

Jihočeská univerzita v Českých Budějovicích

Přírodovědecká fakulta

**Evoluce symbiotických bakterií krevsajícího dvoukřídleho hmyzu**

Diplomová práce

**Bc. Eliška Juhaňáková**

Školitel: Prof. RNDr. Václav Hypša, CSc.

České Budějovice 2020

Juhaňáková, E., 2020: Evoluce symbiotických bakterií dvoukřídleho krevsajícího hmyzu. [Evolution of symbiotic bacteria in blood-sucking dipteran insects. Mgr. Thesis, in Czech.] – 69 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

This study is focused on bacterial symbionts of blood-sucking insects from the family Hippoboscidae (Diptera), namely *Ornithomya biloba* and *Crataerina pallida*. Bacterial symbionts are essential for blood-sucking insects because they provide nutrients absent in blood, particularly vitamin B and cofactors. In this study metagenomic data from *Ornithomya biloba* and *Crataerina pallida* were processed by bioinformatic tools and genome drafts were assembled. Genomes of two bacterial symbionts of the genus *Arsenophonus* were compared with their relatives and levels of genomic modification of these genomes were evaluated. Moreover, the presence of horizontally transferred genes in both *Arsenophonus* genomes was phylogenetically determined. Finally, metabolic pathways for essential B vitamins and cofactors were reconstructed and compared with other species of the genus *Arsenophonus*.

Prohlašuji, že svoji diplomovou 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 diplomové 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, 21. května 2020

Bc. Eliška Juhaňáková

## **Poděkování**

Ráda bych poděkovala svému školiteli Václavu Hypšovi za cenné rady, věcné připomínky a vstřícnost při konzultacích a vypracování diplomové práce. Dále bych chtěla poděkovat Janě Říhové za cenné a užitečné rady při provádění bioinformatických analýz. A v neposlední řadě bych chtěla poděkovat ostatním členům laboratoře za užitečné rady a podporu při zpracování mé diplomové práce.

|         |  |        |
|---------|--|--------|
| 1       | Úvod .....   | - 1 -  |
| 1.1     | Symbiotické bakterie hmyzu.....  | - 1 -  |
| 1.2     | Krevsající hmyz .....  | - 3 -  |
| 1.3     | Nadčeleď Hippoboscoidea.....   | - 4 -  |
| 1.4     | <i>Arsenophonus</i> spp. ....  | - 6 -  |
| 1.5     | Změny v genomech symbiotických bakterií .....  | - 6 -  |
| 1.6     | Horizontální přenos genetické informace (HGT) .....                                  | - 7 -  |
| 1.7     | B vitamíny a kofaktory .....   | - 8 -  |
| 2       | Cíle .....   | - 11 - |
| 3       | Materiál a metody.....   | - 12 - |
| 3.1     | Vstupní metagenomická data.....  | - 12 - |
| 3.2     | Assembly genomu.....   | - 12 - |
| 3.2.1   | Bakterie z kloše <i>Ornithomya biloba</i> .....                                      | - 12 - |
| 3.2.1.1 | Primární assembly v A5 pipeline.....   | - 13 - |
| 3.2.1.2 | Primární assembly v MetaSPAdes .....   | - 15 - |
| 3.2.2   | Bakterie z kloše <i>Crataerina pallida</i> .....                                     | - 16 - |
| 3.3     | Fylogeneze horizontálně přenesených genů .....                                       | - 17 - |
| 3.4     | Rekonstrukce metabolických drah.....   | - 18 - |
| 4       | Výsledky.....  | - 19 - |
| 4.1     | Genomické assembly .....   | - 19 - |
| 4.1.1   | Metagenomická data z kloše <i>Ornithomya biloba</i> .....                            | - 19 - |
| 4.1.2   | Metagenomická data z kloše <i>Crataerina pallida</i> .....                           | - 19 - |
| 4.2     | Anotace .....  | - 21 - |
| 4.3     | Fylogeneze horizontálně přenesených genů (HGT) .....                                 | - 22 - |
| 4.3.1   | HGT geny z bakterie rodu <i>Arsenophonus</i> z kloše <i>Ornithomya biloba</i> .....  | - 24 - |
| 4.3.2   | HGT geny z bakterie rodu <i>Arsenophonus</i> z kloše <i>Crataerina pallida</i> ..... | - 27 - |
| 4.4     | Porovnání genomů .....   | - 28 - |

|         |   |        |
|---------|---|--------|
| 4.5     | Metabolické dráhy .....   | - 29 - |
| 5       | Diskuze .....   | - 32 - |
| 5.1     | Výsledky assembly .....   | - 32 - |
| 5.2     | Redukce a degradace genomů.....   | - 32 - |
| 5.3     | HGT a mobilní elementy .....  | - 34 - |
| 5.3.1   | Mobilní elementy.....   | - 34 - |
| 5.3.2   | Horizontálně přenesené geny.....  | - 34 - |
| 5.3.2.1 | HGT geny z bakterie rodu <i>Arsenophonus</i> z kloše <i>Ornithomya biloba</i> ..  | - 36 - |
| 5.3.2.2 | HGT geny z bakterie rodu <i>Arsenophonus</i> z kloše <i>Crataerina pallida</i> .. | - 37 - |
| 5.4     | Metabolické dráhy B vitamínů a kofaktorů .....                                    | - 37 - |
| 6       | Závěr.....  | - 41 - |
| 7       | Použitá literatura.....   | - 42 - |
| 8       | Přílohy .....   | - 52 - |

# 1 Úvod

## 1.1 Symbiotické bakterie hmyzu

Hmyz vytváří široké spektrum symbiotických vztahů s bakteriemi od parazitismu až po obligátní mutualismus (Kikuchi, 2009). V nejširším slova smyslu všechny bakteriální komunity v těle hmyzu tvoří mikrobiom, jehož podstatná část se nachází v lumenu střeva. Mikrobiom hraje klíčovou roli při růstu, vývoji a adaptaci hmyzího hostitele na určité podmínky prostředí (Krishnan et al., 2014). Některé bakterie vytvářejí se svým hostitelem specializované typy symbiotických vztahů. Příkladem specializovaného typu symbiotického vztahu mezi hmyzem a bakterií může být vnitrobuněčná endosymbioza, při které se bakterie nachází uvnitř buněk svého hmyzího hostitele. V některých případech mohou tyto buňky v hostiteli vytvářet samostatné specializované orgány zvané bakteriomy. Symbiotické bakterie hmyzu jsou obecně rozdělovány na obligátní a fakultativní symbionty (Moran et al., 2008). V jednom hmyzím hostiteli je často možné pozorovat soužití obligátních a fakultativních symbiontů (Nováková et al., 2015).

Obligátní neboli primární symbionti vytvářejí mutualistické vztahy, které mohou mít dlouhou koevoluční historii v řádu až stovek milionů let (Moran et al., 1993). Obligátní symbionti jsou schopni vytvářet a dodávat svému hostiteli různé metabolické látky. Proto se obligátní symbionti nacházejí u hmyzu, který je závislý na nutričně omezeném zdroji potravy. Z tohoto důvodu jsou obligátní symbionti pro svého hostitele nepostradatelní. V genomu bakterie dochází v průběhu koevoluce s hostitelem k zásadním a typickým změnám, především k degradaci genomu a ztrátě řady genů, v jejichž důsledku dochází k neschopnosti obligátního symbionta žít samostatně bez podpory hostitelské buňky (Weiss and Aksoy, 2011). Z toho důvodu vytvářejí hostitelé pro své obligátní symbionty již výše zmíněné bakteriomy, které jsou tvořeny specializovanými buňkami. Tyto buňky se nazývají bakteriocyty a mohou to být původem například buňky tukového tělíska nebo buňky střevní stěny hmyzu (Moran et al., 2008). Obligátní symbionti jsou přenášeni na dalšího hostitele vertikálně, tedy ze samiček na potomstvo, obvykle transovariální cestou. V důsledku vertikálního přenosu symbiontů vzniká typická kongruentní fylogeneze s hostitelem. Typickým příkladem primárního symbionta může být bakterie *Buchnera aphidicola* u mšic (nadčeleď Aphidoidea, Breandle et al., 2003) nebo bakterie rodu *Wigglesworthia* žijící v bakteriocytech mouchy tsetse (čeleď Glossinidae; Aksoy, 1995). Tyto symbiotické bakterie

vzhledem k degradaci genomu nemohou samostatně infikovat nového hostitele a jsou v tomto ohledu zcela závislé na hostitelských mechanismech přenosu (Breandle et al., 2003).

Fakultativní neboli sekundární symbionti vytvářejí se svými hostiteli různé symbiotické vztahy, které nemají obvykle tak dlouhou koevoluční historii jako obligátní symbionti. Fakultativní symbionti mohou být získáváni z prostředí a přenášeni vertikálně nebo horizontálně, tedy přenosem mezi jedinci jedné generace nebo jedinci odlišného druhu (Weiss and Aksoy, 2011). Fakultativní symbionti nejsou pro svého hostitele esenciální, avšak často ovlivňují život hostitele výrazným způsobem. Fakultativní symbionti mohou být velmi prospěšní pro svého hostitele a pak jsou označováni jako fakultativní mutualisté nebo komenzálové. Část fakultativních symbiontů negativně ovlivňuje fitness svého hostitele, což je považováno za jeden z rysů parazitismu. Typickým příkladem takových fakultativních symbiontů jsou reprodukční manipulátoři. Z důvodu maternálního přenosu těchto bakterií probíhá většinou manipulace s reprodukcí hostitele ve prospěch samičích zárodků (Moran et al., 2008). Mezi mechanismy reprodukční manipulace patří cytoplazmatická inkompatibilita, řízené zabíjení synů (son-killer trait), feminizace synů nebo indukované partenogeneze. Hojně studovaná bakterie *Wolbachia pipientis*, symbiotická bakterie hmyzu, korýšů, roztočů a fíliárních nematodů manipuluje s reprodukcí svého hostitele prostřednictvím cytoplazmatické inkompatibility (Stouthamer et al., 1999). Cytoplazmatická inkompatibilita postupně vytlačuje z populace jedince, kteří nejsou infikováni bakterií rodu *Wolbachia*. Toho je docíleno úmrtím embryí, která vznikla z neinfikované samice (Sinkins et al., 1995). Další reprodukční manipulátor, bakterie *Arsenophonus nasoniae*, ovlivňuje reprodukci svého hostitele, parazitické vosičky *Nasonia vitripennis*, manipulací označovanou jako son-killer trait, který zabíjí haploidní, tedy samčí embrya (Darby et al., 2010; Gherna et al., 1991). Intracelulární bakterie rodu *Cardinium* manipuluje reprodukcí hostitele partenogenezí, feminizací synů a již zmíněnou cytoplazmatickou inkompatibilitou (Weeks et al., 2001).

V některých bakteriálních skupinách lze pozorovat jak obligátní symbionty, tak i fakultativní symbionty (Moran et al., 2008). Příkladem může být rod *Sodalis*, který je pro mouchu tsetse z čeledi Glossinidae fakultativním symbiontem (*S. glossinidius*, Geiger et al. 2007) a pro veš *Proechinophthirus fluctus* (řád Phthiraptera: Anoplura) nebo kloše *Crataerina melbae* (čeleď Hippoboscidae) obligátním symbiontem (Boyd et al., 2014; Nováková and Hypša, 2007). Také u bakterie rodu *Wolbachia* se kromě parazitismu může objevovat i mutualismus. Například pro filariální hlístice je bakterie rodu *Wolbachia* obligátním symbiontem (Taylor et al., 2005). Bakterie rodu *Wolbachia* může být i příkladem pro genomové adaptace provázející změnu životní strategie symbiotické bakterie z parazita na

mutualistu. Tento jev byl konkrétně pozorován u bakterie rodu *Wolbachia* ve štěnici *Cimex lectularius*. K této změně došlo horizontálním přenosem biotinového operonu do genomu bakterie rodu *Wolbachia* (Nikoh et al., 2014; Gerth and Bleidorn, 2017).

Symbiotické bakterie poskytují svým hostitelům velké množství výhod, mezi které patří například ochrana před stresem nebo zachování homeostázy. Další zásadní výhodou je kompenzace nutričních nedostatků potravních zdrojů fytofágního a hematofágního hmyzu (Dale and Moran, 2006). Například již zmíněná bakterie rodu *Wigglesworthia* produkuje esenciální B vitamíny a kofaktory stěžejní pro život svého hostitele, mouchu tsetse (čeleď Glossinidae), a podílí se na správném vývoji imunitního systému svého hostitele (Rio et al. 2012). Bakterie rodu *Baumannia* a *Sulcia* jsou obligátními symbionty, kteří pro svého hostitele z podčeledi křískovitých (Cicadellinae) produkuje esenciální aminokyseliny a kofaktory (Wu et al., 2006). Bakterie *Buchnera aphidicola* zásobuje fytofágní mšice z řádu Hemiptera také esenciálními aminokyselinami a vitamíny (Gündüz and Douglas, 2009). Další výhodou poskytovanou symbiotickými bakteriemi je ochrana před patogeny a parazity. Bakterie rodu *Wolbachia* chrání svého hostitele mouchu *Drosophila* (čeleď *Drosophilidae*) proti RNA virům (Teixeira et al., 2008). Fakultativní bakterie *Reigella insecticola* zvyšuje rezistenci mšice proti některým patogenním houbám (Parker et al., 2013). Fakultativní symbionti *Serratia symbiotica* a *Halmintonella defensa* se podílejí na ochraně svého hostitele, mšiči *Acyrtosiphon pisum*, před parazitoidy *Aphidius cervi* a *A. eadyi* (Ferrari et al., 2004). Některé symbiotické bakterie chrání svého hostitele před predátory produkcí jedovatých látek. Tato obranná strategie byla pozorována u symbiotických bakterií brouků rodu *Paederidus*, které produkuje jedovatý pederin a chrání tak svého hostitele před pavouky (Piel et al., 2004).

## 1.2 Krevsající hmyz

Jedním z potravních zdrojů hmyzu je krev obratlovců. Mezi skupiny krevsajícího hmyzu patří blechy (Siphonaptera), některé ploštice (Heteroptera), vši (Phthiraptera: Anoplura) a některé skupiny dvoukřídlého hmyzu (Diptera) jako komárovití (Culicidae) nebo klošovití (Hippoboscoidea; Lehane, 2005). Mezi zástupci krevsajícího hmyzu najdeme vektory významných lidských parazitů a patogenů. Mezi tyto vektory patří například komáři rodu *Anopheles* přenášející *Plasmodium* způsobující malárii nebo vektory trypanosomy, mouchu tsetse a ploštici z čeledi Reduviidae (Amino et al., 2006; Maia Da Silva et al., 2007; Van Den Abbeele et al., 1999). Vektorová kapacita hmyzu neboli schopnost přenášet patogeny může být inhibována některými bakteriálními komunitami, které jsou součástí mikrobiomu hmyzu. Příkladem může být bakterie rodu *Wolbachia*, která inhibuje přenos parazita *Plasmodium*



komárem *Aedes aegypti* (Sinkins et al., 2013). Z výsledků studií také vyplývá, že vektorová kapacita hmyzu závisí i na rovnováze zastoupení bakteriálních druhů, jelikož přítomnost některých bakterií může inhibovat výskyt jiných bakterií. Tento případ byl pozorován u bakterií rodu *Wolbachia* a *Asaia* (Hughes et al., 2014).

Některé skupiny krevsajícího hmyzu se živí jak krví, tak i další potravou. Zdroj výživy se může lišit mezi samci a samicemi, ale i mezi jednotlivými vývojovými stádii. Příkladem mohou být zástupci komárovitých (Culicidae). Larvy a kukly se živí organickým materiálem z vody. Samci komárovitých sají rostlinný nektar. Samice se mohou živit stejně jako samec rostlinným nektarem, avšak z důvodu správného vývoje vajíček je nezbytné, aby samice sála krev. Krev totiž obsahuje větší množství potřebných bílkovin (Foster, 1995). Jiné skupiny krevsajícího hmyzu se živí výhradně krví. Mezi tyto skupiny patří například zástupci klošovitých (čeleď Hippoboscidae). Samci a samice zástupců klošovitých se živí výhradně krví svých hostitelů. V případě larev dochází k příjmu potravy od matky, ve které probíhá jejich vývoj (viz níže; Baker, 1967; Krafsur, 2009). Krev jako potravní zdroj neobsahuje dostatečné množství všech vitamínů a kofaktorů, a z toho důvodu najdeme ve střevech hmyzu, který se živí výhradně krví, obligátní symbionty, kteří tyto nedostatky kompenzují (Dale and Moran, 2006; Nováková et al., 2015; Pais et al., 2008; Rio et al. 2012).

### 1.3 Nadčeleď Hippoboscoidea

Nadčeleď Hippoboscoidea, z řádu dvoukřídlého hmyzu (Diptera), sestává ze čtyř čeledí, jmenovitě Hippoboscidae (anglicky louse flies nebo deer keds), Glossinidae (anglicky tsetse flies), Streblidae a Nycteribiidae (anglicky bat flies). Fylogeneze nadčeledi Hippoboscoidea není úplně jasná. Čeleď Glossinidae, Hippoboscidae a Nycteribiidae jsou označovány za monofyletické skupiny. U čeledi Streblidae nebyla monofyletičnost potvrzena (Šochová et al., 2017). Zástupci nadčeledi Hippoboscoidea jsou výhradně krevsající ektoparazité s velkým množstvím unikátních a výrazných fyziologických a morfologických znaků (Petersen et al., 2007). Jedním ze znaků charakterizující tuto skupinu je unikátní způsob reprodukce v podobě adenotrofické viviparie. Larva se vyvíjí v reprodukčním systému samice, kde se živí sekretem produkovaným přídatnými pohlavními žlázami. Larva zůstává v reprodukčním systému samice do konce vývoje (Krafsur, 2009). V důsledku tohoto způsobu vývoje larvy dochází k získávání potravy v průběhu celého životního cyklu jedince výhradně z krve. Tudíž zástupci nadčeledi Hippoboscoidea patří mezi zástupce krevsajícího hmyzu, který má typické obligátní vnitrobuněčné symbionty, kteří musejí poskytovat svému hostiteli některé metabolické látky chybějící v krvi.

Mouchy tsetse, čeleď Glossinidae neboli bodalkovití, jsou významnými vektory závažného lidského onemocnění spavé nemoci způsobeného prvokem z rodu *Trypanosoma* (Kennedy, 2013). Výzkumy mikrobiomu mouchy tsetse odhalily přítomnost některých symbiotických bakterií ve střevě těchto much, které hrály v organismu svého hostitele různé role. Nutriční nedostatky krve kompenzuje mouše tsetse bakterie rodu *Wigglesworthia* jako obligátní mutualista vázaný v bakteriomech. Bakterie rodu *Sodalis* je pro mouchu tsetse fakultativním symbiontem přítomným ve všech tkáních svého hostitele, která pravděpodobně ovlivňuje účinnost přenosu parazita *Trypanosoma* z mouchy tsetse na člověka. Obě tyto bakterie jsou přenášeny vertikálně přes mléčné žlázy (Dale and Maudlin, 1999; Dale and Welburn, 2001; Greiger et al., 2007). Další bakterií pozorovanou v mikrobiomu mouchy tsetse je bakterie rodu *Wolbachia*, jejíž funkce v mikrobiomu nebyla odhalena (Shaw and Mooloo, 1991).

Čeleď Hippoboscidae neboli klošovití je kosmopolitní skupinou hmyzu s výrazně odlišnými morfologickými znaky a životními strategiemi (Petersen et al., 2007). Zástupci čeledi Hippoboscidae jsou krevsajícími ektoparazity savců a ptáků. Přičemž krev sají jak samci, tak i samice (Small, 2005). Čeleď Hippoboscidae je tvořena třemi podčeledmi, jmenovitě Hippoboscinae, Lipopteninae a Ornithomyinae. Podčeledi Hippoboscinae a Lipopteninae jsou monofyletickými skupinami. Podčeleď Ornithomyinae je parafiletická skupina, do které patří dvě monofyletické skupiny, konkrétně Olfersini a Ornithomyini (Petersen et al., 2007). Pro zástupce čeledi Hippoboscidae je typické dorsoventrální zploštění těla, prognátní hlava a krabovitě uspořádané nohy. Většina klošů má velmi dobře vyvinutá a funkční křídla, jako například zástupci rodu *Ornithomya*. Avšak některé druhy, jako například *Lipoptena cervi*, ztrácí křídla poté, co jedinec najde svého hostitele. Někteří zástupci čeledi Hippoboscidae jsou dokonce apterní jako například *Melophagus ovinus* (Askew, 1971; Baker, 1967; Mustonen et al., 2015; Petersen et al., 2007). Zástupci čeledi Hippoboscidae vytvářejí symbiotické vztahy s různými druhy bakterií. Studie zkoumající mikrobiom kloše *Melophagus ovinus* odhalila přítomnost bakterií rodu *Wolbachia*, *Bartonella*, fakultativního symbionta, bakterii rodu *Sodalis*, a obligátního symbionta, bakterii *Arsenophonus* (Nováková et al., 2015). Bakterie *Sodalis* byla přítomna i v mikrobiomu dalšího zástupce čeledi Hippoboscidae, konkrétně kloše *Craterina melbae* (Nováková and Hypša, 2007). V mikrobiomu dalšího zástupce z čeledi klošovitých, kloše *Lipoptena cervi*, byla prokázána přítomnost fakultativní bakterie rodu *Bartonella*, bakterie rodu *Acinetobacter* a obligátního endosymbionta *Canidatus Arsenophonus* lipopteni (Dehio et al., 2004; Nováková et al., 2016; Regier et al., 2018).

#### 1.4 *Arsenophonus* spp.

*Arsenophonus*, Gammaproteobacteria z čeledi Enterobacteriaceae, je bakterie vytvářející velmi široké spektrum vztahů s rozmanitými hostiteli. Nejčastějšími hostiteli bakterie rodu *Arsenophonus* jsou zástupci hmyzu, jako například včely, plošnice z podčeledi Triatominae, parazitické vosičky, kloši z čeledi Hippoboscidae a Streblidae nebo vši, ale najdeme mezi nimi i zástupce vytvářející vztahy s rostlinami (Babendreier et al., 2007; Dale et al., 2006; Gherna et al., 1991; Hypša and Dale, 1997; Trowbridge et al., 2006; Zreik et al., 1998). Někteří zástupci bakterie rodu *Arsenophonus* jsou pro svého hostitele obligátními symbionty jako například *Arsenophonus lipopteni* pro krevsajícího kloše *Lipoptena cervi* nebo bakterie rodu *Riesia* (*Arsenophonus*) obývající bakteriomy lidí, šimpanzů a goril (Boyd et al., 2014; Nováková et al., 2016). Jiní jsou pro svého hostitele fakultativními symbionty jako například *Arsenophonus triatominarum* pro krevsající plošnici *Triatoma infestans* (Hypša and Dale, 1997). A někteří zástupci jsou pro svého hostitele parazity jako například *Arsenophonus nasoniae* pro parazitickou vosičku *Nilaparvata lugens* (viz výše). U volně žijící bakterie rodu *Proteus* byla prokázána blízká příbuznost k bakterii *Arsenophonus nasoniae* (Gherna et al., 1991).

#### 1.5 Změny v genomech symbiotických bakterií

Bakteriální genomy lze porovnávat na základě velikosti genomu a obsahu GC bazí v genetické informaci bakterií. Velikosti genomů mezi jednotlivými druhy volně žijících bakterií se značně liší. Rozmezí doposud osekvenovaných genomů volně žijících bakterií se pohybuje od 0,6 Mbp až po 10 Mbp. Nejmenší genomy volně žijících bakterií (ve smyslu ne-symbiotických bakterií) o velikosti okolo 530 Kbp byly osekvenovány u bakterie rodu *Mycoplasma* a jejich příbuzných druhů (McCutcheon and Moran, 2012). Naopak největší osekvenovaný genom o velikosti 9,7 Mbp byl osekvenován u bakterie *Burkholderia xenovorans* (Binnewies et al., 2006). Obsah GC určuje procentuální zastoupení nukleotidových bazí guaninu a cytozinu v genetické informaci. Mutace genetické informace bakterií probíhá ve prospěch nukleotidových bazí adeninu a thyminu (A-T), čímž klesá procentuální zastoupení nukleotidových bazí guaninu a cytozinu (G-C), jejichž vazba je energeticky náročnější. Z toho plyne snižování energetických nákladů bakterií při degradaci jejich genomu. Obsah GC udává míru degradace bakteriálního genomu. Obecně dosahují obligátní symbionti nižších hodnot obsahu GC než volně žijící bakterie (McCutcheon and Moran, 2012). Hodnota obsahu GC je mezi jednotlivými druhy bakterií značně variabilní. Průměrná hodnota obsahu GC u volně žijících bakterií je 49% (Bentley and Parkhill, 2004).

Nejvyšší hodnota obsahu GC byla zjištěna u volně žijící bakterie *Anaeromyxobacter dehalogens*, jejíž genom obsahoval 75% GC vazeb (Hildebrand et al., 2010).

Genomy symbiotických bakterií se odlišují od volně žijících bakterií a vyznačují se některými unikátními charakteristikami, ke kterým dochází v průběhu soužití bakterie a jejího hostitele. Mezi tyto charakteristiky patří výrazná degradace genomu, jak nukleotidovým složením, tak i velikostí, vysoká míra exprese chaperonů podílejících se na skládání bílkovin, pravděpodobně z důvodu odolnosti bílkovin k nepříznivým podmínkám prostředí, anebo rychlá evoluce proteinů, způsobená častými mutacemi genetické informace symbiotických bakterií. Genomy obligátních symbiontů mohou redukovat svou velikost až pod 1 Mbp. Mezi nejmenší bakteriální genomy symbiotických bakterií dosahujících velikosti menší než 300 Kbp patří například *Candidatus Carsonella ruddii*, *Candidatus Hodgkinia cicadicola* nebo *Candidatus Tremblaya princeps*. Průměrná hodnota obsahu GC obligátních symbiontů je 38% (Bentley and Parkhill, 2004). Nejnižší hodnota obsahu GC, 16,5%, byla zjištěna u endosymbiotické bakterie *Carsonella ruddii* (Hildebrand et al., 2010).

Všechny výše zmíněné charakteristiky poskytují informace o délce trvání a míře evolučního vývoje vztahu mezi bakterií a hostitelem (Nováková et al., 2015). Při zmenšování genomu symbiotických bakterií dochází ke ztrátě genů, které nejsou nutné pro život bakterie v hostiteli. Mezi tyto geny patří i geny pro ochranné mechanismy, virulenci, regulaci genové exprese nebo geny účastnící se rekombinací a oprav DNA (Gosalbes et al., 2010; McCutcheon and Moran, 2012). Symbiotické bakterie se postupnou degradací genomu stávají závislé na svém hostiteli. Hostitel poskytuje symbiotické bakterii bezpečné prostředí a esenciální složky jejich potravy, které si bakterie nedokáže produkovat sama. Tyto esenciální složky jsou většinou meziprodukty metabolismu hostitele (Moran, 2002).

Degradací genomu bakterie postupně ztrácí i geny kompenzující metabolické nedostatky hostitele. V takovém případě může dojít k zisku jiného symbionta, který doplní nebo nahradí biosyntetické schopnosti původního symbionta. Často však dochází k úplnému zániku původního symbionta a jeho nahrazení jinou symbiotickou bakterií (McCutcheon and Moran, 2010). Degradace genomu bakterie může být také kompenzována vertikálním nebo horizontálním přenosem genetické informace (Boto, 2010).

## **1.6 Horizontální přenos genetické informace (HGT)**

Horizontální přenos genetické informace (dále jen HGT) je významným evolučním faktorem u bakterií. Horizontální přenos genetické informace probíhá i mezi jedinci jedné generace a je realizován různými molekulární mechanismy (Boto, 2010). Mezi tyto

mechanismy patří například transformace, při které dochází k absorpci cizorodé genetické informace z okolního prostředí bakterií. Dalším mechanismem je konjunkce, při které dochází k přenosu mobilních genetických elementů mezi dvěma dočasně spojenými bakteriemi. Transdukce je mechanismus horizontálního přenosu genů virových částic, ke kterému dochází nejčastěji při infekci bakteriofágem (Thomas and Nielsen, 2005). Horizontální přenos genetické informace je adaptivní a inovativní proces vylepšující genetickou informaci bakterie. Dopad HGT na evoluci bakterií je závislý na několika faktorech. Záleží na množství přenesených genů nebo množství genetické informace následně úspěšně začleněných do bakteriálního genomu. Dále záleží na rozsahu horizontálního přenosu z časového hlediska, tedy zda se jedná o starší či nedávný HGT, a na fylogenetické vzdálenosti mezi organismy (Boto, 2010). Zajímavým faktem zůstává, že většina HGT není bakterií přijata a je v průběhu jejího života ztracena. V bakteriální populaci zůstávají zachovány pouze horizontálně přenesené úseky DNA a geny, které jsou pro bakterii prospěšné a výhodné (Berg and Kurland, 2002). V genomu bakterie často dochází ke kumulaci genů pro stejné metabolické procesy do jednoho místa, označovaného jako operon. Mezi bakteriemi může docházet k horizontálnímu přenosu jednotlivých genů, ale také celých operonů. Přenos genů v operonu zvyšuje pravděpodobnost přenosu kompletní a funkční dráhy pro různé mechanismy (Lawrence and Roth, 1996). Horizontální přenos genů se týká i kompletních operonů pro důležité B vitamíny. Přenos biotinového operonu byl například pozorován u obligátní symbiotické bakterie *Legionella polyplacis* (Říhová et al., 2017).

Horizontální přenos genů, který obohacuje genom bakterií o geny, které dělají bakterii pro hostitele užitečnější, částečně kompenzuje redukci a degradaci genomu symbiotické bakterie. Možnost odhalení HGT genů nebo sekvence se zvyšuje s množstvím osekvenovaných a assemblovaných genomů různých druhů bakterií. Dříve bylo možné odhalit a odhadnout množství HGT genů jen u velmi malého množství bakterií. Jednou z nich byla například bakterie *Escherichia coli* (Ochman et al., 2000). Vzhledem ke stále se zvyšujícímu počtu osekvenovaných bakteriálních genomů je možné odhalit a prověřit přítomnost HGT genů porovnáním ortologních sekvencí genů.

## **1.7 B vitamíny a kofaktory**

Symbiotické bakterie kompenzují metabolické nedostatky krevsajícího hmyzu syntézou chybějících metabolických látek, z nichž nejtypičtějšimi jsou B vitamíny a kofaktory (Moran et al., 2008). Vitamíny jsou definovány jako komplex organických složek, které jsou v nepatrném množství přítomné v potravě a jsou esenciální pro normální fungování

metabolismu. Nedostatek vitamínů v potravě způsobuje různá onemocnění. Pojem vitamín byl použit poprvé v roce 1911 polským biochemikem Casimirem Funkem a od té doby bylo popsáno a identifikováno 15 vitamínů (McDowell, 2000).

Jedním z nejdůležitějších B vitamínů je biotin, označován také jako vitamín B7. Aktivní forma biotinu je prostetickou skupinou enzymu, která přenáší oxid uhličitý mezi různými karboxylačními, dekarboxylačními a transkarboxylačními reakcemi. Biotin se jako enzymatická složka podílí na syntéze mastných kyselin, metabolismu mastných kyselin a glukoneogenezi (Jitrapakdee and Wallace, 2003; Knowles, 1989; Lynen, 1967). Biotin produkují pouze bakterie, houby a rostliny (Lin and Cronan, 2011).

Thiamin, označovaný také jako vitamín B1, je dalším životně důležitým B vitamínem především pro přežití buněk. Defosforylovaná forma thiaminu, thiamin pyrofosfát, slouží jako kofaktor pro enzymy zapojené do buněčných procesů zahrnujících jejich energetický metabolismus a biosyntézu prekurzorů pro nukleotidy, mastné kyseliny a další buněčné složky (Sannino et al., 2018). Z výsledků nedávných studií vyplývá, že hmyz získává thiamin z potravy. Pomocí ABC transportéru hostitel poskytuje thiamin symbiotickým bakteriím. Produkce thiaminu tedy s největší pravděpodobností není prováděna symbiotickými bakteriemi, jak je tomu u ostatních B vitamínů a kofaktorů (Husník, 2018).

Kyselina listová, jinak také vitamín B9 (anglicky folic acid), zahrnuje i sloučeninu folátu. Tento vitamín je jako kofaktor zapojený do syntézy purinu a pyrimidinu, látek stěžejních pro výrobu nukleových kyselin, a také do přeměny mastných kyselin obsahujících sulfur. Další funkcí tohoto vitamínu je přenos jednouhlíkatého radikálu mezi molekulami folátu nebo mezi folátem a dalšími sloučeninami (Blatch et al., 2010). Význam vitamínu B9 potvrzuje i fakt, že nedostatek nebo poruchy v jeho metabolismu mohou způsobit u zvířat rakovinu nebo poškození nervové trubice (Bender, 2003). Hmyz není schopen tento vitamín sám produkovat. Jak již bylo zmíněno, v případě krevsajícího hmyzu je nedostatek vitamínu B9 kompenzován symbiotickými bakteriemi (Nováková et al., 2015).

Kyselina pantotenová, označovaná také jako vitamín B5 (anglicky pantothenic acid), je prekurzorem pro biosyntézu koenzymu A (CoA) a proteinů obsahujících fosfopanteteinovou prostetickou skupinu, která je zapojena do syntézy mastných kyselin, neribosomálních peptidů (polyketidů). Koenzym A je důležitým kofaktorem podílejícím se na růstu buňky. Dále je zapojen do různých metabolických reakcí zahrnujících syntézu fosfolipidů, syntézu a degradaci mastných kyselin nebo reakce citrátového cyklu (Miller and Schlesinger, 1993).

Riboflavin, známý také jako vitamín B2, je důležitou složkou základního metabolismu. Tento vitamín je prekurzorem koenzymů flavinadenindinukleotid (FAD) a flavinmononukleotid (FMN), které se účastní dýchacího řetězce (Babor et al., 2002). Vitamín B2 je produkován mikroorganismy, rostlinami a houbami (Vitreschak et al., 2002).

Pyridoxin, označovaný také jako vitamín B6, funguje jako kofaktor pro mnoho enzymů. Aktivní forma vitamínu B6, pyridoxal 5'-fosfát (PLP), je univerzálním kofaktorem enzymů zapojených do metabolismu složek aminokyselin. Vitamín B6 syntetizují jednobuněčné organismy a rostliny (Tanaka et al., 2005).

Kyselina lipoová (anglicky lipoic acid) je koenzym obsahující síru, jehož aktivní forma se váže na proteiny prostřednictvím amidové vazby na  $\epsilon$ -aminoskupinu specifických lyzinových reziduí. Kyselina lipoová je jako kofaktor zapojena do oxidativně dekarboxylačních reakcí alfa-keto kyselin a štěpení glycinu (Reed and Cronan, 1993).

Koenzym Q10, známý také jako ubiquinone, je přírodní složka s vlastnostmi podobnými vitamínům. Koenzym Q10 hraje zásadní roli v buněčné bioenergetice jako kofaktor zapojený do dýchacího řetězce a je tedy esenciální pro produkci ATP. Redukovaná forma koenzymu Q10, známá též jako ubiquinol, je silným lipofilním antioxidantem, který recykluje a regeneruje antioxidanty. Koenzym Q10 se dále podílí na signalizaci buňky a na expresi jejích genů (Bhagavan and Chopra, 2006).

Protoheme IX, známý též jako heme b nebo fe-protoporfyrin IX, hraje zásadní roli v základních buněčných procesech, především v transportu a skladování plynných molekul, úspoře energie nebo katalytických procesech (Baureder and Hederstedt, 2013).

Nikotinamid může být konvertován transaminací z kyseliny nikotinové, označované též jako vitamín B3, která je hlavním prekurzorem nikotinamidu adenin dinukleotidu (NAD<sup>+</sup>; Ma et al., 2016).

## 2 Cíle

Cílem této práce je rozšířit znalost symbiotických bakterií dvoukřídlého hmyzu (Diptera) na příkladu genomové evoluce obligátních symbiontů dvou druhů klošů z čeledi Hippoboscidae. Konkrétně u kloše *Ornithomya biloba* a *Crataerina pallida* byla již dříve na fylogenetické úrovni potvrzena přítomnost bakterie rodu *Arsenophonus*. V této práci byly rekonstruovány genomy bakterií rodu *Arsenophonus*, jejich metabolických kapacit a podíl horizontálního přenosu genů.

Proto byly pro tuto práci stanoveny následující cíle:

1. Zpracovat metagenomická data z *Ornithomya biloba* a *Crataerina pallida* a složení genomu symbiotické bakterie *Arsenophonus* spp.
2. Odvodit stupeň genomické modifikace a porovnat genomy bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba* a *Crataerina pallida*.
3. Otestovat přítomnost horizontálně přenesených genů v genomech bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba* a *Crataerina pallida*.
4. Rekonstruovat metabolické dráhy pro důležité B vitamíny a kofaktory bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba* a *Crataerina pallida*.



### 3 Materiál a metody

#### 3.1 Vstupní metagenomická data

Výchozími daty pro mou práci byly metagenomické soubory pair-end readů vygenerované na HiSeq Illumina platformě (GeneCore, Heidelberg) pro dva druhy dvoukřídlého hmyzu čeledi Hippoboscidae, z jednoho jedince kloše *Ornithomya biloba* a jednoho jedince kloše *Crataerina pallida*. Kvalita metagenomických dat byla prověřována pomocí programu FastQC (Andrews, 2010) a na základě výsledků byla hrubá metagenomická data otrimována na 104 bp (tj. tak, aby kvalita readů měřená pomocí Phred score neklesla pod hodnotu 20) pomocí trimovacího nástroje BBduk (<https://jgi.doe.gov/data-and-tools/bbtools/>, viz Tab.1).

Tabulka 1: Základní informace o osekvenovaných readech z kloše *Ornithomya biloba* a *Crataerina pallida* na HiSeq Illumina platformě.

|  | Délka readů (bp) | Obsah GC (%) | Počet readů |
|--|------------------|--------------|-------------|
| Dataset forwardových readů z kloše <i>Ornithomya biloba</i>  | 93-104           | 39           | 113 673 020 |
| Dataset reverzních readů z kloše <i>Ornithomya biloba</i>    |                  |              |             |
| Dataset forwardových readů z kloše <i>Crataerina pallida</i> | 91 - 104         | 36           | 117 556 819 |
| Dataset reverzních readů z kloše <i>Crataerina pallida</i>   |                  |              |             |

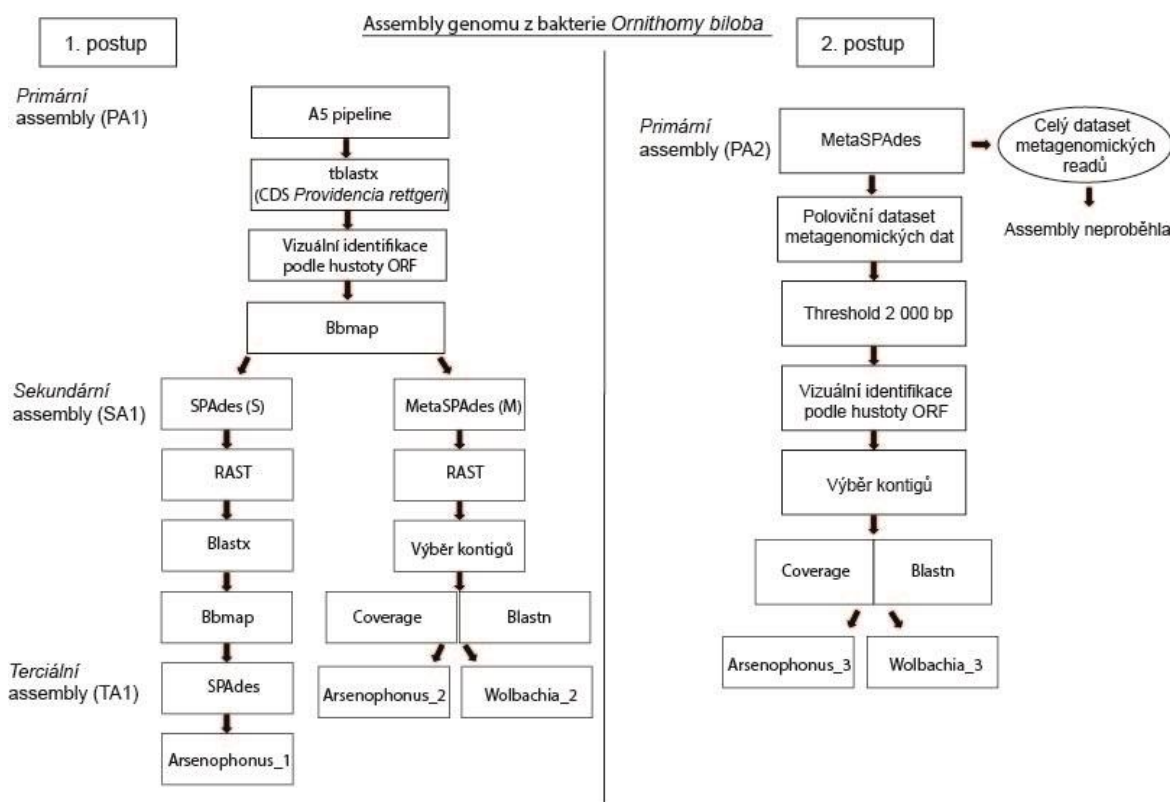
#### 3.2 Assembly genomu

Pro lepší přehlednost použitého postupu při assembly genomu bakterie z kloše *Ornithomya biloba* a bakterie z kloše *Crataerina pallida* bylo vytvořeno přehledné schéma (viz Obr. 1 a Obr. 3).

##### 3.2.1 Bakterie z kloše *Ornithomya biloba*

Assembler MetaSPAdes (Nurk et al., 2017) nebyl schopen zpracovat celý dataset metagenomických readů, patrně z důvodu vysoké komplexity dat. Proto byly pro další zpracování metagenomických dat z kloše *Ornithomya biloba* použity dva alternativní postupy (porovnání a kombinace dvou odlišných postupů zajistilo minimalizaci a kompenzaci případných artefaktů). První postup zahrnoval vytvoření jednoduché tzv. *primární* assembly,

“vychytání” readů odpovídajících symbiotickým bakteriím (pomocí mapování na *primární* assembly) a jejich konečné zpracování v *sekundární* a *terciární* assembly. Druhý postup spočíval v náhodném filtrování vstupního datasetu readů na polovinu a jeho assembly programem MetaSpades. Oba postupy jsou podrobněji popsány níže.

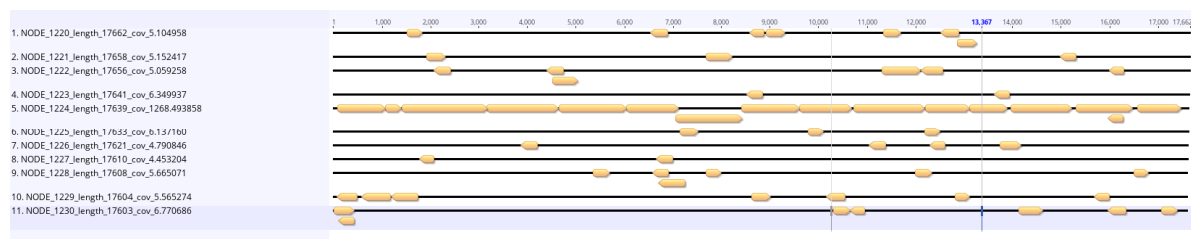


Obrázek 1: Schéma assembly genomu bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba*.

### 3.2.1.1 Primární assembly v A5 pipeline

V prvním postupu skládání metagenomických dat z kloše *Ornithomya biloba* byl pro *primární* assembly kontigů použit assembler A5 pipeline (Coil, Jospin and Darling, 2015; viz Obr. 1 PA1). Kontigy odpovídající genomu bakterie rodu *Arsenophonus* byly identifikovány blastováním (tblastx, e-value e-10) kódujících oblastí (CDS) bakterie *Providencia rettgeri* (NZ\_CP017671.1, [www.ncbi.nlm.nih.gov/genome/](http://www.ncbi.nlm.nih.gov/genome/)), blízkce příbuzné rodu *Arsenophonus*, proti databázi, která byla vytvořena z kontigů získaných při *primární* assembly. Při kontrole těchto kontigů byly odhaleny kontigy patřící kloši *Ornithomya biloba* (patrně v důsledku homologie některých genů rozeznávaných v průběhu blastu i mezi hostitelem a bakterií), které byly následně z datasetu vyřazeny. Eukaryotické kontigy lze od prokaryotických kontigů identifikovat vizuálně podle hustoty uspořádání ORF (open reading frames) v kontizích předpovězených programem Geneious (Kearse et al., 2012; viz Obr. 2). Hustota uspořádání

ORF na prokaryotickém kontigu je vyšší než u eukaryotického kontigu. Každý ORF zpravidla kóduje jeden gen. Jednotlivé geny jsou odděleny krátkými mezigenovými oblastmi. ORF v eukaryotických kontizích jsou kratší než u prokaryotických kontigů, jelikož se nejedná o celé geny. Eukaryotické geny jsou tvořeny kodujícími oblastmi tzv. exony a nekodujícími oblastmi tzv. introny. Na rozdíl od eukaryotických genů prokaryotické geny neobsahují nekodující introny. Na set vybraných prokaryotických kontigů byly pomocí nástroje bmap (<https://jgi.doe.gov/data-and-tools/bbtools/>) namapovány originální ready z původního metagenomického datasetu. Cílem bylo získat ready, které mohou potenciálně prodlužovat jednotlivé kontigy, popřípadě je navzájem spojit. Z výsledného datasetu namapovaných readů byly v programu Geneious vytvořeny datasey forwardových a reverzních readů, ze kterých byly dále vytvořeny *sekundární* assembly. Pro *sekundární* assembly byly použity dva různé assembly, konkrétně SPAdes (Bankevich et al., 2012) a MetaSPAdes. Dva assembly byly zvoleny na základě odlišností jejich algoritmu používaného v přístupu k datům a ověření jejich funkčnosti a úspěšnosti při assembly metagenomických dat. Algoritmus assembleru SPAdes implicitně nepředpokládá přítomnosti více genomů v datasetu. Naopak algoritmus assembleru MetaSPAdes byl navržen pro assembly komplexních metagenomických dat, ve kterých se objevuje větší množství genomů různých organismů.



Obrázek 2: Ukázka vizuální identifikace bakteriálních kontigů podle hustoty předpovězených ORF v programu Geneious. Kontig s názvem HALF\_NODE\_1224\_length\_17639\_cov\_1268.493858 patří bakterii. Všechny ostatní kontigy patří eukaryotickým organismům, v tomto případě pravděpodobně kloši, hostiteli kloše nebo jeho mnohobuněčným parazitům.

**SPAdes** (viz Obr. 1 SA1, S): složené kontigy byly anotovány pomocí serveru Rapid Annotation using Subsystem Technology (Aziz et al., 2008; dále jen Rast server). Pro identifikaci jednotlivých bakteriálních kontigů byl použit blast. Pro rychlejší průběh analýzy byly použity individuální sekvence genů získané anotací v Rast serveru. Sekvence genů byly blastovány proti nukleotidové nr/nt databázi z NCBI pomocí blastovacího nástroje tblastx. Z výsledků blastu jednotlivých genů bylo možné identifikovat původ celých kontigů. Pro další

analýzy byly vybrány pouze kontigy, jejichž blastem nalezené homology pocházely z bakterií rodu *Arsenophonus* a jejich blízkých příbuzných, rodu *Providencia*, *Proteus*, *Xenorhabdus*, *Photorhabdus* a *Riesia* (soubor těchto homologních sekvencí byl použit pro blast, viz níže). Na vybrané kontigy byly namapovány originální metagenomické ready pomocí nástroje bbmap. Z výsledného datasetu namapovaných readů byly v programu Geneious vytvořeny soubory forwardových a reverzních readů, které byly následně použity pro *terciální assembly* pomocí assembleru SPAdes (viz Obr. 1 TA1). Výsledný dataset takto vytvořených kontigů byl pojmenován jako **Arsenophonus\_1** (viz Tab 2).

**MetaSPAdes** (viz Obr. 1 SA1, M): Anotace assemblovaných kontigů proběhla na Rast serveru. Identifikace jednotlivých bakteriálních kontigů proběhla kombinovaně dvěma způsoby. Při prvním způsobu byly kontigy na základě výrazně odlišných hodnot coverage rozděleny do dvou skupin, jedna s rozmezím coverage 6 - 41 a druhá s rozmezím > 1 000 (viz Tab. 2). Přítomnost dvou skupin s výrazně odlišnými hodnotami coverage naznačovala přítomnost dvou bakterií v datasetu (narozdíl od assembly z programu SPAdes). Druhým způsobem identifikace jednotlivých bakteriálních kontigů byl blastn. Oba postupy odhalily přítomnost dvou bakterií v datasetu kontigů. Kontigy byly rozděleny do dvou datasetů, pojmenovaných **Wolbachia\_2** a **Arsenophonus\_2** (viz Tab. 2).

### 3.2.1.2 Primární assembly v MetaSPAdes

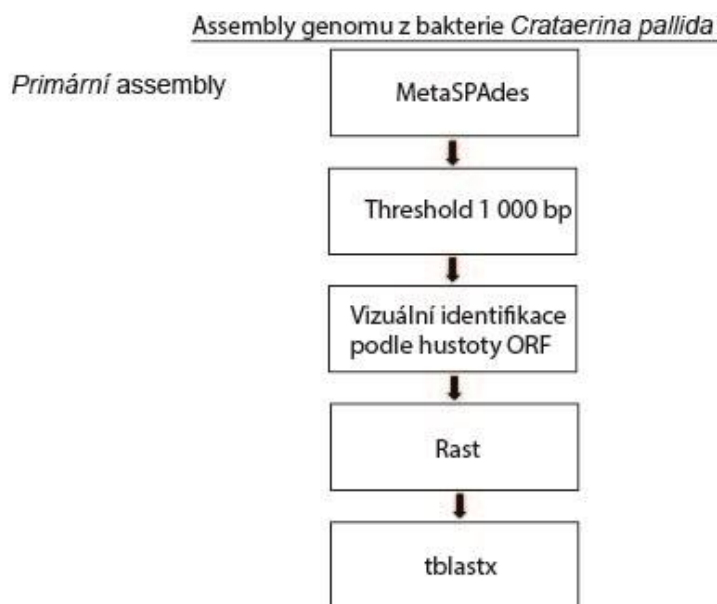
V druhém postupu byla pro *primární assembly* použita polovina z datasetu originálních readů (viz Obr. 1 PA2). Ready byly z celého datasetu metagenomických readů vybrány náhodně pomocí nástroje Fastq-tools (Droop, 2016; příkazem: fastq-sample ornitomya\_bbduk\_fwd.fastq ornitomya\_bbduk\_rev.fastq -p 0.5 -o half\_reads). Z výsledného souboru 715 tisíc kontigů byly pro další analýzy použity pouze kontigy od velikosti 2 000 bp. Výběr kontigů proběhl v programu Geneious a jeho výsledkem byl soubor 173 tisíc kontigů. Anotace assemblovaných kontigů byla provedena na Rast serveru. V této fázi obsahoval dataset jak bakteriální kontigy, tak i kontigy patřící kloši *Ornithomya biloba*, popřípadě ptačímu hostiteli kloše nebo jeho krevním parazitům. Z datasetu kontigů bylo tedy nutné pro další analýzy vybrat pouze bakteriální kontigy. Bakteriální kontigy byly vizuálně identifikovány podle hustoty uspořádání ORF (open reading frames, viz Obr. 1) předpovězených programem Geneious. Dalším krokem byla identifikace vybraného datasetu bakteriálních kontigů. Toho bylo docíleno kombinací dvou postupů. Zaprvé byly kontigy rozděleny na základě výrazně odlišných hodnot coverage do dvou skupin, jedna s rozmezím

coverage 14 - 25 a druhá s rozmezím  $> 100$  (viz Tab. 2), což naznačovalo výskyt dvou odlišných bakterií v datasetu. Zadruhé byl pro identifikaci vybraných bakteriálních kontigů použit blast, který odhalil, ze které bakterie daný kontig pochází. Z vybraných bakteriálních kontigů byla vytvořena databáze. Jako vstupní soubor pro tento blastn v programu Geneious byl použit soubor sekvencí získaných anotací kontigů ze *sekundární* assembly pomocí assembleru SPAdes (viz Primární assembly v A5 pipeline, SPAdes). Výsledek blastu potvrdil přítomnost dvou bakterií v datasetu kontigů, a to konkrétně bakterie *Wolbachia* a *Arsenophonus*. Kontigy byly rozděleny do dvou datasetů pojmenovaných **Wolbachia\_3** a **Arsenophonus\_3** (viz Tab. 2).

Pro následné analýzy byly použity jen datasety kontigů bakterie rodu *Arsenophonus*. V průběhu assembly genomu za použití assembleru SPAdes (Obr. 1 TA1, SPAdes) nebyla odhalena přítomnost dvou bakteriálních genomů v metagenomických datech, patrně z důvodu použitého assembleru. Proto byl dataset Arsenophonus \_1 vyřazen z dalších analýz. Dataset Arsenophonus\_2 a Arsenophonus\_3 byly navzájem porovnávány pomocí blastovacího nástroje blastn v programu Geneious. Unikátní kontigy použité pro další analýzy byly vybrány podle výsledku de novo assembly, která odhalil shody mezi kontigy z jednotlivých datasetů. Vybírány byly delší z duplicitních kontigů. Kontigy byly poté anotovány pomocí Rast serveru. Výsledkem tohoto postupu byl draft genomu bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba* (viz Tab. 3).

### 3.2.2 Bakterie z kloše *Crataerina pallida*

Pro *primární* assembly byl použit assembler MetaSPAdes (viz Obr. 3). Výsledný soubor kontigů z assembly obsahoval okolo 600 tisíc sekvencí, a nebylo možné takto velké množství kontigů zpracovat, proto byl stanoven threshold kontigů na 1 000 bp. Dataset kontigů delších než 1 000 bp obsahoval jak bakteriální kontigy, tak i kontigy kloše *Crataerina pallida*. Bakteriální kontigy byly vizuálně identifikovány podle hustoty uspořádání ORF (open reading frames) předpovězených programem Geneious a pro následné analýzy byly vybrány pouze bakteriální kontigy. Anotace kontigů byla provedena pomocí Rast serveru. Anotované sekvence bakteriálních kontigů byly blastovány proti nr/nt databázi z NCBI pomocí blastovacího nástroje tblastx. Kontigy bakterie *Arsenophonus* byly použity pro rekonstrukci metabolických drah.



Obrázek 3: Schéma assembly genomu bakterie rodu *Arsenophonus* z kloše *Crataerina pallida*.

### 3.3 Fylogeneze horizontálně přenesených genů

Anotací kontigů pro bakterii rodu *Arsenophonus* z klošů *Ornithomya biloba* (dále jen *A.\_O.b.*) a *Crataerina pallida* (dále jen *A.\_C.p.*) pomocí Rast serveru byly identifikovány jednotlivé geny a bylo zjištěno, zda se jedná o geny se známou funkcí, nebo o hypotetické geny, u kterých funkci neznáme. K přibližnému taxonomickému zařazení genů byl použit blastp. Jednotlivé geny byly blastovány proti nr/nt databázi z NCBI. Pro další posuzování potenciálního zisku genů horizontálním přenosem byly geny pracovníě rozděleny do následujících skupin: *Ribozomální geny*: nebyly prověřovány na horizontální přenos genů, jelikož jsou velmi konzervativní, neobsahují tedy téměř žádný fylogenetický signál a jejich fylogenetické analýzy jsou často nevěrohodné. *Geny pro fágové proteiny, mobilní elementy, viry a virové částice*: nebyly prověřovány na horizontální přenos, jelikož z jejich podstaty je zřejmé, že jsou fylogeneticky nestabilní a dochází u nich k častému přenosu mezi organismy. *Vlastní geny*: do této skupiny byly zahrnuty všechny geny, u kterých výsledek blastu přesvědčivě ukázal původ v rámci rodu *Arsenophonus*, popřípadě jeho blízkých příbuzných, jmenovitě rodu *Providencia*, *Proteus*, *Xenorhabdus*, *Photorhabdus* and *Riesia*. Z toho důvodu byly tyto geny považovány za vertikálně zděděné a nebyly fylogeneticky testovány, jelikož jsou geny těchto bakterií natolik příbuzné, že neobsahují dostatečné množství fylogenetické informace pro odhalení případného horizontálního přenosu. *Potenciální HGT*: tyto geny podle výsledku blastu pocházely potenciálně z jiných skupin bakterií. Blast je však pouze orientační nástroj a nemůže sloužit k odhalení horizontálního přenosu genů, proto byl původ genů

testován pomocí fylogenetické analýzy. Výběr ortologů pro skupinu genů označených jako *potenciální* HGT do matice pro fylogenetickou analýzu probíhal za využití blastovacího nástroje blastp na serveru NCBI Blast (Altschul et al., 1990). Zprvce bylo provedeno blastování proti vybraným rodům bakterií, konkrétně rodu *Arsenophonus*, *Proteus*, *Providencia*, *Xenorhabdus*, *Photorhabdus*, *Riesia* a *Escherichia* (s cílem ověřit, zda je homolog daného genu přítomen u některé blízké příbuzné bakterie a získat zároveň outgroup), a zadruhé bylo blastování provedeno proti kompletní nr/nt databázi (s cílem získat set nejbližších sekvencí bez ohledu na jejich fylogenetické postavení). V programu Geneious byl dále proveden alignment pomocí MAFFT alignment tool v 7.450 (Kato et al., 2002) s nastavením default parametrů. Tento alignment byl dále upravován pomocí programu GBlocks s nastavením volné selekce eliminující nepřesně alignované pozice s případnými divergentními úseky (Catresana, 2000). Fylogenetické stromy byly vytvořeny s využitím maximum-likelihood analýzy v programu online PhyML v. 3.3 (Guindon et al., 2010), kde byl vybrán specifický evoluční model pro každý gen pomocí SMS (smart model selection) funkce na základě Akaikeova informačního kritéria (AIC; Lefort et al., 2017). Konkrétní použité evoluční modely jsou zmíněny u jednotlivých fylogenetických stromů.

### 3.4 Rekonstrukce metabolických drah

Pro rekonstrukci metabolických drah byl použit anotační nástroj BlastKOALA (<https://www.kegg.jp/blastkoala/>) a mapovací nástroj KEGG Mapper ([https://www.genome.jp/kegg/tool/map\\_pathway.html](https://www.genome.jp/kegg/tool/map_pathway.html)). BlastKOALA porovnával jednotlivé sekvence z našeho datasetu se setem KEGG genů a přiřadila každé sekvenci z našeho datasetu K číslo. Soubor těchto K čísel Kegg mapper namapoval do schémat jednotlivých metabolických drah. Na výsledném schématu je tedy možné pozorovat přítomnost jednotlivých genů z našeho datasetu v metabolických drahách. Absence některých genů v jednotlivých metabolických drahách, detekovaná pomocí této metody, byla dodatečně ověřována. Sekvence ověřovaných genů byly nalezeny v některém z dostupných genomů bakterií rodu *Arsenophonus* (v databázi NCBI) a následně byly blastovány proti databázi vytvořené z genomu bakterie A.\_O.b. a A.\_C.p. Pro porovnávání byly do analýzy přidány další dva genomy bakterií rodu *Arsenophonus*, a to konkrétně *Arsenophonus melophagi* (IMG databáze, ID: 2510065004, <https://img.jgi.doe.gov/>) a *Candidatus Arsenophonus lipopteni* (NCBI databáze, ID: NZ\_CP013920.1, <https://www.ncbi.nlm.nih.gov/genome/>). Genomy těchto dvou bakterií byly staženy z NCBI databáze a zpracovány stejným způsobem jako bakterie rodu *Arsenophonus* z klošů *Ornithomya biloba* a *Crataerina pallida*.

## 4 Výsledky

### 4.1 Genomické assembly

#### 4.1.1 Metagenomická data z kloše *Ornithomya biloba*

V rámci *primární* assembly bylo pomocí assembleru A5 pipeline získáno 363 915 kontigů (viz Obr. 1 PA1). Následným blastem (tblastx, e-value e-10) bylo vybráno 1 031 kontigů. Vizuální identifikace podle hustoty uspořádání ORF (viz Obr. 2) odhalila 338 bakteriálních kontigů, na které se namapovalo 30 milionů readů z původní metagenomické assembly za použití nástroje bmap.

V rámci *sekundární* assembly bylo pomocí assembleru SPAdes získáno 491 kontigů (viz Obr. 1 SA1 S). Následným blastem bylo vybráno 192 kontigů, na které se namapovalo 28 milionů readů z původní metagenomické assembly. Zpracováním tohoto datasetu readů v rámci *terciální* assembly pomocí assembleru SPAdes vznikl dataset **Arsenophonus\_1** obsahující 201 kontigů. Základní genomové údaje o tomto datasetu jsou uvedeny v tabulce 2.

V rámci *sekundární* assembly (viz Obr. 1 SA1 M) bylo pomocí assembleru MetaSPAdes a následného výběru kontigů na základě hodnoty coverage (viz Tab. 2) a výsledků blastn získáno 363 kontigů, ze kterých byly vytvořeny dva datasety, konkrétně **Arsenophonus\_2** a **Wolbachia\_2**. Základní genomové údaje o těchto datasetech jsou uvedeny v tabulce 2.

V rámci *primární* assembly pomocí assembleru MetaSPAdes z polovičního datasetu readů bylo získáno 715 764 kontigů (viz Obr. 1 PA2). Po vyřazení kontigů menších než 2 000 bp obsahoval dataset 172 930 kontigů. Vizuální identifikace podle hustoty uspořádání ORF (viz Obr. 2) odhalila 363 bakteriálních kontigů, ze kterých byly vytvořeny dva datasety, konkrétně **Arsenophonus\_3** a **Wolbachia\_3**. Základní genomové údaje o těchto datasetech jsou uvedeny v tabulce 2.

#### 4.1.2 Metagenomická data z kloše *Crataerina pallida*

Při primární assembly pomocí assembleru MetaSPAdes bylo získáno 600 042 kontigů (viz Obr. 3). Po vyřazení kontigů menších než 1 000 bp obsahoval dataset 167 705 kontigů. Vizuální identifikace podle hustoty uspořádání ORF (viz Obr. 2) odhalila 766 kontigů. Základní genomové údaje o výsledném datasetu kontigů bakterie rodu *Arsenophonus* z kloše *Crataerina pallida* jsou uvedeny v tabulce 3.



Tabulka 2: Základní genomové údaje o datasetech kontigů získaných při procesu skládání genomu bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba* pro všechny použité přístupy. Názvy datasetů *Arsenophonus\_1*, *Arsenophonus\_2* a *Arsenophonus\_3* byly zkráceny na *Arseno\_1*, *Arseno\_2* a *Arseno\_3*.

|                              | <i>Ornithomya biloba</i> |             |          |             |          |
|------------------------------|--------------------------|-------------|----------|-------------|----------|
|                              | 1. postup                |             |          | 2. postup   |          |
|                              | SPAdes                   | MetaSPAdes  |          | MetaSPAdes  |          |
|                              | Arseno_1                 | Wolbachia_2 | Arseno_2 | Wolbachia_3 | Arseno_3 |
| Počet kontigů                | 201                      | 127         | 185      | 57          | 121      |
| Nejdelší kontig (Kbp)        | 18,2                     | 29,7        | 18,2     | 37,9        | 17,6     |
| Nejkratší kontig (bp)        | 85                       | 70          | 555      | 2938        | 2003     |
| Konkatenovaná sekvence (Mbp) | 0,69                     | 1,01        | 0,66     | 0,64        | 0,53     |
| Obsah GC (%)                 | 24,4                     | 35,3        | 24,7     | 35,3        | 25,8     |
| Predikované geny (CDS)       | 562                      | 679         | 553      | 630         | 443      |
| Coverage                     | >1000                    | 6 - 41      | >1000    | 14 - 25     | >100     |

Dataset kontigů bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba* (dále jen *A.\_O.b.*) obsahoval 190 kontigů. Dataset kontigů bakterie rodu *Arsenophonus* z kloše *Crataerina pallida* (dále jen *A.\_C.p.*) obsahoval 161 kontigů. Základní informace o obou datasetech jsou uvedeny v tabulce 3. Dataset kontigů bakterie *A.\_O.b.* obsahuje větší počet kontigů než dataset kontigů bakterie *A.\_C.p.*, ale celková délka konkatenované sekvence bakterie *A.\_C.p.* je zhruba čtyřikrát větší než celková délka konkatenované sekvence bakterie *A.\_O.b.* (2,41 Mbp vs 0,68 Mbp). Bakterie *A.\_C.p.* má i větší počet genů a vyšší obsah GC.

Tabulka 3: Základní informace o výsledných datasetech kontigů bakterií rodu *Arsenophonus* z kložů *Ornithomya biloba* a *Crataerina pallida*.

|                              | <i>Arsenophonus</i><br>z <i>Ornithomya biloba</i> | <i>Arsenophonus</i><br>z <i>Crataerina pallida</i> |
|------------------------------|---|--|
| Počet kontigů                | 190   | 161  |
| Nejdelší kontig (Kbp)        | 18,2  | 69,3   |
| Nejkratší kontig (bp)        | 796   | 1405   |
| Konkatenovaná sekvence (Mbp) | 0,68  | 2,41   |
| Obsah GC (%)                 | 24,7  | 38,3   |
| Celkový počet genů           | 615   | 2306   |

## 4.2 Anotace

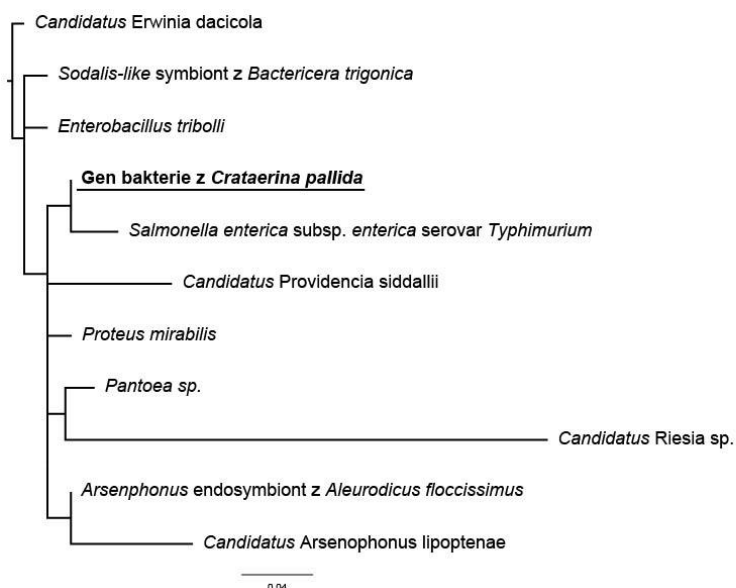
Výsledkem anotace kontigů na Rast serveru bylo 615 anotovaných genů bakterie *A.\_O.b.* a 2 306 anotovaných genů bakterie *A.\_C.p.* Genom bakterie *A.\_O.b.* tedy obsahoval zhruba čtyřikrát menší počet genů než genom bakterie *A.\_C.p.* Geny byly z důvodu posuzování potenciálního horizontálního přenosu genů rozděleny do skupin zmíněných výše (viz. kapitola *Fylogeneze horizontálně přenesených genů*). Počty genů pro zmíněné skupiny byly zaneseny do tabulky 4. V genomu bakterie *A.\_O.b.* nebyly identifikované žádné fágové geny, mobilní elementy, viry ani virové částice.

Tabulka 4: Přehled jednotlivých kategorií anotovaných genů pro bakterie rodu *Arsenophonus* z kložů *Ornithomya biloba* a *Crataerina pallida* a anotovaných genů prověřovaných na horizontální přenos (‘ geny se známou funkcí, ‘‘ hypotetické geny).

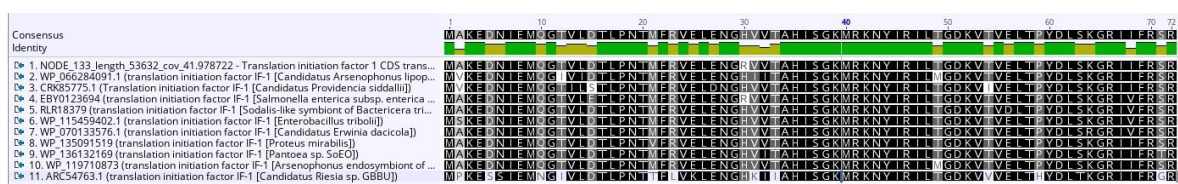
|                          | <i>Arsenophonus</i><br>z <i>Ornithomya biloba</i> | <i>Arsenophonus</i><br>z <i>Crataerina pallida</i> |
|--------------------------|---|--|
| Celkový počet genů       | 615   | 2306   |
| Vlastní geny             | 506‘+ 50‘‘  | 1799‘+ 308‘‘                                       |
| Ribozomální geny         | 45  | 52   |
| Geny pro fágové proteiny | 0   | 72   |
| Mobilní elementy         | 0   | 16   |
| Viry, virové částice     | 0   | 9  |
| Potenciální HGT          | 12‘ + 2‘‘   | 29‘ + 21‘‘   |

### 4.3 Fylogeneze horizontálně přenesených genů (HGT)

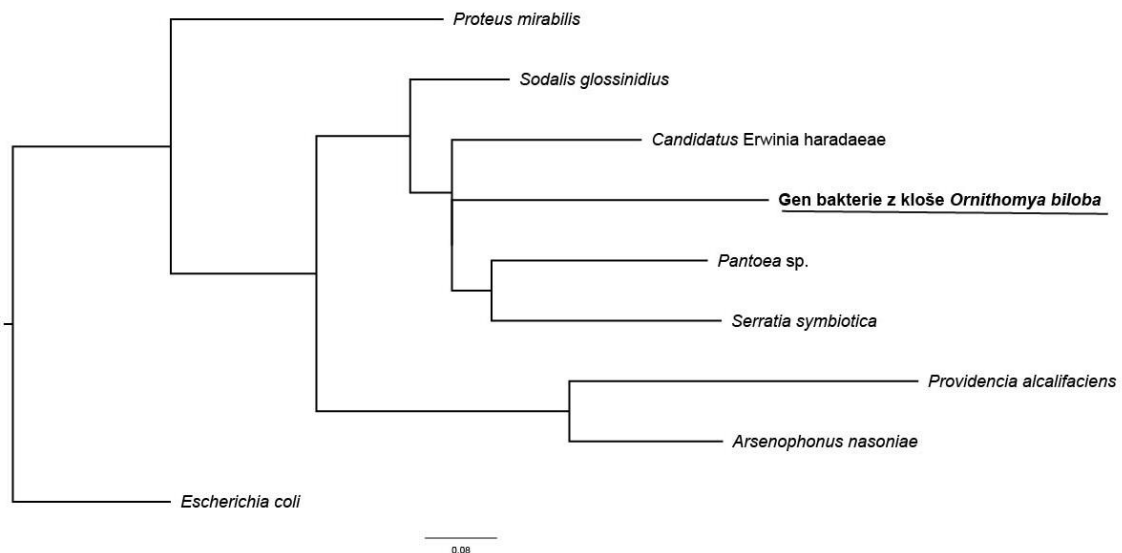
Horizontální přenos genů byl prokázán u 4 genů z bakterie *A.\_O.b.* ze 14 ověřovaných (1 gen hypotetický a 3 geny s konkrétní určenou funkcí). U dalších 6 genů byl prokázán původ z bakterie rodu *Arsenophonus*, u dalších 4 genů nebylo možné horizontální přenos ani potvrdit ani vyvrátit (viz Tab. 4). U genů byl pro jednoznačnost zachován anglický název. Bakteriální taxony ve fylogenetických stromech těchto nejistých genů nevytvářejí očekávané (tedy známé) fylogenetické vztahy. Příkladem může být gen pro translation initiation factor 1 (viz Obr. 4 a 5) nebo gen pro phospholipase D (viz Obr. 6). Ve fylogenetických stromech genů symbiotických bakterií je v některých případech možné pozorovat výskyt dlouhých větví, které by mohly podléhat metodickým artefaktům. Příkladem může být gen pro integration host factor alpha (viz Obr. 7 a 8).



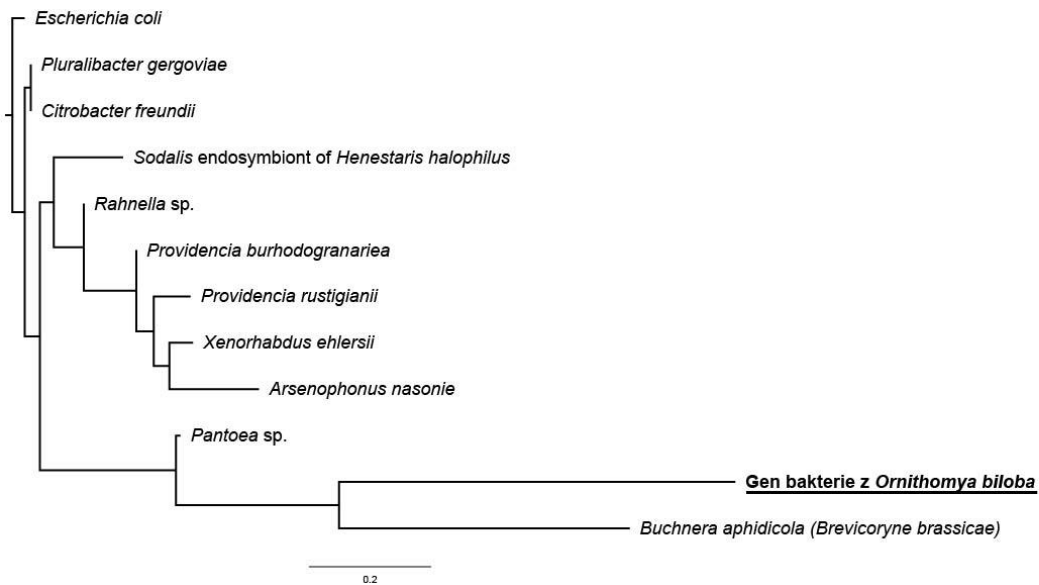
Obrázek 4: Fylogenetický strom pro gen translation initiation factor 1, který vznikl za použití evolučního modelu Flu+G v programu PhyML 3.3.



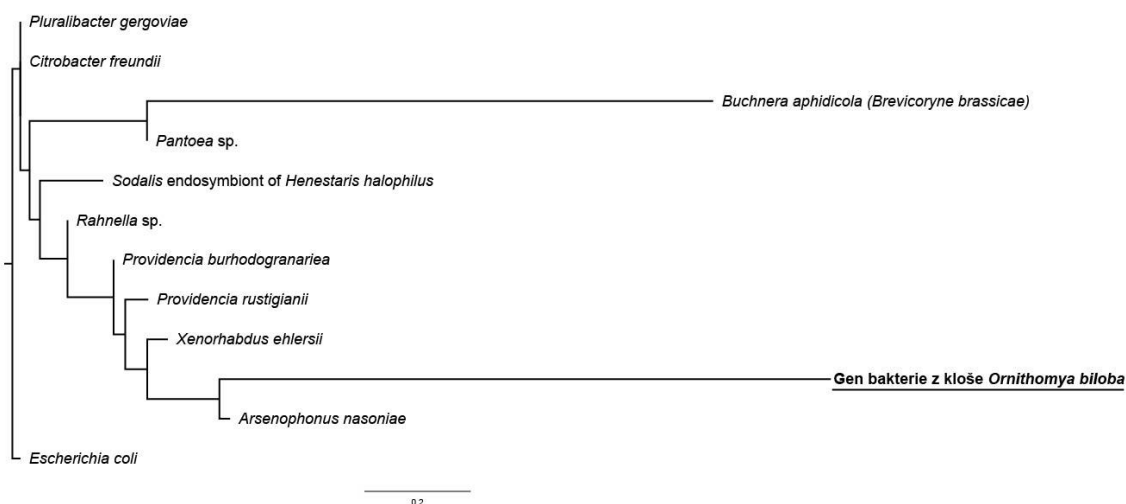
Obrázek 5: Alignment pro gen translation initiation factor 1.



Obrázek 6: Fylogenetický strom genu pro phospholipase D, který vznikl za použití evolučního modelu LG+G v programu PhyML 3.3.



Obrázek 7: Fylogenetický strom pro gen integration host factor alpha, který vznikl za použití evolučního modelu LG v programu PhyML 3.3.



Obrázek 8: Fylogenetický strom pro gen integration host factor alpha, který vznikl za použití evolučního modelu MtMa +G+ +F v programu PhyML 3.3.

U genů z bakterie *A.\_C.p.* byl horizontální přenos prokázán u 6 genů z 50 ověřovaných (4 geny hypotetické a 2 geny s konkrétní určenou funkcí). U 31 genů byl prokázán původ z bakterie rodu *Arsenophonus*, u 13 genů nebylo možné horizontální přenos ani potvrdit, ani vyvrátit (viz Tab. 5). Výsledné fylogenetické stromy genů pro hypotetické proteiny v této práci nejsou uvedeny - vzhledem k tomu, že funkce těchto genů není známa, nebyly součástí dalších analýz a interpretací. Výsledné fylogenetické stromy pro geny s konkrétní funkcí jsou uvedeny v následujících kapitolách. Všechny sekvence použité při fylogenezi genů jsou uvedeny v příloze Tabulka I a II.

Tabulka 5: Přehled výsledků analýzy horizontálního přenosu genů (‘ geny se známou funkcí, ‘‘ hypotetické geny).

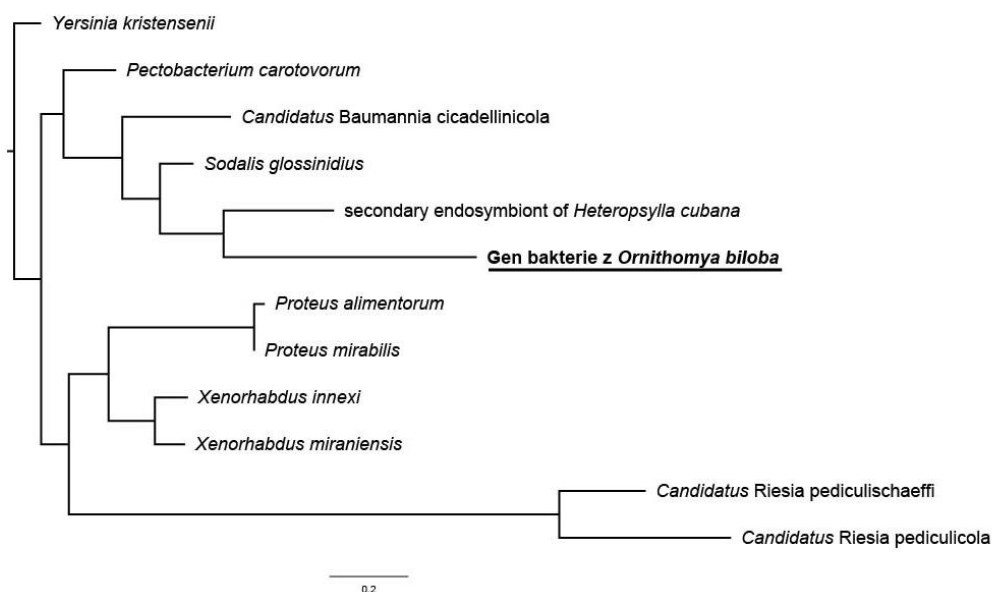
|                                  | <i>Arsenophonus</i><br>z <i>Ornithomya biloba</i> | <i>Arsenophonus</i><br>z <i>Crataerina pallida</i> |
|----------------------------------|---|--|
| Celkový počet prověřovaných genů | 14  | 50   |
| Horizontálně přenesené geny      | 3‘+1‘‘  | 2‘+ 4‘‘  |
| HGT neurčitelné geny             | 4‘  | 7‘+ 6‘‘  |
| Geny <i>Arsenophonus</i>         | 5‘ + 1‘‘  | 20‘ + 11‘‘   |

#### 4.3.1 HGT geny z bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba*

Horizontální přenos byl prokázán u 3 genů s konkrétní funkcí. Jedná se o geny pro pantoate-beta-alanine ligase (viz Obr. 9), 3-methyl-2-oxobutanoate hydroxymethyltransferase

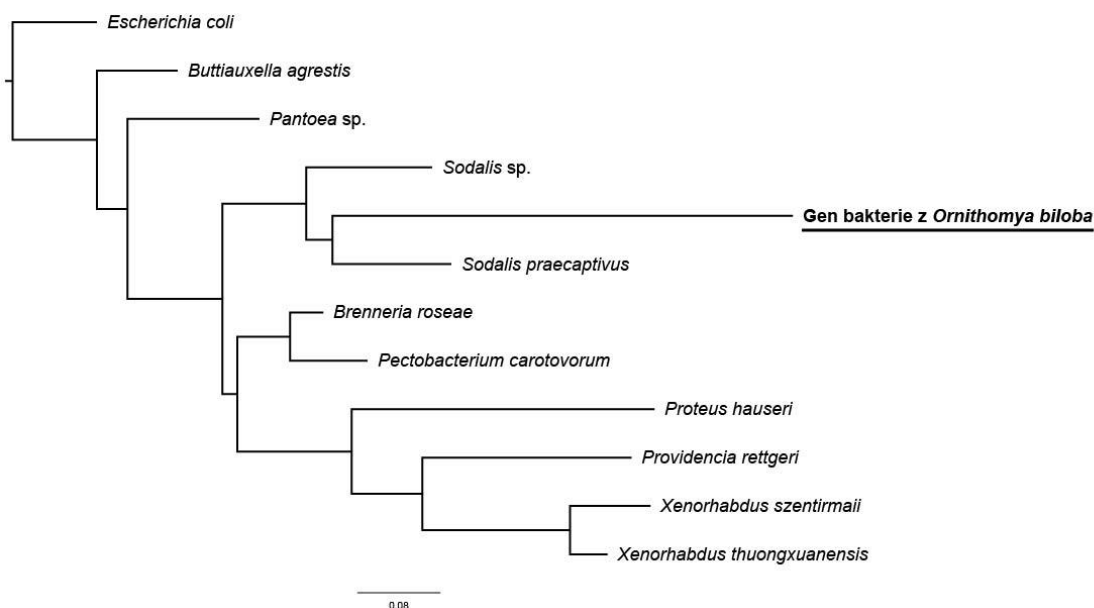
(viz Obr. 10) a 2-dehydropantoate 2-reductase (viz Obr. 11). Tyto 3 geny jsou zapojeny v biosyntéze pantotenátu a CoA (viz níže kapitola Metabolické dráhy).

Pantoate-beta-alanine ligase (viz Obr. 9) je aktivační enzym zapojený do metabolismu beta-alaninu (Cronan, 1980). Nejbližší příbuzný tohoto genu, pocházejícího z našich metagenomických dat, nalezený v NCBI, je homolog ze sekundárního endosymbionta ze zástupce čeledi merovitých (Psyllidae) *Heteropsylla cubana*. Pro vytvoření fylogenetického stromu pro tento gen byl použit evoluční model LG+G+I.



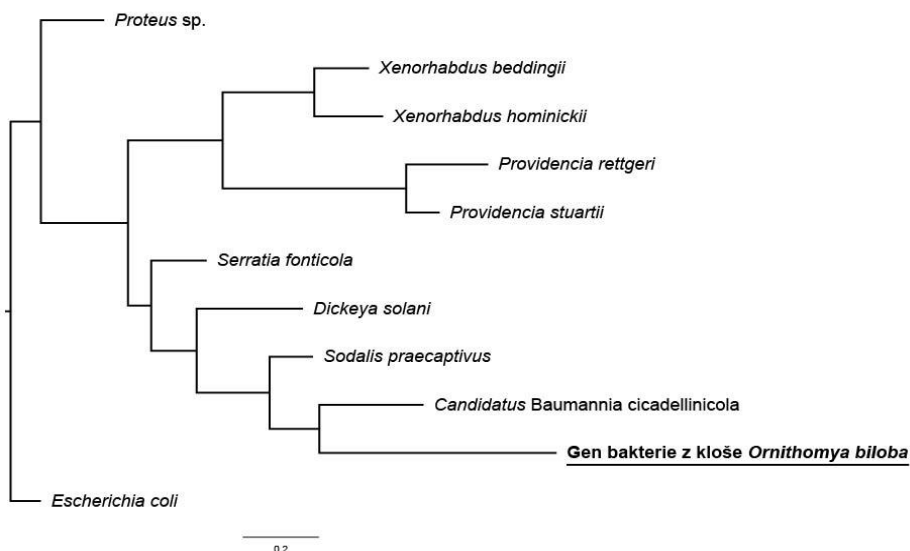
Obrázek 9: Fylogenetický strom genu pro pantoate-beta-alanine ligase, který vznikl za použití evolučního modelu LG+G+I v programu PhyML 3.3.

3-methyl-2-oxobutanoate hydroxymethyltransferase (viz Obr. 10) je katalytický enzym zapojený do metabolismu kyseliny pantotenové, konkrétně v reverzibilní syntéze ketapantoátu (Teller et al., 1976). Nejbližší příbuzný genu pocházejícího z našich metagenomických dat, je homolog ze symbiotické bakterie rodu *Sodalis*. Pro vytvoření fylogenetického stromu pro tento gen byl použit evoluční model LG+G+I.



Obrázek 10: Fylogenetický strom genu pro 3-methyl-2-oxobutanoate hydroxymethyltransferase, který vznikl za použití evolučního modelu LG+G+I v programu PhyML 3.3.

2-dehydropantoate 2-reductase (viz Obr. 11) je katalytický enzym zapojený do metabolismu kyseliny pantotenové, konkrétně do NADPH-závislé produkce pantoátu (Sanchez et al., 2015). Nejbližší příbuzný genu pocházejícího z našich metagenomických dat je homolog z bakterie *Candidatus Baumannia cicadellinicola*. Pro vytvoření fylogenetického stromu pro tento gen byl použit evoluční model LG+G+I+F.

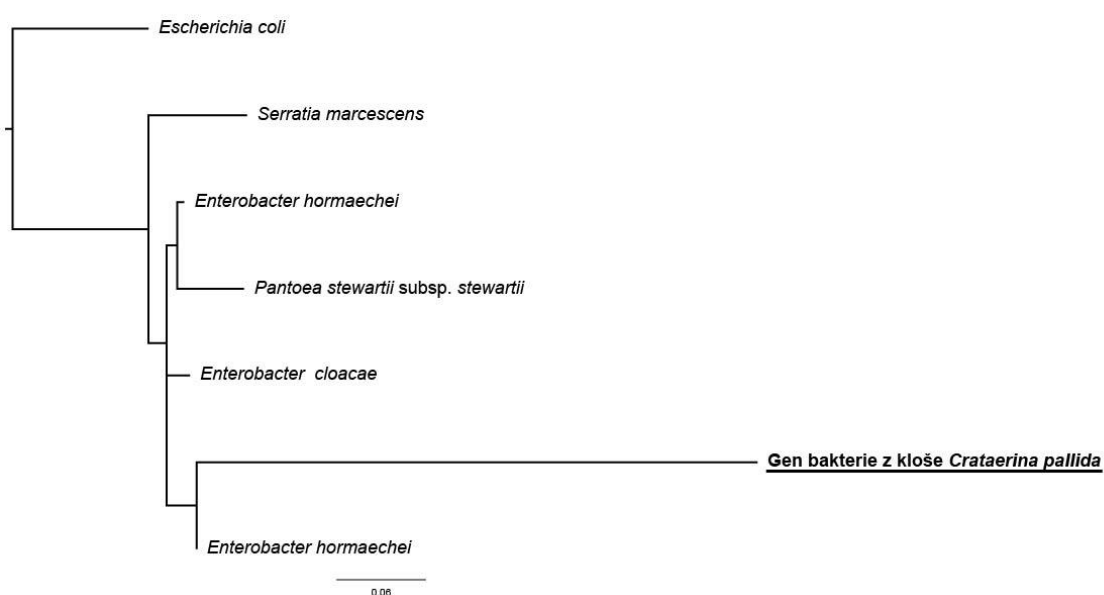


Obrázek 11: Fylogenetický strom genu pro 2-dehydropantoate 2-reductase, který vznikl za použití evolučního modelu LG+G+I+F v programu PhyML 3.3.

#### 4.3.2 HGT geny z bakterie rodu *Arsenophonus* z kloše *Crataerina pallida*

Horizontální přenos byl prokázán u 2 genů s konkrétní funkcí. Jedná se o geny pro YcfA family protein, predicted periplasmic or secreted lipoprotein (viz Obr. 12) a restriction endonuclease, subunit S (viz Obr. 13).

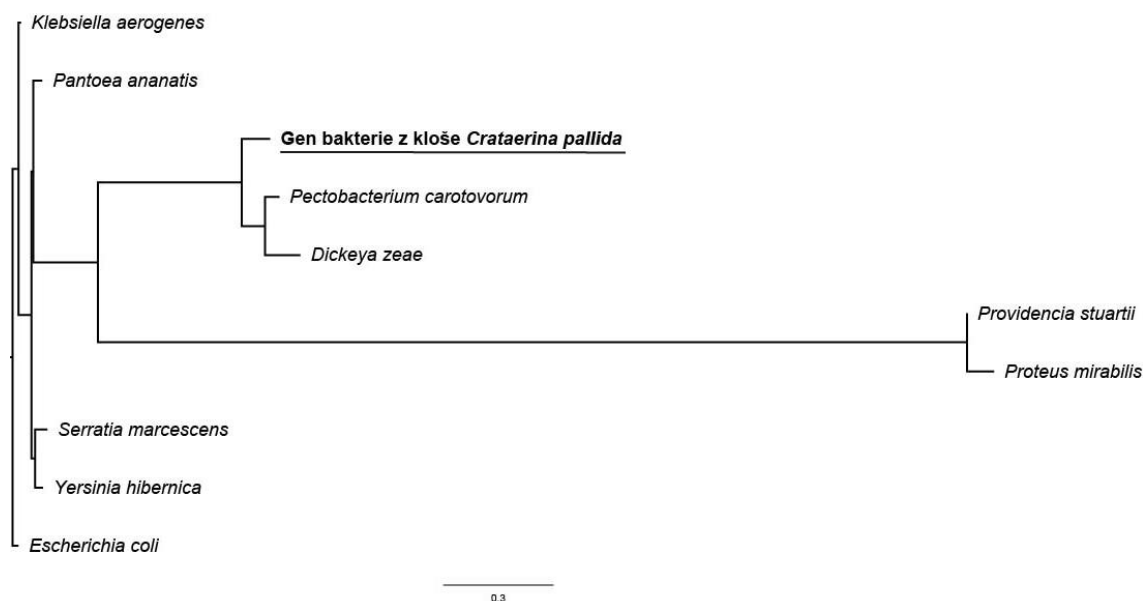
YcfA family protein, predicted periplasmic or secreted lipoprotein (viz Obr. 12) je gen pro periplazmatický nebo sekreční protein, který váže mRNA (<https://www.uniprot.org/uniprot/B8F4F4>). Nejbližší příbuzný tohoto genu pocházejícího z našich metagenomických dat je homolog z bakterie *Enterobacter hormaechei*. Pro vytvoření fylogenetického stromu pro tento gen byl použit evoluční model GTR+I.



Obrázek 12: Fylogenetický strom genu pro YcfA-like protein, predicted periplasmic or secreted lipoprotein, který vznikl za použití evolučního modelu GTR+I v programu PhyML 3.3.

Restriction endonuclease, subunit S (viz Obr. 13) je gen zapojený do štěpení a methylace DNA. Podjednotka S tohoto genu je zodpovědná za označení štěpené DNA (<http://www.ebi.ac.uk/interpro/entry/InterPro/IPR004546/#PUB00035705>). Nejbližší příbuzné sekvence tohoto genu, pocházejícího z našich metagenomických dat, jsou homology z bakterie *Dickeya zae* a *Pectobacterium carotovorum*. Pro vytvoření fylogenetického stromu pro tento gen byl použit evoluční model HIVb+G.





Obrázek 13: Fylogenetický strom genu pro restriction endonuclease (subunit S), který vznikl za použití evolučního modelu HIVb+G v programu PhyML 3.3.

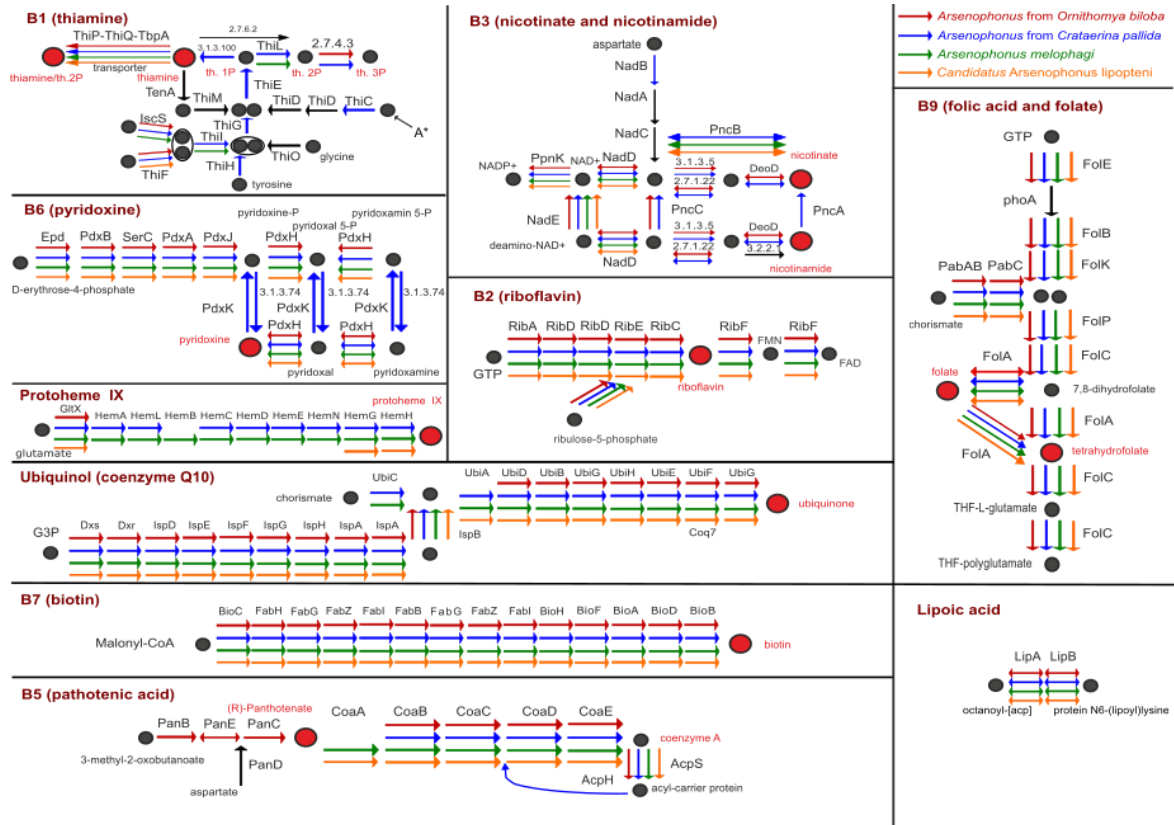
#### 4.4 Porovnání genomů

Genomy bakterií *A.\_O.b.* a *A.\_C.p.* byly porovnávány s genomy dalších dvou bakterií rodu *Arsenophonus*, které jsou symbionty ektoparazitických zástupců z čeledi Hippoboscidae. První porovnávanou bakterií byl *Arsenophonus melophagi*, který je symbiontem kloše *Melophagus ovinus*, krevsajícího ektoparazita ovčí. Druhou porovnávanou bakterií byl *Candidatus Arsenophonus lipopteni*, který je symbiontem kloše *Lipoptena cervi*, který saje krev na zástupcích čeledi jelenovitých (*Cervidae*). Z tabulky základních genomických informací o jednotlivých genech (viz Tab. 6) vyplývá, že bakterie *A.\_O.b.* je nejmenším z porovnávaných genomů a má i nejnižší počet genů. Největším genomem v našem porovnávání je *A.\_C.p.* a ze všech námi porovnávaných genomů obsahuje *A.\_C.p.* také největší počet genů a zároveň má i nejvyšší GC obsah.

Tabulka 6: Základní genomické informace o čtyřech porovnávaných genomech bakterií rodu *Arsenophonus*.

|                               | <i>Arsenophonus</i><br>z <i>Crataerina</i><br><i>pallida</i> | <i>Arsenophonus</i><br><i>melophagi</i> | <i>Candidatus</i><br><i>Arsenophonus</i><br>lipopteni | <i>Arsenophonus</i><br>z <i>Ornithomya</i><br><i>biloba</i> |
|-------------------------------|--|---|---|---|
| Konkatovaná<br>sekvence (Mbp) | 2,41   | 1,16                                    | 0,84  | 0,68  |
| Obsah GC (%)                  | 38,3   | 32,2                                    | 24,9  | 24,7  |
| Počet genů                    | 2306   | 765                                     | 633   | 615   |

## 4.5 Metabolické dráhy



Obrázek 14: Přehled rekonstruovaných metabolických drah pro esenciální B vitamíny a kofaktory poskytované bakterií rodu *Arsenophonus* pro čtyři zástupce tohoto rodu bakterie, konkrétně *Arsenophonus melophagi*, *Candidatus Arsenophonus lipopteni* a dva genomy izolované z klošů *Ornithomya biloba* a *Crataerina pallida*, jejichž assembly proběhla v této práci (A\* zastupuje název 5-amino-1-(5-phospho-D-ribosyl)imidazole).

Z výsledků rekonstrukce metabolických drah pro esenciální B vitamíny a kofaktory bylo možné vyčíst míru kompletnosti a funkčnosti těchto vybraných drah. Dále byla zjištěna přítomnost či absence jednotlivých konkrétních genů v metabolických drahách. Porovnání vybraných metabolických drah (Obr. 14) ukázalo rozdíly mezi čtyřmi zástupci rodu *Arsenophonus* z různých druhů hostitelů čeledi Hippoboscidae.

Metabolické dráhy pro biotin (vitamín B7), kyselinu lipoovou, riboflavin (vitamín B2) a folát s kyselinou listovou (vitamín B9) jsou pro všechny čtyři bakterie kompletní a s největší pravděpodobností tedy funkční. Metabolická dráha pro produkci ubiquinolu z glyceraldehyd-3-fosfátu je funkční pro 3 ze čtyř porovnávaných genomů: v genomu *A.\_O.b.* chybí gen pro 4-hydroxybenzát polyprenyltransferázu (viz Obr. 14, gen *UbiA*). Bakterie *A.\_C.p.* a *Arsenophonus melophagi* jsou jediné dvě bakterie, které dokážou vytvářet ubiquinol i z chorismátu, což je alternativní způsob produkce ubiquinolu.

Metabolická dráha pro thiamin (vitamín B1) je úplná a pravděpodobně funkční pouze u *A.\_C.p.* U ostatních bakterií je tato metabolická dráha neúplná a nefunkční, jelikož u všech 3 bakterií najdeme pouze 2 až 4 geny pro tuto metabolickou dráhu. Ve všech případech je však přítomen gen pro ABC transportér (thiP, thiQ, tbpA).

Metabolická dráha pro kyselinu pantotenovou (vitamín B5) není kompletní pro žádnou námi pozorovanou bakterii. Nejúplnější je tato dráha pro bakterii *A.\_O.b.*, které chybí v genomu pro tuto dráhu pouze gen type I pantothenate kinase (viz Obr. 14, gen CoaA). Z fylogenetických výsledků bylo zjištěno, že tři geny z genomu bakterie *A.\_O.b.*, které jsou součástí metabolické dráhy pro kyselinu pantotenovou, jsou horizontálně přenesené geny (viz výše kapitola HGT geny z bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba*), konkrétně se jedná o geny 3-methyl-2-oxobutanoate hydroxymetyltransferase (viz Obr. 14, gen PanB), 2-dehydropantoate 2-reductase (viz Obr. 14, gen E) a pantoate-beta-alanine ligase (viz Obr. 14, gen PanC).

Metabolická dráha pro protoheme IX je v případě genomu bakterie *Arsenophonus melophagi* kompletní a s největší pravděpodobností i funkční. V případě bakterie *A.\_C.p.* je dráha pro protoheme IX téměř kompletní, chybí pouze gen pro porfobilinogen syntázu (viz Obr. 14, gen HemB). Genomy bakterií *Candidatus Arsenophonus lipopteni* a *A.\_O.b.* neobsahují geny potřebné pro fungování této metabolické dráhy. Obsahují pouze jediný gen, což naznačuje úplnou nefunkčnost této dráhy.

Metabolická dráha pro pyridoxin (vitamín B6) je pro 3 ze 4 námi pozorovaných genomů bakterie rodu *Arsenophonus* nekompletní, a proto s největší pravděpodobností nefunkční, přestože v těchto 3 genomech (*Arsenophonus melophagi*, *Candidatus Arsenophonus lipopteni*, *A.\_O.b.*) chybí pouze 2 geny, konkrétně gen pro pyridoxal fosfatázu (viz Obr. 14, gen 3.1.3.74) a pyridoxin kinázu (viz Obr. 14, gen PdxK). Metabolická dráha pro pyridoxin je pro bakterii *A.\_C.p.* kompletní a s největší pravděpodobností funkční.

Metabolická dráha pro nikotinát s nikotinamidem (vitamín B3) je nekompletní a pravděpodobně nefunkční pro všechny 4 námi pozorované bakterie rodu *Arsenophonus*, přestože v genomu bakterie *A.\_C.p.* chybí pouze dva geny pro tuto metabolickou dráhu, konkrétně geny pro quinolinate synthase (viz Obr. 14, gen NadA) a nicotinate-nucleotide pyrophosphorylase (viz Obr. 14, gen NadC).

V souhrnu je patrné, že bakterie *A.\_C.p.* má kompletní největší počet metabolických drah pro esenciální B vitamíny a kofaktory, konkrétně 7 metabolických drah (biotin, thiamin, pyridoxin, ubiquinol, riboflavin, kyselinu lipoovou a folát s kyselinou listovou). Tato

skutečnost koreluje také s největší velikostí genomu bakterie *A.\_C.p.* (velikost konkatenované sekvence je 2,41 Mbp) a také nejvyšším počtem genů (2306 genů) z porovnávaných genomů. Bakterie *Arsenophonus melophagi* má kompletních 6 metabolických drah (biotin, ubiquinol, protoheme IX, riboflavin, kyselinu lipoovou, folát s kyselinou listovou). Bakterie *Candidatus Arsenophonus lipopteni* má kompletních 5 metabolických drah (biotin, ubiquinol, riboflavin, kyselinu lipoovou, folát a kyselinu listovou).

Bakterie *A.\_O.b.* má kompletní 4 metabolické dráhy (biotin, riboflavin, kyselinu lipoovou, folát s kyselinou listovou). Metabolické dráhy pro ubiquinol a kyselinu pantotenovou jsou však také téměř kompletní, chybí v nich 1 až 2 geny. Genom této bakterie je značně redukován, což se projevilo na počtu funkčních a nefunkčních metabolických drah.

## 5 Diskuze

### 5.1 Výsledky assembly

Výsledkem genomové assembly bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba* (dále jen *A.\_O.b.*) a bakterie rodu *Arsenophonus* z kloše *Crataerina pallida* (dále jen *A.\_C.p.*) nebyly kompletní uzavřené genomy. Genomy nebylo možné složit do uzavřené formy ze dvou možných příčin. První příčinou mohla být přítomnost repetitivních sekvencí v metagenomických datech, které algoritmus assemblerů nedokáže správně rozlišit a rekonstruovat, a proto v takovém místě assembler ukončí kontig a nedojde ke složení kompletního uzavřeného genomu. Repetitivní sekvence způsobují problémy především při assembly velkých genomů. U menších genomů obvykle repetitivní sekvence nedělají problém. Tato příčina by se snad mohla projevit u bakterie *A.\_C.p.*, ale u menší bakterie *A.\_O.b.* je tato příčina nepravděpodobná. Druhou příčinou neúspěšného složení kompletních uzavřených genomů mohla být příliš velká komplexnost metagenomických dat a přítomnost DNA více organismů, mezi nimi i několika bakterií v metagenomických datech. V případě bakterie *A.\_O.b.* se komplexnost metagenomických dat projevila při *primární* assembly, kdy musel být dataset metagenomických readů zmenšen na polovinu, aby použitý assembler zvládl zpracovat tato data. Navíc byla v původním datasetu metagenomických readů odhalena přítomnost druhé bakterie, konkrétně bakterie *Wolbachia*. Od toho se dále odvíjel další proces zpracování dat. V případě genomu bakterie *A.\_C.p.* nebyl proces assembly genomu tak složitý jako u bakterie *A.\_O.b.*, ale i tak nebylo možné složit kompletní uzavřený genom. U genomů bakterie *A.\_O.b.* a bakterie *A.\_C.p.* nelze jednoznačně rozhodnout, co je příčinou neúspěšného složení kompletních uzavřených genomů. Odhalit příčinu by bylo pravděpodobně možné až po získání kompletního uzavřeného genomu.

### 5.2 Redukce a degradace genomů

Genom bakterie *A.\_O.b.* (0,68 Mbp, GC obsah 24,7%) spadá spíše mezi degenerovanější bakteriální genomy, které jsou v porovnání s ostatními genomy malé a mají nízký obsah GC. Obvykle se předpokládá, že míra redukce a degradace genomu (obsah GC) jsou navzájem spřažené a odráží míru adaptace bakterie k symbióze v hostiteli (patrně v jisté míře i dobu trvání symbiózy) (Nováková et al., 2015). Z tohoto pohledu je malý bakteriální genom s nižším obsahem GC výsledkem delšího soužití a větší míry vzájemné adaptace bakterie s hostitelem. V průběhu tohoto soužití se stává bakterie pro svého hostitele obligátním endosymbiontem. Mezi zástupci bakterie *Arsenophonus* byla obligátní endosymbioza u

zástupců dvoukřídlého hmyzu z čeledi klošovitých (Hippoboscidae) již dříve pozorována například u bakterie *Arsenophonus melophagi*, jejíž genom o velikosti 1,16 Mbp a obsahem GC 32,2% spadá mezi redukovanější genomy (Nováková et al., 2015), nebo bakterie *Candidatus Arsenophonus lipopteni*, jejíž genom o velikosti 0,89 Mbp a obsahu GC 24,9% spadá mezi velmi redukované genomy (Nováková et al., 2016).

Výsledky genomové assembly metagenomických dat z kloše *Ornithomya biloba* odhalily přítomnost druhé bakterie, konkrétně bakterie rodu *Wolbachia*. Genom bakterie rodu *Wolbachia* je větší a méně degenerovaný (vyšší obsah GC) než genom bakterie *A.\_O.b.* Z toho je možné usuzovat, že bakterie rodu *Wolbachia* je pro kloše *Ornithomya biloba* fakultativním symbiontem. Této interpretaci nasvědčuje i fakt, že bakterie rodu *Wolbachia* je pro svého hostitele fakultativním symbiontem v naprosté většině případů, často by se dalo říci reprodukčním manipulátorem, který může působit různé defekty v reprodukci svého hostitele (Sinkins et al., 1995; Stouthamer et al., 1999; Teixeira et al., 2008). Navíc přítomnost obligátního a fakultativního symbionta v jednom hostiteli byla již dříve také prokázána. Jedním z mnoha případů takového soužití může být bakterie *A. melophagi*, jako obligátní symbiont, a bakterie rodu *Wolbachia* a *Sodalis* jako fakultativní symbionti u zástupce čeledi klošovitých (*Hippoboscidae*), konkrétně *Melophagus ovinus* (Nováková et al., 2015).

Výsledek genomové assembly bakterie *A.\_C.p.* (2,41 Mbp, GC obsah 38,3%) naznačuje, že velikostí a mírou degradace genomu, ze které lze usuzovat pravděpodobně kratší soužití bakterie s hostitelem, by se bakterie *A.\_C.p.* mohla řadit mezi fakultativní symbionty. V metagenomických datech z kloše *Crataerina pallida* však nebyla zaznamenána přítomnost obligátního symbionta. Z toho vyplývá, že bakterie *A.\_C.p.* bude s největší pravděpodobností pro kloše *Crataerina pallida* obligátním symbiontem. To potvrzuje i fakt, že genom bakterie *A.\_C.p.* i přes jeho poměrně silnou redukci obsahuje geny zapojené do produkce esenciálních B vitamínů a kofaktorů, čímž se stává bakterie *A.\_C.p.* pro svého hostitele nepostradatelnou. Je vysoce pravděpodobné, že bakterie *A.\_C.p.* nahradila nějakého staršího symbionta kloše *Crataerina pallida* poměrně recentně a stala se obligátním symbiontem pro svého hostitele. Bakterie *A.\_O.b.* s největší pravděpodobností také v minulosti nahradila nějakého staršího symbionta svého hostitele kloše *Ornithomya biloba*, ale podle všeho se tak stalo před delší dobou. Takovou dynamiku výměny symbiontů ukazuje i fylogenetická studie zaměřená na bakterie rodu *Arsenophonus* a rodu *Sodalis* v kloších z čeledi Hippoboscidae (Šochová et al., 2017).

## 5.3 HGT a mobilní elementy

### 5.3.1 Mobilní elementy

Absence mobilních elementů, fágových genů, virů a virových částic v genomu bakterie *A.\_O.b.* je v souladu s velikostí a degradací genomu této bakterie. To je pravděpodobně dáno silnější redukcí genomu a mírou přizpůsobení hostiteli. Přítomnost fágových genů, mobilních elementů, virů a virových částic poskytuje bakterii schopnost odolávat nepříznivým podmínkám a přirozeným nepřítelům ve volné přírodě (Oliver et al., 2009). Z toho vyplývá, že bakterie *A.\_O.b.* je na svém hostiteli plně závislá a mimo buňku hostitele by pravděpodobně nepřežila.

V genomu bakterie *A.\_C.p.* byly identifikovány některé mobilní elementy, fágové geny, viry a virové částice. Tento výsledek naznačuje menší míru přizpůsobení bakterie *A.\_C.p.* ke svému hostiteli. Mobilní elementy, fágové geny, viry a virové částice jsou pravděpodobně pozůstatkem z období volně žijící bakterie *A.\_C.p.* a dosud nedošlo k jejich úplné ztrátě při postupné redukcí genomu.

### 5.3.2 Horizontálně přenesené geny

Množství genů prověřovaných na potenciálních HGT bylo v případě genů z bakterie *A.\_O.b.* a *A.\_C.p.* v porovnání s jinými bakteriálními symbionty velmi nízké, jednalo se pouze o desítky genů. Například u symbionta příbuzného bakterii rodu *Neisseria* ze vší *Hoplopleura acanthopus* byly na horizontální přenos prověřovány až stovky genů (Říhová et al., 2019). Většina z celkového setu genů bakterie *A.\_O.b.* a *A.\_C.p.* byla blastem identifikována jako geny vertikálně zděděné, jelikož nejpodobnější geny nalezené blastem patřily bakteriím rodu *Arsenophonus*, nebo některé z blízkce příbuzných bakterií, konkrétně bakterie rodu *Proteus*, *Providencia*, *Xenorhabdus*, *Photorhabdus*, *Riesia*. Mezi těmito geny mohou však teoreticky být geny, které bakterie *A.\_O.b.* nebo *A.\_C.p.* získala horizontálním přenosem z jiného druhu bakterie rodu *Arsenophonus*, v situaci, kdy se v jednom hostiteli setkají dvě bakterie rodu *Arsenophonus*. Tato teorie je v souladu s výsledky studie E. Šochové, ze které vyplývá, že symbiotické bakterie rodu *Arsenophonus* a *Sodalis*, mohou být vyměňovány mezi svými hostiteli z čeledi *Hippoboscidae* (Šochová et al., 2017). HGT genů pocházejících z blízkce příbuzných bakterií tedy není možné fylogenetickou analýzou ani blastem odhalit, jelikož sekvence jednoho genu neobsahuje dostatečné množství informace pro tak přesnou fylogenezi. Z tohoto důvodu tedy mohou v setu genů vyhodnocených jako vertikálně zděděné geny (vlastní geny) být i některé neodhalené horizontální přenosy genů.

Geny, které nebyly identifikovány jako geny bakterie rodu *Arsenophonus* nebo jejich nejbližších příbuzných, bylo nutné z hlediska HGT prověřit fylogenetickou analýzou. V některých případech byl však výsledek fylogenetické analýzy nejasný. Geny s nepřesvědčivým výsledkem fylogenetické analýzy byly označeny jako HGT-neurčitelné geny a byly ponechány jako nerozhodnuté.

Bakteriální taxony ve fylogenetických stromech těchto HGT-neurčitelných genů nevytvářejí očekávané (tedy známé) fylogenetické vztahy. Uspořádání těchto taxonů je možné interpretovat buď jako důsledek nedostatku fylogenetické informace, nebo jako důsledek velkého množství přenosů daného genu mezi fylogeneticky příbuznými skupinami bakterií. Konkrétním příkladem může být alignment a fylogenetický strom genu pro translation initiation factor 1 (viz Obr. 4 a 5). V alignmentu tohoto genu je vidět, že se mezi sebou sekvence liší jen malým počtem změn. Ve výsledném fylogenetickém stromu vytvořeném pro tento gen je možné pozorovat polytomické uspořádání pěti skupin. Dalším příkladem genu, u kterého nebylo možné s jistotou prokázat horizontální přenos, je gen pro gen phospholipase D (viz Obr. 6). Fylogenetický strom tohoto genu do jisté míry naznačuje možnost horizontálního přenosu. Avšak vzhledem k postavení bakterie rodu *Proteus* mimo monofyletickou skupinu tvořenou bakteriemi rodu *Arsenophonus* a *Providencia* není možné tento fylogenetický strom brát jako jasný důkaz horizontálního přenosu. Ve fylogenetickém stromu tohoto genu navíc tvoří bakterie *Candidatus Erwinia haradaeae*, bakterie *A.\_O.b.* a dichotomická skupina tvořená bakteriemi *Pantoea* sp. a *Serratia symbiotica* polytomii. To znamená, že fylogenetická analýza nebyla schopna vyřešit přesné příbuznosti mezi geny těchto bakterií. Proto byl tento gen zařazen do skupiny HGT-neurčitelných genů.

V některých fylogenetických stromech genů byly pozorovány extrémně dlouhé větve, u kterých bývá obtížné rozhodnout jejich přesné fylogenetické postavení z důvodu výskytu metodického artefaktu, tzv. long branch attraction (LBA), který chybně seskupuje dvě či více dlouhých fylogenetických větví jako sesterské skupiny (Bergsten, 2005). Tento artefakt je například typický u symbiotických zástupců z Enterobacteriaceae, kteří tvoří dlouhé větve a sdružují se do zdánlivých monofyletických skupin (Husník et al., 2011). Artefaktům spojených s nesprávným sdružováním dlouhých větví nepříbuzných skupin je možné alespoň částečně předejít použitím správné metodologie (např. metoda maximum likelihood je méně náchylná k LBA, než maximum parsimony; důležitou roli hraje také zvolení správného modelu, který odpovídá co nejpřesněji evoluci dané sekvence). V mých datech byl typickým příkladem dlouhých větví gen pro integration host factor alpha (viz Obr. 6 a 7), jehož sekvence byla v rámci alignmentu silně pozměněna u dvou symbiotických zástupců, bakterie *Buchnera*



*aphidicola* z hostitele *Brevicoryne brassicae* a bakterie *A.\_O.b.* Zatímco jednoduchý evoluční model LG (defaultní model programu PhyML 3.3) by klastroval tyto sekvence do monofyletické skupiny (viz Obr. 6), správně zvolený, velmi odvozený evoluční model (MtMam+G+I+F, určený pomocí SMS PhyML) zařadil každou sekvenci do skupiny jejich předpokládaného původu (viz Obr. 7). Tuto topologii lze považovat za poměrně přesvědčivou, protože nepodléhá přitahování dvou extrémně dlouhých větví, ale odpovídá předpokládaným fylogenetickým vztahům (tj. nulové hypotéze).

Poslední skupinou genů byly geny, jejichž fylogenetické vztahy silně podporovaly možnost HGT těchto genů. V případě bakterie *A.\_O.b.* bylo v genomu objeveno 14 potenciálních HGT genů. V genomu bakterie *A.\_C.p.* bylo objeveno 50 potenciálních HGT genů. Jednotlivé skupiny genů byly zaneseny do tabulky 5 a dále budou diskutovány jednotlivé případy HGT genů.

### 5.3.2.1 HGT geny z bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba*

Z fylogenetického stromu genu pro pantoate-beta-alanine ligase z *A.\_O.b.* (viz Obr. 8) je patrné, že tento gen byl horizontálně přenesen, jelikož netvoří monofyletickou skupinu s bakterií rodu *Arsenophonus* ani jejími nejbližšími příbuznými. Geny z bakterií rodu *Xenorhabdus*, *Proteus* a *Riesia* tvoří monofyletickou skupinu a i ostatní bakterie ve fylogenetickém stromu genu pro pantoate-beta-alanine ligase z *A.\_O.b.* tvoří příbuzenské vztahy odpovídající současné klasifikaci. Příbuznost genů v monofyletické skupině, kterou tvoří gen pro pantoate-beta-alanine ligase z *A.\_O.b.* a gen z endosymbionta z hostitele *Heteropsylla cubana*, podporuje poměrně dlouhá společná větev.

Z fylogenetického stromu genu pro 3-methyl-2-oxobutanoate hydroxymetyltransferase z *A.\_O.b.* (viz Obr. 9) je patrné, že tento gen byl horizontálně přenesen. Geny z bakterií rodu *Xenorhabdus*, *Proteus* a *Providencia* tvoří monofyletickou skupinu, avšak gen pro 3-methyl-2-oxobutanoate hydroxymetyltransferase z *A.\_O.b.* nepatří do této monofyletické skupiny. Tento gen spadá do monofyletické skupiny společně s bakteriemi rodu *Sodalis*. U bakterie rodu *Sodalis* byl již dříve pozorován symbiotický vztah se zástupci čeledi klošovitých (Hippoboscidae; Nováková and Hypša, 2007). Horizontální přenos genu tedy mohl být reálně uskutečněn v jednom ze zástupců čeledi klošovitých.

Fylogenetický strom genu pro 2-dehydropantoate 2-reductase z *A.\_O.b.* (viz Obr. 10) naznačuje horizontální přenos. Avšak bakterie rodu *Proteus* netvoří monofyletickou skupinu s bakteriemi rodu *Xenorhabdus* a *Providencia*. Z toho důvodu nelze tento strom považovat za spolehlivý důkaz horizontálního přenosu tohoto genu, přestože ostatní bakteriální skupiny ve

fylogenetickém stromu genu pro 2-dehydropanoate 2-reductase z *A.\_O.b.* tvoří příbuzenské vztahy odpovídající současné klasifikaci.

### 5.3.2.2 HGT geny z bakterie rodu *Arsenophonus* z kloše *Crataerina pallida*

Gen pro YcfA family protein (Obr. 12) z *A.\_C.p.* není pro bakterii nezbytně nutný a objevuje se jen u některých druhů bakterií. Pro tento gen nebyly nalezeny pomocí blastu žádné ortology bakterie rodu *Arsenophonus* ani jejích blízké příbuzných bakterií rodu *Proteus*, *Providencia*, *Xenorhabdus*, *Photorhabdus*, *Riesia*. Již z absence ortologů u bakterií rodu *Arsenophonus* a jejích blízkých příbuzných lze usuzovat možný druhotný zisk tohoto genu horizontálním přenosem. Avšak z topologie fylogenetického stromu genu pro YcfA family protein (Obr. 12) z *A.\_C.p.* nelze jednoznačně potvrdit horizontální přenos tohoto genu, jelikož bakterie *Enterobacter cloacae*, dichotomická skupina tvořená bakterií *Pantoea stewartii* subsp. *stewartii* a dichotomická skupina tvořená bakterií *A.\_C.p.* a bakterií *Enterobacter hormaaechei*, tvoří polytomickou skupinu, z čehož nelze určit přesné příbuznosti mezi těmito bakteriemi.

Z fylogenetického stromu genu pro restriction endonuclease, subunit S (viz Obr. 13) z *A.\_C.p.* lze usuzovat horizontální přenos tohoto genu ze dvou důvodů: i) bakterie *Providencia stuartii* a *Proteus mirabilis*, blízké příbuzní bakterie rodu *Arsenophonus*, tvoří monofyletickou skupinu, ii) ostatní bakterie tvoří smysluplné příbuzenské vztahy. Gen z *A.\_C.p.* je nejvíce příbuzný genům z bakterií *Pectobacterium carotovorum* a *Dickeya zeae* (Gammaproteobacteria).

## 5.4 Metabolické dráhy B vitamínů a kofaktorů

Při rekonstrukci vybraných metabolických drah B vitamínů a kofaktorů byla mezi čtyřmi genomy bakterií rodu *Arsenophonus* porovnávána kompletnost a z toho vyplývající předpokládaná funkčnost těchto vybraných drah. Výsledky ukázaly, že na kompletnost a funkčnost metabolických drah má vliv velikost genomu a množství genů v genomu bakterie. Čím větší bakteriální genom je, tím větší počet genů obsahuje a tím větší počet metabolických drah je kompletní a pravděpodobně funkční. Genom bakterie *A.\_C.p.* je největším z porovnávaných genomů, obsahuje největší množství genů a také má největší počet kompletních metabolických drah, což je v souladu s nejmenší mírou redukce a degradace genomu této bakterie s ohledem na pravděpodobně nejnižší míru evolučního přizpůsobení bakterie ke svému hostiteli. Se zmenšujícím se genomem a zmenšováním množství genů se postupně snižuje počet kompletních a funkčních metabolických drah u bakterie *Arsenophonus*

*melophagi*, *Candidatus Arsenophonus lipopteni* i *A.\_O.b.* V případě genomu bakterie *A.\_O.b.* je možné pozorovat až přibližně čtyřnásobně menší velikost genomu a počet genů než u bakterie *A.\_C.p.* Přesto má bakterie *A.\_O.b.* funkční pravděpodobně téměř polovinu z vybraných metabolických drah pro esenciální B vitamíny a kofaktory, konkrétně 4, což je o 3 metabolické dráhy méně než bakterie *A.\_C.p.* V případě genů zapojených do metabolických drah B vitamínů a kofaktorů proběhla menší redukce než v případě celkového genomu. Hlavním důvodem bude pravděpodobně pozitivní selekce genů, při které dochází k zachování prospěšných genů na úkor méně potřebných genů. Selektce genů není jediným faktorem ovlivňující přítomnost genů v genomu bakterie. V některých případech může být ztráta genů kompenzována horizontálním přenosem genů (Boto, 2010). Konkrétní příklad horizontálně přenesených genů zapojených do metabolických drah pro B vitamíny a kofaktory u bakterie *A.\_O.b.* je možné pozorovat v metabolické dráze pro kyselinu pantotenovou, ve které jsou tři geny tvořící část této dráhy horizontálně přenesené z jiných bakterií. Konkrétně gen pro 3-methyl-2-oxobutanoate hydroxymethyltransferase je s největší pravděpodobností horizontálně přenesen z bakterie rodu *Sodalis* (viz Obr. 9). K přenosu mohlo dojít v některém ze zástupců čeledi klošovitých (Hippoboscidae), jelikož bakterie rodu *Sodalis* byla již dříve v klošovitých pozorována. Horizontální přenos genů, konkrétně celého operonu pro biotin, byl pozorován například také u bakterie *Legionella polyplacis* ze vši *Polyplax serrata* (čeleď Polyplacidae). Horizontální přenos biotinového operonu byl pravděpodobně i faktorem, který změnil životní strategii bakterie rodu *Legionella* z intracelulárního symbionta améb a původce závažných lidských onemocnění na obligátního symbionta vši *Polyplax serrata* a *P. spinulosa* (Říhová et al., 2017). Dalším příkladem změny životní strategie bakterie horizontálním přenosem genů je bakterie rodu *Wolbachia*, známá především jako parazitická bakterie hmyzu ovlivňující reprodukční systém svého hostitele. Bakterie rodu *Wolbachia* se získáním biotinového operonu horizontálním přenosem stala nutričním symbiontem štěnice *Cimex lectularis* (Gerth and Bleidorn, 2017; Nikoh et al., 2014). U krevsajícího hmyzu je nedostatek B vitamínů a kofaktorů kompenzován symbiotickými bakteriemi (Hosokawa et al., 2010; Nikoh et al., 2014; Nogge, 1981; Rio et al., 2016).

Metabolická dráha pro protoheme IX je kompletní pouze u bakterie *Arsenophonus melophagi* a téměř kompletní u bakterie *A.\_C.p.*, jelikož postrádá jeden gen (vysvětleno níže). Tyto dvě bakterie mají ze čtyř porovnávaných bakterií největší genomy a největší množství genů, což značí menší redukci a degradaci genomu a tím pádem i menší evoluční přizpůsobení ke svému hostiteli. Silně redukovaní obligátní symbionté v tomto případě konkrétně bakterie *A.\_O.b.* a *Candidatus Arsenophonus lipopteni* nejsou schopni protoheme IX produkovat.

Ztrátu schopnosti produkovat protoheme IX z glutamátu bakteriemi *A.\_O.b.* a *Candidatus Arsenophonus lipopteni* v průběhu soužití s hostitelem lze usuzovat z přítomnosti pouze jednoho genu z deseti v genomu bakterie *A.\_O.b.* a tří genů v genomu bakterie *Candidatus Arsenophonus lipopteni* pro metabolickou dráhu pro protoheme IX.

Metabolická dráha pro produkci ubiquinolu z glycerinaldehyd-3-phosphate je kompletní a pravděpodobně funkční pro 3 ze 4 porovnávaných bakteriálních genomů. Pouze u bakterie *A.\_O.b.* nebylo možné přímo určit funkčnost této dráhy, jelikož postrádá jeden gen (vysvětleno níže). Bakterie *A.\_C.p.* a *Arsenophonus melophagi* jsou jediné dvě bakterie, které dokážou vytvářet ubiquinol i z chorismátu, což je alternativní způsob produkce ubiquinolu. To je pravděpodobně dáno menší mírou redukce a degradace genomu těchto dvou bakterií, které jsou dvěma největšími porovnávanými bakteriemi, a s tím spojeným největším množstvím genů a tedy i větší komplexností metabolických drah.

V našich výsledcích byla ve dvou případech pozorována absence pouze jediného genu v jinak kompletní a pravděpodobně funkční metabolické dráze, konkrétně se jednalo o metabolickou dráhu pro protoheme IX u bakterie *A.\_C.p.* a o metabolickou dráhu pro ubiquinol u bakterie *A.\_O.b.* Ve dvou případech byla pozorována absence pouze dvou genů v jinak kompletní metabolické dráze, konkrétně v metabolické dráze pro nikotinát a nikotinamid u bakterie *A.\_C.p.* a v metabolické dráze pro pyridoxin u bakterií *Arsenophonus melophagi*, *Candidatus Arsenophonus lipopteni* a *A.\_O.b.* Přítomnost těchto chybějících genů v genomech bakterie *A.\_O.b.* a *A.\_C.p.* byla prověřována pomocí nástroje blastn, žádný z nich však touto metodou nebyl nalezen. Ve všech případech byly tedy tyto geny pravděpodobně ztraceny při redukci genomu, kterému podléhají tyto bakterie ve svém hostiteli. V případě absence jednoho či dvou genů v metabolické dráze by mohlo dojít ke kompenzaci tohoto nedostatku ze strany hostitele nebo jiné bakterie. Ke kompenzaci by mohlo dojít horizontálním přenosem tohoto genu, jak již bylo zmíněno výše, nebo by mohl být gen přítomen v genomu hostitele.

Výsledek rekonstrukce metabolické dráhy pro thiamin potvrzuje, že symbiotické bakterie většinou neprodukují tento vitamín pro svého hostitele, jelikož pro 3 ze 4 porovnávaných bakterií je metabolická dráha pro thiamin nekompletní a s největší pravděpodobností nefunkční. Pouze jediná bakterie *A.\_C.p.* měla ve svém genomu geny potřebné pro produkci thiaminu. To může být dáno kratší dobou soužití této bakterie s hostitelem, během které nedošlo k redukci a degradaci genomu v takové míře, aby se dráha pro thiamin stala nefunkční. Všechny porovnávané genomy však obsahovaly gen pro ABC transportér thiaminu, což může znamenat, že tyto symbiotické bakterie získávají thiamin od

svého hostitele. Stejná situace s metabolickou dráhou pro thiamin byla pozorována i u bakterie rodu *Riesia* nebo bakterie rodu *Legionella*, jejichž genom obsahoval z genů pro metabolickou dráhu thiaminu pouze ABC transportér pro thiamin (Husník, 2018; Říhová et al., 2017). Druhým možným vysvětlením může být produkce thiaminu jinou symbiotickou bakterií. Tento jev byl pozorován u bakterie rodu *Wigglesworthia*, která poskytovala thiamin nejen hostiteli, mouše tsetse (čeleď Glossinidae), ale také bakterii rodu *Sodalis*, která neměla funkční metabolickou dráhu pro thiamin (Rio et al., 2012).

Stejně jako bakterie rodu *Riesia*, *Wolbachia* sp. Cle a rodu *Wigglesworthia* mají i všechny 4 porovnávané bakterie rodu *Arsenophonus* kompletní a tedy pravděpodobně funkční metabolické dráhy pro riboflavin a folát s kyselinou listovou (Husník, 2018).

Metabolická dráha pro pyridoxin není kompletní u žádné z porovnávaných bakterií. Nejkompletnější metabolickou dráhu pro pyridoxin vlastní bakterie *A.\_C.p.*, které chybí pouze dva geny, což pravděpodobně souvisí s velikostí genomu a množstvím genů této bakterie.

## 6 Závěr

Výskyt symbiotických bakterií rodu *Arsenophonus* byl identifikován v metagenomických datech dvou zástupců krevsajícího hmyzu z čeledi Hippoboscidae, konkrétně *Ornithomya biloba* a *Crataerina pallida*. Genomy obou bakterií rodu *Arsenophonus* nebylo možné z důvodu komplexnosti metagenomických dat složit do podoby kompletních uzavřených genomů. Obě bakterie rodu *Arsenophonus* byly zkoumány a porovnány z hlediska genomických charakteristik, na jejichž základě byl odvozen stupeň jejich genomické modifikace, evoluce a přizpůsobení k životu v hostiteli. Genom bakterie *Arsenophonus* z kloše *Ornithomya biloba* (0,68 Mbp, GC obsah 24,7%) byl redukovánější než genom bakterie *Arsenophonus* z kloše *Crataerina pallida* (2,41 Mbp, GC obsah 38,3%). Obě bakterie rodu *Arsenophonus* jsou pro svého hostitele s největší pravděpodobností obligátními symbionty. U obou zástupců bakterie rodu *Arsenophonus* byla testována a v některých případech i potvrzena přítomnost horizontálně přenesených genů. Dále byla v této práci porovnávána kompletnost a funkčnost jednotlivých metabolických drah esenciálních B vitamínů a kofaktorů mezi čtyřmi zástupci rodu *Arsenophonus* žijících v symbiotickém vztahu s různými druhy čeledi Hippoboscidae. Výsledky ukazují kompletnost a pravděpodobnou funkčnost poměrně velkého množství metabolických drah B vitamínů a kofaktorů i přes silnou redukci genomů bakterií rodu *Arsenophonus*.

## 7 Použitá literatura

- Aksoy, S. (1995). *Wigglesworthia* gen. nov. and *Wigglesworthia glossinidia* sp. nov., taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse flies. *International Journal of Systematic and Evolutionary Microbiology*, 45(4), 848-851.
- Allen, J. M., Reed, D. L., Perotti, M. A., & Braig, H. R. (2007). Evolutionary relationships of “*Candidatus* Riesia spp.,” endosymbiotic Enterobacteriaceae living within hematophagous primate lice. *Applied. Environmental. Microbiology.*, 73(5), 1659-1664.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.
- Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F., & Ménard, R. (2006). Quantitative imaging of Plasmodium transmission from mosquito to mammal. *Nature medicine*, 12(2), 220-224.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
- Askew, R. R. (1971). Parasitic insects. *Parasitic insects*.
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., & Meyer, F. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9(1), 75.
- Babendreier, D., Joller, D., Romeis, J., Bigler, F., & Widmer, F. (2007). Bacterial community structures in honeybee intestines and their response to two insecticidal proteins. *FEMS microbiology ecology*, 59(3), 600-610.
- Babor, M., Sobolev, V., & Edelman, M. (2002). Conserved positions for ribose recognition: importance of water bridging interactions among ATP, ADP and FAD-protein complexes. *Journal of molecular biology*, 323(3), 523-532.
- Baker, J. R. (1967). A review of the role played by the Hippoboscidae (Diptera) as vectors of endoparasites. *The Journal of parasitology*, 412-418.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology: A journal of computational molecular cell biology*, 19(5), 455-477.

- Baureder, M., & Hederstedt, L. (2013). Heme proteins in lactic acid bacteria. *In Advances in microbial physiology* (Vol. 62, pp. 1-43). Academic Press.
- Bender, D. A. (2003). *Nutritional biochemistry of the vitamins*. Cambridge university press.
- Bentley, S. D., & Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Annual Review of Genetics.*, 38, 771-791.
- Berg, O. G., & Kurland, C. G. (2002). Evolution of microbial genomes: sequence acquisition and loss. *Molecular biology and evolution*, 19(12), 2265-2276.
- Bergsten, J. (2005). A review of long-branch attraction. *Cladistics*, 21(2), 163-193.
- Bhagavan, H. N., & Chopra, R. K. (2006). Coenzyme Q10: absorption, tissue uptake, metabolism and pharmacokinetics. *Free radical research*, 40(5), 445-453.
- Binnewies, T. T., Motro, Y., Hallin, P. F., Lund, O., Dunn, D., La, T., Hampson, D. J., Bellgard, M., Wassenaar, T. M. & Ussery, D. W. (2006). Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Functional & integrative genomics*, 6(3), 165-185.
- Blatch, S., Meyer, K. W., & Harrison, J. F. (2010). Effects of dietary folic acid level and symbiotic folate production on fitness and development in the fruit fly *Drosophila melanogaster*. *Fly*, 4(4), 312-319.
- Bordenstein, S. R., & Reznikoff, W. S. (2005). Mobile DNA in obligate intracellular bacteria. *Nature Reviews Microbiology*, 3(9), 688-699.
- Boto, L. (2010). Horizontal gene transfer in evolution: facts and challenges. *Proceedings of the Royal Society B: Biological Sciences*, 277(1683), 819-827.
- Boyd, B. M., Allen, J. M., de Crécy-Lagard, V., & Reed, D. L. (2014). Genome sequence of *Candidatus* Riesia pediculischaeffi, endosymbiont of chimpanzee lice, and genomic comparison of recently acquired endosymbionts from human and chimpanzee lice. *G3: Genes, Genomes, Genetics*, 4(11), 2189-2195.
- Boyd, B. M., Allen, J. M., Koga, R., Fukatsu, T., Sweet, A. D., Johnson, K. P., & Reed, D. L. (2016). Two bacterial genera, *Sodalis* and *Rickettsia*, associated with the seal louse *Proechinophthirus fluctus* (Phthiraptera: Anoplura). *Applied and Environmental Microbiology*, 82(11), 3185-3197.
- Braendle, C., Miura, T., Bickel, R., Shingleton, A. W., Kambhampati, S., & Stern, D. L. (2003). Developmental origin and evolution of bacteriocytes in the aphid–*Buchnera* symbiosis. *PLoS Biology*, 1(1).
- Catresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17, 540-552.



- Coil, D., Jospin, G., Darling, A. (2015). A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics*, 31(4), 587–589.
- Cronan J. E., Jr (1980). Beta-alanine synthesis in *Escherichia coli*. *Journal of Bacteriology*, 141(3), 1291–1297.
- Dale, C., & Maudlin, I. (1999). *Sodalis* gen. nov. and *Sodalis glossinidius* sp. nov., a microaerophilic secondary endosymbiont of the tsetse fly *Glossina morsitans morsitans*. *International Journal of Systematic and Evolutionary Microbiology*, 49(1), 267-275.
- Dale, C., & Moran, N. A. (2006). Molecular interactions between bacterial symbionts and their hosts. *Cell*, 126(3), 453-465.
- Dale, C., & Welburn, S. C. (2001). The endosymbionts of tsetse flies: manipulating host–parasite interactions. *International Journal for Parasitology*, 31(5-6), 628-631.
- Darby, A. C., Choi, J. H., Wilkes, T., Hughes, M. A., Werren, J. H., Hurst, G. D. D., & Colbourne, J. K. (2010). Characteristics of the genome of *Arsenophonus nasoniae*, son-killer bacterium of the wasp *Nasonia*. *Insect Molecular Biology*, 19, 75-89.
- Darmon, E., & Leach, D. R. (2014). Bacterial genome instability. *Microbiology and Molecular Biology Reviews.*, 78(1), 1-39.
- Dehio, C., Sauder, U., & Hiestand, R. (2004). Isolation of *Bartonella schoenbuchensis* from *Lipoptena cervi*, a blood-sucking arthropod causing deer ked dermatitis. *Journal of Clinical Microbiology*, 42(11), 5320-5323.
- Dick, C. W. (2006). Checklist of world Hippoboscidae (Diptera: Hippoboscoidea). *Department of Zoology, Field Museum Natural History, Chicago, IL*.
- Diederer, B. M. W. (2008). *Legionella* spp. and Legionnaires' disease. *Journal of Infection*, 56(1), 1-12.
- Douglas, A. E. (1998). Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. *Annual Review of Entomology*, 43(1), 17-37.
- Droop, A. P. (2016). fqtools: an efficient software suite for modern FASTQ file manipulation. *Bioinformatics*, 32(12), 1883-1884.
- Ferrari, J., Darby, A. C., Daniell, T. J., Godfray, H. C. J., & Douglas, A. E. (2004). Linking the bacterial community in pea aphids with host-plant use and natural enemy resistance. *Ecological Entomology*, 29(1), 60-65.
- Foster, W. A. (1995). Mosquito sugar feeding and reproductive energetics. *Annual review of entomology*, 40(1), 443-474.

- Geiger, A., Ravel, S., Mateille, T., Janelle, J., Patrel, D., Cuny, G., & Frutos, R. (2007). Vector competence of *Glossina palpalis* gambiensis for *Trypanosoma brucei* sl and genetic diversity of the symbiont *Sodalis glossinidius*. *Molecular Biology and Evolution*, 24(1), 102-109.
- Gerth, M., & Bleidorn, C. (2016). Comparative genomics provides a timeframe for *Wolbachia* evolution and exposes a recent biotin synthesis operon transfer. *Nature Microbiology*, 2(3), 1-7.
- Gherna, R. L., Werren, J. H., Weisburg, W., Cote, R., Woese, C. R., Mandelco, L., & Brenner, D. J. (1991). *Arsenophonus nasoniae* gen. nov., sp. nov., the causative agent of the son-killer trait in the parasitic wasp *Nasonia vitripennis*. *International Journal of Systematic and Evolutionary Microbiology*, 41(4), 563-565.
- Gosalbes, M. J., Latorre, A., Lamelas, A., & Moya, A. (2010). Genomics of intracellular symbionts in insects. *International Journal of Medical Microbiology*, 300(5), 271-278.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), 307-321.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), 307-321.
- Gündüz, E. A., & Douglas, A. E. (2009). Symbiotic bacteria enable insect to use a nutritionally inadequate diet. *Proceedings of the Royal Society B: Biological Sciences*, 276(1658), 987-991.
- Hildebrand, F., Meyer, A., & Eyre-Walker, A. (2010). Evidence of selection upon genomic GC-content in bacteria. *PLoS genetics*, 6(9).
- Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X. Y., & Fukatsu, T. (2010). *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proceedings of the National Academy of Sciences*, 107(2), 769-774.
- Hughes, G. L., Dodson, B. L., Johnson, R. M., Murdock, C. C., Tsujimoto, H., Suzuki, Y., Patt, A. A., Cui, L., Nossa, C. W., Barry, R. M., Sakamoto, J. M., Hornett, E. A., & Rasgon, J. L. (2014). Native microbiome impedes vertical transmission of *Wolbachia* in *Anopheles* mosquitoes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(34), 12498–12503.
- Husník, F. (2018). Host–symbiont–pathogen interactions in blood-feeding parasites: nutrition, immune cross-talk and gene exchange. *Parasitology*, 145(10), 1294-1303.

- Husník, F., Chrudimský, T., & Hypša, V. (2011). Multiple origins of endosymbiosis within the Enterobacteriaceae ( $\gamma$ -Proteobacteria): convergence of complex phylogenetic approaches. *BMC biology*, 9(1), 87.
- Hypša, V., & Dale, C. (1997). In vitro culture and phylogenetic analysis of “*Candidatus Arsenophonus triatominarum*,” an intracellular bacterium from the triatomine bug, *Triatoma infestans*. *International Journal of Systematic and Evolutionary Microbiology*, 47(4), 1140-1144.
- 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.
- Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059-3066.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Ashton, B., Meintjes, P., Drummond, A., & Thierer, T. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Kennedy, P. G. (2013). Clinical features, diagnosis, and treatment of human African trypanosomiasis (sleeping sickness). *The Lancet Neurology*, 12(2), 186-194.
- Kikuchi, Y. (2009). Endosymbiotic bacteria in insects: their diversity and culturability. *Microbes and Environments*, 0908180109-0908180109.
- Kim, J. S., DeGiovanni, A., Jancarik, J., Adams, P. D., Yokota, H., Kim, R., & Kim, S. H. (2005). Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proceedings of the National Academy of Sciences*, 102(9), 3248-3253.
- Knowles, J. R. (1989). The mechanism of biotin-dependent enzymes. *Annual Review of Biochemistry*, 58(1), 195-221.
- Knowles, T. J., Scott-Tucker, A., Overduin, M., & Henderson, I. R. (2009). Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nature Reviews Microbiology*, 7(3), 206-214.
- Krafsur, E. S. (2009). Tsetse flies: genetics, evolution, and role as vectors. *Infection, Genetics and Evolution*, 9(1), 124-141.
- Krishnan, M., Bharathiraja, C., Pandiarajan, J., Prasanna, V. A., Rajendhran, J., & Gunasekaran, P. (2014). Insect gut microbiome—An unexploited reserve for biotechnological application. *Asian Pacific Journal of Tropical Biomedicine*, 4, S16-S21.

- Lawrence, J. G., & Roth, J. R. (1996). Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics*, *143*(4), 1843-1860.
- Lefort, V., Longueville, J. E., & Gascuel, O. (2017). SMS: smart model selection in PhyML. *Molecular Biology and Evolution*, *34*(9), 2422-2424.
- Lehane, M. J. (2005). *The biology of blood-sucking in insects* Cambridge University Press. *New York*.
- Lin, S., & Cronan, J. E. (2011). Closing in on complete pathways of biotin biosynthesis. *Molecular BioSystems*, *7*(6), 1811-1821.
- Lynen, F. (1967). The role of biotin-dependent carboxylations in biosynthetic reactions. *Biochemical Journal*, *102*(2), 381.
- Ma, Y., Bao, Y., Wang, S., Li, T., Chang, X., Yang, G., & Meng, X. (2016). Anti-inflammation effects and potential mechanism of saikosaponins by regulating nicotinate and nicotinamide metabolism and arachidonic acid metabolism. *Inflammation*, *39*(4), 1453-1461.
- Maia Da Silva, F., Junqueira, A. C. V., Campaner, M., Rodrigues, A. C., Crisante, G., Ramirez, L. E., Caballero, Z. C. E., Monteiro, F. A., Coura, J. R., Añez, N. & Teixeira, M. M. G. (2007). Comparative phylogeography of *Trypanosoma rangeli* and *Rhodnius* (Hemiptera: Reduviidae) supports a long coexistence of parasite lineages and their sympatric vectors. *Molecular Ecology*, *16*(16), 3361-3373.
- McCutcheon, J. P., & Moran, N. A. (2010). Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biology and Evolution*, *2*, 708-718.
- McCutcheon, J. P., & Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, *10*(1), 13-26.
- McDowell, L. R. (2000). Reevaluation of the metabolic essentiality of the vitamins-review. *Asian-Australasian Journal of Animal Sciences*, *13*(1), 115-125.
- Miller, S. L., & Schlesinger, G. (1993). Prebiotic syntheses of vitamin coenzymes: II. Pantoic acid, pantothenic acid, and the composition of coenzyme A. *Journal of Molecular Evolution*, *36*(4), 308-314.
- Moran, N. A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. *Cell*, *108*(5), 583-586.
- Moran, N. A., Degnan, P. H., Santos, S. R., Dunbar, H. E., & Ochman, H. (2005). The players in a mutualistic symbiosis: insects, bacteria, viruses, and virulence genes. *Proceedings of the National Academy of Sciences*, *102*(47), 16919-16926.
- Moran, N. A., McCutcheon, J. P., & Nakabachi, A. (2008). Evolution and genomics of heritable bacterial symbionts. *Annual Review of Genetics*, *42*, 165-190.

- Moran, N. A., Munson, M. A., Baumann, P., & Ishikawa, H. (1993). A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 253(1337), 167-171.
- Mustonen, A. M., Käkälä, R., Paakkonen, T., & Nieminen, P. (2015). Life stage-related differences in fatty acid composition of an obligate ectoparasite, the deer ked (*Lipoptena cervi*)—Influence of blood meals and gender. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 179, 62-70.
- 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. (1981). Significance of symbionts for the maintenance of an optional nutritional state for successful reproduction in hematophagous insects. *Parasitology*, 82.
- Nováková, E., & Hypša, V. (2007). A new *Sodalis* lineage from bloodsucking fly *Craterina melbae* (Diptera, Hippoboscoidea) originated independently of the tsetse flies symbiont *Sodalis glossinidius*. *FEMS microbiology letters*, 269(1), 131-135.
- Nováková, E., Husník, F., Šochová, E., & Hypša, V. (2015). *Arsenophonus* and *Sodalis* symbionts in louse flies: an analogy to the *Wigglesworthia* and *Sodalis* system in tsetse flies. *Applied and Environmental Microbiology*, 81(18), 6189-6199.
- Nováková, E., Hypša, V., Nguyen, P., Husník, F., & Darby, A. C. (2016). Genome sequence of *Candidatus Arsenophonus lipopteni*, the exclusive symbiont of a blood sucking fly *Lipoptena cervi* (Diptera: Hippoboscidae). *Standards in Genomic Sciences*, 11(1), 72.
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, 27(5), 824-834.
- Ochman, H., Lawrence, J. G., & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405(6784), 299.
- Oliver, K. M., Degnan, P. H., Hunter, M. S., & Moran, N. A. (2009). Bacteriophages encode factors required for protection in a symbiotic mutualism. *Science*, 325(5943), 992-994.
- Pais, R., Lohs, C., Wu, Y., Wang, J., & Aksoy, S. (2008). The obligate mutualist *Wigglesworthia glossinidia* influences reproduction, digestion, and immunity processes of its host, the tsetse fly. *Applied and Environmental Microbiology*, 74(19), 5965-5974.
- Parker, B. J., Spragg, C. J., Altincicek, B., & Gerardo, N. M. (2013). Symbiont-mediated protection against fungal pathogens in pea aphids: a role for pathogen specificity. *Applied and Environmental Microbiology*, 79(7), 2455-2458.
- Petersen, F. T., Meier, R., Kutty, S. N., & Wiegmann, B. M. (2007). The phylogeny and evolution of host choice in the Hippoboscoidea (Diptera) as reconstructed using four molecular markers. *Molecular Phylogenetics and Evolution*, 45(1), 111-122.

- Piel, J., Höfer, I., & Hui, D. (2004). Evidence for a symbiosis island involved in horizontal acquisition of pederin biosynthetic capabilities by the bacterial symbiont of *Paederus fuscipes* beetles. *Journal of Bacteriology*, 186(5), 1280-1286.
- Reed, K. E., & Cronan, J. E. (1993). Lipoic acid metabolism in *Escherichia coli*: sequencing and functional characterization of the lipA and lipB genes. *Journal of Bacteriology*, 175(5), 1325-1336.
- Rio, R. V., Attardo, G. M., & Weiss, B. L. (2016). Grandeur alliances: symbiont metabolic integration and obligate arthropod hematophagy. *Trends in Parasitology*, 32(9), 739-749.
- Rio, R. V., Symula, R. E., Wang, J., Lohs, C., Wu, Y. N., Snyder, A. K., Bjornson, R. D., Oshima, K., Biehl, B. S., Perna, N. T., Aksoy, S. & Hattori, M. (2012). Insight into the transmission biology and species-specific functional capabilities of tsetse (Diptera: Glossinidae) obligate symbiont *Wigglesworthia*. *MBio*, 3(1), e00240-11.
- Ruby, E. G. (2008). Symbiotic conversations are revealed under genetic interrogation. *Nature Reviews Microbiology*, 6(10), 752-762.
- Říhová, J., Batani, G., Rodríguez-Ruano, S. M., Martinů, J., Nováková, E., & Hypša, V. (2019). A new symbiotic lineage related to *Neisseria* and *Snodgrassella* arises from the dynamic and diverse microbiomes in sucking lice. *bioRxiv*, 867275.
- Říhová, J., Nováková, E., Husník, F., & Hypša, V. (2017). Legionella becoming a mutualist: Adaptive processes shaping the genome of symbiont in the louse *Polyplax serrata*. *Genome Biology and Evolution*, 9(11), 2946-2957.
- Sanchez, J. E., Gross, P. G., Goetze, R. W., Walsh Jr, R. M., Peeples, W. B., & Wood, Z. A. (2015). Evidence of kinetic cooperativity in dimeric ketopantoate reductase from *Staphylococcus aureus*. *Biochemistry*, 54(21), 3360-3369.
- Sannino, D. R., Dobson, A. J., Edwards, K., Angert, E. R., & Buchon, N. (2018). The *Drosophila melanogaster* gut microbiota provisions thiamine to its host. *MBio*, 9(2), e00155-18.
- Shaw, M. K., & Mooloo, S. K. (1991). Comparative study on *Rickettsia*-like organisms in the midgut epithelial cells of different *Glossina* species. *Parasitology*, 102(2), 193-199.
- Sinkins, S. P. (2013). *Wolbachia* and arbovirus inhibition in mosquitoes. *Future Microbiology*, 8(10), 1249-1256.
- Sinkins, S. P., Braig, H. R., & O'Neill, S. L. (1995). *Wolbachia* superinfections and the expression of cytoplasmic incompatibility. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 261(1362), 325-330.
- Small, R. W. (2005). A review of *Melophagus ovinus* (L.), the sheep ked. *Veterinary Parasitology*, 130(1-2), 141-155.

- Stouthamer, R., Breeuwer, J. A., & Hurst, G. D. (1999). *Wolbachia pipientis*: microbial manipulator of arthropod reproduction. *Annual Reviews in Microbiology*, 53(1), 71-102.
- Šochová, E., Husník, F., Nováková, E., Halajian, A., & Hypša, V. (2017). *Arsenophonus* and *Sodalis* replacements shape evolution of symbiosis in louse flies. *PeerJ*, 5, e4099.
- 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.
- Taylor, M. J., Bandi, C., & Hoerauf, A. (2005). *Wolbachia*. Bacterial endosymbionts of filarial nematodes. *Advances in Parasitology*, 60, 245-284.
- Teixeira, L., Ferreira, Á., & Ashburner, M. (2008). The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. *PLoS biology*, 6(12).
- Teller, J. H., Powers, S. G., & Snell, E. E. (1976). Ketopantoate hydroxymethyltransferase. I. Purification and role in pantothenate biosynthesis. *Journal of Biological Chemistry*, 251(12), 3780-3785.
- Thomas, C. M., & Nielsen, K. M. (2005). Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, 3(9), 711.
- Trowbridge, R. E., Dittmar, K., & Whiting, M. F. (2006). Identification and phylogenetic analysis of *Arsenophonus*- and *Photorhabdus*-type bacteria from adult Hippoboscidae and Streblidae (Hippoboscoidea). *Journal of Invertebrate Pathology*, 91(1), 64-68.
- Van Den Abbeele, J., Claes, Y., Van Bockstaele, D., Le Ray, D., & Coosemans, M. (1999). *Trypanosoma brucei* spp. development in the tsetse fly: characterization of the post-mesocyclic stages in the foregut and proboscis. *Parasitology*, 118(5), 469-478.
- Van Der Heijden, M. G., Bakker, R., Verwaal, J., Scheublin, T. R., Rutten, M., Van Logtestijn, R., & Staehelin, C. (2006). Symbiotic bacteria as a determinant of plant community structure and plant productivity in dune grassland. *FEMS Microbiology Ecology*, 56(2), 178-187.
- Vitreschak, A. G., Rodionov, D. A., Mironov, A. A., & Gelfand, M. S. (2002). Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Research*, 30(14), 3141-3151.
- Weeks, A. R., Marec, F., & Breeuwer, J. A. (2001). A mite species that consists entirely of haploid females. *Science*, 292(5526), 2479-2482.
- Weiss, B., & Aksoy, S. (2011). Microbiome influences on insect host vector competence. *Trends in Parasitology*, 27(11), 514-522.
- Wu, D., Daugherty, S. C., Van Aken, S. E., Pai, G. H., Watkins, K. L., Khouri, H., Tallon, L. J., Zaborsky, J. M., Dunbar, H. E., Tran, P. L., Moran, N. A., & Eisen, J. A. (2006).

Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biology*, 4(6).

Zreik, L., Bove, J. M., & Garnier, M. (1998). Phylogenetic characterization of the bacterium-like organism associated with marginal chlorosis of strawberry and proposition of a Candidatus taxon for the organism, 'Candidatus Phlomobacter fragariae'. *International Journal of Systematic and Evolutionary Microbiology*, 48(1), 257-261.



## 8 Přílohy

Tabulka I: Přístupová čísla (accession numbers) sekvencí stažených z NCBI pro ověření HGT genů bakterie rodu *Arsenophonus* z kloše *Ornithomya biloba* pomocí fylogenetických analýz.

| <b>Horizontálně přenesené geny</b>                      |   |
|---|---|
| <b>Pantoate--beta-alanine ligase</b>                    |   |
| WP_010299569  | pantoate--beta-alanine ligase [ <i>Pectobacterium carotovorum</i> ]                     |
| WP_011410460  | pantoate--beta-alanine ligase [ <i>Sodalis glossinidius</i> ]                           |
| WP_013034929.1  | pantoate--beta-alanine ligase [ <i>Candidatus Riesia pediculicola</i> ]                 |
| WP_014888919  | pantoate--beta-alanine ligase [secondary endosymbiont of <i>Heteropsylla cubana</i> ]   |
| WP_049558854  | pantoate--beta-alanine ligase [ <i>Yersinia kristensenii</i> ]                          |
| WP_080626418.1  | pantoate--beta-alanine ligase [ <i>Candidatus Riesia pediculischaeffi</i> ]             |
| WP_086955473.1  | pantoate--beta-alanine ligase [ <i>Xenorhabdus innexi</i> ]                             |
| WP_088493609.1  | pantoate--beta-alanine ligase [ <i>Proteus mirabilis</i> ]                              |
| WP_099074281.1  | pantoate--beta-alanine ligase [ <i>Proteus alimentorum</i> ]                            |
| WP_099113405.1  | pantoate--beta-alanine ligase [ <i>Xenorhabdus miraniensis</i> ]                        |
| WP_101626882  | pantoate--beta-alanine ligase [ <i>Candidatus Baumannia cicadellinicola</i> ]           |
| <b>3-methyl-2-oxobutanoate hydroxymethyltransferase</b> |   |
| WP_000805501.1  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Escherichia coli</i> ]            |
| WP_025423436  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Sodalis praecaptivus</i> ]        |
| WP_034493440  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Buttiauxella agrestis</i> ]       |
| WP_038236150.1  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Xenorhabdus szentirmaii</i> ]     |
| WP_039360885  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Pectobacterium carotovorum</i> ]  |
| WP_042848196.1  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Providencia rettgeri</i> ]        |
| WP_064719126.1  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Proteus hauseri</i> ]             |
| WP_109055838  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Brenneria roseae</i> ]            |
| WP_128179501  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Pantoea</i> sp. LMG 27579]        |
| WP_132926066.1  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Sodalis</i> sp. 159R]             |
| WP_139292432.1  | 3-methyl-2-oxobutanoate hydroxymethyltransferase [ <i>Xenorhabdus thuongxuanensis</i> ] |
| <b>2-dehydropantoate 2-reductase</b>                    |   |
| NZ_CP016176   | 2-dehydropantoate 2-reductase [ <i>Xenorhabdus hominickii</i> ]                         |
| NZ_CP017671   | 2-dehydropantoate 2-reductase [ <i>Providencia rettgeri</i> ]                           |

|   |   |
|---|---|
| WP_016244311.1                            | 2-dehydropantoate 2-reductase [ <i>Escherichia coli</i> ]   |
| WP_021805488                              | 2-dehydropantoate 2-reductase [ <i>Serratia fonticola</i> ]   |
| WP_022632545                              | 2-dehydropantoate 2-reductase [ <i>Dickeya solani</i> ]   |
| WP_025423156                              | 2-dehydropantoate 2-reductase [ <i>Sodalis praecaptivus</i> ]   |
| WP_038498392                              | 2-dehydropantoate 2-reductase [ <i>Candidatus Baumannia cicadellincola</i> ]                          |
| WP_070596451.1                            | 2-dehydropantoate 2-reductase [ <i>Proteus</i> sp. HMSC10D02]   |
| WP_070926789.1                            | 2-dehydropantoate 2-reductase [ <i>Providencia stuartii</i> ]   |
| WP_086111428.1                            | 2-dehydropantoate 2-reductase [ <i>Xenorhabdus beddingii</i> ]  |
| <b>Hypothetical protein</b>               |   |
| CBL47416                                  | hypothetical protein pPAA3_0025 [ <i>Photorhabdus asymbiotica</i> ]                                   |
| PAU99117.1                                | hypothetical protein CBG25_20535 [ <i>Arsenophonus</i> sp. ENCA]                                      |
| WP_010617335                              | hypothetical protein [ <i>Plautia stali</i> symbiont]   |
| WP_081700702.1                            | hypothetical protein [ <i>Arsenophonus nasoniae</i> ]   |
| WP_087757553.1                            | hypothetical protein [ <i>Aeromonas salmonicida</i> ]   |
| WP_100191410                              | hypothetical protein [ <i>Yersinia kristensenii</i> ]   |
| WP_119797778                              | hypothetical protein [ <i>Candidatus Fukatsuia symbiotica</i> ]                                       |
| <b>HGT neurčitelné geny</b>               |   |
| <b>Phospholipase D family protein</b>     |   |
| ASU04962.1                                | Endonuclease [ <i>Proteus mirabilis</i> ]   |
| CUR53405                                  | PLD-like domain-containing protein [ <i>Serratia symbiotica</i> ]                                     |
| VFP83473                                  | PLD-like domain-containing protein [ <i>Candidatus Erwinia haradaeae</i> ]                            |
| WP_011279209                              | phospholipase D family protein [ <i>Sodalis glossinidius</i> ]  |
| WP_026823965.1                            | phospholipase D family protein [ <i>Arsenophonus nasoniae</i> ]                                       |
| WP_036962334.1                            | phospholipase D family protein [ <i>Providencia alcalifaciens</i> ]                                   |
| WP_097324293.1                            | phospholipase D family protein [ <i>Escherichia coli</i> ]  |
| WP_136132114                              | phospholipase D family protein [ <i>Pantoea</i> sp. SoEO]   |
| <b>Translation initiation factor IF-1</b> |   |
| QCI18856.1                                | translation initiation factor IF-1 [ <i>Buchnera aphidicola</i> ( <i>Aphis nerii</i> )]               |
| RLR18379                                  | translation initiation factor IF-1 [ <i>Sodalis</i> -like symbiont of <i>Bactericera trigonica</i> ]  |
| VFP78596                                  | Translation initiation factor IF-1 [ <i>Candidatus Erwinia haradaeae</i> ]                            |
| WP_033552083                              | translation initiation factor IF-1 [ <i>Escherichia coli</i> ]  |
| WP_041855224.1                            | translation initiation factor IF-1 [ <i>Candidatus Riesia pediculicola</i> ]                          |
| WP_069316402.1                            | translation initiation factor IF-1 [ <i>Xenorhabdus hominickii</i> ]                                  |
| WP_096148880                              | translation initiation factor IF-1 [ <i>Enterobacter hormaechei</i> ]                                 |
| WP_115459402.1                            | translation initiation factor IF-1 [ <i>Enterobacillus tribolii</i> ]                                 |
| WP_135091519                              | translation initiation factor IF-1 [ <i>Proteus mirabilis</i> ]                                       |
| WP_136131543                              | translation initiation factor IF-1 [ <i>Pantoea</i> sp. SoET]   |
| <b>ATP synthase subunit C F0F1</b>        |   |
| ECC6117832                                | F0F1 ATP synthase subunit C [ <i>Salmonella enterica</i> ]  |
| KYP97068                                  | ATP F0F1 synthase subunit C [ <i>Sodalis</i> -like endosymbiont of <i>Proechinophthirus fluctus</i> ] |

|  |   |
|--|---|
| WP_001667966.1                             | F0F1 ATP synthase subunit C [ <i>Escherichia coli</i> ]   |
| WP_006890850.1                             | F0F1 ATP synthase subunit C [ <i>Salmonella enterica</i> ]  |
| WP_009873965.1                             | F0F1 ATP synthase subunit C [ <i>Buchnera aphidicola</i> ]  |
| WP_011126301.1                             | F0F1 ATP synthase subunit C [ <i>Candidatus Blochmannia floridanus</i> ]                                |
| WP_012986655.1                             | F0F1 ATP synthase subunit C [ <i>Xenorhabdus bovienii</i> ]   |
| WP_029094800                               | F0F1 ATP synthase subunit C [ <i>Budvicia aquatica</i> ]  |
| WP_036942415.1                             | F0F1 ATP synthase subunit C [ <i>Providencia stuartii</i> ]   |
| WP_042047671.1                             | F0F1 ATP synthase subunit C [ <i>Aeromonas simiae</i> ]   |
| WP_129108803.1                             | F0F1 ATP synthase subunit C [ <i>Arsenophonus</i> endosymbiont of <i>Bemisia tabaci</i> ]               |
| <b>SsrA-binding protein SmpB</b>           |   |
| KLN46507.1                                 | SsrA-binding protein [ <i>Providencia rettgeri</i> ]  |
| PAU99992.1                                 | SsrA-binding protein [ <i>Arsenophonus</i> sp. ENCA]  |
| WP_008916271.1                             | SsrA-binding protein SmpB [ <i>Providencia sneebia</i> ]  |
| WP_026821676.1                             | SsrA-binding protein SmpB [ <i>Arsenophonus nasoniae</i> ]  |
| WP_032114781.1                             | SsrA-binding protein SmpB [ <i>Arsenophonus</i> endosymbiont of <i>Nilaparvata lugens</i> ]             |
| WP_044705885                               | SsrA-binding protein SmpB [ <i>Escherichia coli</i> ]   |
| WP_064720628.1                             | SsrA-binding protein SmpB [ <i>Proteus hauseri</i> ]  |
| WP_086113201.1                             | SsrA-binding protein SmpB [ <i>Xenorhabdus beddingii</i> ]  |
| WP_099118085.1                             | SsrA-binding protein SmpB [ <i>Xenorhabdus ishibashii</i> ]   |
| WP_105466934.1                             | SsrA-binding protein SmpB [ <i>Escherichia coli</i> ]   |
| WP_109407355.1                             | SsrA-binding protein SmpB [ <i>Proteus</i> genomsp. 4]  |
| <b>Geny Arsenophonus</b>                   |   |
| <b>Tetraacyldisaccharide 4'-kinase</b>     |   |
| CBA73144                                   | tetraacyldisaccharide 4'-kinase [ <i>Arsenophonus nasoniae</i> ]  |
| WP_004918145                               | tetraacyldisaccharide 4'-kinase [ <i>Providencia stuartii</i> ]   |
| WP_026822374.1                             | tetraacyldisaccharide 4'-kinase [ <i>Arsenophonus nasoniae</i> ]  |
| WP_036774185                               | tetraacyldisaccharide 4'-kinase [ <i>Photorhabdus australis</i> ]                                       |
| WP_046337195.1                             | tetraacyldisaccharide 4'-kinase [ <i>Xenorhabdus bovienii</i> ]   |
| WP_066284194                               | tetraacyldisaccharide 4'-kinase [ <i>Candidatus Arsenophonus lipopteni</i> ]                            |
| WP_099114325.1                             | tetraacyldisaccharide 4'-kinase [ <i>Xenorhabdus miraniensis</i> ]                                      |
| WP_112836465.1                             | tetraacyldisaccharide 4'-kinase [ <i>Providencia rustigianii</i> ]                                      |
| WP_119710886.1                             | tetraacyldisaccharide 4'-kinase [ <i>Arsenophonus</i> endosymbiont of <i>Aleurodicus floccissimus</i> ] |
| XP_023297146                               | uncharacterized protein LOC111679748 [ <i>Lucilia cuprina</i> ]   |
| <b>RNA polymerase-binding protein DksA</b> |   |
| WP_004389064                               | RNA polymerase-binding protein DksA [ <i>Yersinia kristensenii</i> ]                                    |
| WP_006659320.1                             | RNA polymerase-binding protein DksA [ <i>Providencia alcalifaciens</i> ]                                |
| WP_013200823                               | RNA polymerase-binding protein DksA [ <i>Erwinia billingiae</i> ]                                       |
| WP_032116626                               | RNA polymerase-binding protein DksA [ <i>Arsenophonus</i> endosymbiont of <i>Nilaparvata lugens</i> ]   |
| WP_040173129                               | RNA polymerase-binding protein DksA [ <i>Klebsiella pneumoniae</i> ]                                    |

|   |   |
|---|---|
| WP_076932495  | RNA polymerase-binding protein DksA [ <i>Salmonella enterica</i> ]                                    |
| WP_086108532.1  | RNA polymerase-binding protein DksA [ <i>Xenorhabdus vietnamensis</i> ]                               |
| WP_094029331  | RNA polymerase-binding protein DksA [ <i>Escherichia coli</i> ]                                       |
| WP_099136112.1  | RNA polymerase-binding protein DksA [ <i>Xenorhabdus budapestensis</i> ]                              |
| WP_119712400.1  | RNA polymerase-binding protein DksA [Arsenophonus endosymbiont of <i>Aleurodicus floccissimus</i> ]   |
| <b>NAD-dependent glyceraldehyde-3-phosphate dehydrogenase</b> |   |
| WP_004920333.1  | glyceraldehyde-3-phosphate dehydrogenase [ <i>Providencia stuartii</i> ]                              |
| WP_006658762.1  | glyceraldehyde-3-phosphate dehydrogenase [ <i>Providencia alcalifaciens</i> ]                         |
| WP_023655220  | glyceraldehyde-3-phosphate dehydrogenase [ <i>Erwinia piriflorinigra</i> ]                            |
| WP_026821958.1  | glyceraldehyde-3-phosphate dehydrogenase [ <i>Arsenophonus nasoniae</i> ]                             |
| WP_037405813  | glyceraldehyde-3-phosphate dehydrogenase [ <i>Chania multitudinisentens</i> ]                         |
| WP_038188954.1  | glyceraldehyde-3-phosphate dehydrogenase [ <i>Xenorhabdus bovienii</i> ]                              |
| WP_039300086  | glyceraldehyde-3-phosphate dehydrogenase [ <i>Pectobacterium betavasculorum</i> ]                     |
| WP_088499706  | glyceraldehyde-3-phosphate dehydrogenase [ <i>Serratia odorifera</i> ]                                |
| WP_099114842.1  | glyceraldehyde-3-phosphate dehydrogenase [ <i>Xenorhabdus miraniensis</i> ]                           |
| WP_129109510.1  | glyceraldehyde-3-phosphate dehydrogenase [Arsenophonus endosymbiont of <i>Bemisia tabaci</i> ]        |
| <b>Integration host factor subunit alpha</b>                  |   |
| QCI19710.1  | integration host factor subunit alpha [ <i>Buchnera aphidicola</i> ( <i>Brevicoryne brassicae</i> )]  |
| WP_006813286.1  | integration host factor subunit alpha [ <i>Providencia rustigianii</i> ]                              |
| WP_008911506.1  | integration host factor subunit alpha [ <i>Providencia burhodogranariae</i> ]                         |
| WP_026821649.1  | integration host factor subunit alpha [ <i>Arsenophonus nasoniae</i> ]                                |
| WP_081354998.1  | integration host factor subunit alpha [ <i>Pluralibacter gergoviae</i> ]                              |
| WP_086589325.1  | integration host factor subunit alpha [ <i>Escherichia coli</i> ]                                     |
| WP_097032363  | integration host factor subunit alpha [Sodalis endosymbiont of <i>Henestaris halophilus</i> ]         |
| WP_099132831.1  | integration host factor subunit alpha [ <i>Xenorhabdus ehlersii</i> ]                                 |
| WP_114193972  | integration host factor subunit alpha [ <i>Rahnella</i> sp. DSM 105170]                               |
| WP_115454648  | integration host factor subunit alpha [ <i>Citrobacter freundii</i> ]                                 |
| WP_136129973  | integration host factor subunit alpha [ <i>Pantoea</i> sp. SoEL]                                      |
| <b>Outer membrane protein assembly factor BamE</b>            |   |
| WP_006814814  | outer membrane protein assembly factor BamE [ <i>Providencia rustigianii</i> ]                        |
| WP_013316584  | outer membrane protein assembly factor BamE [ <i>Dickeya dadantii</i> ]                               |
| WP_023187381.1  | outer membrane protein assembly factor BamE [ <i>Salmonella enterica</i> ]                            |
| WP_032114779  | outer membrane protein assembly factor BamE [Arsenophonus endosymbiont of <i>Nilaparvata lugens</i> ] |

|                             |  |
|-----------------------------|--|
| WP_049218895.1              | outer membrane protein assembly factor BamE [ <i>Proteus mirabilis</i> ]                             |
| WP_066747277.1              | outer membrane protein assembly factor BamE [ <i>Cosenzaea myxofaciens</i> ]                         |
| WP_089570189.1              | outer membrane protein assembly factor BamE [ <i>Escherichia coli</i> ]                              |
| WP_097032457                | outer membrane protein assembly factor BamE [ <i>Sodalis endosymbiont of Henestaris halophilus</i> ] |
| WP_099113796.1              | outer membrane protein assembly factor BamE [ <i>Xenorhabdus miraniensis</i> ]                       |
| WP_117262918.1              | outer membrane protein assembly factor BamE [ <i>Klebsiella pneumoniae</i> ]                         |
| <b>Hypothetical protein</b> |  |
| SUC16805                    | Inner membrane protein ybbJ [ <i>Proteus vulgaris</i> ]  |
| WP_004704950                | NfeD family protein [ <i>Yersinia aldovae</i> ]  |
| WP_004922378.1              | NfeD family protein [ <i>Providencia stuartii</i> ]  |
| WP_026821504.1              | NfeD family protein [ <i>Arsenophonus nasoniae</i> ]   |
| WP_050108856                | NfeD family protein [ <i>Yersinia frederiksenii</i> ]  |
| WP_066283908                | NfeD family protein [ <i>Candidatus Arsenophonus lipopteni</i> ]                                     |
| WP_072070199                | NfeD family protein [ <i>Proteus vulgaris</i> ]  |
| WP_099123925.1              | NfeD family protein [ <i>Xenorhabdus stockiae</i> ]  |
| WP_115826368.1              | NfeD family protein [ <i>Xenorhabdus cabanillasii</i> ]  |
| WP_132496721.1              | NfeD family protein [ <i>Providencia alcalifaciens</i> ]   |

Tabulka II: Přístupová čísla (accession numbers) sekvencí stažených z NCBI pro ověření HGT genů bakterie rodu *Arsenophonus* z kloše *Crataerina pallida* pomocí fylogenetických analýz.

| <b>Horizontálně přenesené geny</b>        |  |
|---|--|
| <b>Restriction endonuclease subunit S</b> |  |
| WP_012883144                              | restriction endonuclease subunit S [ <i>Dickeya zeae</i> ]   |
| WP_036941067.1                            | restriction endonuclease subunit S [ <i>Providencia stuartii</i> ]   |
| WP_039522310                              | restriction endonuclease subunit S [ <i>Pectobacterium carotovorum</i> ]   |
| WP_043140826                              | restriction endonuclease subunit S [ <i>Serratia marcescens</i> ]  |
| WP_047037302.1                            | restriction endonuclease subunit S [ <i>Klebsiella aerogenes</i> ]   |
| WP_096982818                              | restriction endonuclease subunit S [ <i>Escherichia coli</i> ]   |
| WP_105091702                              | restriction endonuclease subunit S [ <i>Pantoea ananatis</i> ]   |
| WP_126656755.1                            | restriction endonuclease subunit S [ <i>Proteus mirabilis</i> ]  |
| WP_145507534                              | restriction endonuclease subunit S [ <i>Yersinia hibernica</i> ]   |
| <b>YcfA-like protein</b>                  |  |
| CP010384                                  | hypothetical protein [ <i>Enterobacter hormaechei</i> ]  |
| CP011584                                  | hypothetical protein [ <i>Enterobacter hormaechei</i> ]  |
| CP017581                                  | addiction module toxin, HicA family [ <i>Pantoea stewartii</i> subsp. <i>Stewartii</i> ]                           |
| CP042512                                  | type II toxin-antitoxin system HicA family toxin [ <i>Serratia marcescens</i> ]                                    |
| LK999954                                  | hypothetical protein [ <i>Escherichia coli</i> ]   |
| LT840187                                  | predicted RNA-binding protein YcfA [ <i>Enterobacter cloacae</i> ]   |
| <b>hypothetical protein</b>               |  |
| WP_004253755                              | hypothetical protein [ <i>Providencia rettgeri</i> ]   |
| WP_036953334                              | hypothetical protein [ <i>Providencia alcalifaciens</i> ]  |
| WP_048822414.1                            | hypothetical protein [ <i>Morganella morganii</i> ]  |
| WP_060706601                              | hypothetical protein [ <i>Serratia marcescens</i> ]  |
| WP_064290092                              | hypothetical protein [ <i>Klebsiella oxytoca</i> ]   |
| WP_088375657.1                            | hypothetical protein [ <i>Escherichia coli</i> ]   |
| WP_120065682                              | hypothetical protein [ <i>Escherichia coli</i> ]   |
| WP_142814486                              | hypothetical protein [ <i>Serratia liquefaciens</i> ]  |
| <b>hypothetical protein</b>               |  |
| CCJ99304                                  | hypothetical protein BN130_1963 [ <i>Cronobacter malonaticus</i> 507]  |
| CDK69956.1                                | hypothetical protein [ <i>Klebsiella pneumoniae</i> IS22]  |
| CNU19656                                  | Uncharacterised protein [ <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Bovismorbificans</i> ]      |
| CQQ92379.1                                | Uncharacterised protein [ <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i> str. DT104] |
| EJI88629                                  | hypothetical protein Y71_0921 [ <i>Kosakonia radicincitans</i> DSM 16656]  |
| GDE08777                                  | hypothetical protein HmCmsJML288_03711 [ <i>Escherichia coli</i> ]   |
| GDU83066.1                                | hypothetical protein BvCmsSIP076_04669 [ <i>Escherichia coli</i> ]   |

|  |   |
|--|---|
| SKZ58482.1   | Uncharacterised protein [ <i>Mycobacteroides abscessus</i> subsp. abscessus]              |
| SPW25903   | Uncharacterised protein [ <i>Edwardsiella tarda</i> ]                                     |
| WP_064719094.1   | hypothetical protein [ <i>Proteus hauseri</i> ]   |
| <b>hypothetical protein</b>  |   |
| CBA75315.1   | inhibitor of septum formation [ <i>Arsenophonus nasoniae</i> ]                            |
| WP_024105308   | hypothetical protein [ <i>Dickeya dianthicola</i> ]                                       |
| WP_040001101   | hypothetical protein [ <i>Dickeya chrysanthemi</i> ]                                      |
| WP_040031284   | hypothetical protein [ <i>Pectobacterium carotovorum</i> ]                                |
| WP_083163092   | hypothetical protein [ <i>Yersinia enterocolitica</i> ]                                   |
| WP_108088136   | hypothetical protein [ <i>Yersinia massiliensis</i> ]                                     |
| WP_127156476   | hypothetical protein [ <i>Pectobacterium polaris</i> ]                                    |
| WP_129705265   | hypothetical protein [ <i>Pectobacterium zantedeschiae</i> ]                              |
| <b>hypothetical protein</b>  |   |
| BAN95919   | hypothetical protein E05_11530 [ <i>Plautia stali</i> symbiont]                           |
| SAY44554   | Uncharacterised protein [ <i>Serratia marcescens</i> ]                                    |
| SAY44554.1   | Uncharacterised protein [ <i>Serratia marcescens</i> ]                                    |
| WP_026823821.1   | hypothetical protein [ <i>Arsenophonus nasoniae</i> ]                                     |
| WP_059108315.1   | hypothetical protein [ <i>Candidatus Hamiltonella defensa</i> ]                           |
| WP_129109584.1   | hypothetical protein [ <i>Arsenophonus</i> endosymbiont of <i>Bemisia tabaci</i> ]        |
| WP_139276943   | hypothetical protein [ <i>Serratia marcescens</i> ]                                       |
| <b>HGT neurčitelné geny</b>  |   |
| <b>Type I restriction-modification system, specificity subunit S</b> |   |
| MHN58610   | restriction endonuclease subunit S [ <i>Escherichia coli</i> ]                            |
| WP_005344795   | restriction endonuclease subunit S [ <i>Aeromonas allosaccharophila</i> ]                 |
| WP_012986644.1   | restriction endonuclease subunit S [ <i>Xenorhabdus bovienii</i> ]                        |
| WP_036941072.1   | restriction endonuclease subunit S [ <i>Providencia stuartii</i> ]                        |
| WP_051391304   | restriction endonuclease subunit S [ <i>Gluconobacter oxydans</i> ]                       |
| WP_071841598   | restriction endonuclease subunit S [ <i>Serratia</i> sp. FS14]                            |
| WP_094962249.1   | restriction endonuclease subunit S [ <i>Providencia rettgeri</i> ]                        |
| WP_107403062.1   | restriction endonuclease subunit S [ <i>Shewanella baltica</i> ]                          |
| WP_138611663   | restriction endonuclease subunit S [ <i>Pseudoalteromonas</i> sp. S983]                   |
| WP_145607938.1   | hypothetical protein [ <i>Yersinia kristensenii</i> ]                                     |
| EBR3717936   | hypothetical protein [ <i>Salmonella enterica</i> ]                                       |
| <b>Type I restriction endonuclease subunit R</b>                     |   |
| MHN58617.1   | type I restriction endonuclease subunit R [ <i>Escherichia coli</i> ]                     |
| SLM61600   | Type I restriction-modification system, restriction subunit R [ <i>Dickeya aquatica</i> ] |
| WP_033929741   | type I restriction endonuclease subunit R [ <i>Vibrio cholerae</i> ]                      |
| WP_074480794.1   | type I restriction endonuclease subunit R [ <i>Escherichia coli</i> ]                     |
| WP_088498739.1   | type I restriction endonuclease subunit R [ <i>Providencia stuartii</i> ]                 |
| WP_107111803   | type I restriction endonuclease subunit R [ <i>Shewanella algae</i> ]                     |
| WP_114502416   | type I restriction endonuclease subunit R [ <i>Klebsiella pneumoniae</i> ]                |

|   |   |
|---|---|
| WP_116820368.1  | type I restriction endonuclease subunit R [ <i>Providencia stuartii</i> ]   |
| WP_126656751.1  | type I restriction endonuclease subunit R [ <i>Proteus mirabilis</i> ]  |
| WP_130309315  | type I restriction endonuclease subunit R [ <i>Vibrio vulnificus</i> ]  |
| <b>Type I restriction-modification system subunit M</b> |   |
| WP_026823684.1  | type I restriction-modification system subunit M [ <i>Arsenophonus nasoniae</i> ]                                   |
| WP_038219398.1  | type I restriction-modification system subunit M [ <i>Xenorhabdus nematophila</i> ]                                 |
| WP_039522313  | type I restriction-modification system subunit M [ <i>Pectobacterium carotovorum</i> ]                              |
| WP_044030206  | type I restriction-modification system subunit M [ <i>Serratia</i> sp. FS14]  |
| WP_061584396  | type I restriction-modification system subunit M [ <i>Salmonella enterica</i> ]                                     |
| WP_063867267  | type I restriction-modification system subunit M [ <i>Enterobacter cloacae</i> ]                                    |
| WP_094888002.1  | type I restriction-modification system subunit M [ <i>Escherichia coli</i> ]  |
| WP_126656754.1  | type I restriction-modification system subunit M [ <i>Proteus mirabilis</i> ]                                       |
| WP_139397146  | type I restriction-modification system subunit M [ <i>Klebsiella aerogenes</i> ]                                    |
| WP_145607937  | type I restriction-modification system subunit M [ <i>Yersinia kristensenii</i> ]                                   |
| <b>Translation initiation factor 1</b>                  |   |
| ARC54763.1  | translation initiation factor IF-1 [ <i>Candidatus Riesia</i> sp. GBBU]   |
| CRK85775.1  | Translation initiation factor IF-1 [ <i>Candidatus Providencia siddallii</i> ]                                      |
| EBY0123694  | translation initiation factor IF-1 [ <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Typhimurium</i> ] |
| RLR18379  | translation initiation factor IF-1 [ <i>Sodalis</i> -like symbiont of <i>Bactericera trigonica</i> ]                |
| WP_066284091.1  | translation initiation factor IF-1 [ <i>Candidatus Arsenophonus lipopteni</i> ]                                     |
| WP_070133576.1  | translation initiation factor IF-1 [ <i>Candidatus Erwinia dacicola</i> ]   |
| WP_115459402.1  | translation initiation factor IF-1 [ <i>Enterobacillus tribolii</i> ]   |
| WP_119710873  | translation initiation factor IF-1 [ <i>Arsenophonus</i> endosymbiont of <i>Aleurodicus floccissimus</i> ]          |
| WP_135091519  | translation initiation factor IF-1 [ <i>Proteus mirabilis</i> ]   |
| WP_136132169  | translation initiation factor IF-1 [ <i>Pantoea</i> sp. SoEO]   |
| <b>Ribonuclease HI</b>                                  |   |
| PAV01705  | ribonuclease HI [ <i>Arsenophonus</i> sp. ENCA]   |
| WP_006957886.1  | ribonuclease HI [ <i>Vibrio coralliilyticus</i> ]   |
| WP_032116701  | ribonuclease HI [ <i>Arsenophonus</i> endosymbiont of <i>Nilaparvata lugens</i> ]                                   |
| WP_041855226.1  | ribonuclease HI [ <i>Candidatus Riesia pediculicola</i> ]   |
| WP_063215520.1  | ribonuclease HI [ <i>Proteus mirabilis</i> ]  |
| WP_097341772.1  | ribonuclease HI [ <i>Escherichia coli</i> ]   |
| WP_099113528.1  | ribonuclease HI [ <i>Xenorhabdus miraniensis</i> ]  |



|  |   |
|--|---|
| WP_102435058   | ribonuclease HI [ <i>Vibrio lentus</i> ]  |
| WP_104872987.1   | ribonuclease HI [ <i>Providencia stuartii</i> ]   |
| WP_129109735.1<br>1  | ribonuclease HI [ <i>Arsenophonus</i> endosymbiont of <i>Bemisia tabaci</i> ]                     |
| <b>Putative DNA-binding protein in cluster with Type I restriction-modification systém</b> |   |
| CDH24420.1   | conserved hypothetical protein [ <i>Xenorhabdus bovienii</i> str. kraussei Becker Underwood]      |
| WP_025160287.1   | virulence RhuM family protein [ <i>Serratia marcescens</i> ]                                      |
| WP_039522319   | virulence RhuM family protein [ <i>Pectobacterium carotovorum</i> ]                               |
| WP_041164245.1   | virulence RhuM family protein [ <i>Dickeya zaeae</i> ]  |
| WP_047048191   | virulence RhuM family protein [ <i>Klebsiella aerogenes</i> ]                                     |
| WP_050104085.1   | virulence RhuM family protein [ <i>Yersinia pseudotuberculosis</i> ]                              |
| WP_051876002.1   | hypothetical protein [ <i>Xenorhabdus bovienii</i> ]  |
| WP_109912289.1   | virulence RhuM family protein [ <i>Providencia rettgeri</i> ]                                     |
| WP_113374063   | virulence RhuM family protein [ <i>Escherichia coli</i> ]   |
| WP_129623271.1   | virulence RhuM family protein [ <i>Proteus mirabilis</i> ]  |
| <b>ATP synthase F0, C subunit</b>  |   |
| EHU54472.1   | ATP synthase F0, C subunit [ <i>Escherichia coli</i> DEC3B]                                       |
| OAT50616.1   | ATP synthase subunit C [ <i>Proteus hauseri</i> ATCC 700826]                                      |
| OZI14963   | ATP F0F1 synthase subunit C [ <i>Sodalis</i> -like symbiont of <i>Philaenus spumarius</i> ]       |
| WP_006708948   | F0F1 ATP synthase subunit C [ <i>Serratia symbiotica</i> ]  |
| WP_016192413.1   | F0F1 ATP synthase subunit C [ <i>Erwinia tracheiphila</i> ]                                       |
| WP_036942415.1   | F0F1 ATP synthase subunit C [ <i>Providencia stuartii</i> ]                                       |
| WP_075433078.1   | F0F1 ATP synthase subunit C [ <i>Buchnera aphidicola</i> ]  |
| WP_086954828.1   | F0F1 ATP synthase subunit C [ <i>Xenorhabdus innexi</i> ]   |
| WP_129108803   | F0F1 ATP synthase subunit C [ <i>Arsenophonus</i> endosymbiont of <i>Bemisia tabaci</i> ]         |
| <b>Putative membrane protein</b>   |   |
| OZI13927   | DUF2190 domain-containing protein [ <i>Sodalis</i> -like symbiont of <i>Philaenus spumarius</i> ] |
| PAV02113.1 1   | DUF2190 domain-containing protein [ <i>Arsenophonus</i> sp. ENCA]                                 |
| WP_010616702   | DUF2190 family protein [ <i>Plautia stali</i> symbiont]   |
| WP_016858136   | DUF2190 family protein [ <i>Candidatus Hamiltonella defensa</i> ]                                 |
| WP_032116663   | DUF2190 family protein [ <i>Arsenophonus</i> endosymbiont of <i>Nilaparvata lugens</i> ]          |
| WP_071045926.1   | DUF2190 family protein [ <i>Escherichia coli</i> ]  |
| WP_074019694   | DUF2190 family protein [ <i>Xenorhabdus thuongxuanensis</i> ]                                     |
| WP_105397942.1   | DUF2190 family protein [ <i>Photorhabdus luminescens</i> ]  |
| WP_133813569   | DUF2190 family protein [ <i>Photorhabdus temperata</i> ]  |
| <b>Hypothetical protein</b>  |   |
| WP_034938334   | hypothetical protein [ <i>Erwinia mallotivora</i> ]   |
| WP_044172449   | hypothetical protein [ <i>Metakosakonia massiliensis</i> ]  |

|  |  |
|--|--|
| WP_074561456.1   | hypothetical protein [Proteus mirabilis]   |
| WP_077294410   | hypothetical protein [Yersinia enterocolitica]   |
| WP_094121730   | hypothetical protein [Pantoea conspicua]   |
| WP_094488182   | hypothetical protein [Klebsiella pneumoniae]   |
| WP_099062651   | hypothetical protein [Serratia sp. BW106]  |
| WP_144242669   | hypothetical protein [Providencia stuartii]  |
| <b>Hypothetical protein</b>  |  |
| AID90856   | hypothetical protein KONIH1_17760 [Klebsiella oxytoca KONIH1]                              |
| EBQ6113621   | hypothetical protein DAZ36_25830 [Salmonella enterica subsp. enterica serovar Enteritidis] |
| PAV06455.1   | hypothetical protein CBG25_05975 [Arsenophonus sp. ENCA]                                   |
| WP_015833348   | hypothetical protein [Photorhabdus asymbiotica]  |
| WP_019213350.1   | hypothetical protein [Yersinia massiliensis]   |
| WP_037430651   | hypothetical protein [Serratia plymuthica]   |
| WP_038193527.1   | hypothetical protein [Xenorhabdus bovienii]  |
| WP_046894990   | hypothetical protein [Morganella morganii]   |
| WP_063086065.1   | hypothetical protein [Escherichia coli]  |
| WP_099121320.1   | hypothetical protein [Xenorhabdus sp. KK7.4]   |
| <b>Hypothetical protein</b>  |  |
| CRL65248   | hypothetical protein BN1804_03403 [Proteus vulgaris]                                       |
| WP_001411752.1   | hypothetical protein [Escherichia coli]  |
| WP_024157164   | hypothetical protein [Salmonella enterica]   |
| WP_047964788   | hypothetical protein [Xenorhabdus khoisanae]   |
| WP_053110609.1   | hypothetical protein [Providencia stuartii]  |
| WP_067398246   | hypothetical protein [Morganella psychrotolerans]  |
| WP_080981838.1   | hypothetical protein [Morganella morganii]   |
| WP_135522249   | hypothetical protein [Escherichia sp. E4742]   |
| <b>Hypothetical protein</b>  |  |
| BAN99909.1   | hypothetical protein E05_51430 [Plautia stali symbiont]                                    |
| CBA71366.1   | hypothetical protein ARN_00260, partial [Arsenophonus nasoniae]                            |
| WP_013583091   | hypothetical protein [Yersinia pestis]   |
| WP_013583091.1   | hypothetical protein [Yersinia pestis]   |
| <b>Hypothetical protein</b>  |  |
| EAP8457973   | hypothetical protein EO847_15125 [Salmonella enterica]                                     |
| EBY6257850   | hypothetical protein D5O62_03390 [Salmonella enterica subsp. enterica serovar Warnow]      |
| OIV47833   | hypothetical protein BK025_01130 [Sodalis sp. TME1]  |
| WP_045268335   | hypothetical protein [Enterobacter kobei]  |
| WP_069602290   | hypothetical protein [Enterobacter kobei]  |
| WP_115244438   | hypothetical protein [Klebsiella pneumoniae]   |
| YP_009287751   | hypothetical protein BI095_gp19 [Enterobacter phage Tyrion]                                |
| <b>Geny Arsenophonus</b>   |  |
| <b>Type II toxin-antitoxin system death-on-curing family toxin</b> |  |

|   |  |
|---|--|
| SFO12436.1  | death on curing protein [ <i>Xenorhabdus japonica</i> ]  |
| WP_032115810  | type II toxin-antitoxin system death-on-curing family toxin [ <i>Arsenophonus endosymbiont of Nilaparvata lugens</i> ] |
| WP_036895893.1                                      | type II toxin-antitoxin system death-on-curing family toxin [ <i>Proteus mirabilis</i> ]                               |
| WP_038180622  | type II toxin-antitoxin system death-on-curing family toxin [ <i>Xenorhabdus bovienii</i> ]                            |
| WP_059279199.1                                      | type II toxin-antitoxin system death-on-curing family toxin [ <i>Escherichia albertii</i> ]                            |
| WP_065821856  | type II toxin-antitoxin system death-on-curing family toxin [ <i>Photorhabdus australis</i> ]                          |
| WP_070974897  | type II toxin-antitoxin system death-on-curing family toxin [ <i>Photorhabdus temperata</i> ]                          |
| WP_099117343.1                                      | type II toxin-antitoxin system death-on-curing family toxin [ <i>Xenorhabdus ishibashii</i> ]                          |
| WP_129109080  | type II toxin-antitoxin system death-on-curing family toxin [ <i>Arsenophonus endosymbiont of Bemisia tabaci</i> ]     |
| <b>Aspartate aminotransferase</b>                   |  |
| WP_008915241.1                                      | aspartate/tyrosine/aromatic aminotransferase [ <i>Providencia sneebia</i> ]  |
| WP_026822279  | aspartate/tyrosine/aromatic aminotransferase [ <i>Arsenophonus nasoniae</i> ]  |
| WP_036778895.1                                      | aspartate/tyrosine/aromatic aminotransferase [ <i>Photorhabdus luminescens</i> ]                                       |
| WP_046893860  | aspartate/tyrosine/aromatic aminotransferase [ <i>Morganella morgani</i> ]   |
| WP_066283859.1                                      | aspartate/tyrosine/aromatic aminotransferase [ <i>Candidatus Arsenophonus lipopteni</i> ]                              |
| WP_069316368.1                                      | aspartate/tyrosine/aromatic aminotransferase [ <i>Xenorhabdus hominickii</i> ]   |
| WP_112860687  | aspartate/tyrosine/aromatic aminotransferase [ <i>Providencia alcalifaciens</i> ]                                      |
| WP_115464336.1                                      | aspartate/tyrosine/aromatic aminotransferase [ <i>Escherichia coli</i> ]   |
| <b>Mrr restriction system protein</b>               |  |
| PLK98006  | restriction endonuclease [ <i>Klebsiella pneumoniae</i> ]  |
| WP_000217932  | restriction endonuclease [ <i>Escherichia coli</i> ]   |
| WP_026823221.1                                      | restriction endonuclease [ <i>Arsenophonus nasoniae</i> ]  |
| WP_045957564.1                                      | restriction endonuclease [ <i>Xenorhabdus poinarii</i> ]   |
| WP_052901526  | restriction endonuclease [ <i>Erwinia iniecta</i> ]  |
| WP_088493941  | restriction endonuclease [ <i>Proteus mirabilis</i> ]  |
| WP_098941324  | restriction endonuclease [ <i>Proteus vulgaris</i> ]   |
| WP_102138159.1                                      | restriction endonuclease [ <i>Providencia sp. WCHPr000369</i> ]  |
| <b>Methyl-directed repair DNA adenine methylase</b> |  |
| OZI15153  | restriction endonuclease subunit M [ <i>Sodalis-like symbiont of Philaenus spumarius</i> ]                             |

|   |  |
|---|--|
| PAV02875  | restriction endonuclease subunit M [Arsenophonus sp. ENCA]   |
| RXK32791  | restriction endonuclease subunit M, partial [Arsenophonus endosymbiont of Bemisia tabaci Asia II 3]              |
| WP_016858111  | DNA adenine methylase [Candidatus Hamiltonella defensa]  |
| WP_026821882  | DNA adenine methylase [Arsenophonus nasoniae]  |
| WP_032116673  | DNA adenine methylase [Arsenophonus endosymbiont of Nilaparvata lugens]  |
| WP_111763633  | DNA adenine methylase [Salmonella enterica]  |
| <b>Magnesium and cobalt efflux protein CorC</b>     |  |
| PLP31576.1  | hypothetical protein CWM88_14920, partial [Klebsiella pneumoniae]  |
| RLR17502  | DUF21 domain-containing protein, partial [Sodalis-like symbiont of Bactericera trigonica]                        |
| SUC78052.1 1  | Putative Mg <sup>2+</sup> and Co <sup>2+</sup> transporter CorB [Providencia stuartii]                           |
| VTP60512.1  | Putative Mg <sup>2+</sup> and Co <sup>2+</sup> transporter CorB [Serratia rubidaea]                              |
| WP_026822680.1                                      | HlyC/CorC family transporter [Arsenophonus nasoniae]   |
| WP_032115486  | HlyC/CorC family transporter [Arsenophonus endosymbiont of Nilaparvata lugens]                                   |
| WP_036782341.1                                      | HlyC/CorC family transporter [Photorhabdus luminescens]  |
| WP_047961639.1                                      | HlyC/CorC family transporter [Xenorhabdus khoisaniae]  |
| WP_077883540.1                                      | HlyC/CorC family transporter [Escherichia sp. B1147]   |
| WP_088493487.1                                      | HlyC/CorC family transporter [Proteus mirabilis]   |
| Prevent host death protein, Phd antitoxin           |  |
| HBO24704  | TPA: type II toxin-antitoxin system prevent-host-death family antitoxin [Providencia sp.]                        |
| WP_023583377.1                                      | type II toxin-antitoxin system Phd/YefM family antitoxin [Proteus hauseri]                                       |
| WP_032115809  | type II toxin-antitoxin system Phd/YefM family antitoxin [Arsenophonus endosymbiont of Nilaparvata lugens]       |
| WP_045967830.1                                      | type II toxin-antitoxin system Phd/YefM family antitoxin [Xenorhabdus doucetiae]                                 |
| WP_047686540.1                                      | type II toxin-antitoxin system Phd/YefM family antitoxin [Xenorhabdus sp. NBAII XenSa04]                         |
| WP_059256823.1                                      | type II toxin-antitoxin system Phd/YefM family antitoxin [Escherichia albertii]                                  |
| WP_088494133.1                                      | type II toxin-antitoxin system Phd/YefM family antitoxin [Proteus mirabilis]                                     |
| WP_119712639.1                                      | type II toxin-antitoxin system Phd/YefM family antitoxin [Arsenophonus endosymbiont of Aleurodicus floccissimus] |
| <b>Putative predicted metal-dependent hydrolase</b> |  |
| CND42006.1  | metal-dependent hydrolase [Yersinia frederiksenii]   |
| PAV10513.1  | hypothetical protein CBG25_01730 [Arsenophonus sp. ENCA]   |
| WP_012986642.1                                      | M48 family metallopeptidase [Xenorhabdus bovienii]   |
| WP_019455552.1                                      | M48 family metallopeptidase [Serratia marcescens]  |

|   |  |
|---|--|
| WP_036941079.1                                | M48 family metallopeptidase [ <i>Providencia stuartii</i> ]  |
| WP_049221203.1                                | M48 family metallopeptidase [ <i>Proteus mirabilis</i> ]   |
| WP_088543516.1                                | M48 family metallopeptidase [ <i>Escherichia coli</i> ]  |
| <b>Flagellar hook protein FlgE</b>            |  |
| QBY42942.1                                    | Flagellar hook protein FlgE [ <i>Arsenophonus nasoniae</i> ]   |
| WP_006657511.1                                | flagellar hook protein FlgE [ <i>Providencia alcalifaciens</i> ]   |
| WP_014499488.1                                | flagellar hook protein FlgE [ <i>Buchnera aphidicola</i> ]   |
| WP_047256578                                  | flagellar hook protein FlgE [ <i>Moellerella wisconsensis</i> ]  |
| WP_099141651.1                                | flagellar hook protein FlgE [ <i>Xenorhabdus kozodoii</i> ]  |
| WP_109373798.1                                | flagellar hook protein FlgE [ <i>Proteus genomsp. 6</i> ]  |
| WP_110591519.1                                | flagellar hook protein FlgE [ <i>Providencia rettgeri</i> ]  |
| WP_119710925.1                                | flagellar basal-body rod protein FlgG [ <i>Arsenophonus endosymbiont of Aleurodicus floccissimus</i> ]                           |
| WP_129041717                                  | flagellar hook protein FlgE [ <i>Proteus hauseri</i> ]   |
| <b>Flagellar hook-associated protein FlgK</b> |  |
| EBQ6150789                                    | flagellar hook-associated protein FlgK, partial [ <i>Salmonella enterica</i> subsp. <i>enterica</i> serovar <i>Enteritidis</i> ] |
| PAV05783.1                                    | hypothetical protein CBG25_06235 [ <i>Arsenophonus</i> sp. ENCA]   |
| PGG07604.1                                    | flagellar hook-associated protein FlgK, partial [ <i>Escherichia coli</i> ]  |
| WP_004248337.1                                | flagellar hook-associated protein FlgK [ <i>Proteus mirabilis</i> ]  |
| WP_009874301.1                                | flagellar hook-associated protein 1 [ <i>Buchnera aphidicola</i> ]   |
| WP_026823266                                  | flagellar hook-associated protein FlgK [ <i>Arsenophonus nasoniae</i> ]  |
| WP_099134832.1                                | flagellar hook-associated protein FlgK [ <i>Xenorhabdus budapestensis</i> ]  |
| WP_119710929                                  | hypothetical protein [ <i>Arsenophonus endosymbiont of Aleurodicus floccissimus</i> ]  |
| WP_123382090                                  | flagellar hook-associated protein FlgK [ <i>Providencia alcalifaciens</i> ]  |
| <b>Putative insecticidal toxin complex</b>    |  |
| ACY62102                                      | insecticidal toxin [ <i>Yersinia pestis</i> D182038]   |
| PHM46893.1                                    | insecticidal toxin, SepC/Tcc class [ <i>Xenorhabdus miraniensis</i> ]  |
| PHM61563.1                                    | insecticidal toxin [ <i>Xenorhabdus ishibashii</i> ]   |
| SUP86319.1                                    | toxin protein [ <i>Yersinia pseudotuberculosis</i> ]   |
| TCT27931.1                                    | insecticidal toxin complex protein TccC [ <i>Providencia alcalifaciens</i> ]   |
| WP_052044900.1                                | RHS repeat protein [ <i>Arsenophonus endosymbiont of Nilaparvata lugens</i> ]  |
| WP_087803361.1                                | RHS repeat protein [ <i>Proteus mirabilis</i> ]  |
| WP_103005122.1                                | RHS repeat protein [ <i>Proteus vulgaris</i> ]   |
| <b>Putative insecticidal toxin complex</b>    |  |
| CBA71829.1                                    | insecticidal toxin complex protein, partial [ <i>Arsenophonus nasoniae</i> ]   |
| WP_016261049.1                                | toxin, partial [ <i>Yersinia pestis</i> ]  |
| WP_046336363.1                                | RHS repeat-associated core domain-containing protein [ <i>Xenorhabdus bovienii</i> ]   |
| WP_139655805.1                                | RHS repeat protein [ <i>Photorhabdus luminescens</i> ]   |
| WP_146136358.1                                | RHS repeat protein, partial [ <i>Yersinia pseudotuberculosis</i> ]   |

|   |   |
|---|---|
| <b>Polyphenol oxidase</b>                             |   |
| WP_015835312.1  | polyphenol oxidase [Photorhabdus asymbiotica]   |
| WP_026822667  | polyphenol oxidase [Arsenophonus nasoniae]  |
| WP_032115473  | polyphenol oxidase [Arsenophonus endosymbiont of Nilaparvata lugens]  |
| WP_038220174.1  | polyphenol oxidase [Xenorhabdus nematophila]  |
| WP_054305789.1  | polyphenol oxidase [Serratia rubidaea]  |
| WP_109491536.1  | polyphenol oxidase [Brenneria roseae]   |
| WP_137525611.1  | polyphenol oxidase [Escherichia coli]   |
| WP_140172705  | polyphenol oxidase [Providencia rettgeri]   |
| <b>Autolysin sensor kinase</b>                        |   |
| PAV02252.1  | sensor histidine kinase [Arsenophonus sp. ENCA]   |
| WP_004915834.1  | sensor histidine kinase [Providencia stuartii]  |
| WP_023045004  | sensor histidine kinase [Photorhabdus temperata]  |
| WP_038184685  | sensor histidine kinase [Xenorhabdus bovienii]  |
| WP_038261524  | sensor histidine kinase [Xenorhabdus cabanillasii]  |
| WP_088493975.1  | sensor histidine kinase [Proteus mirabilis]   |
| WP_097339296.1  | two-component regulatory system sensor histidine kinase BtsS [Escherichia coli]                             |
| WP_126462312.1  | sensor histidine kinase [Providencia rustigianii]   |
| <b>IncF plasmid conjugative transfer protein TraD</b> |   |
| MIA85980.1  | type IV conjugative transfer system coupling protein TraD [Escherichia coli]                                |
| WP_026823554.1  | type IV conjugative transfer system coupling protein TraD [Arsenophonus nasoniae]                           |
| WP_051297015.1  | type IV conjugative transfer system coupling protein TraD [Arsenophonus nasoniae]                           |
| WP_052044982  | type IV conjugative transfer system coupling protein TraD [Arsenophonus endosymbiont of Nilaparvata lugens] |
| WP_080335494  | type IV conjugative transfer system coupling protein TraD [Serratia marcescens]                             |
| WP_082170566  | type IV secretion system DNA-binding domain-containing protein [Pantoea vagans]                             |
| WP_099119146  | type IV conjugative transfer system coupling protein TraD [Xenorhabdus ishibashii]                          |
| WP_135023939  | type IV conjugative transfer system coupling protein TraD [Proteus mirabilis]                               |
| <b>Terminase B protein, putative</b>                  |   |
| WP_036953309  | hypothetical protein [Providencia alcalifaciens]  |
| WP_087696910.1  | terminase [Morganella morganii]   |
| WP_112854914.1  | terminase [Escherichia coli]  |
| WP_115113506.1  | terminase [Providencia stuartii]  |
| WP_115370893.1  | terminase [Proteus vulgaris]  |

|   |   |
|---|---|
| WP_117162261.1                                  | terminase [Providencia rettgeri]  |
| WP_126616505                                    | terminase [Morganella morganii]   |
| WP_130060306.1                                  | terminase [Escherichia coli]  |
| WP_141060576.1                                  | terminase [Proteus mirabilis]   |
| <b>Autoinducer synthesis protein LuxI</b>       |   |
| CBA75574.1                                      | N-acylhomoserine lactone synthase [Arsenophonus nasoniae]                       |
| WP_026821910.1                                  | acyl-homoserine-lactone synthase [Arsenophonus nasoniae]                        |
| WP_032116572.1                                  | hypothetical protein [Arsenophonus endosymbiont of Nilaparvata lugens]          |
| WP_039307030                                    | acyl-homoserine-lactone synthase [Pectobacterium betavascularum]                |
| WP_039498397.1                                  | acyl-homoserine-lactone synthase [Pectobacterium carotovorum]                   |
| WP_119872627                                    | acyl-homoserine-lactone synthase [Pectobacterium carotovorum]                   |
| <b>Mu-like prophage FluMu protein gp36</b>      |   |
| BAE74117.1                                      | hypothetical phage protein [Sodalis glossinidius str. 'morsitans']              |
| OKP07306.1                                      | hypothetical protein Xentx_01583 [Xenorhabdus thuongxuanensis]                  |
| SAY44548  | Uncharacterised protein [Serratia marcescens]                                   |
| WP_029490412                                    | DUF1320 domain-containing protein [Escherichia coli]                            |
| WP_059108317                                    | DUF1320 domain-containing protein [Candidatus Hamiltonella defensa]             |
| WP_060555602.1                                  | DUF1320 domain-containing protein [Proteus mirabilis]                           |
| WP_113411351                                    | DUF1320 domain-containing protein [Escherichia coli]                            |
| WP_129109587                                    | DUF1320 domain-containing protein [Arsenophonus endosymbiont of Bemisia tabaci] |
| WP_138650879                                    | DUF1320 domain-containing protein [Salmonella enterica]                         |
| <b>Translation elongation factors (GTPases)</b> |   |
| WP_032186307.1                                  | hypothetical protein [Escherichia coli]   |
| WP_048608734                                    | hypothetical protein [Providencia rettgeri]                                     |
| WP_049256809.1                                  | hypothetical protein [Proteus mirabilis]  |
| WP_050144309                                    | hypothetical protein [Yersinia enterocolitica]                                  |
| WP_060441863.1                                  | hypothetical protein [Serratia marcescens]                                      |
| WP_062773527                                    | hypothetical protein [Morganella morganii]                                      |
| WP_069318014.1                                  | hypothetical protein [Xenorhabdus hominickii]                                   |
| WP_092512240.1                                  | hypothetical protein [Xenorhabdus mauleonii]                                    |
| WP_112860229.1                                  | hypothetical protein [Providencia alcalifaciens]                                |
| WP_115370625.1                                  | hypothetical protein [Proteus vulgaris]   |
| <b>Arc family DNA-binding protein</b>           |   |
| EAR6708283                                      | Arc family DNA-binding protein [Salmonella enterica]                            |
| WP_048981614                                    | Arc family DNA-binding protein [Enterobacter hormaechei]                        |
| WP_071992024.1                                  | Arc family DNA-binding protein [Providencia alcalifaciens]                      |
| WP_086112289                                    | Arc family DNA-binding protein [Xenorhabdus beddingii]                          |
| WP_096043092                                    | Arc family DNA-binding protein [Proteus mirabilis]                              |
| WP_105881111                                    | Arc family DNA-binding protein [Providencia rettgeri]                           |
| WP_107131117.1                                  | Arc family DNA-binding protein [Escherichia coli]                               |

|                             |  |
|-----------------------------|--|
| WP_115224736                | Arc family DNA-binding protein [ <i>Klebsiella pneumoniae</i> ]                        |
| WP_135024822.1              | Arc family DNA-binding protein [ <i>Proteus mirabilis</i> ]                            |
| <b>Peptidase</b>            |  |
| WP_036407714                | peptidase [ <i>Morganella morganii</i> ]   |
| WP_036953317.1              | peptidase [ <i>Providencia alcalifaciens</i> ]   |
| WP_060555918                | peptidase [ <i>Proteus mirabilis</i> ]   |
| WP_071425340                | peptidase [ <i>Proteus mirabilis</i> ]   |
| WP_097491472.1              | peptidase [ <i>Escherichia coli</i> ]  |
| WP_109912477                | peptidase [ <i>Providencia rettgeri</i> ]  |
| WP_140179420.1              | peptidase [ <i>Providencia stuartii</i> ]  |
| <b>Hypothetical protein</b> |  |
| EAX3734500                  | hypothetical protein AYD19_18550 [ <i>Salmonella enterica</i> ]                        |
| NZ_CP017671                 | hypothetical protein CDS translation [ <i>Providencia rettgeri</i> ]                   |
| WP_049207418                | hypothetical protein [ <i>Serratia marcescens</i> ]                                    |
| WP_053294727.1              | hypothetical protein [ <i>Escherichia coli</i> ]                                       |
| WP_072063454                | hypothetical protein [ <i>Proteus vulgaris</i> ]                                       |
| WP_115271827                | hypothetical protein [ <i>Klebsiella pneumoniae</i> ]                                  |
| WP_144242759.1              | hypothetical protein [ <i>Proteus mirabilis</i> ]                                      |
| <b>Hypothetical protein</b> |  |
| EAR2441462                  | hypothetical protein [ <i>Salmonella enterica</i> ]                                    |
| PAV08338.1                  | hypothetical protein CBG25_04320 [ <i>Arsenophonus</i> sp. ENCA]                       |
| WP_032116661.1              | hypothetical protein [ <i>Arsenophonus</i> endosymbiont of <i>Nilaparvata lugens</i> ] |
| WP_036843632.1              | hypothetical protein [ <i>Photorhabdus kharii</i> ]                                    |
| WP_063862470.1              | hypothetical protein [ <i>Enterobacter cloacae</i> ]                                   |
| WP_080492024                | hypothetical protein [ <i>Serratia marcescens</i> ]                                    |
| WP_087816488.1              | hypothetical protein [ <i>Yersinia intermedia</i> ]                                    |
| WP_105397940.1              | hypothetical protein [ <i>Photorhabdus luminescens</i> ]                               |
| WP_133813571.1              | hypothetical protein [ <i>Photorhabdus temperata</i> ]                                 |
| <b>Hypothetical protein</b> |  |
| WP_032340947.1              | hypothetical protein [ <i>Escherichia coli</i> ]                                       |
| WP_036407749                | hypothetical protein [ <i>Morganella morganii</i> ]                                    |
| WP_036417626                | hypothetical protein [ <i>Morganella</i> sp. EGD-HP17]                                 |
| WP_036953344                | hypothetical protein [ <i>Providencia alcalifaciens</i> ]                              |
| WP_049196342                | hypothetical protein [ <i>Proteus mirabilis</i> ]                                      |
| WP_115113521.1              | hypothetical protein [ <i>Providencia stuartii</i> ]                                   |
| WP_115349916                | hypothetical protein [ <i>Proteus vulgaris</i> ]                                       |
| WP_135147384.1              | hypothetical protein, partial [ <i>Escherichia coli</i> ]                              |
| WP_144141231                | hypothetical protein [ <i>Providencia rettgeri</i> ]                                   |
| <b>Hypothetical protein</b> |  |
| SST02788                    | Uncharacterised protein [ <i>Acinetobacter baumannii</i> ]                             |
| SST02788.1                  | Uncharacterised protein [ <i>Acinetobacter baumannii</i> ]                             |
| WP_129467040.1              | hypothetical protein [ <i>Providencia rettgeri</i> ]                                   |



|                             |  |
|-----------------------------|--|
| WP_140186764.1              | hypothetical protein [Providencia stuartii]  |
| <b>Hypothetical protein</b> |  |
| CBA74993.1                  | conserved hypothetical protein [Arsenophonus nasoniae]                                 |
| WP_026822566.1              | hypothetical protein [Arsenophonus nasoniae]   |
| WP_094193066                | hypothetical protein [Klebsiella cf. planticola B43]                                   |
| WP_097407798                | hypothetical protein [Kosakonia sacchari]  |
| WP_128483029                | hypothetical protein [Kosakonia cowanii]   |
| WP_139567865                | hypothetical protein [Kosakonia cowanii]   |
| <b>Hypothetical protein</b> |  |
| ACR68705                    | hypothetical protein NT01EI_1519 [Edwardsiella ictaluri 93-146]                        |
| SPP32403                    | hypothetical protein ARAF_2566 [Arsenophonus endosymbiont of Aleurodicus floccissimus] |
| WP_025154948                | hypothetical protein [Morganella morgani]  |
| WP_040056149                | hypothetical protein [Candidatus Hamiltonella defensa]                                 |
| WP_051297068                | hypothetical protein [Arsenophonus nasoniae]   |
| WP_071882213                | hypothetical protein [Candidatus Sodalis pierantonius]                                 |
| WP_081164863                | hypothetical protein [Edwardsiella ictaluri]   |
| WP_099404451                | hypothetical protein [Chitinimonas sp. BJB300]   |
| <b>Hypothetical protein</b> |  |
| WP_014656100.1              | hypothetical protein [Providencia stuartii]  |
| WP_044172449.1              | hypothetical protein [Metakosakonia massiliensis]                                      |
| WP_094488182.1              | hypothetical protein [Klebsiella pneumoniae]   |
| WP_117258060.1              | hypothetical protein [Klebsiella pneumoniae]   |
| <b>Hypothetical protein</b> |  |
| ASU04999.1                  | Hypothetical protein [Proteus mirabilis]   |
| PAV05649.1                  | addiction module [Arsenophonus sp. ENCA]   |
| WP_006816348.1              | hypothetical protein [Providencia rustigianii]   |
| WP_040360331                | hypothetical protein [Capnocytophaga gingivalis]                                       |
| WP_049211226.1              | hypothetical protein [Proteus mirabilis]   |
| WP_077791038                | acetyltransferase [Escherichia coli]   |
| WP_113589074.1              | acetyltransferase [Escherichia coli]   |
| WP_115286364.1              | hypothetical protein [Proteus mirabilis]   |
| WP_145892157                | hypothetical protein [Pantoea dispersa]  |
| <b>Hypothetical protein</b> |  |
| GCP95386                    | hypothetical protein BvCmsHHNP023_04738 [Escherichia coli]                             |
| MDF22637                    | cag pathogenicity island Cag12 family protein [Salmonella enterica]                    |
| PAV02456.1                  | hypothetical protein CBG25_10810 [Arsenophonus sp. ENCA]                               |
| WP_013572665                | hypothetical protein [Yersinia pestis]   |
| WP_032116057.1              | hypothetical protein [Arsenophonus endosymbiont of Nilaparvata lugens]                 |
| WP_115195437                | cag pathogenicity island Cag12 family protein [Klebsiella pneumoniae]                  |
| <b>Hypothetical protein</b> |  |

|                             |   |
|-----------------------------|---|
| WP_006658824.1              | hypothetical protein [Providencia alcalifaciens]  |
| WP_036953323                | hypothetical protein [Providencia alcalifaciens]  |
| WP_067424738.1              | hypothetical protein [Morganella psychrotolerans] |
| WP_104208894.1              | hypothetical protein [Escherichia coli]           |
| WP_112550160                | hypothetical protein [Morganella morganii]        |
| WP_135000028.1              | hypothetical protein [Proteus mirabilis]          |
| WP_140179426                | hypothetical protein [Providencia stuartii]       |
| <b>Hypothetical protein</b> |   |
| WP_087696913                | hypothetical protein [Morganella morganii]        |
| WP_115065428                | hypothetical protein [Proteus penneri]            |
| WP_115065428.1              | hypothetical protein [Proteus penneri]            |
| WP_117162256.1              | hypothetical protein [Providencia rettgeri]       |
| WP_135147470.1              | hypothetical protein, partial [Escherichia coli]  |
| WP_143475682                | hypothetical protein [Proteus mirabilis]          |