
EUB 16S R	CCATTGTAGCACGTGT	
SP6	ATTTAGGTGACACTATAGAA	Sekvence produktů při klonování
T7	TAATACGACTCACTATAGGG	
DF	CACACTGGAAGTGGAYACG	Legionella 16S rDNA (Hypša & Křížek)
DR	CRACACGAGCTGACGACA	
JRF 50	GAAAGCTTGCTTTCTTGTCG	N-symb (nově navržený)
JRR	ACCCAGTCATGAAACATAC	

3.1.3 Gelová elektroforéza

Přibližně 2-3 μ l PCR produktu bylo smícháno se stejným množstvím směsi 6x DNA Loading Dye (Thermo Scientific) + fluorescenční barvivo SYBR Green (Invitrogen) a napipetováno do připraveného 1% agarózového gelu. Pro přibližné určení délky fragmentů DNA byl použit 1kb DNA Ladder (O'GeneRuler 1kb, Thermo Scientific). Po elektroforetické separaci byly fragmenty DNA vizualizovány pomocí UV záření v dokumentačním zařízení UVITEC. Při naamplifikování více fragmentů DNA (v případě legionelly), způsobeným nespecifickým nasedáním primerů byla provedena extrakce požadovaného PCR produktu z gelu ručně vymačkáním přes parafilm.

3.1.4 Klonování a transformace

Ke klonování byl použit kit s vektorem pGEM®-T Easy (Promega), do kterého byl zaligován PCR produkt. Ligační reakce byla namíchána do 5 μ l – 0,5 μ l T4 ligáza (Promega), 1,5 μ l PCR produkt, 0,5 μ l vektor, 2,5 μ l 2x ligační pufr (Promega). Tento roztok byl nechán přes noc ve 4°C.

K transformaci plazmidů s vloženým inzertem do komplementárních buněk bylo smícháno (na ledu) 2,5 μ l ligační reakce s 25 μ l kompetentních buněk. Tato směs byla nechána 30 min. na ledu a následně byl proveden heat shock (42°C na 30s), kdy buňky přijmou plazmid s vloženým inzertem. Poté bylo do směsi přidáno 300 μ l SOC (Super Optimal Broth) média a následovně byla směs umístěna na 60 min. při 37°C do rotační třepačky. Takto připravený roztok byl rozetřen na připravené gelové plotny (120 μ l) s ampicilinem a 40 μ l substrátu X-Gal, jehož štěpením vzniká modrý produkt. Tyto plotny byly ponechány přes noc na 37°C v inkubátoru. Následně byla provedena modro-bílá selekce, kdy právě bílé kolonie obsahovaly vložený inzert. Bílé kolonie byly sesbírány a ponechány v LB (Lysogeny Broth) médiu s antibiotikem (ampicilin) přes noc na 37°C v inkubátoru.

Izolace plazmidů z kompetentních buněk byla provedena pomocí kitu (Plasmid Miniprep Spin Kit, Jetquick), dle manuálu výrobce. Poté došlo k izolaci inzertu z plazmidu – reakční směs: 1 μ l vyizolovaných plazmidů, 0,4 μ l Eco RI (restrikční enzym, Fermentas), 1 μ l pufru (10x, Fermentas), 7,5 H₂O (destilovaná, deionizovaná). Ponecháno 1 hod. ve 37°C a vzniklý produkt vizualizován na agarózovém gelu.

3.1.5 Amplifikace

Z důvodu malého množství DNA izolátu z jednotlivých vší a nízké koncentraci obsažené DNA (změřeno na nanophotometru - NanoPhotometer P-Class, IMPLLEN) bylo rozhodnuto o celkové amplifikaci některých zkušebních vzorků. Amplifikace byla provedena podle návodu výrobce pomocí kitu (Illustra GenomiPhi V2 DNA Amplification Kit).

3.1.6 Práce se sekvencemi a fylogenetické analýzy

Získané nukleotidové sekvence byly zpracovány v programu Geneious (<http://www.geneious.com>, Kearse et al., 2012). Orientační taxonomická příslušnost osekvenovaných bakterií byla ověřena pomocí Standard Nucleotide BLAST Search nukleotidové databáze instituce NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). V programu Geneious byly u takto vybraných sekvencí odstraněny nepřesně osekvenované konce sekvencí, sekvence primerů a vektorů. Sekvence zachycené primerem forward a reverse byly asemblovány pomocí funkce De Novo Assemble (Geneious). Pomocí BLAST search byly vyhledány nejpodobnější bakteriální sekvence malé ribozomální podjednotky 16S (500 pro každou z obou bakterií) a alignovány pomocí programu MAFFT Alignment (Kato & Standley, 2013), kde byl použit E-INS-i algoritmus. Alignment byl v programu GBlocks (Castresana, 2000) zbaven nepřesně alignovaných pozic a divergentních oblastí, které mohou později způsobovat problémy a artefakty při fylogenetických analýzách.

V této fázi obsahovaly matice vzorek 501 sekvencí pro Legionellales a 525 sekvencí pro Neisseriales vybraných tak, aby při vyhledávání pozice symbiontů nemohlo dojít k artefaktu v důsledku nedostatečného samplingu. Pro první ověření přibližných fylogenetických pozic byl použit program FastTree (Price et. al, 2009) v programu Geneious, který využívá analýzu maximum-likelihood pro sestavení rychlého a přibližného fylogenetického stromu z velkých datasetů. Pro naše datasety bylo ve FastTree ponecháno defaultní nastavení. Podle těchto stromů byly datasety redukovány tak, aby pomocí několika desítek taxonů poskytly reprezentativní fylogenetické pokrytí obou čeledí (Neisseriales i Legionellales).

K sestavení fylogenetických stromů z obou takto upravených datasetů byly použity dva typy analýzy – maximum-likelihood a (ML) v programu PhyML (Guindon et. al, 2010) a výpočet posteriorních pravděpodobností pomocí metody Bayesian inference (BI) v programu MrBayes v3.1.2 (Ronquist & Huelsenbeck, 2003). Optimální modely pro obě matice byly vyhledány v programu jModeltest (Darriba et al., 2012). Výsledné modely (GTR+G+I v případě neisserií a TVM +G+I v případě legionel) byly použity v následné ML

analýze v programu PhyML (frekvence nukleotidů, poměr tranzicí/trasverzí a podíl konzervativních míst byly ponechány jako proměnné odhadované programem přímo z dat). Pro rozložení substituční rychlosti byla nastavena aproximace pomocí gamma distribuce ve 4 kategoriích a parametr určující zakřivení gamma distribuce α byl opět odhadnut programem z dat. Počáteční strom byl sestaven distanční metodou BIONJ. Pro prohledávání topologií stromů byla použita metoda SPR (Subtree Pruning and Regrafting). Počet opakování analýz byl nastaven na hodnotu 10. Pro ověření vlivu modelu na analyzovaná data byl tentýž postup zopakován s nastavením základního modelu HKY85 (Hasegawa et al. 1985). V programu MrBayes byl použit jako základní model GTR (General Time Reversible - Rodriguez et al., 1990) s gama distribucí aproximovanou ve 4 kategoriích a ostatními parametry ponechanými jako proměnné parametry proposal distribution. Analýza proběhla ve 2 běžích na 5 milionech generací u datasetu legionell a na 10 milionech generací u neisserií (s ohledem na velikost a povahu matic). Dosažení konvergence bylo odečteno přímo z parametrů analýzy. Burn-in byl stanoven po kontrole parametrů arbitrárně na 20%. Výsledné fylogenetické stromy byly upravovány ve dvou grafických editorech FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>) a Inkscape (<http://www.inkscape.org/>).

3.2 Genomické analýzy

Vstupní data pro následující genomické analýzy byla poskytnuta ve formě „pair-end-reads“, která byla zbavena nekvalitních sekvencí a značených adaptorů. O předchozí přípravě dat jsem byla informována následujícím způsobem: DNA byla izolována z 20 jedinců jedné populace vší rodu *Polyplax* sebraných na území Německa. Směs izolované DNA byla odeslána na sekvenaci s využitím sekvenátoru na platformě Illumina HiSeq 2000, která využívá metody NGS (Next-Generation Sequencing) a poskytuje tak celogenomové sekvenování (Whole-Genome sequencing - WGS).

Tato vstupní data jsem dále zpracovala následujícím způsobem:

- 1) Assembly vstupních „reads“ – (tj. sestavení kontigů a skafoldů)
- 2) Fylogenetické roztřídění jednotlivých kontigů
- 3) Extrakce kontigů dané bakterie
- 4) Anotace a blast genů dané bakterie
- 5) Rekonstrukce metabolických drah
- 6) Mapování původních readů na sestavené kontigy

ad 1) Assembly vstupních „reads“ (dále *Input assembly*)

Input assembly byla provedena pomocí programu SPAdes (Bankevich, Nurk et al., 2012), s nastavením parametru *careful*, který snižuje počty mismatchů a krátkých indelů. Spouští zároveň i MismatchCorrector, nástroj, který opracovává data a pro jejich zpracování používá BWA tools (viz. dále). Výsledkem této analýzy byly kontigy (*Complete contigs*) a skafoldy (*Complete scaffolds*) zahrnující sekvence všech obsažených organismů.

Input assembly byla zkontrolována pomocí programu Quast (Gurevich et al., 2013), který slouží jako nástroj pro zhodnocení a porovnání kvality genomové assembly.

ad 2) Fylogenetické rozřídění *Complete contigs*

Fylogenetická příslušnost *Complete contigs* byla zjištěna pomocí metody Phylotyping v programu PhylaAMPHORA (Wang & Wu, 2013), který na kontigy mapuje markery z vlastní bakteriální databáze jednokopiových genů z kompletních bakteriálních genomů zastupujících 20 bakteriálních kmenů. Výsledkem této analýzy byl seznam kontigů (*Nodes*) s jejich příslušným fylogenetickým zařazením.

Od následujícího bodu byl další postup prováděn zvlášť pro každou z vybraných bakterií.

ad 3) Extrakce kontigů dané bakterie (vytvoření *Targeted contigs*)

Nodes byly převedeny do excelové tabulky a zde byla provedena extrakce identifikačních kódů požadovaných *Targeted contigs*, tedy kontigů, které vykazují fylogenetickou příslušnost k čeledi Neisseriales a Legionellales. Pomocí identifikačních kódů byla následně provedena fitrace těchto *Targeted contigs* z *Input assembly*.

ad 4) Anotace a blast genů dané bakterie

Anotace *Targeted contigs* byla provedena pomocí programu PROKKA (Seemann, 2014), což je softwarový nástroj pro rychlou anotaci bakteriálních genomů. Výsledkem této analýzy byl seznam anotovaných genů přítomných v *Targeted contigs* dané bakterie.

Následně byla prohledána proteinová databáze nr (non-redundant) pomocí BLASTP search (protein blast) a k jednotlivým proteinům bylo vyhledáno maximálně 50 nejpodobnějších proteinů z celé databáze. Takto vytvořený seznam proteinů byl převeden do

excelové tabulky a ručně prohledán na potenciální horizontální přenos genů z jiných skupin bakterií.

ad 5) Rekonstrukce metabolických drah

V tomto kroku byly anotované geny z programu PROKKA zpracovány v databázi KEGG (Kyoto Encyclopedia of Genes and Genomes, Kanehisa et al., 2014). KEGG je databázový zdroj, který obsahuje katalogy kompletně osekvenovaných genomů. Nejprve byly anotované geny z programu PROKKA zaslány do interního anotovacího nástroje BlastKOALA (KEGG Orthology And Links Anotation) pro přidělení K čísel uživatelským sekvenčním datům pomocí BLAST vyhledávače proti neredundantnímu setu KEGG GENES.

Výstupní soubor s přidělenými K čísly byl zaslán do programu KEGG Mapper, který je propojen se serverem BlastKOALA. KEGG Mapper je soubor nástrojů, který mapuje elementární datasey (geny, proteiny, malé molekuly) na integrované datasey dostupné přes internet (KEGG pathway maps). K naší potřebě byla využita operace KEGG pathway mapping s volbou Reconstruct Pathway a s její pomocí byly zrekonstruovány metabolické dráhy pro jednotlivé bakterie. Vybraná metabolická dráha pro biotin (viz. Výsledky pro důvod k výběru této dráhy) byla porovnána mezi oběma bakteriemi navzájem (rozdíly N-symb a L-symb) a mezi symbiontem a jeho fylogeneticky nejbližšími příbuznými (např. L-symb vs. *Legionella pneumophila* a *Legionella elongata*) dostupnými v databázi metabolických drah KEGG Pathway.

ad 6) Mapování původních readů na *Targeted contigs*

Smyslem tohoto kroku bylo získání:

- a) souborů „reads” pro studované symbionty (výchozí soubor pro finální assembly – plánovaný krok, který již není součástí této práce),
- b) alignmentů *Targeted contigs* s namapovanými vstupními „reads” pro posouzení kvality assembly.

Nízkodivergentní sekvence (vstupní „reads”) byly mapovány na *Targeted contigs* pomocí softwarových balíčků BWA (Li & Durbin, 2009) a Bowtie2 (Langmead & Salzberg, 2012). Kontigy s namapovanými „reads“ byly vizualizovány v programu Tablet a byla ověřena kvalita jejich assembly (viz. Obr. 10).

3.3 FISH

FISH je cytogenetická metoda, která detekuje určité úseky DNA pomocí fluorescenčně značených specifických oligonukleotidů, které mají komplementární sekvenci k detekovanému úseku DNA. Pro vyzkoušení této metody jsem použila 1 jedince vší rodu *Polyplax* z lokality Hlinsko (Česká republika). Hned po sběru byla veš fixována v acetonu pro uchování struktury tkáně. Po 24 hodinách byl vzorek přemístěn na dalších 24 hodin do 6% peroxidu vodíku v ethanolu, kde došlo k dehydrataci a odbarvení tkáně. Dále byl odpipetován peroxid vodíku a tkáň byla rehydratována alkoholovou řadou. Ethanolová řada byla míchána do 100ml a to v následujících koncentracích: 1. 100%, 2. 95%, 3. 80%, 4. 70%. Rehydratace byla započata napipetováním 100 μ l 100% ethanolu ke vzorku, po 5 min. byl ethanol odebrán a postupovalo se stejným způsobem sestupně ethanolovou řadou.

V další fázi byl připraven hybridizační pufr smícháním 29,25g NaCl, 10ml 1M Tris-HCl (pH = 7,2) a 5 ml 10% SDS. Takto připravený roztok byl doplněn do 100ml destilovanou H₂O a sterilizován. Rehydratovaný vzorek byl omýván 15 min. v hybridizačním pufru. Následně byl vzorek prehybridizován spolu s hybridizačním pufrům v inkubátoru ve 45°C po 30 min. Dále bylo smícháno 0,2 μ l 100 μ M eubakteriální sondy s navázaným barvivem fluoresceinem (specifikace: EUB338: 5'-[Flc] GCTGCCTCCCGTAGGA, SIGMA, převzato z práce: Amann et al., 1990) s 50 μ l ukotvovacího média Vectashield (Vector Laboratories) s přimíchaným fluorescenčním barvivem DAPI (49,6-diamidino-2-phenylindole), které se pevně váže na A-T bohaté oblasti DNA. Tento roztok byl doplněn hybridizačním pufrům do 100 μ l.

Přehybridizovaný vzorek byl po inkubaci přemístěn do roztoku se sondou a následně znovu inkubován (hybridizován) po 3 hod. ve 45°C. Po 3 hodinách byl roztok krátce stočen na centrifuze a supernatant odpipetován. Při pokojové teplotě byl vzorek 3x po 10 min. omyt ve 100 μ l hybridizačního pufru předehřátého na 46°C. Dále byl roztok krátce stočen na centrifuze a vzorek byl spolu s 20 μ l média Vectashield přemístěn na sklíčko a okraje krycího sklíčka byly připevněny lakem na nehty. Preparát byl ponechán 3 hod. ve 4°C a následně prohlížen pomocí konfokálního mikroskopu. (Olympus FV1000).

4 VÝSLEDKY

4.1 Screening bakterií

Ze screeningu 24 jedinců rodu *Polyplax* byla získána 1 sekvence 16S rDNA N-symb z amplifikovaného vzorku S4 (Slovensko) a 2 sekvence 16S rDNA L-symb z amplifikovaných vzorků S5 (Slovensko) a 24 (Srbsko). U ostatních jedinců se mi nepodařilo získat sekvence genu pro 16 S ribozomální podjednotku ani sekvence genu GroEL. Amplifikace daného úseku DNA pomocí klonování a transformace se nezdařila. Další sekvence 16S rDNA obou symbiotů byly vyextrahovány z genomických dat.

4.2 Fylogenetické analýzy

4.2.1 Legionellales

Výsledky obou analýz (BI, ML) vykazují podobné fylogenetické uspořádání (viz. Obr. 2 a 3). Pro interpretaci a další diskusi získaných fylogenetických stromů v nich vymezují tři hlavní linie. Z důvodu snazší orientace jsou použity následující zkratky pro jednotlivé linie – UNCBACT (větev s jedním zástupcem *Uncultured bacterium* – AB930519), LEGIO (větev s převážným zastoupením bakterií rodu *Legionella*), L-SYMB (větev s bakteriálními endosymbionty vši rodu *Polyplax*). UNCBACT se klastruje jako sesterská linie k liniím LEGIO a L-SYMB. Ve výsledcích BI analýzy lze zaznamenat polytomii 8 linií bakterií rodu *Legionella* uvnitř linie LEGIO. Endosymbionti vši rodu *Polyplax* se klastrují mimo linii LEGIO, ale zároveň je zde patrná nízká posteriorní pravděpodobnost pro monofylii linie LEGIO. Jako outgroup byly použity bakterie *Thiobacillus plumbophilus* (Betaproteobacteria), *Beggiatoa alba* (Gammaproteobacteria) a *Pseudomonas* sp. (Gammaproteobacteria). Při použití vyhledaného modelu TVM v ML analýze se výsledek shodoval s fylogenetickým stromem vytvořeným BI analýzou.

4.2.2 Neisseriales

Z výsledků BI analýzy (viz. Obr. 3 a 4) lze jasně vytyčit 4 monofyletické linie. Pro lepší přehlednost jsou zde použity opět zkratky pro jednotlivé linie: SNOD (větev s převážným zastoupením bakterií rodu *Snodgrassella*), N-SYMB (větev, kde byla fylogeneticky umístěna skupina mnou sledovaných symbiotických bakterií), VITREO (větev s výhradním a úplným zastoupením bakterií rodu *Vitreoscilla*) a NEIS (větev s převážným zastoupením bakterií rodu *Neisseria*). Linie SNOD a NEIS jsou kromě bakterií rodu

Neisseria a Snodgrassella narušeny ještě dalšími rody bakterií. Poslední linie N-SYMB je tvořena skupinou složenou ze dvou sekvencí zástupců N-symb a k ní na dlouhé větvi umístěnou skupinou *Uncultured bacteria* (podrobněji diskuze). N-symb vytváří jasně monofyletickou skupinu s maximální posteriorní pravděpodobností 1 a klastrují se uvnitř řádu Neisseriales společně s liniemi SNOD a VITREO. Bezprostředně fylogeneticky nejbližší bakterií k N-symbiontům je bakterie EU137419 izolovaná z blechy *Oropsylla hirsuta*, sebrané z psouna prériového *Cynomys ludovicianus* (Jones et al., 2008). Jako outgroup byla použity sekvence 16S rDNA bakterií *Enterobacter soli* (Gammaproteobacteria), *Methylophilus luteus* (Betaproteobacteria), *Burkholderia humi* (Betaproteobacteria) a *Chromobacterium violaceum* (Betaproteobacteria). Ve výsledcích BI analýzy sekvence bakterie *Enterobacter soli* způsobovala artefakty, proto byla z matice odstraněna a jako outgroup byly použity pouze sekvence bakterií *Methylophilus luteus*, *Burkholderia humi* a *Chromobacterium violaceum*.

Výsledek ML analýzy s použitým modelem HKY85 je v celkové struktuře kompatibilní s výsledkem BI analýzy. Nejvýraznější změna oproti BI analýze je, že linie NEIS je zde jasně parafyletická oproti ostatním liniím. Při použití vyhledaného modelu GTR v ML analýze je výsledný fylogenetický strom kompatibilní s výsledky BI analýzy.

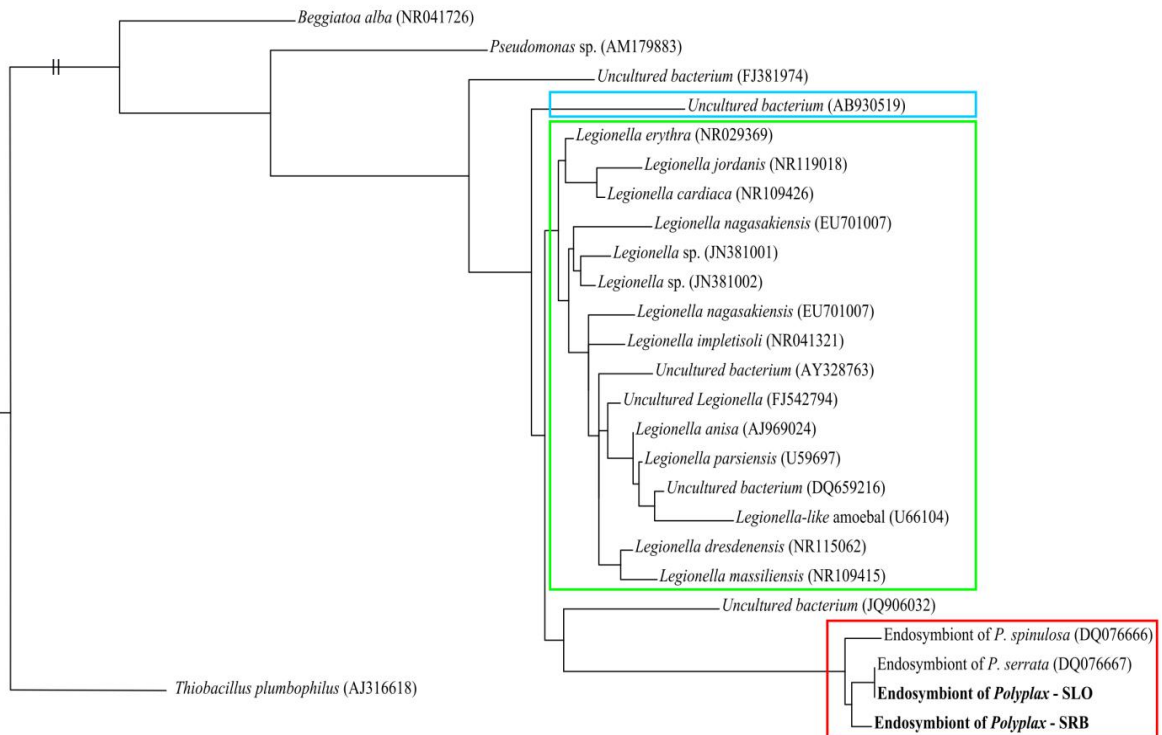
4.3 Genomické analýzy

Pomocí bioinformatických analýz byla vstupní data složena do celkového počtu 124985 kontigů (viz. Tab. V). Nejdelší kontigy byly přiřazeny hostitelské vši rodu *Polyplax*. Extrakcí *Targeted contigs* bylo získáno 23 kontigů pro N-symb a 112 kontigů pro L-symb (viz. Tab. V). Pro zjištění zda se jedná o bakteriální endosymbionty bylo provedeno porovnání velikosti genomu, počtu genů a CG obsahu L-symb a N-symb s jejich volně žijícími příbuznými, jejichž charakteristiky byly vyhledány pomocí genové databáze poskytované institucí NCBI (viz. Tab. VI a VII). Při prohlížení výsledků genomických analýz je zjevné, že *Targeted contigs* obou bakterií jsou dostatečně dlouhé a při sledování coverage jednotlivých kontigů se zjistilo, že mají poměrně vysoké pokrytí mapovanými ready. Následně byly genomy obou bakterií prohledány na hypotetický HGT u obou bakterií. Ve dvou kontizích genomu L-symb byl objeven operon složený z 6 genů, kódující biosyntézu vitamínu B7 (biotin), jehož geny vykazují nejvyšší sekvenční shodu s homologickými geny endosymbionta *Wolbachia* ze štěnice *Cimex lectularius*. Geny operonu jsou v kontigu seřazeny v tomto pořadí: BioD, BioH, BioF, BioB, BioA. Ani jeden

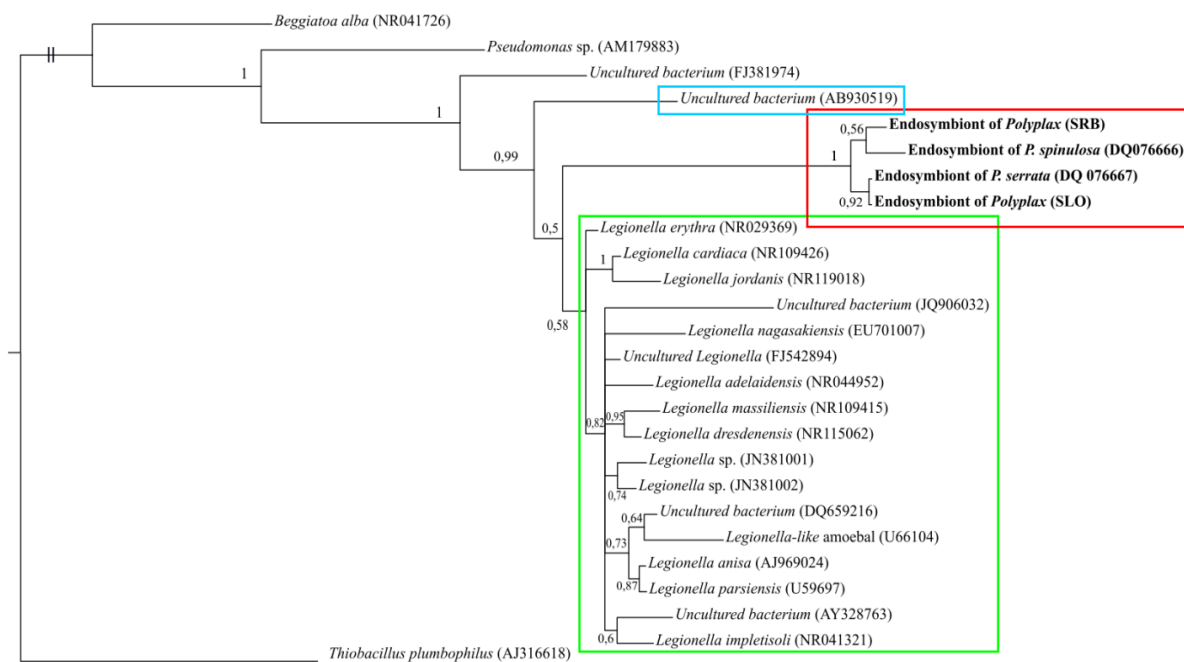
z kontigů L-symb obsahujících biotinový operon nevykazuje známky chimérní assembly (viz. Obr. 10). K těmto získaným kontigům byly pomocí BLAST search (Geneious) vyhledány nejpodobnější bakteriální sekvence (viz. Obr. 6 a 7). Pro bližší pochopení zapojení biotinového operonu v metabolické dráze biosyntézy biotinu, byly sestaveny tyto metabolické dráhy pro L-symb i N-symb a jejich základní struktura byla porovnána v rámci nejpříbuznějších bakterií z čeledí Legionellales a Neisseriales (viz. Obr. 8 a 9). Dále byly porovnány počty genů pro jednotlivé funkční skupiny genů a srovnány mezi L-symb a N-symb (viz. Obr. 10). V příloze jsou umístěny tabulky s přehledem počtu genů kódující proteiny identifikované v genomu L-symb (Tabulka I) a tabulka porovnávající počet genů L-symb a N-symb pro skupiny definované databází KEGG (Tabulka II).

4.4 FISH

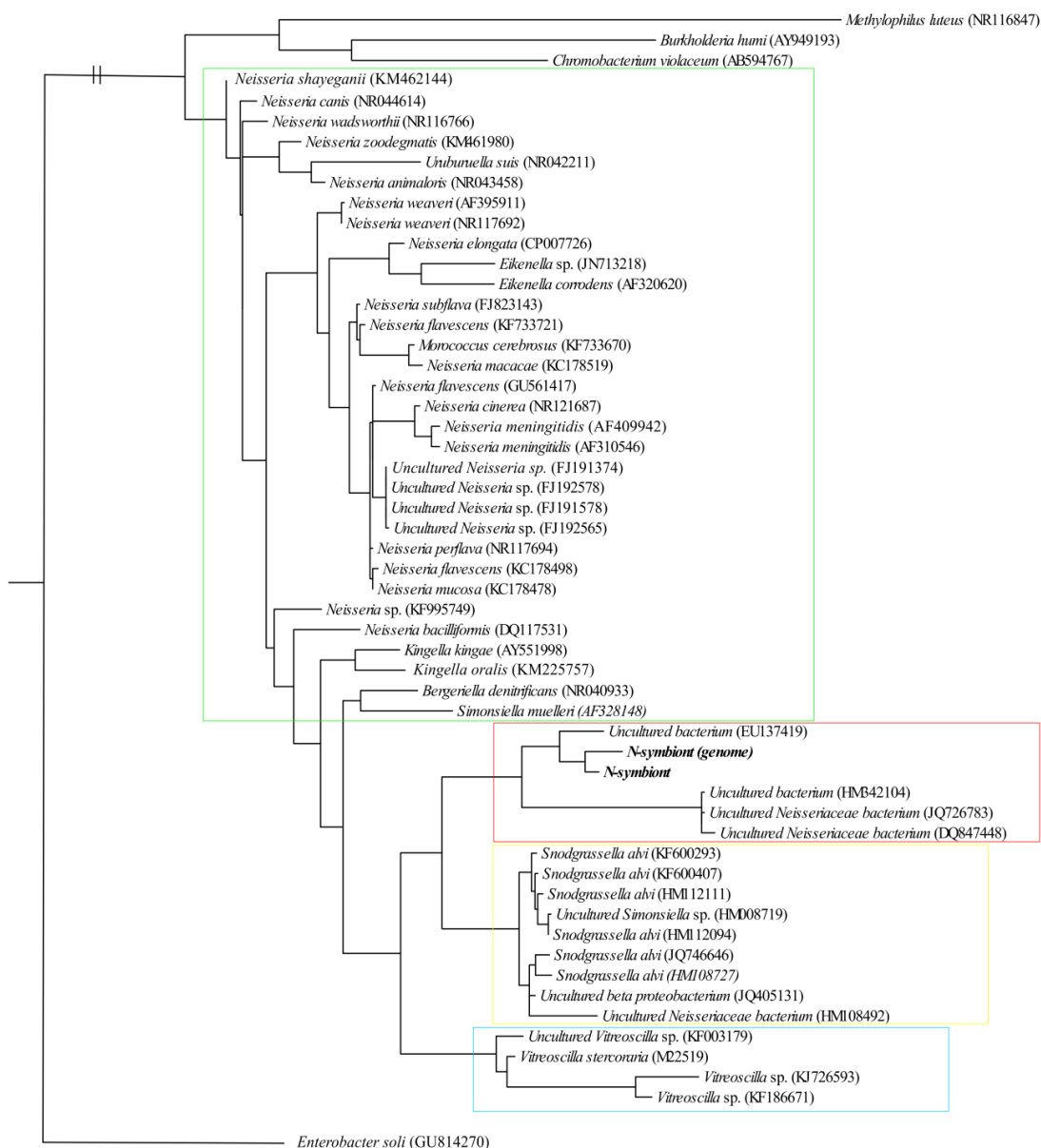
S využitím metody FISH bylo zjištěno přibližné umístění bakteriálních endosymbiontů uvnitř těla vši rodu *Polyplax* (viz Obr. 12). Pro vyzkoušení této metody byly použity 2 vzorky vši rodu *Polyplax* (sesbírané v lokalitě Hlinsko) a pro závěrečnou vizualizaci byl vybrán ten vzorek, u kterého byla přítomnost bakteriálních endosymbiontů zřetelnější.



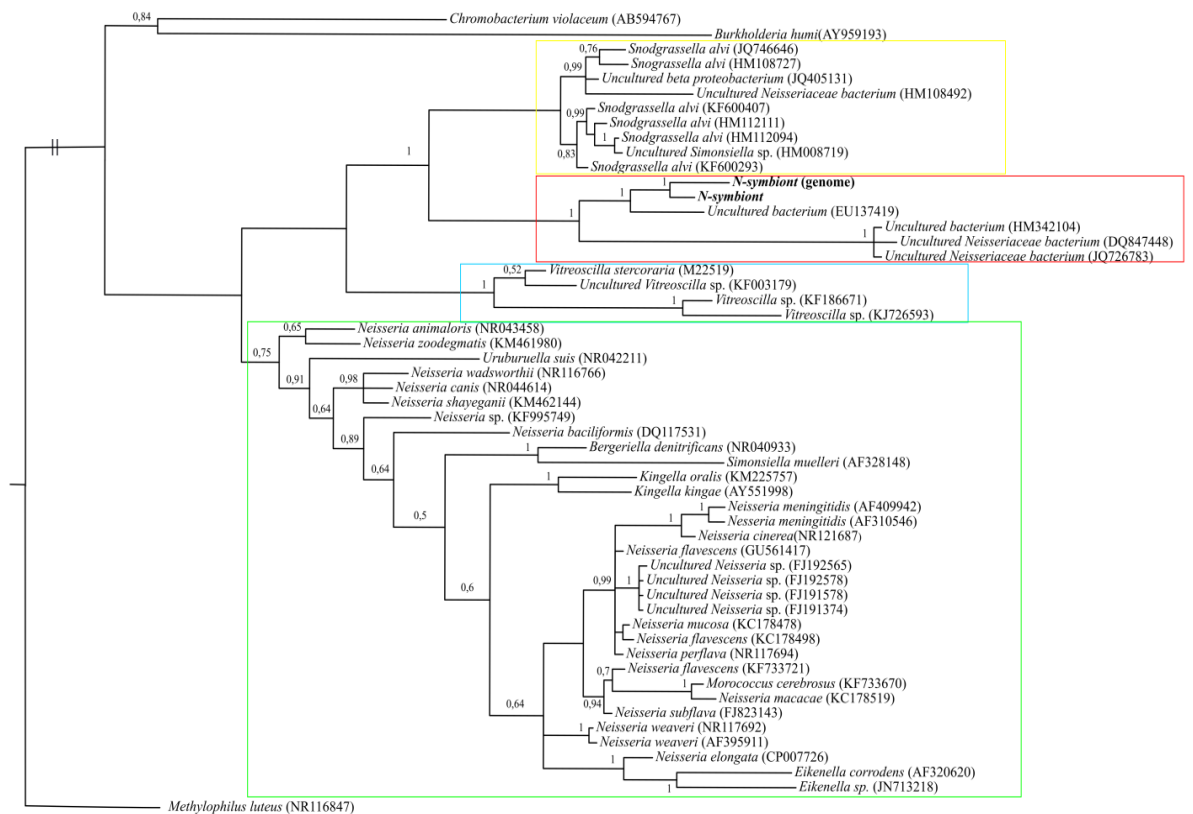
Obr. 2: Fylogeneze L-symb v rámci bakterií z čeledi Legionellales (Gammaproteobacteria) sestavená ML analýzou pomocí 16S rDNA. Jednotlivé linie barevně vyznačeny: L-SYMB (červeně), LEGIO (zeleně), UNCBACT (modře).



Obr. 3: Fylogeneze L-symb v rámci bakterií z čeledi Legionellales (Gammaproteobacteria) sestavená BI analýzou pomocí 16S rDNA. Hodnoty posteriorních pravděpodobností jsou vyznačeny u odpovídajících uzlů. Jednotlivé linie barevně vyznačeny: L-SYMB (červeně), LEGIO (zeleně), UNCBACT (modře).



Obr. 4: Fylogeneze N-symb v rámci bakterií z čeledi Neisseriales (Betaproteobacteria) sestavená ML analýzou pomocí 16S rDNA. Jednotlivé linie barevně vyznačeny: N-SYMB (červeně), NEIS (zeleně), VITREO (modře) a SNOD (žlutě).



Obr. 5: Fylogeneze N-symb v rámci bakterií z čeledi Neisseriales (Betaproteobacteria) sestavená BI analýzou pomocí 16S rDNA. Hodnoty posteriorních pravděpodobností jsou vyznačeny u odpovídajících uzlů. Jednotlivé linie barevně vyznačeny: N-SYMB (červeně), NEIS (zeleně), VITREO (modře) a SNOD (žlutě).

Tab. V: Tabulka základní charakteristiky získaných genomických dat (výstup z programu Quast)

	Input assembly	L-symb	N-symb
Celkový počet skafoldů	113560	–	–
Celkový počet kontigů	124985	112	23
Počet kontigů (> 1000 bp)	19195	92	23
Celková velikost	≈ 177 Mb	≈ 812 Kb	≈ 1,69 Mb
Počet genů		854	1663
CG obsah (%)	38,83	22,96	33,77
Coverage	–	30-250x	28-32x

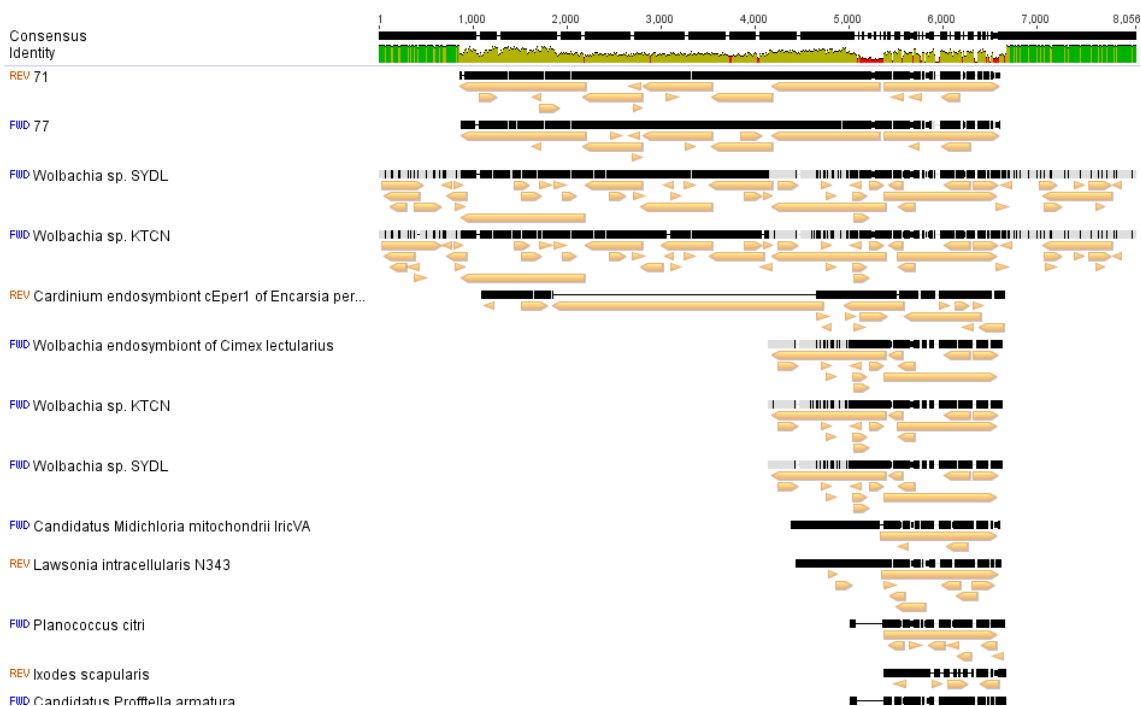
Tab. VI: Tabulka základní genomické charakteristiky hlavních zástupců čeledi Legionellales ve srovnání se základní charakterizací genomu bakterie L-symb

Název bakterie	Přístupový kód (NCBI)	Velikost genomu (Mb)	Počet genů	CG obsah (%)	Výskyt
<i>Legionella pneumophila</i>	NC_002942.5	3,4	3003	38,3	Patogen člověka
<i>Legionella longbeachae</i>	NC_013861.1	4,08	3488	37,1	Patogen člověka
<i>Legionella massiliensis</i>	NZ_AJGC000000000.1	1,73	1810	38,9	Patogen člověka
<i>Legionella anisa</i>	NZ_CANP000000000.1	4,32	3821	38,2	Patogen člověka
<i>Legionella dracourtii</i>	NZ_ACUL000000000.2	4,07	3646	39,2	Patogen améb
<i>Legionella norrlandica</i>	NZ_JNCF000000000.1	3,07	2710	37,5	Vodní bakterie
<i>Legionella moravica</i>	NZ_AUHS000000000.1	3,76	3199	40	Vodní bakterie
<i>Legionella fallonii</i>	NZ_LN614827.1	4,18	3492	38,3	Vodní bakterie
L-symb	–	0,812	854	22,96	Endosymbiont vší rodu <i>Polyplox</i>

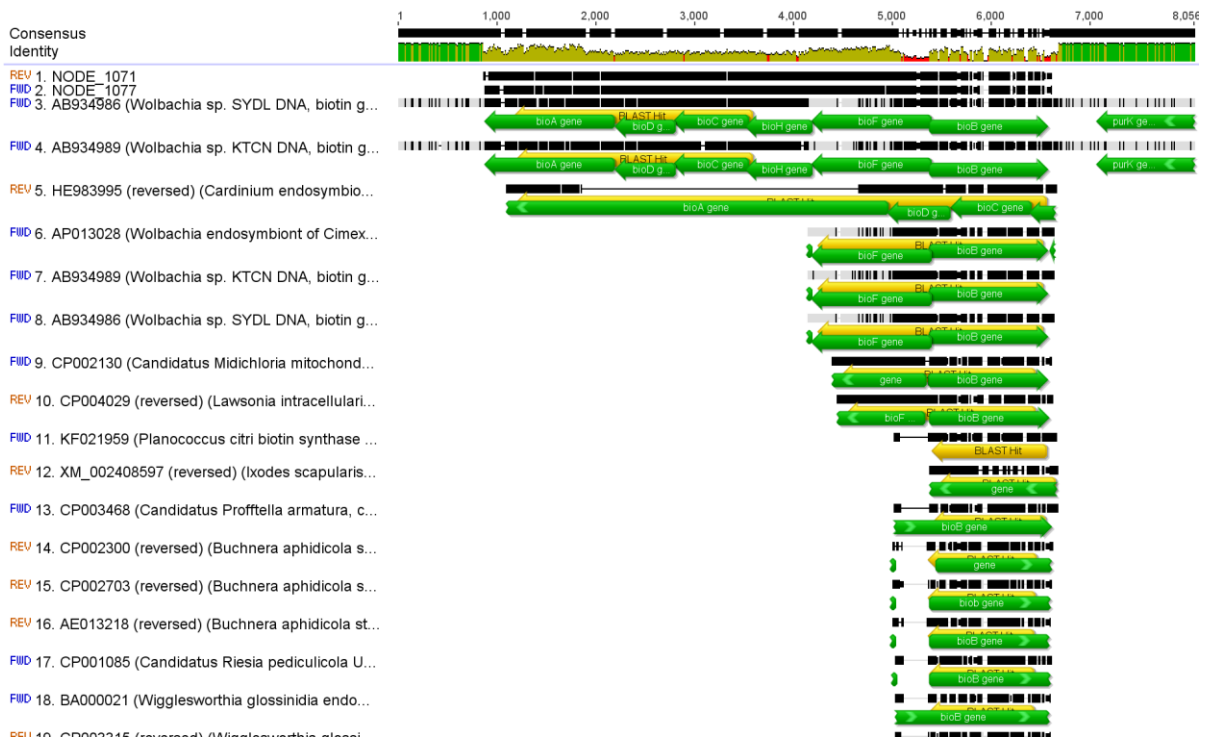
Tab. VII: Tabulka základní genomické charakteristiky hlavních zástupců čeledi Neisseriales ve srovnání se základní charakterizací genomu bakterie N-symb

Název bakterie	Přístupový kód (NCBI)	Velikost genomu (Mb)	Počet genů	CG obsah (%)	Výskyt
<i>Snodgrassella alvi</i>	CP007446.1	2,53	2370	41,3	Střevo včel/čmeláků
<i>Vitreoscilla stercoraria</i>	NZ_ARNN000000000.1	2,58	2372	43,9	Aerobní bakterie
<i>Kingella kingae</i>	NZ_AFHS000000000.1	1,92	2157	46,8	Patogen člověka
<i>Neisseria gonorrhoeae</i>	NC_002946.2	2,15	1975	52,7	Patogen člověka
<i>Neisseria meningitidis</i>	NC_003112.2	2,27	2114	51,5	Patogen člověka
<i>Neisseria elongata</i>	NZ_ADBF000000000.1	2,28	3089	53,5	Patogen člověka
<i>Neisseria lactamica</i>	NC_014752.1	2,22	2137	52,3	Nepatogenní komenzální bakterie člověka
<i>Neisseria cinerea</i>	NZ_ACDY000000000.2	1,87	2191	50,8	Nepatogenní komenzální

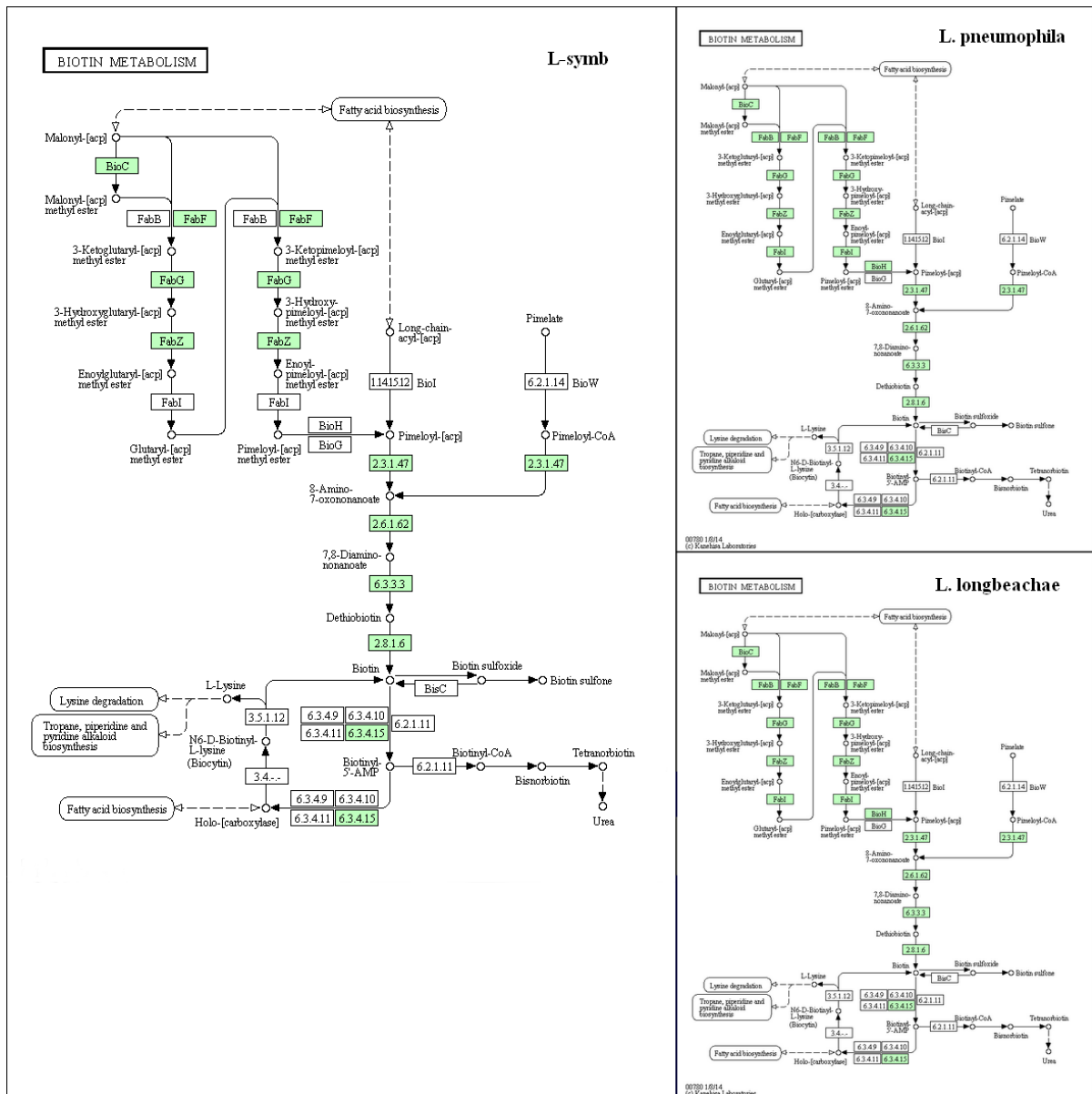
Název bakterie	Přístupový kód (NCBI)	Velikost genomu (Mb)	Počet genů	CG obsah (%)	Výskyt
					bakterie člověka
<i>Neisseria mucosa</i>	NZ_ACRG000000000.1	2,16	2084	49,6	Aerobní bakterie a příležitostný patogen člověka
<i>Neisseria flavescens</i>	NZ_ACQV000000000.1	2,2	2417	49,2	Aerobní bakterie a příležitostný patogen člověka
<i>Simonsiella muelleri</i>	NZ_ADCY000000000.2	2,34	2340	41,5	Aerobní bakterie
N-symb	–	1,69	1663	33,77	Endosymbiont vši rodu <i>Polyplax</i>



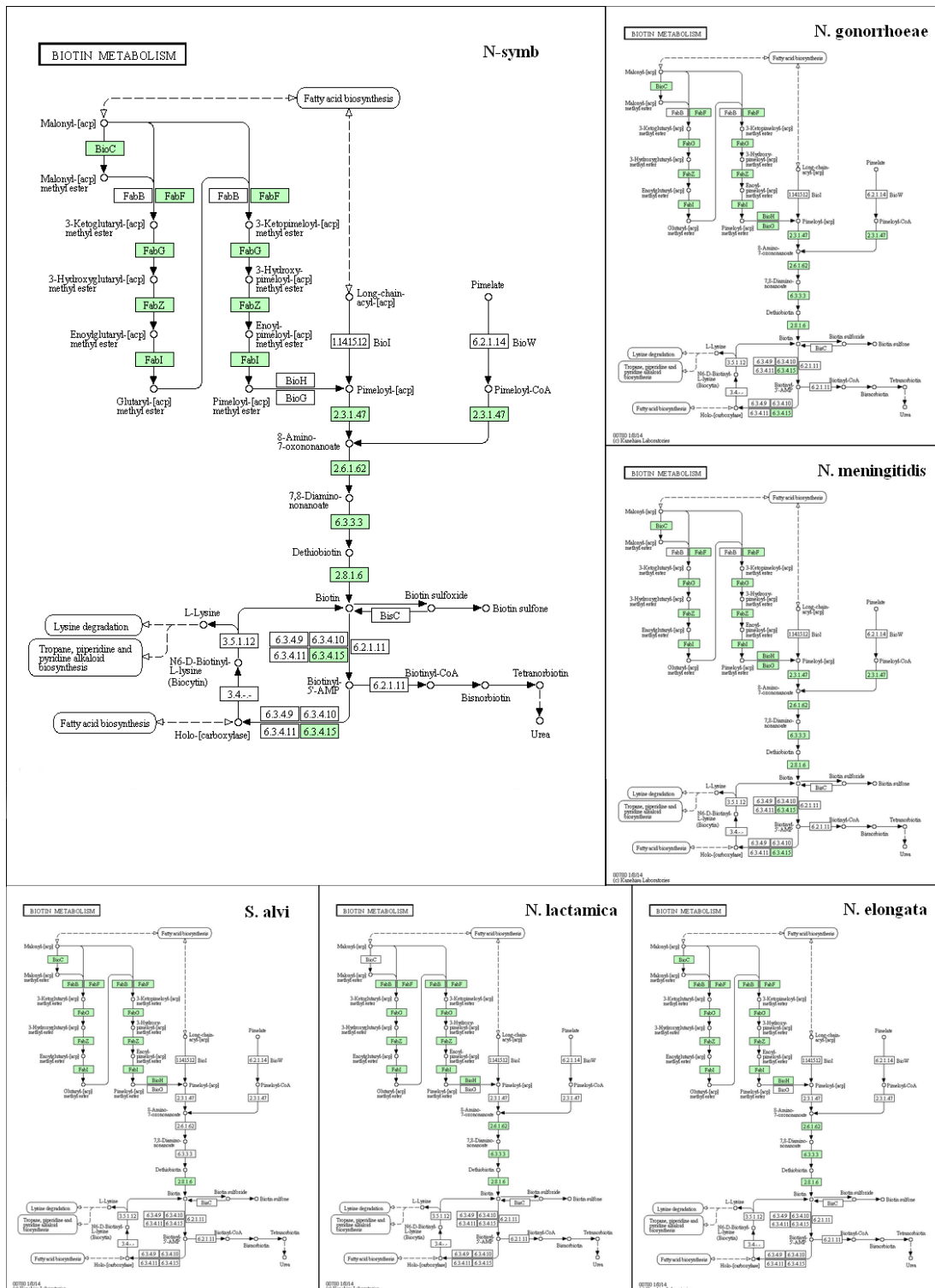
Obr. 6: Automatická identifikace ORF (Open reading frames) pomocí programu Geneious v úseku obsahujícím biotinový operon. Sekvence 77 a 71 jsou části dvou kontigů extrahovaných pro bakterii L-symb, pocházející pravděpodobně, ze dvou linií L-symb (viz. diskuze). Dále jsou zde mapovány kontigy s nejpodobnějším genovým uspořádáním – *Wolbachia* sp. SYDL a *Wolbachia* KTCN jsou kontigy izolované z endosymbiotických bakterií štěnice *Cimex lectularius*.



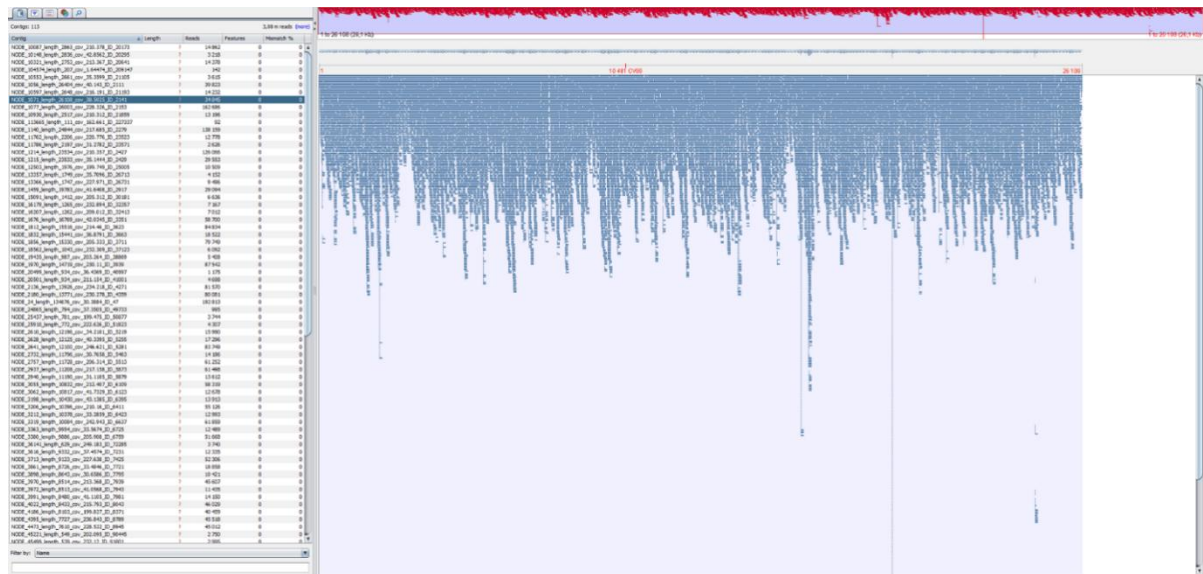
Obr. 7: Genové uspořádání biotinového operonu s vyznačenými ORF a jejich zařazením pomocí BLAST (Geneious). NODE_1071 a NODE_1077 jsou části dvou kontigů extrahovaných z L-symb, pocházející pravděpodobně ze dvou linií L-symb (viz. diskuze). Dále jsou zde mapovány kontigy s nejpodobnějším genovým uspořádáním – *Wolbachia* sp. SYDL a *Wolbachia* KTCN jsou kontigy izolované z endosymbiotických bakterií štěnice *Cimex lectularius*.



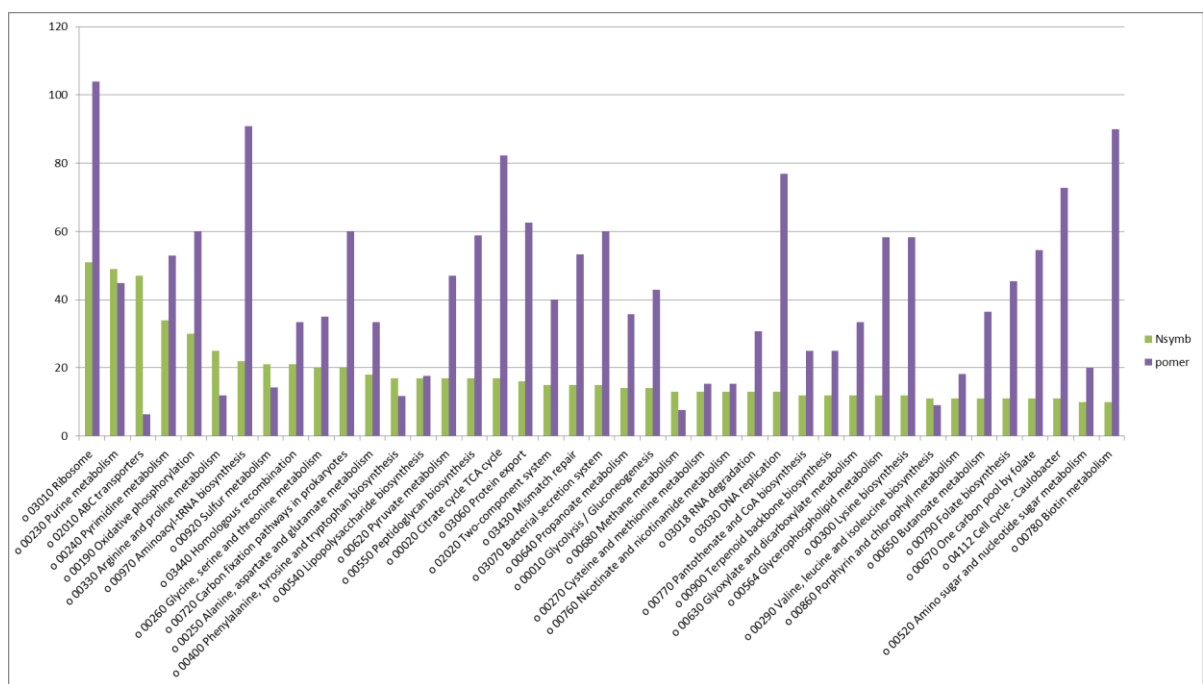
Obr. 8: Přehled biotinových metabolických drah pro všechny dostupné zástupce z čeledi Legionellales ve srovnání s biotinovou metabolickou dráhou *L-symb*. Model této metabolické dráhy je pro všechny bakterie univerzální, mění se zde pouze přítomnost jednotlivých genů obsažených v analyzovaném genomu. Přítomné geny jsou zde zvýrazněny zeleně.



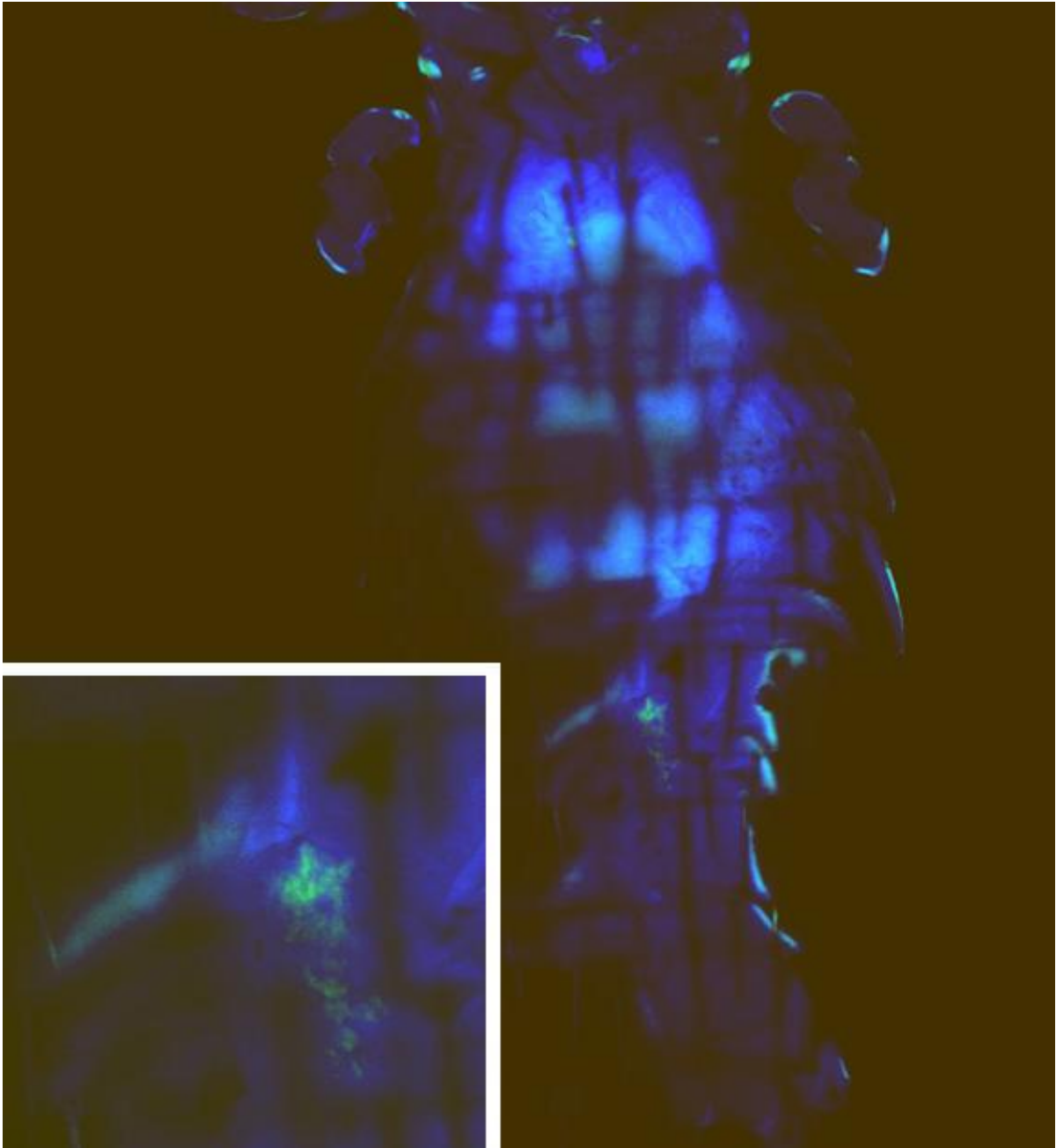
Obr. 9: Přehled biotinových metabolických drah pro všechny dostupné zástupce z čeledi Neisseriales ve srovnání s biotinovou metabolickou dráhou N-symb. Model této metabolické dráhy je pro všechny bakterie univerzální, mění se zde pouze zastoupení jednotlivých genů obsažených v analyzovaném genomu. Přítomně geny označeny zeleně.



Obr. 10: Vizualizace jednoho z dvojice kontigů, jehož součástí je biotinový operon. V tomto kroku byly použity namapované „reads“ mapované na již složené kontigy pro vyloučení chimérně složených kontigů. Zobrazeno v programu Tablet 1.14.10.20 (Milne et al., 2013).



Obr. 11: Graf poměru počtu genů pro funkční skupiny definované databází KEGG: porovnání genomů L-symb (fialově) a N-symb (zeleně).



Obr. 12: Snímek vši rodu *Polyplax* prohlížený pomocí fluorescenčního mikroskopu. Fluorescenčním barvivem DAPI je obarvena jaderná DNA (modře). Fluorescenční barvivo flourescein navázaný na sondě detekuje bakteriální DNA (zeleně). Ve výřezu je přiblížena oblast s výskytem vizualizovaných bakterií.

5 DISKUZE

Tato práce ukazuje, že obě zkoumané bakterie ze vší rodu *Polyplax* (N-symb i L-symb) jsou obligátními endosymbionty. Jedním z důkazů je fylogeneticky velmi blízká příbuznost bakteriálních sekvencí získaných z různých populací vší rodu *Polyplax*, která dokládá, že každá ze symbiotických linií vytváří monofyletickou skupinu procházející koevolucí s hostitelem a není pouze náhodnou kontaminací nebo infekcí. Z výsledků jejich genomického uspořádání lze navíc vyvodit některé předpokládané adaptace pro symbiózu a degenerativní změny typické pro bakteriální endosymbionty.

5.1 Fylogeneze

V případě obou symbiontů může být jejich fylogenetické uspořádání ukázané v této práci zatíženo určitými artefakty, které jsou typické pro fylogenetické analýzy zahrnující endosymbionty hmyzu. V důsledku vysoké rychlosti akumulace mutací, vytvářejí symbiotické linie často dlouhé větve tzv. LBA (long branch attraction), které zvyšují nejistotu jejich fylogenetické příslušnosti (Moran, 1996). Fylogeneze postavená na jednoduché analýze podjednotky 16S rDNA navíc zřejmě neobsahuje dostatek informací pro řešení vztahů v rámci čeledi Neisseriales a Legionellales. (Hypša & Křížek, 2007).

5.2 L-symb

Identifikování obligátní endosymbionti z čeledi Legionellales vytváří monofyletickou linii s nejasnou příbuzností s dalšími bakteriemi této čeledi. Tato nejistota odráží umístění linie L-SYMB na dlouhé větvi, v důsledku kterého je její pozice v jednotlivých analýzách nestabilní. Ve fylogenetickém stromu sestaveném ML analýzou vykazuje k L-symb nejbližší sekvenční příbuznost *Uncultured bacterium* (JQ906032) izolovaná jako potenciálně patogenní bakterie z vodovodní vody (Perkins et al., 2009), zatímco ve fylogenetickém stromu vytvořeném pomocí BI analýzy se tato bakterie klastruje uvnitř linie LEGIO. V rámci celkového srovnání výsledných fylogenezí tedy nelze rozhodnout o vzájemném fylogenetickém postavení linií L-SYMB a LEGIO a nelze vyloučit, že monofyletická linie L-SYMB je ve skutečnosti součástí rodu *Legionella* (viz. Hypša & Křížek, 2007) „odtahovanou“ k sekvencím outgroups v důsledku LBA. Podpurným argumentem může být i nízká posteriorní pravděpodobnost linie LEGIO (0,5). Z dlouhodobých diskuzí v literatuře je patrné, že výběr modelu není vždy věrohodný. Použitím vyhledaného modelu TMV byla pomocí ML analýzy potvrzena monofylie L-symb a fylogeneticky blízká příbuznost

k bakteriím rodu *Legionella*. Použitím více modelů bylo zároveň prokázáno stabilní fylogenetické postavení symbiontů při změně modelu.

5.3 N-symb

Z výsledků fylogenetických analýz čeledi Neisseriales je zřejmé, že obligátní endosymbiont N-symb není blízce příbuzný žádnému známému rodu z této čeledi. N-symbionti vytváří monofyletickou linii N-SYMB spolu s *Uncultured bacterium* (EU137419), bakterií která byla popsána u blechy *Oropsylla hirsuta* (Jones et al., 2008). K linii N-SYMB je fylogeneticky nejpříbuznější skupina *Uncultured bacterium*, která je umístěná na dlouhé větvi. Do této skupiny patří dvě bakterie (HM342104, DQ847448) izolované z kůže člověka (Kong et al., 2012; Gao et al., 2007) a jedna bakterie (JQ726783) izolovaná ze střeva plošnice *Nysius plebeius* (Heteroptera, Matsuura et al., 2012). Velká rozdílnost prostředí, ze kterých byly tyto bakterie izolovány, může poukazovat na nedostatek dat pro tuto skupinu bakterií a tedy absenci přechodných článků. Další fylogeneticky příbuzné linie k N-symbiontům jsou tvořeny bakteriemi rodu *Vitreoscilla* a *Snodgrassella*. Rod *Vitreoscilla* byl popsán jako skupina bezbarvých, aerobních, vláknitých mikroorganismů s klouzavým pohybem (Pringsheim, 1949). Zástupci rodu *Snodgrassella* jsou známi jako mutualističtí symbionti včel (*Apis* spp.) a čmeláku (*Bombus* spp.) (Kwong et al., 2014). Přestože výsledná fylogenetická analýza neumožňuje věrohodnou evoluční interpretaci, je zajímavé, že v příbuzných klastrech se objevují hned dva případy symbiontů hmyzu. Při použití vybraného modelu GTR v ML analýze byl výsledek téměř identický s výsledným fylogenetickým postavením získaným BI analýzou při použití GTR modelu. Porovnáním tří použitých modelů se zásadní fylogenetické postavení všech tří linií nemění.

Problémy s artefakty fylogenetického zařazení obou zkoumaných obligátních endosymbiontů v rámci jejich čeledi by bylo teoreticky možné řešit např. nehomogenními a nestacionárními modely jako je nhPhyML, které počítají s variabilitou v evoluci, minimalizují kompoziční vychýlení a tím zlepšují fylogenetické rekonstrukce (Galtier and Gouy, 1998; Tarrío et al., 2001; Herbeck et al., 2005). Na druhou stranu, použití těchto metod je vhodné a smysluplné pro zpracovávání větších multigenových matic. Jednogenové analýzy v mnoha případech nemůžou dát jistou odpověď o správném fylogenetickém zařazení daných organismů. V poslední době se navíc ukazuje, že v případě fylogenetických artefaktů spojených se symbiotickými bakteriemi, se nejedná jen o problematiku nehomogenosti matice, ale spíše o zařazení konkrétních problematických pozic (Husník et

al., 2011). Fylogenetické analýzy mojí práce jsou provedeny na základě 1 genu (16S ribozomální podjednotka) a slouží pouze ke zjištění přibližného fylogenetického postavení bakteriálních endosymbiontů v rámci čeledi Neisseriales a Legionellales. Z těchto důvodů je v budoucnu jediným možným řešením pro zjištění správného fylogenetického postavení daných bakterií sestavení multigenové matice a použití „protiartefaktových metod“ (Philippe & Roure, 2011).

5.4 Genomika

V dnešní době se běžně využívají metody, které umožňují *de novo* složení genomů různých organismů (Schatz et al., 2010). Pomocí těchto metod se podařilo složit genomy endosymbiotických bakterií do určitého počtu kontigů nikoli však do kompletního genomu, tedy do podoby kruhové molekuly. Ačkoliv genomové sekvenování je v dnešní době rychlé a poměrně levné, hlavním problémem v genomových projektech je vyplnění všech mezer mezi kontigy. Tyto mezery vznikají v důsledku chyb zaváděných již při přípravě vzorku a následně z důvodu přeskoků DNA polymerázy při amplifikaci DNA různých organismů obsažených v jednom vzorku (Gundry & Vijg, 2012), popřípadě v důsledku repetitivních sekvencí. Kompletního složení genomu může být dosaženo určením pořadí a orientace všech kontigů a získáním dalších sekvencí, které vyplní mezery mezi danými kontigy (Delcher et al., 2002).

Při porovnávání genomů bakterií navzájem je zřejmé, že L-symb má menší velikost genomu než N-symb. Tato rozdílnost může být v principu způsobena horší assembly kontigů L-symb nebo redukováností genomu typickou pro bakteriální endosymbionty (McCutcheon & Moran, 2012). Charakteristika získaných dat podporuje spíše druhou variantu, tedy výraznou degeneraci genomu L-symb, typickou pro P-symbionty. Například při porovnání GC obsahu dostupných genomů bakterií rodu *Legionella* s genomem L-symb je zřejmé, že L-symb má výrazně nižší GC obsah oproti svým volně žijícím příbuzným (viz. Tab. V). Při porovnání redukce obou symbiontů se navíc ukázalo, že v genomu L-symb jsou zachovány geny podílející se na základních funkcích buňky (transkripce, translace, replikace) v obdobné míře jako v genomu N-symb, zatímco další funkční skupiny genů jsou podstatně redukovánější (viz. Obr. 11). Tyto výsledky poskytují důkaz, že se v tomto případě nejedná o špatnou assembly daných kontigů L-symb, ale že genom této bakterie vykazuje vlastnosti charakteristické pro genomy obligátních endosymbiontů.

Při důkladnějším studiu kvality kontigů L-symb bylo zjištěno, že se jednotlivé úseky se shodným pořadím genů opakují po dvojicích. Při porovnání aligmentu v programu Geneious je zřejmé, že tyto dvojice kontigů, nebo jejich částí, jsou homologickými úseky genomu, které se odlišují sekvenčně, ale jsou shodné v uspořádání genů (viz. Obr. 6). Z toho lze vyvodit, že se patrně jedná o genomy dvou různých L-symb, patrně ze dvou hostitelských linií rodu *Polyplax* (viz. Tabulka I). S jistotou se dá říci, že při rozdělení těchto dvou genomů se velikost jednoho genomu L-symb zmenší přibližně na polovinu, stejně tak se dá očekávat, že počet výsledných kontigů a genů bude pro jeden genom zhruba poloviční. Díky tomuto poznatku bude otázkou dalšího výzkumu vyhledat vhodné markery vši a pomocí fylogenetických analýz zjistit do kterých linií se klastrují vši rodu *Polyplax* obsažené ve vzorku. Dále na základě porovnání genomických sekvencí těchto dvou L-symb bude možné určit, které geny podléhají selekci u těchto bakteriálních endosymbiontů. Dynamika molekulární evoluce dvou symbiontů se určuje porovnáním celogenomových rychlostí synonymních a nesynonymních substitucí (Yang & Nielsen, 2000).

Při analýze genomických dat endosymbiontů jsem se zaměřila hlavně na vyhledávání genů spojených s produkcí B-vitaminů, které jsou často poskytovány bakteriálními endosymbionty krevsajícího hmyzu (Baumann, 2005). V genomu L-symb jsem našla biotinový operon, který se skládá z 6 genů kódujících syntézu tohoto esenciálního vitamínu. Uspořádání genů operonu je identické s uspořádáním genů biotinového operonu nalezeného v genomu obligátního endosymbionta *Wolbachia* u štěnice *Cimex lectularius* (Nikoh et al., 2014). Zároveň je zde patrná velká sekvenční příbuznost těchto operonů. Bakterie *Wolbachia* je běžná a velmi rozšířená skupina bakterií vyskytující se v reprodukčních orgánech arthropod (Zug & Hammerstein, 2012), proto není zvláštností, že se pravděpodobně předci bakterií L-symb a *Wolbachia* mohli setkat ve stejném hostiteli a operon byl přenesen pomocí HGT z genomu předka bakterie *Wolbachia* do genomu předka bakterie L-symb. Dále byla posouzena ucelenost metabolické dráhy pro syntézu biotinu u L-symb. Zde byl objeven artefakt způsobený programem KEGG, který do předlohy této metabolické dráhy nenamapoval gen BioH, který byl již v předešlých analýzách identifikován v programu Geneious (viz. Obr. 6 a 7). Doplněním genu BioH, můžeme usoudit, že její začátek probíhá od BioC do BioH stejně jako u biotinové biosyntetické dráhy bakterie *Escherichia coli* (Lin & Cronan, 2011). V této dráze od BioC-BioH nebyl namapován gen FabI (enoyl-[acyl-carrier protein] reductase I), který je důležitý pro celkovou komplexitu metabolické dráhy a funkční zapojení biotinového operonu. Možná přítomnost

toho genu v genomu L-symb bude posuzována v rámci dalšího výzkumu. Při porovnání metabolické biosyntetické dráhy L-symb s těmito metabolickými drahami mapovanými u nejbližších příbuzných (*Legionella pneumophila* a *Legionella longbeachae*) je patrné uchování schopnosti syntézy biotinu, která je však u N-symb zajištěna výhradně přeneseným operonem, fylogeneticky vzdáleným genům ostatních legionel.

Uspořádání genů v genomu N-symb neposkytlo zřetelnou informaci o HGT. Tento genom bude podroben detailnější analýze v další fázi výzkumu. Pro srovnání byla analyzována biotinová metabolická dráha N-symb. U N-symb je biotinová dráha také nekompletní. V rámci srovnání této metabolické dráhy mezi nejbližšími příbuznými bakteriemi z čeledi Neisseriales lze posoudit, že u některých bakterií (*Neisseria gonorrhoeae* a *Neisseria meningitidis*) je schopnost biosyntézy biotinu zachována. U těchto dvou druhů bakterií je navíc zajímavý současný výskyt genů BioH a BioG. Pokud by metabolická dráha biotinu u N-symb začínala od BioC dalo by se teoreticky předpokládat, že FabI, který je kódován v této metabolické dráze, by mohl být využíván bakterií L-symb jako esenciální metabolit doplňující biosyntetickou dráhu biotinu. Tato schopnost výměny určitých metabolitů byla teoreticky popsána např. mezi endosymbionty *Tremblaya princeps* a *Moranella endobia* u vlnatky *Planococcus citri* (Husník et al., 2013).

Zatímco přítomnost biotinového operonu, získaného prostřednictvím HGT, se zdá být zjevnou adaptací pro úlohu nutričního symbionta (obdobně, jako v případě wolbachie u štěnic, Nikoh et al., 2014), vytvoření konkrétního evolučního scénáře této události je podstatně komplikovanější.

V principu mohlo k získání tohoto operonu dojít:

1) před vznikem symbiózy „legionel“ s hmyzem, jakožto preadaptaci k symbiotickému způsobu života,

2) po vzniku symbiózy, jakožto způsobu obnovení symbiotické funkce u silně degenerované bakterie. K odlišení těchto dvou alternativ bude zapotřebí taxonomicky širšího vzorku L-symbiontů.

5.5 FISH

Z výsledků této analýzy se s jistotou se nedá říci, kde se zachycení endosymbionti v těle nacházejí, nicméně podle histologických řezů provedených již v pracích Riese (Ries, 1931) by mohlo jít o část střeva či bakteriomu. Pro identifikaci umístění konkrétní endosymbiotické bakterie v těle hostitele bude nutné použít specifické sondy komplementární k vybranému úseku DNA dané bakterie. Tyto sondy bude možné navrhnout až po zdárné optimalizaci podmínek pro bezproblémový screening bakteriálních endosymbiontů. Pro zjištění přesnějšího umístění těchto bakterií bude možné použít příčné řezy vší rodu *Polyplax*. Z hlediska těchto plánovaných kroků je podstatné, že zkušební FISH experiment prokázal funkčnost této metody pro náš vybraný model a její využitelnost v našem dalším výzkumu.

6 ZÁVĚR

Krevsající vší rodu *Polyplax* v sobě hostí dva bakteriální endosymbionty z čeledi Legionellales a Neisseriales. Fylogenetická analýza i hlavní genomické charakteristiky naznačují jejich endosymbiotickou povahu a koevoluční historii společnou s hostitelem. Pro objasnění základních charakteristik a pravděpodobných funkcí těchto endosymbiotických bakterií byla provedena série bioinformatických analýz. Při komplexním studiu genomu L-symb byl objeven operon kódující biosyntézu biotinu. Tento genový klastr byl s největší pravděpodobností přenesen pomocí HGT z genomu nutričního endosymbionta *Wolbachia* popsaného u štěnice *Cimex lectularius*, u kterého byla schopnost biosyntézy biotinu potvrzena. Otázkou dalšího výzkumu je upřesnit fylogenetické postavení L-symb a N-symb v rámci jejich čeledí a na základě podrobnějšího studia genomů posoudit funkční zapojení těchto endosymbiontů v hostitelské evoluci. Z důvodu netypické fylogenetické příslušnosti se daní endosymbionti stali zajímavými objekty pro studium endosymbiózy u hmyzu.

7 POUŽITÁ LITERATURA

- Adeolu M., & Gupta R. S., (2013). Phylogenomics and molecular signatures for the order Neisseriales: proposal for division of the order Neisseriales into the emended family Neisseriaceae and Chromobacteriaceae fam. nov. *Antonie van Leeuwenhoek*, 104: –24.
- Akman, L., Yamashita, A., Watanabe, H., Oshima, K., Shiba, T., Hattori, M., & Aksoy, S. (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nature genetics*, 32(3), 402-407.
- Amann, R. I., Binder B. J., Olson R. J., Chisholm S. W., Devereux R. and Stahl D. A., (1990). Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* 56: 1919-1925.
- Bankevich, A., et al. (2012). *Journal of Computational Biology*, 19(5): 455-477. doi:10.1089/cmb.2012.0021.
- Baumann, P. (2005). Biology of bacteriocyte-associated endosymbionts of plant sap-sucking insects. *Annu. Rev. Microbiol.*, 59, 155-189.
- Baumann, P., Moran, N. A., & Baumann, L. (2006). Bacteriocyte-associated endosymbionts of insects. *The Prokaryotes: Volume 1: Symbiotic associations, Biotechnology, Applied Microbiology*, 403-438.
- Boucher, D.H. (1985) The Idea of Mutualism, Past and Future. *The Biology of Mutualism*
- Buchner, P. (1965). Endosymbiosis of animals with plant microorganisms. *Ecology and Evolution*. London: Croom Helm Ltd. P. 1–28.
- Burke, G. R., & Moran, N. A. (2011). Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome biology and evolution*, 3, 195-208.
- Carvalho, F. R., Nastasi, F. R., Gamba, R. C., Foronda, A. S., & Pellizari, V. H. (2008). Occurrence and diversity of Legionellaceae in polar lakes of the Antarctic peninsula. *Current microbiology*, 57(4), 294-300.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17, 540-552.
- Campillos, M., von Mering, C., Jensen, L. J., & Bork, P. (2006). Identification and analysis of evolutionarily cohesive functional modules in protein networks. *Genome Research*, 16(3), 374-382.

- Cleveland, L.R. (1926). Symbiosis among animal with special reference to termites and their intestinal flagellates. *Quarterly Review of Biology*. 1926. Vol. 1. P. 51–59.
- Costopoulos, K., Kovacs, J. L., Kamins, A., & Gerardo, N. M. (2014). Aphid facultative symbionts reduce survival of the predatory lady beetle *Hippodamia convergens*. *BMC ecology*, *14*(1), 5.
- Dale, C., Beeton, M., Harbison, C., Jones, T., Pontes, M. (2006). Isolation, pure culture, and characterization of “*Candidatus Arsenophonus arthropodicus*,” an intracellular secondary endosymbiont from the hippoboscid louse fly *Pseudolynchia canariensis*. *Applied and environmental microbiology*, *72*(4), 2997-3004.
- Darriba D, Taboada GL, Doallo R, Posada D (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* *9*(8), 772.
- *De Bary A. Die Erscheinung der Symbiose. Strassburg: Verlag von Karl J. Trubner, 1879. 30 S.
- Delcher, A. L., Phillippy, A., Carlton, J., & Salzberg, S. L. (2002). Fast algorithms for large-scale genome alignment and comparison. *Nucleic acids research*, *30*(11), 2478-2483.
- Douglas, A. E. (1989). Mycetocyte symbiosis in insects. *Biological Reviews*, *64*(4), 409-434.
- Engel, P. & Moran N. A. (2013). The gut microbiota of insects- diversity in structure and function. *FEMS Microbiology reviews*. 669-735.
- Engel, P. & Moran N. A. (2013). Functional and evolutionary insights into the simple yet specific gut microbiota of honey bee from metagenomic analysis. *Gut microbes*, 60-65.
- *Frank, A.B., (1877). Uber die biologischen Verhältnisse des Thallus einiger krusten flechten. *Beitrage zur Biologie der Pflanzen*. 1877. Vol. 2. S. 123–200.
- Galtier, N., & Gouy, M. (1998). Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. *Molecular biology and evolution*, *15*(7), 871-879.
- Gao, Z., Tseng, C. H., Pei, Z., & Blaser, M. J. (2007). Molecular analysis of human forearm superficial skin bacterial biota. *Proceedings of the National Academy of Sciences*, *104*(8), 2927-2932.
- Garrity, G. M., Bell, J. A., & Lilburn, T. (2005). Legionellales ord. nov. In *Bergey's Manual® of Systematic Bacteriology* (pp. 210-247). Springer US.

- Garrity, G. M., Bell, J. A., & Lilburn, T. (2007). Family II. Coxiellaceae fam. nov. *Bergey's Manual® of Systematic Bacteriology: Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria*, 2, 237.
- Goodhead, I., & Darby, A. C. (2015). Taking the pseudo out of pseudogenes. *Current opinion in microbiology*, 23, 102-109.
- Guindon S., Dufayard J. F., Lefort V., Anisimova M., Hordijk W., Gascuel O. (2010) *Systematic Biology*, 59(3):307-21.
- Gundry, M., & Vijg, J. (2012). Direct mutation analysis by high-throughput sequencing: from germline to low-abundant, somatic variants. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 729(1), 1-15.
- Gurevich A., Saveliev V., Vyahhi N., Tesler G., (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics*. 29 (8): 1072-1075.
- Haitlinger, R. (2005). Polyplax oxyrrhyncha Cummings, 1915 and P. brachyrrhyncha Cummings, 1915 (Anoplura: Polyplacidae) two species new to the fauna of Greece, collected on Acomys minous Bate, 1906 (Rodentia: Muridae) Acomys minous Bate, 1906 (Rodentia: Muridae) üzerinden Yunanistan faunası için iki yeni kayıt, Polyplax oxyrrhyncha Cummings, 1915 ve P. brachyrrhyncha Cummings, 1915 (Anoplura: Polyplacidae). *Türkiye Entomoloji Dergisi*, 29(4).
- Hasegawa, M., Kishino, H., & Yano, T. A. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of molecular evolution*, 22(2), 160-174.
- Herbeck, J. T., & Wall, D. P. (2005). Converging on a general model of protein evolution. *Trends in biotechnology*, 23(10), 485-487.
- Heuer, H., & Smalla, K. (2007). Horizontal gene transfer between bacteria. *Environmental biosafety research*, 6(1-2), 3-13.
- Hooke, R. (1665) *Micrographia: or some physiological description of minute bodies*. London.
- Husník, F., Chrudimský, T., & Hypša, V. (2011). Multiple origins of endosymbiosis within the Enterobacteriaceae (γ -Proteobacteria): convergence of complex phylogenetic approaches. *BMC biology*, 9(1), 87.
- Husnik, F., et al., (2013). Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealy bug symbiosis. *Cell*, 153(7), 1567-1578.

- 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.
- Jamet et al., (2015). A New Family of Secreted Toxins in Pathogenic Neisseria Species. *PLoS pathogens*, *11*(1), e1004592.
- Jitrapakdee, S., & Wallace, J. C. (2003). The biotin enzyme family: conserved structural motifs and domain rearrangements. *Current Protein and Peptide Science*, *4*(3), 217-229.
- Jones, R. T., McCormick, K. F., & Martin, A. P. (2008). Bacterial communities of Bartonella-positive fleas: diversity and community assembly patterns. *Applied and environmental microbiology*, *74*(5), 1667-1670.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* *42*, D199 – D205 (2014).
- Katoh, Standley, (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability.(outlines version 7) *Molecular Biology and Evolution* *30*:772-780.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28* (12), 1647-1649.
- Kim, K. C. (1985). Evolution and host associations of Anoplura.
- Kim, K. C. (2006). Blood-sucking lice (Anoplura) of small mammals: True parasites. In *Micromammals and Macroparasites* (pp. 141-160). Springer, Japan.
- Knight, R. D., Freeland, S. J., & Landweber, L. F. (2001). Rewiring the keyboard: evolvability of the genetic code. *Nature Reviews Genetics*, *2*(1), 49-58.
- Knowles, J. R. (1989). The mechanism of biotin-dependent enzymes. *Annual review of biochemistry*, *58*(1), 195-221.
- Koch, H., & Schmid-Hempel, P. (2011). Socially transmitted gut microbiota protect bumble bees against an intestinal parasite. *Proceedings of the National Academy of Sciences*, *108* (48), 19288-19292.

- Koga, R., Bennett, G. M., Cryan, J. R., & Moran, N. A. (2013). Evolutionary replacement of obligate symbionts in an ancient and diverse insect lineage. *Environmental microbiology*, *15*(7), 2073-2081.
- Kong, H., et al., (2012). Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. *Genome research*, *22*(5), 850-859.
- Kwong, W. K., Engel, P., Koch, H., & Moran, N. A. (2014). Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proceedings of the National Academy of Sciences*, *111*(31), 11509-11514.
- Kyei-Poku, G. K., Colwell, D. D., Coghlin, P., Benkel, B., & Floate, K. D. (2005). On the ubiquity and phylogeny of Wolbachia in lice. *Molecular ecology*, *14*(1), 285-294.
- Langmead B., Salzberg S., (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*:357-359.
- Lawrence, J. G., & Roth, J. R., (1996). Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics*, *143*(4), 1843-1860.
- Lehane, M. J. (2005). The biology of blood-sucking in insects. *Cambridge University Press*.
- Light, J. E., Smith, V. S., Allen, J. M., Durden, L. A., & Reed, D. L. (2010). Evolutionary history of mammalian sucking lice (Phthiraptera: Anoplura). *BMC evolutionary biology*, *10*(1), 292.
- Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, *25*:1754-60.
- Lin, S., & Cronan, J. E. (2011). Closing in on complete pathways of biotin biosynthesis. *Molecular Bio Systems*, *7*(6), 1811-1821.
- Lin, S., Hanson, R. E., & Cronan, J. E. (2010). Biotin synthesis begins by hijacking the fatty acid synthetic pathway. *Nature chemical biology*, *6*(9), 682-688.
- Lynen, F. (1967). The role of biotin-dependent carboxylations in biosynthetic reactions. *Biochemical Journal*, *102*(2), 381.
- Mampel, J., Spirig, T., Weber, S. S., Haagensen, J. A., Molin, S., & Hilbi, H. (2006). Planktonic replication is essential for biofilm formation by *Legionella pneumophila* in a complex medium under static and dynamic flow conditions. *Applied and environmental microbiology*, *72*(4), 2885-2895.

- Martin, B. D. & Schwab, E. (2012). Symbiosis: "Living together" in chaos. *Studies in the History of Biology*, 4(4), 7-25.
- Matsuura, Y., Kikuchi, Y., Meng, X. Y., Koga, R., & Fukatsu, T. (2012). Novel clade of alphaproteobacterial endosymbionts associated with stinkbugs and other arthropods. *Applied and environmental microbiology*, 78(12), 4149-4156.
- McCutcheon, J. P., & Moran, N. A. (2007). Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proceedings of the National Academy of Sciences*, 104(49), 19392-19397.
- McCutcheon, J. P., & Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, 10(1), 13-26.
- Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD and Marshall D. (2013). Using Tablet for visual exploration of second-generation sequencing data. *Briefings in Bioinformatics*, 14(2), 193-202.
- Moran, N. A. (1996). Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proceedings of the National Academy of Sciences*. 93(7), 2873-2878.
- Moran, N. A., McCutcheon, J. P., Nakabachi, A. (2008). Genomics and evolution of heritable bacterial symbionts. *Annual review of genetics*, 42, 165-190.
- Moter, A., & Göbel, U. B. (2000). Fluorescence in situ hybridization (FISH) for direct visualization of microorganisms. *Journal of microbiological methods*, 41(2), 85-112.
- Nikoh, N., McCutcheon, J. P., Kudo, T., Miyagishima, S. Y., Moran, N. A., & Nakabachi, A. (2010). Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS genetics*, 6(2), e1000827
- Nikoh, N., Hosokawa, T., Moriyama, M., Oshima, K., Hattori, M., & Fukatsu, T. (2014). Evolutionary origin of insect – *Wolbachia* nutritional mutualism. *Proceedings of the National Academy of Sciences*, 111(28), 10257-10262.
- Nogge, G. (1976). Sterility in tsetse flies (*Glossina morsitans* Westwood) caused by loss of symbionts. *Experientia*, 32(8), 995-996.
- Ochman, H., & Davalos, L. M. (2006). The nature and dynamics of bacterial genomes. *Science*, 311(5768), 1730-1733.
- Oliver, K. M., Russell, J. A., Moran, N. A., & Hunter, M. S. (2003). Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. *Proceedings of the National Academy of Sciences*, 100(4), 1803-1807.

- O'Neill, S. L., Giordano, R., Colbert, A. M., Karr, T. L., & Robertson, H. M. (1992). 16S rRNA phylogenetic analysis of the bacterial endosymbionts associated with cytoplasmic incompatibility in insects. *Proceedings of the National Academy of Sciences*, 89(7), 2699-2702.
- Park, M., Yun, S. T., Kim, M. S., Chun, J., & Ahn, T. I. (2004). Phylogenetic characterization of Legionella-like endosymbiotic X-bacteria in *Amoeba proteus*: a proposal for 'Candidatus *Legionella jeonii*' sp. nov. *Environmental microbiology*, 6(12), 1252-1263.
- Perkins, S. D., Mayfield, J., Fraser, V., & Angenent, L. T. (2009). Potentially pathogenic bacteria in shower water and air of a stem cell transplant unit. *Applied and environmental microbiology*, 75(16), 5363-5372.
- Perotti, M. A., Kirkness, E. F., Reed, D. L., & Braig, H. R. (2009). Endosymbionts of lice. *Insect symbiosis*, 3, 205-220.
- Philippe, H., & Roure, B. (2011). Difficult phylogenetic questions: more data, maybe; better methods, certainly. *BMC biology*, 9(1), 91.
- Pizza, M., & Rappuoli, R. (2015). Neisseria meningitidis: pathogenesis and immunity. *Current opinion in microbiology*, 23, 68-72.
- Pratt, H. D., & Karp, H. (1953). Notes on the rat lice *Polyplax spinulosa* (Burmeister) and *Hoplopleura oenomydis* Ferris. *The Journal of parasitology*, 495-504.
- Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) FastTree: Computing Large Minimum-Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* 26:1641-1650, doi:10.1093/molbev/msp077.
- Pringsheim, E. G. (1949). The relationship between Bacteria and Myxophyceae. *Bacteriological reviews*, 13(2), 47.
- *Puchta, O. 1954. Experimentelle Untersuchungen über die Symbiose der Kleiderlaus *Pediculus vestimenti* Burm. *Naturwissenschaften* 41:71-72.
- *Puchta, O. 1955. Experimentelle Untersuchungen über die Bedeutung der Symbiose der Kleiderlaus *Pediculus vestimenti* Burm. *Z. Parasitenkd.* 17:1-40.
- *Ries, E. (1931). Die Symbiose der Läuse und Federlinge. *Zeitschrift für Morphologie und Ökologie der Tiere*, 20(2-3), 233-367.
- Ronquist, F. and J. P. Huelsenbeck. 2003. MrBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572-1574.

- Rowbotham, T. J. (1980). Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. *Journal of clinical pathology*, 33(12), 1179-1183.
- Saffo M. B. (1992). Coming to terms with a field: Words and concepts in symbiosis. *Symbiosis*. Vol. 14.P. 17–31.
- Scarborough, C. L., Ferrari, J., & Godfray, H. C. J. (2005). Aphid protected from pathogen by endosymbiont. *Science*, 310(5755), 1781-1781.
- Schatz, M. C., Delcher, A. L., & Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. *Genome research*, 20(9), 1165-1173.
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 15;30(14): 2068-9. PMID: 24642063.
- Serebriiskii, I. G., Vassin, V. M., & Tsygankov, Y. D. (1996). Two new members of the BioB superfamily: cloning, sequencing and expression of BioB genes of *Methylobacillus flagellatum* and *Corynebacterium glutamicum*. *Gene*, 175(1), 15-22.
- Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y., & Ishikawa, H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature*, 407(6800), 81-86.
- Shigenobu, S., & Wilson, A. C. (2011). Genomic revelations of a mutualism: the pea aphid and its obligate bacterial symbiont. *Cellular and Molecular Life Sciences*, 68(8), 1297-1309.
- Stouthamer, R., Breeuwer, J. A., & Hurst, G. D. (1999). *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annual Reviews in Microbiology*, 53(1), 71-102.
- 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.
- *Šulc, K.. (1910). "Pseudovitellus" und ähnliche Gewebe der Homopteren sind Wohnstätten symbiotischer Saccharomyceten.
- Tanaka, T., Tateno, Y., & Gojobori, T. (2005). Evolution of vitamin B6 (pyridoxine) metabolism by gain and loss of genes. *Molecular biology and evolution*, 22(2), 243-250.

- Tarrío, R., Rodríguez-Trelles, F., & Ayala, F. J. (2001). Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. *Molecular Biology and Evolution*, 18(8), 1464-1473.
- Toh, H., Weiss, B. L., Perkin, S. A., 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 tse-tse host. *Genome Research*, 16(2), 149-156.
- Tønjum, T. (2005). Family I. Neisseriaceae Prevot 1933, 119 AL emend. Dewhirst, Paster and Bright 1989, 265. *Bergey's Manual of Systematic Bacteriology*, 775-776.
- Volf, P. (1991). Postembryonal development of Mycetocytes and symbionts of the spiny rat louse *Polyplax spinulosa*. *Journal of Invertebrate Pathology*, 58(1), 143-146.
- Wallace, D. C., & Morowitz, H. J. (1972). Genome size and evolution. *Chromosoma*, 40(2), 121-126.
- Wang, Z., & Wu, M. (2013). A phylum-level bacterial phylogenetic marker database. *Molecular biology and evolution*, 30(6), 1258-1262.
- Werren, J. H., Baldo, L., & Clark, M. E. (2008). Wolbachia: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, 6(10), 741-751.
- *Wildiers, E. (1901). Nouvelle substance indispensable au developpement de la levure. *La cellule*, 18(3), 3-33.
- Yang, Z., & Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular biology and evolution*, 17(1), 32-43.
- Zug, R., & Hammerstein, P. (2012). Still a host of hosts for Wolbachia: analysis of recent data suggests that 40% of terrestrial arthropod species are infected. *PLoS one*, 7(6), e38544.

* Označené zdroje jsou citovány podle současné literatury, nikoli na základě četby původní práce.

8 PŘÍLOHY

Tabulka I: Přehled počtu kopií genů kódující jednotlivé proteiny identifikované v genomu L-symb.

Název proteinu	Počet genů
(Dimethylallyl)adenosine tRNA methyltransferase MiaB	3
10 kDa chaperonin	1
1-acyl-sn-glycerol-3-phosphate acyltransferase	2
1-deoxy-D-xylulose 5-phosphate reductoisomerase	1
23S rRNA (guanosine-2'-O-)-methyltransferase RlmB	1
2-oxoglutarate dehydrogenase E1 component	1
30S ribosomal protein S1	2
30S ribosomal protein S10	3
30S ribosomal protein S11	3
30S ribosomal protein S12	3
30S ribosomal protein S13	3
30S ribosomal protein S14	3
30S ribosomal protein S15	2
30S ribosomal protein S16	2
30S ribosomal protein S17	3
30S ribosomal protein S18	2
30S ribosomal protein S19	3
30S ribosomal protein S2	3
30S ribosomal protein S20	2
30S ribosomal protein S21	2
30S ribosomal protein S3	3
30S ribosomal protein S4	3
30S ribosomal protein S5	3
30S ribosomal protein S6	2
30S ribosomal protein S7	3
30S ribosomal protein S8	3
30S ribosomal protein S9	2
3-deoxy-D-manno-octulosonate 8-phosphate phosphatase KdsC	3
3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ	3
3-hydroxylaminophenol mutase	2
3-methyl-2-oxobutanoate hydroxymethyltransferase	2
3-oxoacyl-[acyl-carrier-protein] reductase FabG	2
3-oxoacyl-[acyl-carrier-protein] synthase 2	4
3-oxoacyl-[acyl-carrier-protein] synthase 3	2
3-phosphoshikimate 1-carboxyvinyltransferase	1
4-hydroxybenzoate octaprenyltransferase	2
4-hydroxy-tetrahydrodipicolinate reductase	2
4-hydroxy-tetrahydrodipicolinate synthase	2

Název proteinu	Počet genů
4'-phosphopantetheinyl transferase superfamily protein	1
50S ribosomal protein L1	3
50S ribosomal protein L10	3
50S ribosomal protein L11	3
50S ribosomal protein L14	3
50S ribosomal protein L15	3
50S ribosomal protein L16	3
50S ribosomal protein L17	3
50S ribosomal protein L18	3
50S ribosomal protein L19	2
50S ribosomal protein L2	3
50S ribosomal protein L20	2
50S ribosomal protein L21	2
50S ribosomal protein L22	3
50S ribosomal protein L23	3
50S ribosomal protein L24	3
50S ribosomal protein L25	2
50S ribosomal protein L27	2
50S ribosomal protein L28	2
50S ribosomal protein L29	3
50S ribosomal protein L3	3
50S ribosomal protein L3 glutamine methyltransferase	4
50S ribosomal protein L30	3
50S ribosomal protein L31	2
50S ribosomal protein L32	2
50S ribosomal protein L33	2
50S ribosomal protein L34	2
50S ribosomal protein L35	2
50S ribosomal protein L36	2
50S ribosomal protein L4	3
50S ribosomal protein L5	2
50S ribosomal protein L6	3
50S ribosomal protein L7/L12	3
50S ribosomal protein L9	2
5-tetrahydropyridine-2	2
7-dimethyl-8-ribityllumazine synthase	2
8-amino-7-oxononanoate synthase	2
A/G-specific adenine glycosylase	2
ABC transporter ATP-binding protein uup	1
Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	2
Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	2
Aconitate hydratase 1	2

Název proteinu	Počet genů
Acyl carrier protein	2
Acyl-[acyl-carrier-protein]-UDP-N-acetylglucosamine O-acyltransferase	1
Adenosylhomocysteinase	2
Adenosylmethionine-8-amino-7-oxononanoate aminotransferase	2
Adenylate kinase	2
Adenylosuccinate lyase	2
Adenylosuccinate synthetase	2
AhpC/TSA family protein	2
Alanine--tRNA ligase	2
Alpha-ketoglutarate-dependent taurine dioxygenase	1
Aminomethyltransferase	2
Antitoxin CptB	2
Arabinose 5-phosphate isomerase KdsD	1
Arginine--tRNA ligase	2
Argininosuccinate synthase	1
Aspartate-semialdehyde dehydrogenase 2	2
Aspartate--tRNA ligase	2
ATP synthase epsilon chain	2
ATP synthase gamma chain	2
ATP synthase subunit a	2
ATP synthase subunit alpha	2
ATP synthase subunit b	2
ATP synthase subunit beta	2
ATP synthase subunit c	2
ATP synthase subunit delta	2
ATP-dependent Clp protease ATP-binding subunit ClpX	2
ATP-dependent Clp protease proteolytic subunit precursor	2
ATP-dependent dethiobiotin synthetase BioD	2
ATP-dependent helicase/nuclease subunit A	2
ATP-dependent protease ATPase subunit HslU	2
ATP-dependent protease subunit HslV	2
ATP-dependent RNA helicase RhlE	1
ATP-dependent zinc metalloprotease FtsH	2
Beta-lactamase HcpA precursor	1
Bifunctional aspartokinase/homoserine dehydrogenase 1	2
Bifunctional ligase/repressor BirA	2
Bifunctional protein FolC	2
Bifunctional protein FolD protein	1
Biotin synthase	2
Bis(5'-nucleosyl)-tetrphosphatase	2
BolA-like protein	2
Carbon storage regulator	2

Název proteinu	Počet genů
CDP-diacylglycerol--glycerol-3-phosphate phosphatidyltransferase	3- 2
Cell division protein FtsZ	2
Cell division topological specificity factor	2
Cell shape-determining protein MreC	2
Citrate synthase 2	2
Cold shock protein CapB	2
Colicin V production protein	2
Cysteine desulfurase	1
cysteine/glutathione ABC transporter membrane/ATP-binding component	1
Cysteine--tRNA ligase	2
Cytidylate kinase	2
D-alanine--D-alanine ligase A	2
D-alanyl-D-alanine carboxypeptidase DacC precursor	2
Daunorubicin/doxorubicin resistance ATP-binding protein DrrA	2
Deoxycytidine triphosphate deaminase	2
Deoxyuridine 5'-triphosphate nucleotidohydrolase	2
Diaminopimelate epimerase	2
Dihydrofolate reductase	2
Dihydrolipoyl dehydrogenase	2
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	2
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex	1
Dihydroneopterin aldolase	2
Dihydropteroate synthase	2
Dihydroxy-acid dehydratase	1
Ditrans	3
DNA gyrase subunit A	2
DNA gyrase subunit B	2
DNA ligase	2
DNA polymerase III subunit alpha	1
DNA polymerase III subunit beta	2
DNA polymerase III subunit epsilon	2
DNA polymerase III subunit tau	4
DNA primase	1
DNA-directed RNA polymerase subunit alpha	3
DNA-directed RNA polymerase subunit beta	1
DNA-directed RNA polymerase subunit beta'	3
DNA-directed RNA polymerase subunit omega	2
DSBA-like thioredoxin domain protein	2
D-transpeptidase catalytic domain	2
Dual-specificity RNA methyltransferase RlmN	2

Název proteinu	Počet genů
Electron transfer flavoprotein-ubiquinone oxidoreductase	1
Electron transport complex protein rnfB	2
Elongation factor 4	2
Elongation factor G	3
Elongation factor P	2
Elongation factor Ts	3
Elongation factor Tu	4
Endonuclease III	2
Endoribonuclease YbeY	2
Exodeoxyribonuclease	2
Exodeoxyribonuclease III	2
Fatty acid desaturase	2
Ferredoxin	1
Ferredoxin 1	2
Fimbrial protein precursor	1
Fumarate hydratase class II	2
Glutamate racemase	2
Glutamate synthase [NADPH] large chain precursor	1
Glutamate--tRNA ligase	2
Glutamine-dependent NAD(+) synthetase	2
Glutamine--tRNA ligase	1
Glutaredoxin-4	1
Glyceraldehyde-3-phosphate dehydrogenase	2
Glycine cleavage system H protein	2
Glycine--tRNA ligase alpha subunit	2
Glycine--tRNA ligase beta subunit	2
GMP synthase [glutamine-hydrolyzing]	2
GTP cyclohydrolase 1	1
GTPase Der	2
GTPase Obg	2
GTPase ObgE/CgtA	1
Guanylate kinase	2
Haemolysin-III related	2
heat-inducible protein	1
Heme A synthase	2
Histidine--tRNA ligase	2
HIT-like protein	1
Hydroxyacylglutathione hydrolase	1
Hypoxanthine-guanine phosphoribosyltransferase	1
Chaperone protein ClpB	1
Chaperone protein DnaJ	2
Chaperone protein HtpG	1
Chromosomal replication initiator protein DnaA	2

Název proteinu	Počet genů
Inner membrane protein YabI	2
Inner membrane transport permease YadH	2
Inner membrane transport protein YajR	2
Inosine-5'-monophosphate dehydrogenase	2
Inositol 2-dehydrogenase/D-chiro-inositol 3-dehydrogenase	1
Inositol-1-monophosphatase	1
Iron-binding protein IscA	1
Iron-sulfur cluster insertion protein ErpA	2
Isocitrate dehydrogenase [NADP]	2
Isoleucine--tRNA ligase	2
Isonitrile hydratase	2
Isopentenyl-diphosphate delta-isomerase	2
L-asparaginase 1	1
Legionella pneumophila major outer membrane protein precursor	2
Leucine--tRNA ligase	2
Lipase 1 precursor	2
Lipase precursor	1
Lipopolysaccharide export system ATP-binding protein LptB	1
Lipopolysaccharide export system protein LptA precursor	1
Lipopolysaccharide-assembly	1
Lipoprotein signal peptidase	1
Lipoyl synthase	2
L-lysine 2	2
Localization factor PodJL	1
Lon protease	2
Lysine--tRNA ligase	2
Major Facilitator Superfamily protein	1
Malonyl CoA-acyl carrier protein transacylase	2
Malonyl-[acyl-carrier protein] O-methyltransferase	2
Membrane fusogenic activity	2
Membrane protein insertase YidC	2
META domain protein	1
Methane monooxygenase component C	1
Methionine aminopeptidase	2
Methionine--tRNA ligase	2
Methionyl-tRNA formyltransferase	2
Modulator of FtsH protease HflC	2
Modulator of FtsH protease HflK	2
Multidrug export protein EmrA	2
Multidrug export protein EmrB	3
Multifunctional CCA protein	2
Na(+)/H(+) antiporter NhaA	2

Název proteinu	Počet genů
N-acetylmuramoyl-L-alanine amidase AmiA precursor	2
N-acetylmuramoyl-L-alanine amidase AmiC precursor	1
NAD(P)H-quinone oxidoreductase subunit 3	2
NAD-dependent malic enzyme	2
NADH-quinone oxidoreductase chain 2	2
NADH-quinone oxidoreductase subunit B	2
NADH-quinone oxidoreductase subunit C 1	2
NADH-quinone oxidoreductase subunit D	2
NAD-specific glutamate dehydrogenase	2
Nicotinate-nucleotide adenylyltransferase	2
Nucleoid-associated protein	1
Nucleoid-associated protein YbaB	1
Nucleoside diphosphate kinase	2
Octanoyltransferase	2
Oligoribonuclease	2
Outer membrane protein assembly factor BamA precursor	3
Pantothenate synthetase	3
PAP2 superfamily protein	2
Para-nitrobenzyl esterase	1
PD-(D/E)XK nuclease superfamily protein	2
peptidase PmbA	2
Peptide deformylase	2
Peptide chain release factor 1	2
Peptide chain release factor 2	2
Peptidyl-tRNA hydrolase	2
Periplasmic pH-dependent serine endoprotease DegQ precursor	2
Persistence and stress-resistance antitoxin PasI	2
Persistence and stress-resistance toxin PasT	1
Phenylalanine--tRNA ligase alpha subunit	2
Phenylalanine--tRNA ligase beta subunit	2
Phosphate acyltransferase	2
Phosphate transporter family protein	2
Phosphatidate cytidyltransferase	3
Phosphatidylglycerophosphatase A	2
Phosphatidylserine decarboxylase proenzyme	2
Phosphoglycerate kinase	2
Phosphoglycolate phosphatase	1
Phospho-N-acetylmuramoyl-pentapeptide-transferase	2
Phosphopentomutase	2
Phosphoserine aminotransferase	3
Polyribonucleotide nucleotidyltransferase	2
preprotein translocase subunit YajC	2
Proline--tRNA ligase	2

Název proteinu	Počet genů
Prolipoprotein diacylglyceryl transferase	1
Protein GrpE	2
Protein translocase subunit SecA	2
Protein translocase subunit SecE	3
Protein translocase subunit SecY	3
Protein-export membrane protein SecG	2
Protein-export protein SecB	1
Protein-L-isoaspartate O-methyltransferase	2
Protoheme IX farnesyltransferase	1
Purine nucleoside phosphorylase 1	2
putative 5-formyltetrahydrofolate cyclo-ligase	2
putative ABC transporter ATP-binding protein YheS	1
putative ABC transporter ATP-binding protein YjjK	1
Putative aliphatic sulfonates transport permease protein SsuC	1
Putative CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyl-transferase 2	2
putative FAD-linked oxidoreductase	1
putative glycerol-3-phosphate acyltransferase	2
putative glycine dehydrogenase (decarboxylating) subunit 1	2
putative glycine dehydrogenase (decarboxylating) subunit 2	2
Putative GTP cyclohydrolase 1 type 2	4
Putative Holliday junction resolvase	2
putative inner membrane protein	4
putative lipoprotein YiaD precursor	1
putative peroxiredoxin	1
Putative signal peptide peptidase SppA	1
putative sulfoacetate transporter SauU	2
Putative universal stress protein	1
Pyruvate dehydrogenase E1 component	2
Regulator of sigma-E protease RseP	1
Release factor glutamine methyltransferase	3
Replicative DNA helicase	2
Riboflavin biosynthesis protein RibBA	1
Riboflavin biosynthesis protein RibD	2
Riboflavin biosynthesis protein RibF	2
Riboflavin synthase	2
Ribonuclease 3	2
Ribonuclease E	2
Ribonuclease P protein component	2
Ribonuclease R	2
Ribose-5-phosphate isomerase A	2
Ribose-phosphate pyrophosphokinase	2
Ribosomal large subunit pseudouridine synthase C	2

Název proteinu	Počet genů
Ribosomal large subunit pseudouridine synthase D	2
Ribosomal RNA large subunit methyltransferase E	2
Ribosomal RNA large subunit methyltransferase K	1
Ribosomal RNA small subunit methyltransferase A	3
Ribosomal RNA small subunit methyltransferase B	2
Ribosomal RNA small subunit methyltransferase D	1
Ribosomal RNA small subunit methyltransferase H	2
Ribosomal RNA small subunit methyltransferase I	2
Ribosome association toxin RatA	1
Ribosome maturation factor RimP	1
Ribosome-binding factor A	3
Ribosome-recycling factor	3
Ribulose-phosphate 3-epimerase	2
RlpA-like protein precursor	2
RNA polymerase sigma factor RpoD	2
RNA polymerase sigma factor RpoH	2
RNA polymerase-binding transcription factor DksA	1
Rod shape-determining protein MreB	2
rod shape-determining protein MreD	2
Rod shape-determining protein RodA	2
Rubredoxin	1
Segregation and condensation protein A	1
Septum site-determining protein MinC	2
Septum site-determining protein MinD	1
Serine--tRNA ligase	2
Shikimate kinase 1	2
Signal peptidase I	1
Signal recognition particle protein	2
Signal recognition particle receptor FtsY	2
Single-stranded DNA-binding protein	2
Spore maturation protein A	2
SsrA-binding protein	2
Stage V sporulation protein D	2
Succinate dehydrogenase cytochrome b556 subunit	1
Succinate dehydrogenase flavoprotein subunit	1
Succinate dehydrogenase hydrophobic membrane anchor subunit	1
Succinate dehydrogenase iron-sulfur subunit	1
Succinyl-CoA ligase [ADP-forming] subunit alpha	1
Succinyl-CoA ligase [ADP-forming] subunit beta	1
Succinyl-diaminopimelate desuccinylase	2
Superoxide dismutase [Fe]	1
SURF1 family protein	2

Název proteinu	Počet genů
Taurine import ATP-binding protein TauB	1
Taurine-binding periplasmic protein precursor	1
Terephthalate 1	1
Thiamine-monophosphate kinase	3
Thioredoxin-1	2
Thiosulfate sulfurtransferase GlpE	2
Threonine--tRNA ligase	2
Threonylcarbamoyl-AMP synthase	1
Thymidylate kinase	2
Thymidylate synthase	2
Transcriptional regulatory protein YpdB	1
Transketolase 1	2
Translation initiation factor IF-1	3
Translation initiation factor IF-2	2
Translation initiation factor IF-3	2
Trigger factor	2
Triosephosphate isomerase	2
tRNA (guanine-N(1)-)-methyltransferase	2
tRNA (guanine-N(7)-)-methyltransferase	2
tRNA dimethylallyltransferase	2
tRNA modification GTPase MnmE	2
tRNA N6-adenosine threonylcarbamoyltransferase	1
tRNA pseudouridine synthase B	2
tRNA threonylcarbamoyladenine biosynthesis protein TsaB	2
tRNA threonylcarbamoyladenine biosynthesis protein TsaE	3
tRNA threonylcarbamoyladenine dehydratase	1
tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG	2
tRNA(Ile)-lysine synthase	2
tRNA-Arg(acg)	1
tRNA-Arg(cct)	2
tRNA-Arg(tcg)	2
tRNA-Cys(gca)	1
tRNA-Gln(ttg)	1
tRNA-Glu(ttc)	2
tRNA-Gly(gcc)	1
tRNA-Gly(tcc)	3
tRNA-Ile(gat)	2
tRNA-Leu(caa)	1
tRNA-Leu(gag)	2
tRNA-Lys(ttt)	2
tRNA-Met(cat)	4
tRNA-modifying protein YgfZ	2

Název proteinu	Počet genů
tRNA-Phe(gaa)	2
tRNA-Ser(gct)	1
tRNA-Ser(gga)	2
tRNA-specific 2-thiouridylase MnmA	2
tRNA-specific adenosine deaminase	2
tRNA-Thr(cgt)	2
tRNA-Thr(ggt)	3
tRNA-Thr(tgt)	2
tRNA-Trp(cca)	3
tRNA-Tyr(gta)	3
Tryptophan--tRNA ligase	2
Tyrosine recombinase XerD	1
Ubiquinone biosynthesis O-methyltransferase	2
UDP-3-O-acylglucosamine N-acyltransferase	1
UDP-N-acetylenolpyruvoylglucosamine reductase	2
UDP-N-acetylglucosamine 1-carboxyvinyltransferase	2
UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase	2
UDP-N-acetylmuramate--L-alanine ligase	2
UDP-N-acetylmuramoylalanine--D-glutamate ligase	2
UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D-alanine ligase	2
Uracil DNA glycosylase superfamily protein	2
Uridylate kinase	3
Valine--tRNA ligase	2
Zinc-finger domain protein	2

Tabulka II: Tabulka porovnání počtu genů L-symb a N-symb pro skupiny identifikované v databázi KEGG.

	L-symb	N-symb
Metabolic pathways	139	358
Biosynthesis of secondary metabolites	50	157
Microbial metabolism in diverse environments	32	88
Biosynthesis of antibiotics	39	110
Carbon metabolism	29	52
2-Oxocarboxylic acid metabolism	4	22
Fatty acid metabolism	8	13
Biosynthesis of amino acids	21	83
Degradation of aromatic compounds	0	4
Carbohydrate metabolism		
Glycolysis / Gluconeogenesis	6	14
Citrate cycle TCA cycle	14	17
Pentose phosphate pathway	5	8
Pentose and glucuronate interconversions	1	3
Fructose and mannose metabolism	1	4
Galactose metabolism	0	4
Ascorbate and aldarate metabolism	0	2
Starch and sucrose metabolism	0	3
Amino sugar and nucleotide sugar metabolism	2	10
Pyruvate metabolism	8	17
Glyoxylate and dicarboxylate metabolism	4	12
Propanoate metabolism	5	14
Butanoate metabolism	4	11
C5-Branched dibasic acid metabolism	2	7
Inositol phosphate metabolism	2	2
Energy metabolism		
Oxidative phosphorylation	18	30
Photosynthesis	8	8
Carbon fixation in photosynthetic organisms	6	9
Carbon fixation pathways in prokaryotes	12	20
Methane metabolism	1	13
Nitrogen metabolism	2	8
Sulfur metabolism	3	21
Lipid metabolism		
Fatty acid biosynthesis	7	8
Fatty acid degradation	0	10
Synthesis and degradation of ketone bodies	0	1
Glycerolipid metabolism	3	4
Glycerophospholipid metabolism	7	12
Arachidonic acid metabolism	0	1

	L-symb	N-symb
Linoleic acid metabolism	0	1
alpha-Linolenic acid metabolism	0	1
Biosynthesis of unsaturated fatty acids	2	1
Nucleotide metabolism		
Purine metabolism	22	49
Pyrimidine metabolism	18	34
Amino acid metabolism		
Alanine, aspartate and glutamate metabolism	6	18
Glycine, serine and threonine metabolism	7	20
Cysteine and methionine metabolism	2	13
Valine, leucine and isoleucine degradation	1	6
Valine, leucine and isoleucine biosynthesis	1	11
Lysine biosynthesis	7	12
Lysine degradation	2	5
Arginine and proline metabolism	3	25
Histidine metabolism	0	9
Tyrosine metabolism	0	4
Phenylalanine metabolism	0	4
Tryptophan metabolism	1	5
Phenylalanine, tyrosine and tryptophan biosynthesis	2	17
Metabolism of other amino acids		
beta-Alanine metabolism	1	4
Taurine and hypotaurine metabolism	2	5
Phosphonate and phosphinate metabolism	0	1
Selenocompound metabolism	0	6
Cyanoamino acid metabolism	1	4
D-Glutamine and D-glutamate metabolism	3	3
D-Alanine metabolism	1	2
Glutathione metabolism	1	8
Glycan biosynthesis and metabolism		
Lipopolysaccharide biosynthesis	3	17
Peptidoglycan biosynthesis	10	17
Metabolism of cofactors and vitamins		
Thiamine metabolism	2	7
Riboflavin metabolism	5	9
Vitamin B6 metabolism	1	5
Nicotinate and nicotinamide metabolism	2	13
Pantothenate and CoA biosynthesis	3	12
Biotin metabolism	9	10
Lipoic acid metabolism	2	2
Folate biosynthesis	5	11
One carbon pool by folate	6	11
Retinol metabolism	0	1

	L-symb	N-symb
Porphyrin and chlorophyll metabolism	2	11
Ubiquinone and other terpenoid-quinone biosynthesis	2	6
Metabolism of terpenoids and polyketides		
Terpenoid backbone biosynthesis	3	12
Zeatin biosynthesis	1	1
Biosynthesis of ansamycins	1	0
Limonene and pinene degradation	0	1
Geraniol degradation	0	1
Tetracycline biosynthesis	2	3
Polyketide sugar unit biosynthesis	0	3
Biosynthesis of vancomycin group antibiotics	0	1
Genetic Information Processing		
Transcription		
RNA polymerase	4	4
Translation		
Ribosome	53	51
Aminoacyl-tRNA biosynthesis	20	22
RNA transport	1	2
Ribosome biogenesis in eukaryotes	2	2
Folding, sorting and degradation		
Protein export	10	16
Protein processing in endoplasmic reticulum	1	2
Sulfur relay system	2	3
RNA degradation	4	13
Replication and repair		
DNA replication	10	13
Base excision repair	4	8
Nucleotide excision repair	1	7
Mismatch repair	8	15
Homologous recombination	7	21
Environmental Information Processing		
Membrane transport		
ABC transporters	3	47
Phosphotransferase system	0	0
Bacterial secretion system	9	15
Signal transduction		
Two-component system	6	15
MAPK signaling pathway - yeast	0	1
HIF-1 signaling pathway	1	2
FoxO signaling pathway	1	2
Phosphatidylinositol signaling system	2	2
PI3K-Akt signaling pathway	1	1

	L-symb	N-symb
AMPK signaling pathway	1	1
Cellular Processes		
Transport and catabolism		
Peroxisome	2	5
Cell growth and death		
Cell cycle - Caulobacter	8	11
Organismal Systems		
Immune system		
NOD-like receptor signaling pathway	1	1
Antigen processing and presentation	1	1
Endocrine system		
Insulin signaling pathway	0	1
Adipocytokine signaling pathway	0	1
PPAR signaling pathway	1	2
Estrogen signaling pathway	1	1
Progesterone-mediated oocyte maturation	1	1
Thyroid hormone synthesis		1
Nervous system		
Glutamatergic synapse	1	1
GABAergic synapse	1	2
Environmental adaptation		
Plant-pathogen interaction	1	1
Human Diseases		
Cancers: Overview		
Pathways in cancer	2	2
Proteoglycans in cancer	1	1
Chemical carcinogenesis	0	1
Cancers: Specific types		
Renal cell carcinoma	1	1
Prostate cancer	1	1
Neurodegenerative diseases		
Alzheimer's disease	1	1
Amyotrophic lateral sclerosis	0	1
Huntington's disease	1	2
Prion diseases	0	1
Endocrine and metabolic diseases		
Type I diabetes mellitus	0	2
Infectious diseases: Bacterial		
Epithelial cell signaling in Helicobacter pylori infection	0	2
Salmonella infection	0	1
Legionellosis	2	3
Tuberculosis	0	3

	L-symb	N-symb
Drug resistance		
beta-Lactam resistance	1	12
Vancomycin resistance	4	5
Biosynthesis of other secondary metabolites		
Carbapenem biosynthesis	0	2
Streptomycin biosynthesis	1	5
Novobiocin biosynthesis	0	1
Xenobiotics biodegradation and metabolism		
Benzoate degradation	0	3
Aminobenzoate degradation	0	2
Chloroalkane and chloroalkene degradation	0	2
Nitrotoluene degradation	0	2
Ethylbenzene degradation	0	1
Styrene degradation	0	1
Atrazine degradation	0	4
Naphthalene degradation	0	1
Metabolism of xenobiotics by cytochrome P450	0	1
Drug metabolism - cytochrome P450	0	1
Drug metabolism - other enzymes	3	4