

Jihočeská Univerzita v Českých Budějovicích
Biologická fakulta

2007



Bakalářská práce

Vyhledávání a testování nových genů použitelných pro studium
evoluce symbiózy ve skupině Enterobacteriaceae

Michaela Šurganová

Školitel: Prof. RNDr. Václav Hypša, CSc.

Bakalářská diplomová práce

Šurganová, M. 2007. Vyhledávání a testování nových genů použitelných pro studium evoluce symbiózy ve skupině Enterobacteriaceae.

[Identification and evaluation of new genes useful for evolutionary studies of symbiotic Enterobacteriaceae] **32 pp. Faculty of Biological Sciences, University of South Bohemia, České Budějovice, Czech Republic.**

Anotace:

The aim of this project is to identify genes which could be used in phylogenetic reconstructions of symbiotic bacteria within the group Enterobacteriaceae. The candidate genes were looked up in COG database prospective primers were designed according to their conservative regions. The putative primer pairs were screened in silico and the best candidates were tested in PCR reactions with several DNA samples. Phylogenetic analysis was performed to test the information value of the best primer pair, amplifying partial gene for adenylosuccinate lyase (purB). The result demonstrates general suitability of this approach to identification of new phylogenetic markers and highlights several problems to be solved in future research.

Prohlašuji, že jsem tuto práci vypracovala samostatně, pouze s použitím uvedené literatury.

Prohlašuji, že v souladu s § 47b zákona č.111/1998 Sb. v platném znění souhlasím zveřejněním své bakalářské práce, a to v nezkrácené podobě fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v českých Budějovicích na jejích internetových stránkách.

14.5.2007

.....
Michaela Šurganová

Poděkování

Chtěla bych poděkovat svému školiteli Václavu Hypšovi za odborné vedení této práce, všem členům Laboratoře molekulární fylogenetiky a evoluce parazitů, zvláště pak Evě Novákové, Janě Kvičerové, Janu Štefkovi a Lence Štifterové. Dále bych chtěla poděkovat Janu Erhartovi za pomoc při chovu a krmení ploštic a Pavlíně Šorfové za pomoc a dobré rady. Také děkuji své rodině a všem přátelům za podporu a trpělivost.

Obsah

1. Úvod

1.1 Symbiotické bakterie	1
1.2 Molekulární data a fylogeneze symbiotických bakterií	2
1.3 Evoluce genomu bakterií	3
1.3.1 Horizontální přenos genu	4
1.3.2 Duplikace genů	4
1.3.3 Ztráta genů a změna	5
1.3.4 Změna nukleotidového složení a rychlost evoluce	5
1.4. Způsoby řešení a eliminace artefaktů	6
1.4.1 COGs databáze	7
1.4.2 Primery	7

2. Cíle práce

3. Materiál a Metodika

3.1 Vyhledávání potenciálně vhodných genů	10
3.2 Navrhování primeru	10
3.3 Izolace DNA	11
3.4 PCR	11
3.5 Elektroforéza	12
3.6 Klonování PCR produktů	12
3.7 Kultivace a selekce buněk	12
3.8 Izolace plazmidové DNA	13
3.9 Sekvenační reakce	13
3.10 Zpracování sekvencí	13

4. Výsledky	
4.1 Výběr vhodných sekvencí	15
4.2 Výběr primerů	18
4.3 Získané sekvence z nových primerů	21
4.4 Ověřování kvality informačního obsahu genu purB	22
5. Diskuze	
5.1 Identifikace a navrhování primerů	24
5.2 Nasedání primeru na templátovou DNA	25
5.3 Informační hodnota purB	25
6. Závěr	26
7. Přehled použité literatury	27

Úvod

1.1 Symbiotické bakterie

Jakékoli soužití dvou či více organismů se nazývá symbióza. Velmi častým typem symbiózy jsou interakce mezi bakteriemi a jejich hostiteli, které nabývají různých podob. Jeden z nejčastějších typů spojení je vztah mezi bakteriemi a členovci, především hmyzem (Moran a Teleng, 1998). Tyto vztahy mohou být nejrůznějšího typu, od parazitismu po mutualismus.

Obecně jsou symbiotické bakterie hmyzu děleny na dvě umělé, především ekologicky vymezené skupiny: primární (P-) a sekundární (S-) symbionty. P-symbionti jsou přenášeni výhradně vertikálně, z jedné generace hostitele do další, a jejich fylogeneze je paralelní s fylogenezí hostitele. Obvykle mají značně redukovaný genom v důsledku ztráty genů, což je typický znak pro bakterie žijící v konstantním, na živiny bohatém vnitřním prostředí. Na druhou stranu si P-symbionti zachovali geny se specifickými biosyntetickými funkcemi, které jsou důležité pro přežití a reprodukci jejich hostitele (Wernegreen, 2002). Mezi tyto biosyntetické funkce patří například tvorba a dodávání esenciálních aminokyselin hostiteli (*Buchnera aphidocola*, Wernegreen *et al.*, 2003), recyklace kyseliny močové (*Blattabacterium*, Moran a Teleng, 1998; Wernegreen *et al.*, 2003) nebo doplňování vitamínů skupiny B (*Wigglesworthia glossinidia*, Aksoy, 2000; Wernegreen *et al.*, 2003). Na rozdíl od těchto bakterií mají S-symbionti fylogenezi velmi často inkongruentní s hostitelskou linií. Velikost jejich genomu a obsah GC jsou srovnatelné s volně žijícími bakteriemi. Rovněž jsou schopny horizontálního transferu v populaci hostitele (Werren *et al.*, 1995; Aksoy *et al.*, 1995; Hypša a Dale, 1997; Thao a Baumann, 2004). Podle současných studií jsou někteří S-symbionti schopni do určité míry kompenzovat hostiteli ztrátu P-symbionta, čímž zvyšují specializaci hostitele na různé typy potravy (Koga *et al.*, 2003), rezistenci vůči parazitoidům (Olivek *et al.*, 2003) a toleranci k vyšším teplotním rozdílům (Chen *et al.*, 2000).

Studium symbiotických bakterií, jejich evoluce a interakcí s hostitelem je možné jen na základě spolehlivé znalosti jejich fylogenetických vztahů. U bakterií nelze tyto vztahy odvozovat pouze na základě morfologických znaků. Proto se v současné době téměř výhradně používají molekulárně-fylogenetické metody.

1.2 Molekulární data a fylogeneze bakterií

Metody molekulární fylogenetiky jsou založeny na získávání informace srovnáváním struktury a organizace molekul, především DNA. Způsoby, jak odvodit příbuzenské vztahy u prokaryot, jsou analýzy lokálního genového pořadí v genomu, odvozování přítomnosti nebo absence znaků, složení nukleotidů a studium GC obsahu s efektem kodonové preference vyskytující se obvykle v synonymní třetí pozici (Huynen a Bork, 1998; Fitz-Gibbon et al., 1999; Ochman et al., 1999; Suyama a Bork, 2001; Chen a Zhang, 2003; Coenye et al., 2005). Nejčastějším zdrojem informace je však primární struktura DNA tvořená pořadím nukleotidů. Výhodou metody je volnost ve výběru optimální sekvence, a tedy plné využití rozmanitosti genomu. Výběr optimálních sekvencí z genomu závisí na několika kritériích, především na rychlosti evoluce a způsobu dědičnosti.

Současná klasifikace prokaryot je nejčastěji založena na srovnávání genových sekvencí 16S rRNA nebo tzv. housekeeping genů (geny aktivní ve všech buňkách a zajišťující základní funkce buněčného metabolismu), např. gen pro rekombinázu A (*recA*), gen pro RNA polymerázu subjednotky alfa (*rpoA*), gen pro DNA gyrázu subjednotky B (*gyrB*), gen pro RNA polymerázu sigma faktoru (*rpoD*) nebo gen pro chaperon 60 (*groEL*) (Yamamoto, 2000; Pride et al., 2003; Herberk et al., 2004; Moran et al., 2004).

Jaderná 16S rRNA je nezbytný komponent pro syntézu proteinů, tedy nepostradatelná součást života organismů. Díky tomu má vlastnosti, které jsou výhodné pro fylogenetické analýzy - je to všudypřítomná molekula, snadno dostupná, má vhodnou délku pro sekvenaci a obsahuje velké množství informací (Ludwig a Schleifer, 1994; Vandamme et al., 1996; Coenye a Vandamme, 2003). V primární struktuře 16S rRNA se střídají variabilní oblasti s konzervativními úseky. Pomalu vyvíjející se oblasti těchto genů poskytují informaci o fylogenetických vztazích vzdálených skupin, části s vysokou variabilitou jsou používány pro vysvětlení nedávného větvení (Ludwig a Schleifer, 1994; Trowbridge et al., 2005). Přesto 16S rRNA nemůže být používána jako univerzální prostředek pro vyřešení vztahů mezi prokaryotními organismy. Například u skupiny *Enterobacteriaceae*, do které patří většina popsaných symbiontů hmyzu, se ukázalo, že 16S rRNA není dostatečně informativní pro fylogenetické analýzy. Příkladem jsou často studované vztahy primárních symbiontů skupin *Buchnera*, *Wigglesworthia* a *Blochmannia*, u kterých vycházely konfliktní topologie s nejasnou pozicí rodu *Buchnera* (Lerat et al., 2003; Herbeck et al., 2005; Olivier et al., 2005). Analýzy pomocí jednoho genu nedokázaly vyřešit, zda zmíněné skupiny P-symbiontů tvoří monofyletickou skupinu bakterií s výraznou tendencí k vytváření symbiózy nebo je tato

topologie způsobená nějakým metodickým artefaktem. V současné době je vyvíjena snaha hledat alternativní přístupy k řešení jejich vztahů. Vlivy mechanismů způsobujících problémy mohou být redukovány nebo eliminovány metodami, které zohledňují tyto mechanismy ve fylogenetické analýze, jako je například použití nehomogenního modelu sekvenční evoluce obsahujícího informaci o variabilní substituční rychlosti mezi oblastmi genomu a o variabilním GC obsahu mezi druhy (Galtier a Gouy, 1995, 1997; Herbeck et al., 2005).

Na obecné úrovni jsou však fylogenetické studie symbiotických bakterií komplikovány řadou evolučních mechanismů majících vliv na diversitu a celkové uspořádání genomu.

1.3 Evoluce genomu bakterií

Správné používání molekulárních dat vyžaduje porozumění obecným mechanismům molekulární evoluce bakterií, stejně jako specifickým rysům vyplývajícím ze symbiotického způsobu života. Genom bakterií je ve srovnání s eukaryotními organismy jednoduchý, neobsahuje dlouhé nekódující oblasti a přerušované geny. Velikost genomu se u různých bakterií pohybuje v rozsahu 0,6-9 Mb. Volně žijící bakterie např. *Escherichia coli*, *Salmonella* nebo *Bacillus* mají genom o velikosti cca 4,6-9 Mb (Wernegreen et al., 2002; Hoffmeister a Martin, 2003), kódující zhruba 1500-6000 proteinů (Klasson a Anderson, 2003). Obligátní paraziti, např. *Rickettsia prowazekii*, *Rickettsia conorii* a *Mycoplasma genitalium* mají genom okolo 1 Mb obsahující geny, které kódují okolo 500-1000 proteinů (Moran, 2002). Například endosymbionti mšic, kteří se oddělili od společného předka s volně žijícími bakteriemi zhruba před 200 miliony lety, podstoupili výraznou redukci genomu díky dlouhotrvajícímu vnitrobuněčnému životu, došlo k eliminaci až 70-75 % původního genomu (Klasson a Anderson, 2003).

Na dynamiku bakteriálního genomu působí dva protichůdné selekční tlaky. Jeden selekční tlak vede k udržení stability genů nezbytných pro růst a rozmnožování bakterií tj. genů řídících DNA opravy, replikaci, transkripci a translaci. Druhý tlak je zprostředkován proměnlivými vnějšími podmínkami a vede ke změnám v genomu (Wolf et al. 2002). Organizaci a stavbu genomu ovlivňují tyto faktory: 1) horizontální přenos genů (Huyen a Bork, 1998), 2) duplikace částí genomu tvořící redundantní genové sekvence (Hooper a Berg, 2003), 3) ztráta genů a degenerace sekvencí (Andersson a Kurland, 1998), 4) inverze a translokace, které způsobují větší genomové přestavby uvnitř organismu (Moran, 1996; Moran a Telang, 1998; Wernegreen a Moran, 1999)

1.3.1 Horizontální přenos genů

Horizontální přenos genů (HGT, Horizontal Gene Transfer) představuje jednu z hlavních sil formujících prokaryotní genomy. K tomuto přenosu dochází pomocí transformace, transdukce a konjugace. Transformace probíhá většinou mezi jedinci stejného druhu, k transdukci dochází přes viry a fágy, ke konjugaci dochází pomocí plazmidů (Steinert et al., 2000). Vliv HGT na změny v genomu je větší než vliv chromozomových aberací. Zatímco mutace vedou obvykle jen k malým genetickým změnám v buňce, HGT zahrnují větší přestavby, které vedou ke vzniku nových funkcí a schopností adaptovat se na změny v prostředí. Tyto nově získané geny (xenology) mohou vést i k rozdělení druhu a diverzifikaci mikrobiální linie (Lawrence, 2002; Gogarten, 2002). K detekci horizontálně přenesených genů slouží analýzy G+C obsahů, kodonové preference a anomálie v topologii fylogenetických stromů (Garcia-Vallve, 2003; Novichkov, 2004). Horizontální přenos je méně častý mezi geny kódujícími subjednotky makromolekulárních komplexů účastnících se translace, transkripce a replikace. Častěji k němu dochází u genů kódujících metabolické enzymy (Rivera et al., 1998).

Xenologní geny mohou zkreslovat fylogenetické analýzy a deformovat topologii fylogenetického stromu (Lawrence, 2002; Gogarten, 2002; Novichkov, 2004). U druhů s častějšími událostmi HGT se vyskytuje zvýšená heterogenita i u genů používaných pro fylogenetické studie (např. gen 16S rRNA). Tato heterogenita je považována za další potenciální problém u těchto analýz.

1.3.2 Duplikace genů

Duplikace genů je jedním z řady způsobů, jak bakterie zvyšují obsah genomu (Snel et al., 2002). Při tomto procesu se homologické sekvence vzniklé duplikací jednoho původního genu postupně vzdalují a dávají vznik tzv. paralogním genům. Paralogní genomové rodiny tvoří významnou část (až 44 %) bakteriální kódující kapacity (Hooper a Berg, 2003) a značně přispívají k fenotypovým rozdílům mezi bakteriálními liniemi (Gevers et al., 2004). Zahrnutí těchto paralogních sekvencí do fylogenetických analýz však často vede k chybnému určení původu organismů. Výsledný kladogram spíše odráží evoluci daného genu než daných organismů. Pro rekonstrukci fylogenetických vztahů je tedy nezbytné rozpoznání a použití

výhradně ortologních genů, tj. homologických genů vzniklých u všech potomků z identické kopie genu (Tatusov et al., 1997; Tatusov et al., 2000).

1.3.3 Ztráta genů

Pokud není selekční tlak pro udržení některých genů dostatečně silný, pak se tyto geny inaktivují a postupně mizí. Inaktivace a ztráta genů se týká hlavně obligátních parazitů a vnitrobuněčných mutualistů (Moran, 1996; Anderson a Kurland, 1998; Mira, et al., 2001). Bakterie žijící v relativně stabilním prostředí jsou schopny chránit před mutacemi a ztrátami jen omezené množství informace. Většinou jsou to geny nezbytné pro replikaci a přenos bakterií, kromě toho si uchovávají specifické biosyntetické funkce důležité i pro svého hostitele (Wernegreen a Moran, 1999; Zientz, 2001; Klasson a Anderson, 2004). Proces redukce a degradace genomu je charakteristický zrychlenou sekvenční evolucí, genetickým driftem, fixací mírně nevýhodných mutací způsobenou tzv. „bottleneck“ efektem (Moran, 1996; Anderson a Kurland, 1998; Wernegreen a Moran, 1999; Ochmann et al., 1999). Příčiny těchto jevů jsou přímým důsledkem intracelulárního způsobu života charakteristického omezenou možností rekombinací a transovariálním přenosem symbiontů mezi generacemi hostitele (vertikální přenos, Mira a Moran, 2002).

Ztráty genů jsou dalším potenciálním problémem pro fylogenetické analýzy. Geny vyskytující se u volně žijících bakterií mohou mít vhodné vlastnosti pro zkoumání vztahů mezi bakteriemi, ale nemusí být k dispozici u bakterií, ve kterých proběhla značná redukce genomu (např. u symbiotických bakterií).

1.3.4 Změna nukleotidového složení a rychlost evoluce

Kromě ztráty genů vede degenerace genomu ke změně nukleotidového složení. Rozsah obsahu GC je u jednotlivých druhů bakterií obrovský, od 20,1 % (*Buchnera aphidicola* str. *Cc*) po 72 % (*Nocardioides* sp.). Tyto rozdíly jsou často způsobeny odlišnou životní strategií bakterií. Bakterie žijící v extrémních podmínkách (např. termofilní bakterie) mají mnohem větší GC obsah, který zvyšuje odolnost DNA proti denaturaci způsobené vysokými teplotami. Na druhou stranu, u intracelulárních bakterií se vyskytují tendence k degradaci genomu, a tedy zvyšování obsahu AT bází (Wernegreen, 2002; Klasson a Anderson, 2003).

Zrychlená sekvenční evoluce, degradace genomu a rozdíly v nukleotidovém složení ovlivňují výsledky fylogenetických analýz. Způsobovat mohou například jev nazývaný

artefakt přitahování dlouhých větví („long branch attraction“, LBA). Díky zrychlené evoluci u některých sekvencí, které výrazně divergovaly od jejich výchozího stavu, dochází ve výsledném fylogramu ke vzájemnému přitahování těchto sekvencí na základě náhodně vzniklých shod, popřípadě k přitahování takových skupin ke kořeni stromu (Philippe a Laurent, 1998; Bergsten, 2005).

Genomové složení má také vliv na frekvenci výskytu některých kodonů a aminokyselin. AT bohaté genomy u intracelulárních bakterií mohou dokonce silně ovlivňovat strukturu a stabilitu proteinů (Palacios a Wernegreen, 2002; Sharp et al., 2005). Zdá se, že změny AT obsahu u aminokyselin nejsou výhodou. Na druhé straně upřednostňování určitých kodonů redukuje počet genů pro tRNA a snižuje tím energetické náklady na syntézu DNA (Palacios a Wernegreen, 2002; Reis et al., 2004).

1.4 Způsoby řešení a eliminace artefaktů

Spolehlivost fylogenetických analýz je do značné míry ovlivněna popsanými typy evolučních procesů v genomu prokaryot. V současné době je vyvíjena snaha o pochopení těchto jevů, odstranění znaků zodpovědných za anomálie v topologii a identifikaci genů vhodných pro fylogenetické analýzy (ortologní geny). Každá metoda má odlišnou citlivost k různým artefaktům. Kombinací metod a použitím různých analýz se můžeme vyhnout některým anomáliím v topologii fylogenetického stromu. Používání více genů (multigenní analýzy) ve fylogenetických analýzách snižuje nedostatky konkrétního genu zahrnutého v jednogenné analýze (Gontcharov et al., 2004; Devulder et al., 2005). Tato metoda se využívá především v souvislosti s použitím 16S rRNA v kombinaci s jinými geny (Devulder et al., 2005). Tento způsob snižuje efekt vnitrodruhové heterogenity u 16S rRNA a zvyšuje citlivost analýzy u blízce příbuzných druhů (vhodné je vybrat geny s vhodnými variabilními úseky pro zvýšení rozlišovací schopnosti). Přidání více genů do fylogenetických analýz může také výrazně snížit artefakt LBA. Z analýz lze také vyloučit geny s velmi odlišnými sekvencemi nebo geny vzniklé různými evolučními událostmi (HGT, duplikace). K detekci nevhodných sekvencí lze v současné době využít COGs databáze (clusters of orthologous groups database; Tatusov, 1997; Huynen and Bork, 1998).

1.4.1 COGs databáze

Pro fylogenetické analýzy je v první řadě nezbytné určit geny s vhodnými vlastnostmi (homologie, přítomnost u většiny bakterií, aktivita, dostupnost, vhodná míra variability atd). V nedávné době byla zhotovená databáze ortologních skupin (COGs; Tatusov, 1997) sloužící jako prostředek pro identifikaci ortologních sekvencí. COGs databáze je tvořena skupinou genů proteinů nebo domén tříděných do sady ortologů (Tatusov et al., 1997; Tatusov et al., 2000; Yanai et al., 2001). COGs databáze se využívá pro fúzní analýzy (Yanai et al., 2001), odvozování funkcí proteinů a jejich klasifikaci (Tatusov, 2000), identifikaci konzervovaných rodin a pro fylogenetické analýzy (Huynen and Bork, 1998). Studie klastrů ortologů z bakteriálních genů (COGs) umožňuje detekovat ortologní geny, nebo pomáhá odhalit skryté paralogy a jiné geny zkreslující topologii fylogenetického stromu (Bapteste et al., 2005).

1.4.2 Primery

Nejpoužívanější metoda pro amplifikaci studovaných genů je princip polymerázové řetězové reakce (PCR), která byla vynalezena v roce 1985 Kary Mullisem. Jedná se o citlivou metodu sloužící k získání obrovského množství kopií vzorku DNA. Namnožená DNA pak slouží pro sekvenční analýzy. Použití PCR je založeno na tvorbě primerů, které se přesně hodí k cílové sekvenci.

Sekvence primeru musí mít vhodnou délku (obvykle 17-25 bází), GC obsah (40 %-60 %) a jeho sekvence by měla být co nejvíce shodná se sekvencí templátu v místě vazby, zejména na 3'-konci primeru. Zásadním parametrem je nasedací teplota primeru (annealing), která se odvozuje z délky a sekvence primeru (Chen et al., 2003; Gadberry, 2005). Tato teplota je blízká teplotě tání templátu T_m , a současně by měla být shodná pro oba použité primery. K přesnějšímu výpočtu nasedací teploty byla vytvořena řada programů, z nichž některé jsou dostupné v interaktivní formě přes Internet (<http://www.idtdna.com/analyzer/Application/OligoAnalyzer>; <http://www.promega.com/biomath/calc11.htm>). Primery by měly být navrženy tak, aby nevytvářely sekundární struktury (vlásky), netvořily dimery a nerozpoznávaly větší množství nespecifických míst v genomu (Chen et al., 2003).

PCR metodu mohou provázet komplikace. Příkladem jsou odlišné vlastnosti zkoumaných fragmentů a primerů, takže je nezbytné některé parametry PCR reakce optimalizovat. Optimalizace provádíme především změnou koncentrací dNTP, $MgCl_2$ a Taq polymerázy. Koncentrace nukleotidů v reakci je 20 až 200 μ mol, snížením koncentrace zvyšujeme

specifičnost reakce. Koncentrace $MgCl_2$ musí být vyšší než u dNTP, protože Mg^{+2} interaguje se zápornými fosfátovými skupinami nukleotidů. V některých případech se $MgCl_2$ nepřidává, např. pokud chceme zvýšit aktivitu Taq polymerázy a dostat větší výtěžek PCR produktu (Chavali et al., 2005).

Dalším způsobem úpravy je modifikace teploty, při které dochází k navázání primerů (annealing). Vyšší teplota zvyšuje specifičnost primerů ke komplementární sekvenci na DNA, a tím zvyšuje i přesnost PCR reakce, nicméně příliš vysoká teplota může zcela znemožnit průběh reakce. U nižších teplot dochází k nasedání i méně specifických primerů. Na druhou stranu, nízká teplota může způsobit vznik falešně pozitivního signálu.

Jedním z kritických faktorů citlivosti PCR je stupeň degenerace primeru. Sekvence primeru je degenerovaná, pokud obsahuje na jedné pozici různé báze (Gadberry, 2005). Různé stupně degenerace mohou vést ke komplikacím v hledání vhodné teploty a délky primeru. Málo degenerované primery mohou být úspěšnější v izolování sekvencí blízce příbuzných druhů, ale obecně selhávají, pokud má sekvence neurčitou nebo vzdálenou příbuznost. Zvýšení pravděpodobnosti úspěchu se dosáhne pomocí primerů s degenerovanými oblastmi zejména na jeho 3' konci (Jabado et al., 2006). Více degenerované primery se dají použít na širokou škálu organismů i v oblastech sekvencí vykazující vyšší variabilitu tj. úspěšněji amplifikují neznámou sekvenci, která je homologní k sekvencím známých proteinů z různých organismů nebo může být příbuzná k známým proteinovým sekvencím patřících do multigenní rodiny uvnitř organismu. Naproti tomu vyšší degenerační stupeň primeru snižuje jeho specifitu (Rose et al., 1998; Gadberry, 2005; Jabado et al., 2006)

2. Cíle práce

Cílem této práce bylo vyhledat a testovat nové geny, potenciálně využitelné pro řešení fylogenetických vztahů v rámci skupiny *Enterobacteriaceae*. Tyto geny by měly kompenzovat komplikace, které brzdí studium evoluce symbiomy mezi bakteriemi a hmyzem. Mělo by se jednat o ortologní geny, přítomné u většiny bakterií, bez ohledu na míru redukce a degenerace genomu. K dosažení tohoto cíle bylo zvoleno několik dílčích kroků:

1. Identifikovat vhodné geny z databáze COG na základě jejich vlastností a porovnáním s dostupnými kompletními genomy symbiotických a volně žijících bakterií ve skupině *Enterobacteriaceae*
2. Sestavit primery pro vybrané geny a testovat jejich použitelnost na modelové symbiotické bakterii
3. Provést fylogenetickou analýzu a zhodnotit použitelnost navržených primerů a genů pro další studie

3 Materiál a Metodika

3.1 Vyhledávání potenciálně vhodných genů

Vhodné ortologní geny byly vybírány pomocí COG databáze. Podle této databáze vytvořili autoři Lerat et al. (2003) seznam genů vyskytujících se ve většině enterobakterií. Z tohoto seznamu byly vyhledávány geny, které jsou přítomny i ve známých genomech symbiotických bakterií. Ke zjištění tohoto faktu byly použity genomy bakterií *Buchnera* (NC_002528), *Baumannia cicadellinicola* str. Hc (NC_007984), *Candidatus Blochmannia floridanus* (NC_005061), *Sodalis glossinidius* (NC_007712) a *Wigglesworthia* (NC_004344), které byly vyhledány v NCBI databázi (http://www.ncbi.nlm.nih.gov/genomes/static/eub_g.html). Dalším faktorem pro výběr vhodných genů byla délka sekvencí (500-2000 bp). Z vybraných sekvencí byly zhotoveny alignmenty obsahující vždy tři sekvence (*Escherichia coli*, *Pseudomonas aeruginosa* a *Buchnera* spp). Pomocí těchto referenčních alignmentů byly ověřovány některé vlastnosti genů, např. přítomnost vhodných konzervativních úseků potřebných pro sestavení primerů a optimální množství variabilních oblastí nesoucích informaci pro fylogenetické analýzy. Vybrané geny byly dále použity k navržení primerů.

3.2 Navrhování primerů

Pro navržení nových primerů byl použit program PrimeClade (<http://dousta.umsl.edu/cgi-bin/primacade.cgi>), do kterého byl vložen alignment ve formátu Nexus (Maddison et al. 1997; soubor uchovávající systematická data ve formátu použitelného pro počítačové programy). V programu lze manipulovat s parametry, které mohou být specifické pro vytváření primeru. Hlavním parametrem je maximální počet degenerovaných bází, který umožňuje vyhledání primerů obsahující 0 až 5 degenerovaných pozic. Program poskytuje uživateli možnost eliminovat vysoce konzervované oblasti, které by se mohly vyskytovat v paralogních genech (Gadbery et al., 2004) nebo by mohly snižovat citlivost fylogenetických analýz. Vybrané sady primerů byly otestovány podle sekvence daného genu *Escherichia coli* a symbiotických bakterií (*Buchnera*, *Sodalis*, *Blochmania*, *Wigglesworthia*) programy PrimeSelect, *In silico* PCR amplification (<http://insilico.ehu.es/PCR/>) a programem IDT SciTools (<http://www.idtdna.com/analyzer/Application/OligoAnalyzer>). Podle zadaných

parametrů programy zhodnotily jednotlivé primery. Primery, které prokazovaly nejlepší vlastnosti, tj. potenciální schopnost amplifikovat požadovaný gen u testovaných sekvencí, byly vyrobeny firmou Generi-Biotech (Hradec Králové) a použity v dalším testování.

3.3 Izolace DNA

Jako referenční vzorky pro testování nově navržených primerů byly použity tři symbiotické bakterie, bakterie rodu *Arsenophonus* z plošnice *Triatoma infestans*, symbionti vši *Haemotopinus suis* a *Pediculus capitis*, a patogenní bakterie *Proteus vulgaris*.

Pro izolaci DNA byla použita souprava izolačního protokolu DNeasy Tissue Kit (Quiagen). Postup byl proveden podle pokynů v protokolu a izolovaná DNA uskladněna v mrazáku při -20°C.

3.4 PCR

Pro amplifikaci DNA byla použita metoda PCR (polymerázová řetězová reakce). Reakční směs pro PCR obsahuje DNA templát (získaný z předchozího kroku), redestilovanou vodu, směs čtyř nukleotidů dNTP, dvojici primeru, Taq polymerázu, pufr a MgCl₂. Současně byly připravovány kontroly pro detekci kontaminace (negativní kontrola) nebo k rozpoznání izolované DNA (pozitivní kontrola).

Nové primery byly použity k amplifikaci DNA z referenčního vzorku, pro pozitivní kontrolu byly použity univerzální primery F40 a R1060 (Křížek, 2005). PCR reakce proběhla v přístrojích Eppendorff Mastercycler a Eppendorff Termocycler Gradient, v rozmezí teplot 45-60 °C. Pokud k amplifikaci nedošlo nebo produkty byly špatné, byla vyzkoušena řada postupů s cílem eliminovat tyto problémy. Bylo manipulováno s reasociační teplotou a koncentrací hořčnatých iontů (viz kap. 1.4.2.), byly použity přístroje Eppendorff Termocycler Gradient nebo procedura alternativních PCR („touch down“ a Long PCR).

„Touch down“ PCR je varianta standardní PCR, která byla použita k odstranění amplifikace nespecifických sekvencí. Teplota s rostoucím počtem cyklů klesá. Vyšší reasociační teplota při prvních cyklech zajistí, že jsou amplifikovány správné fragmenty, ale s nízkým výtěžkem. Další cykly probíhají při nižší teplotě, a tím se zvýší výkonnost PCR.

Long PCR (Fermentas) byla používána v případech, kdy pomocí běžné PCR nebylo dosaženo požadovaných výsledků. Jedná se o kombinaci dvou polymeráz (Taq a druhá termostabilní polymeráza s 3'→5' exonukleázovou aktivitou) a speciálního pufru, jehož

složení působí proti depurinaci a tvorbě niků během reakce. Tato směs polymeráz, stejně jako HotStar polymeráza, připojuje adenin na 3' konec.

3.5 Elektroforéza

Po izolaci DNA a PCR byla provedena kontrola výsledku pomocí elektroforézy na 0,8% agarovém gelu. Gel byl připraven rozpuštěním agarózy v příslušném množství elektrodového pufru a byl ohřán v mikrovlnné troubě. Po zchladnutí je gel zcela ponořen do 1x TAE pufru (Tris-acetát-EDTA). Do jamek v gelu je nanášen standard (ladder) a ostatní vzorky. Pro zviditelnění DNA bylo použito barvivo Sybr Green, které bylo před nanesením do jamek smícháno se vzorky. Elektroforéza probíhala při 80-110 V po dobu 30 minut. Po skončení elektroforézy byl výsledek zkontrolován v UV světle pomocí transluminátoru. Gely byly zdokumentovány digitálním fotoaparátem. Velikosti PCR produktů byly odečítány podle markeru 100bp DNA Ladder nebo 1kb DNA Ladder.

3.6 Klonování PCR produktů

Pro klonování DNA je nejprve nutno vytvořit rekombinantní DNA, která sestává z klonovaného fragmentu a tzv. klonovacího vektoru. Ligační směs byla sestavena z PCR produktu a ligačního pufru, T4 ligázy a pGEM-T Easy vektoru (Promega). Smíchaná směs se inkubuje 2 hodiny při pokojové teplotě nebo přes noc při 4 °C.

3.7 Kultivace a selekce buněk

Rekombinantní DNA byla následně vnesena do buňky hostitele, bakterie *Escherichia coli* (kmen neprodukující restriční endonukleázy), kde je DNA při růstu buněčné kultury mnohonásobně replikována.

Kompetentní buňky *Escherichia coli* byly rozmrazeny na ledu a smíchány s 2 µl ligační směsí. Inkubace trvala 20 minut. Potom byly buňky vystaveny teplotnímu šoku (30-40 sekund při 42° C). Do ochlazené směsi bylo přidáno 125 µl SOC média a následovala inkubace na třepačce po dobu 1,5 hodiny při teplotě 37° C.

Po inkubaci byly transformované bakteriální buňky vysety na agarové plotny. Tyto plotny byly připraveny rozlitím do 25 ml Petriho misek rozehřátého 250 ml LB média (6.25 g LB od firmy Serva, 5g bactoagar od firmy ICN Biomedicals, 250 ml H₂O) smíchaného s

ampicilinem (teplota média nesmí být vyšší než 65 °C). Ještě před výsevem bylo na každé plotně rozetřeno po 40ml X-Gal (5-bromo-4-chloro-3-inodyl-beta-D-galaktosid)

Misky obrácené dnem vzhůru byly inkubovány přes noc při teplotě 37 °C. Pozitivní kolonie byly přeneseny do nového tekutého LB média s ampicilinem a inkubovány na třepače přes noc při 200 rpm a 37° C. Suspenzní kultura bakteriálních buněk byla zakoncentrována pomocí centrifugace trvající po dobu 5 minut při 5000 rpm.

3.8 Izolace plazmidové DNA

Plazmidová DNA byla izolována pomocí komerčního kitu JETQuick Plasmid Miniprep Spin (Genomed) podle návodu. Po izolaci byla provedena restriční kontrolní reakce s cílem ověřit, zda se námi požadovaný úsek DNA začlenil do izolovaného plazmidu. Pro restrikcii byl použit enzym EcoR1. Směs pro restriční reakci byla inkubována po 1-2 hodiny při 37° C. Poté došlo k vyhodnocení na agarovém gelu (1%).

3.9 Sekvenační reakce

Sekvenační reakce byla provedena pomocí sekvenačního kitu Big Dye Terminator v3,1 Cycle Sequencing Kit (Applied Biosystems). Vzorek plazmidové DNA (4 µl vzorek + 11 µl ddH₂O) byl inkubován 3 minuty při 96° C (denaturace). Po skončení denaturace se do reakční směsi přidal ABI mix, sekvenační pufr a primery T7 (5'-TAATACGACTCACTATAGGG-3') a SP6 (5'-ATTTAGGTGACACTATAG-3'). Byl připraven STOP roztok, který se dodal po skončení reakčního cyklu do reakční směsi. Vzorky byly přečištěny a sekvenovány pomocí automatického sekvenátoru ABI Prism 3130 (Ústav molekulární biologie rostlin AVČR).

3.10 Zpracování sekvencí

Výsledné sekvence byly zpracovány programem DNASTAR (Lasergene, INC., 2002). Nově získané sekvence byly taxonomicky zařazeny pomocí identifikace nejbližších příbuzných sekvencí v internetovém programu Blast Search databáze NCBI (<http://www.ncbi.nih.gov/BLAST/>). Čtecí rámeček sekvence byl vyhledán pomocí programu ORF finder (<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>). Na základě tohoto předběžného určení byla sestavena matice pro následné fylogenetické analýzy. Matice byla sestavena programem Megalign s užitím metody ClustalW (Thompson et al., 1994). Do matice byly

dány sekvence z kompletně osekvenovaných genomů enterobakterií z databáze NCBI (http://www.ncbi.nlm.nih.gov/genomes/static/eub_g.html). Jako outgroup byly použity sekvence z druhů *Haemophilus* a *Pasteurella*. Sekvence byly alignovány v aminokyselinové podobě. Výsledný alignment by analyzován programem PAUP* (Swofford, 1998). Fylogenetické analýzy byly prováděny v aminokyselinové, nukleotidové a kombinované (nukleotidy + aminokyseliny) podobě pomocí metody maximální parsimonie (MP) s heuristickým vyhledáváním (heuristic search, hs) pomocí TBR (tree bisection-reconnection) algoritmu se 30 opakováními při náhodném přidávání sekvencí. Analýzy v nukleotidové podobě byly provedeny také ještě metodou maximální likelihood (ML) s výchozím (default) nastavením parametrů, které odpovídá evolučnímu modelu HKY 85.

4. Výsledky

4.1 Výběr vhodných sekvencí

Geny potenciálně vhodné pro rozšíření setu fylogenetických markerů byly vybrány ze seznamu (obsahujícího 205 genů), který byl zhotoven autory Lerat et al. (2003). Z tohoto seznamu byly nejprve odebrány geny, které se nevyskytovaly u některých symbiotických bakterií např. *ygjD*, *smpB* (*Blochmania*), *secE* (*Wigglesworthia*), *riM*, *rbfA* (*Baumannia*) atd. Pomocí alignmentu byly vyloučeny geny s vysoce variabilními sekvencemi neumožňující sestavení primerů (např. *apaH*, *dnaJ*, *lgt*, *rplB*, *rplC*) a byl vytvořen seznam genů s vhodnými konzervovanými oblastmi (viz. tabulka č.1). Tento seznam obsahoval 48,6% genů z původního setu.

Tab1. Seznam genů s vhodnými konzervovanými oblastmi

Název genu	Funkce genu	Velikost genu u <i>Buchnera</i> (pb)	Velikost genu u <i>E. coli</i> (bp)	Obsah GC (%) <i>Buchnera</i>	Obsah GC (%) <i>E. coli</i>
<i>adk</i>	Nukleotidy a nukleosidy	648	645	28,65	51,78
<i>alaS</i>	tRNA aminoacylace	2637	2631	24,76	53,63
<i>carB</i>	Syntéza pyrim. Nukleotidů	3240	2640	32,10	56,36
<i>cca</i>	Transkripce	1245	1239	26,51	56,26
<i>cysS</i>	tRNA aminoacylace	1395	1386	25,23	54,47
<i>ffh</i>	Kat. proteinů a peptidů	1356	1362	25,96	54,19
<i>frr</i>	Translační faktor	558	558	26,88	50,90
<i>glyS</i>	tRNA aminoacylace	2073	2070	23,40	55,41

glnS	tRNA aminoacylace	1716	1665	25,35	52,49
htpG	Degradace proteinů	1875	1876	25,17	52,83
ileS	tRNA aminoacylace	2803	2817	26,60	55,77
infA	Translační faktor	219	219	30,14	48,40
infB	Translační faktor	2595	2673	28,52	53,42
nusA	Transkripční faktor	1491	1488	33,00	52,76
nusG	Transkripční faktor	546	546	33,77	51,65
pheS	tRNA aminoacylace	990	984	25,35	53,86
pheT	tRNA aminoacylace	2388	2388	24,54	54,36
pnp	Degradace RNA	2124	2205	32,77	53,70
purA	Syntéza purin. Nukleotidů	1302	1299	28,96	53,89
purB	Syntéza purin. Nukleotidů	1371	1371	26,40	52,30
purH	Syntéza purin. Nukleotidů	1578	1590	27,12	56,29
pyrE	Syntéza pyrim. Nukleotidů	642	642	27,26	53,12
rplA	Syntéza proteinů	696	705	29,74	51,63
rplC	Syntéza proteinů	630	603	32,54	51,43
rplD	Syntéza proteinů	606	606	32,67	52,97

rplF	Syntéza proteinů	558	534	31,36	47,94
rplJ	Syntéza proteinů	498	498	28,71	52,41
rplK	Syntéza proteinů	429	429	29,60	52,21
rplM	Syntéza proteinů	429	429	28,67	50,58
rplN	Syntéza proteinů	369	372	34,42	51,08
rplQ	Syntéza proteinů	393	384	30,53	53,65
rplT	Syntéza proteinů	357	357	32,21	48,18
rplU	Syntéza proteinů	327	312	25,69	48,40
rplX	Syntéza proteinů	315	315	26,03	46,03
rpmA	Syntéza proteinů	255	258	32,55	50,00
rpsA	Syntéza proteinů	1677	1674	30,41	51,19
rpsB	Syntéza proteinů	735	726	28,71	51,38
rpsE	Syntéza proteinů	504	504	34,72	50,20
rpsJ	Syntéza proteinů	312	312	36,86	55,45
rpsK	Syntéza proteinů	396	390	36,87	48,46
rpsL	Syntéza proteinů	375	375	35,73	53,33

rpsM	Syntéza proteinů	357	357	30,81	51,82
rpsN	Syntéza proteinů	306	306	35,62	52,61
secY	Degradace proteinů	1314	1332	28,39	49,70
serS	tRNA aminoacylace	1281	1293	27,80	52,67
thrS	tRNA aminoacylace	1929	1929	27,11	49,56
trmU	Transkripce	1107	1107	26,20	53,75
valS	tRNA aminoacylace	2868	2856	25,17	54,59

Z tohoto seznamu byly vybírány geny s vhodnou celkovou délkou a vhodnou délkou variabilních a konzervativních oblastí v daném alignmentu. Konzervativní oblasti, které umožňují univerzálnější nasedání primeru, obsahovaly minimálně 20bp pro nasedání primeru a mezi nimi obsahovaly variabilní oblasti v rozmezí 500-1500bp.

4.2 Výběr primerů

Pomocí internetového programu PrimeClade byla získaná sada primerů (15% z původního setu) ze sekvencí genů *alaS*, *adk*, *carB*, *cysS*, *ffh*, *ileS*, *purA*, *purB*, *rho*, *rplA*, *secY*, *thrS*, *trmU*. Programem *In silico* PCR amplification (<http://insilico.ehu.es/PCR/>) bylo zjištěno, jaké primery budou pravděpodobně nasedat na templátovou DNA během PCR reakce. Pomocí tohoto programu byly vyloučeny další nevhodné primery (13% z původního setu), které nasedaly nespecificky nebo nenedaly vůbec, a byly vybrány primery pro geny *carB*, *ileS* a *purB*, které podle simulace nasedaly specificky na templátovou DNA. Jejich vlastnosti jsou popsány v tabulkách číslo 2 a 3. Dále byly geny *carB*, *ileS* a *purB* testovány programem IDT SciTools (<http://www.idtdna.com/Home/Home.aspx>). Pomocí tohoto programu byly ověřovány jednotlivé vlastnosti primerů (tj. tvorba dimerů a vlásenkových struktur, T_m , obsah GC, délka atd.). Primery byly schopny vytvářet sekundární struktury, ale ty se daly odstranit

vyšší teplotou. Sekundární struktury se rozpadaly při nižší teplotě než byla T_m . Ostatní vlastnosti byly popsány v následující tabulce.

Tab2. Vlastnosti získaných primerů

Název primeru	Délka primeru (bp)	T_m (°C)	GC obsah (%)
purBr	17	52,4	57,8
purBf	20	50,3	39,2
carBr	17	51,8	53,9
carBf	18	48,1	42,6
ileSr	20	56,0	54,2
ileSf	18	57,2	62,0
ile2Sr	20	51,3	40,0
ile2Sf	20	56,3	54,2

Tab3. PCR primery použity pro amplifikaci (univerzální i nově navržené)

Název primeru	Sekvence	Zdroj	Použití dvojice primerů	Velikost produktu
23SR	5'-TTRGAGGATGGYCCCCCAT -3'	Šofrová, 2005	Amplifikace celého 16S-23S rDNA spaceru	cca 1200 bp
F40	5'-GCGGCAAGCCTAACACAT -3'	Křížek, 2005		
R1060	5'-CTAACCCAACATTTCTCAACACGAG -3'	Křížek, 2005	S primerem F40 nespecifická amplifikace 16s rDNA	cca 1200 bp
purBr	5'-GGCATRRYVGADGADCC -3'	vlastní	Amplifikace genu pro enzym katalyzující syntézu purinů	cca 635 bp
purBf	5'-ACYAAHCAYGAYRTNAAAGC -3'	vlastní		
carBr	5'-AYBCCCATVACTTCHCC -3'	vlastní	Amplifikace genu pro enzym Carbomyl-fosfát syntázu	cca 900 bp
carBf	5'-BAAYTGyAAyCCKGAAAC -3'	vlastní		
ileSr	5'-GTCCANRKMGRGTDGTCCA -3'	vlastní	Amplifikace genu pro enzym signální peptidázy	cca 500 bp
ileSf	5'-TGGACNACNACDMYNTGGAC -3'	vlastní		
ile2Sr	5'-TTNCCNABNGAYTTDGACAT -3'	vlastní		
Ile2Sf	5'-CCDGGHTGGGAYTGYCAYG -3'	vlastní	Amplifikace genu pro enzym signální peptidázy	cca 1100 bp

4.3 Získané sekvence z nových primerů

Geny *carB*, *ileS* a *purB* byly použity k dalším analýzám. K úspěšné amplifikaci docházelo u rodu *Arsenophonus* pro gen *purB* při 57° C. Ostatní primery amplifikovaly špatný úsek DNA. Primery *purB* byly použity k PCR amplifikaci DNA ze symbiontů vší (*Haemotopinus suis*, *Pediculus capitis*) a patogenní bakterie *Proteus vulgaris*. U symbiontů vší *Haemotopinus suis* a bakterie *Proteus vulgaris* nevznikl výsledný PCR produkt. Sekvence získané pomocí primerů pro *purB* z druhu *Pediculus capitis* neodpovídaly hledané sekvenci *purB*.

Pro sekvenci (*purB*) získanou z genu symbionta rodu *Arsenophonus* byla vytvořena matice z dostupných sekvencí pro protein *purB* u enterobakterií. Výsledný alignment (dlouhý 570 nukleotidových pozic, tj. 190 aminokyselinových) byl jednoznačný a nevyžadoval úpravy. Jediná inzerce v aminokyselinové sekvenci se vyskytovala na 27. pozici u rodu *Blochmannia*.

Tab4. Seznam enterobakterií použitých do matice pro gen *purB*

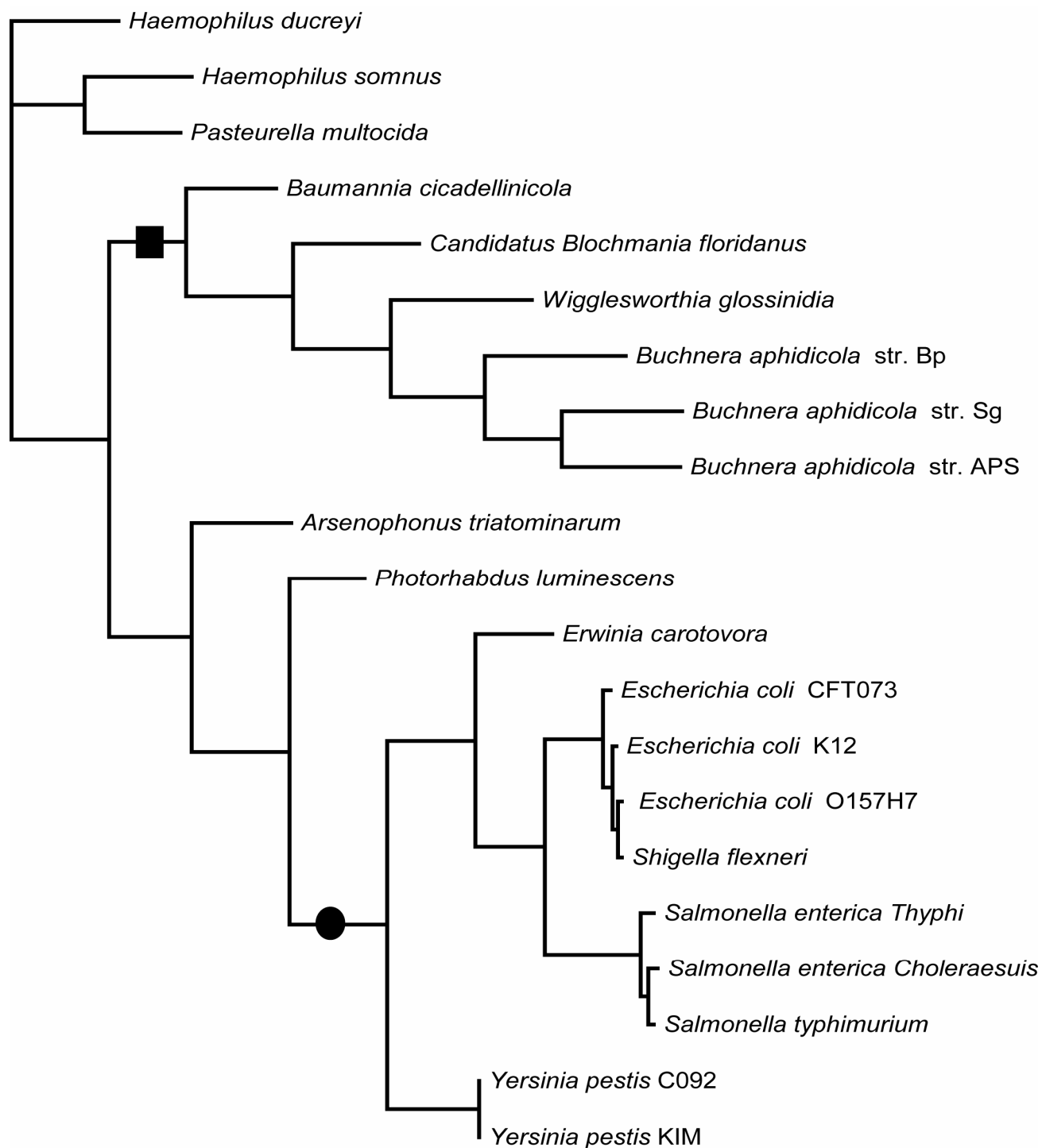
Přístupová čísla	Druh bakterie	Délka sekvence
NC_007984	<i>Baumannia cicadellincola</i> str. Hc	1392
NC_002528	<i>Buchnera ahidicola</i> str. APS	1371
AE016826	<i>Buchnera aphidicola</i> str. Bp	1383
NC_004061	<i>Buchnera aphidicola</i> str. Sg	1371
NC_005061	<i>Candidatus Blochmannia floridanus</i>	1386
NC_002695	<i>Escherichia coli</i> 0157:H7	1371
NC_004431	<i>Escherichia coli</i> CFT073	1371
NC_000913	<i>Escherichia coli</i> K12	1371
NC_004547	<i>Ervinia carotovora</i> subs. <i>atroseptica</i> SCRI1043	1371
NC_002940	<i>Haemophilus ducreyi</i> 35000HP	1368
NC_008309	<i>Haemophilus somnus</i> 129PT	1368
NC_002663	<i>Pasteurella multocida</i> subsp. <i>multocida</i> str. Pm70	1371
NC_005126	<i>Photobacterium luminescens</i> subs. <i>laumondii</i> TT01	1385
NC_006905	<i>Salmonella enterica</i> Choleraesuis	1370
NC_003198	<i>Salmonella enterica</i> Thyphi	1371
NC_003197	<i>Salmonella typhimurium</i> LT2	1371

NC_004741	<i>Shigella flexneri</i>	1371
NC_004344	<i>Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis</i>	1395
NC_003143	<i>Yersinia pestis</i> CO92	1371
NC_004088	<i>Yersinia pestis</i> KIM	1380

4.4 Ověřování kvality informačního obsahu genu *purB*

Fylogenetické analýzy v nukleotidové podobě byly provedeny pomocí metod MP a ML. Pomocí MP metody byl získán jeden maximálně parsimonální strom o délce 1423 kroků. *Consistency index* (CI) tohoto kladogramu byl 0,49, *rescaled consistency index* (RC) byl 0.30. Výsledný kladogram znázorňoval rozdělení enterobakterií na dvě hlavní větve. Jednu tvořili primární symbionti a druhou volně žijící enterobakterie. Na stejné bázi s volně žijícími enterobakteriemi se vyskytoval bakterie rodu *Photorhabdus* a *Arsenophonus* (obr. č1.). Stejný výsledek byl získán metodou ML.

Metoda MP byla použita i pro fylogenetické analýzy v aminokyselinové podobě. Byly získány čtyři fylogenetické stromy (CI = 0,74 a RC = 0,48). Kombinovaná analýza vyprodukovala jeden fylogenetický strom s topologií shodnou s analýzou v nukleotidové podobě. V kombinované analýze byla hodnota konzistenčního indexu rovna 0,53 a rescale index roven 0,33.



■ skupiny volně žijících bakterií nebo patogenů obratlovců

● skupiny P-symbiontů

Obr.1 Kladogram pro gen *purB* získaný metodou MP. Matice zahrnovala všechny sekvence tohoto genu získané z databáze GenBank. Všechny sekvence byly zkráceny na délku genu *purB* (570 bp) získaného z bakterie rodu *Arsenophonus*.

5. Diskuze

5.1 Identifikace a navrhování primerů

Pro identifikaci a následné použití byly kandidátní geny vybírány ze seznamu autorů Lerat et al. (2003) vytvořeného na základě databáze COG. Důvodem použití tohoto seznamu bylo zjednodušení vyhledávání, protože seznam obsahoval už prověřené ortologní geny z COG databáze. Převážně se jednalo o geny účastnících se základních metabolických funkcí nezbytných pro růst a rozmnožování bakterií (např. geny kódující ribosomální proteiny nebo syntézu proteinů). U těchto genů se předpokládá, že nebyly ovlivněny různými evolučními událostmi, které by mohly zkreslovat topologii fylogenetického stromu (viz. kapitola 1.3 a 1.4). Výběr genů byl postupně snižován na základě několika parametrů. Řada genů byla vyloučena kvůli nevhodné délce, nebo nepřítomnosti u referenčních bakterií tj. u druhů, u kterých došlo k velké redukci genomu a je zde možnost ztráty genů, které jsou na seznamu (Moran, 1996; Anderson a Kurland, 1998). Dále byla prověřována množství konzervativních a variabilních oblastí v hledaných genech. Vysoká variabilita byla prověřována u intracelulárních symbiontů (*Buchnera*, *Wigglesworthia* atd). Právě u nich docházelo k rozsáhlým změnám v genomu (Mira, et al., 2001; Klasson a Anderson, 2003).

Výběr vhodných genových sekvencí byl postupně omezen na geny, ze kterých byla získána sada primerů za pomoci programu PrimerClade (Gadberry, 2005). Dále byly primery testovány pomocí internetových programů, podle kterých byly získány sety odpovídající všem požadavkům. Z celkového výchozího počtu genů byly nakonec použity tři, ale jen jedna sekvence byla úspěšně amplifikována u rodu *Arsenophonus*. Nicméně tento výsledek nemusí být konečný, seznam kandidátních genů použitelných pro fylogenetické analýzy může být rozšířen o další geny z COG databáze. Dalším způsobem je úprava nebo aplikace jiných metod než těch, které byly použity pro vyhledávání a testování vhodných sekvencí a následné úpravě získaných primerů. Jedna z možností je úprava vyřazených primerů, které nesplňovaly zadané parametry. Je zapotřebí získat vhodné primery s určitým počtem degenerovaných pozic (viz kap. 1.4.2). Možnosti získávání degenerovaných primerů jsou dvě. Můžeme změnit počet degenerovaných pozic nebo pro zvýšení stupně degenerace použijeme inosin. Primery byly sestavovány pomocí programu PrimerClade (Gadberry, 2005), který umožňoval maximální degeneraci na pěti pozicích. V současné době existuje několik programů pomáhajících sestavit primery s různými stupni degenerace např. CODEHOP (Rose et al., 1998;

<http://blocks.fhrc.org/codehop.html>). Naopak mohou být primery méně degenerovány díky vlivu kodonové preference. Určité kodony jsou upřednostňovány (dva nebo jen jeden) a tím odpadá nutnost využívání primerů, které jsou více degenerovány (Palacios a Wernegreen, 2002; Sharp et al., 2005).

Dále můžeme vlastnosti primeru také ovlivnit změnou jejich délky. Jiný způsob je výběr primeru přímo z vytvořeného alignmentu, tím můžeme volit různé pozice daných primerů, které mohly být vyloučeny na základě parametrů v programu požitého pro identifikaci primerů .

Vlastnosti primerů, které se neamplifikovaly nebo s nimi byl získán špatný pozitivní výsledek (nespecifická sekvence), mohou být ovlivněny prostředím během PCR reakce. V této práci bylo využito jen několik možností, jak ovlivnit výsledek PCR reakce (změna koncentrace různých komponent, teploty, „Touch down“ a Long PCR). Použitím jiných analogických PCR reakcí by mohlo vést k úspěšné amplifikaci těchto primerů.

5.2 Nasedání primeru na templátovou DNA

Ze sady primerů (primery pro gen *purB*, *carB* a dva páry primerů pro *ileS*) byl získán pozitivní výsledek PCR pouze s primery pro *purB*. Faktorů ovlivňujících úspěšnost amplifikace může být několik. Každý krok mohl být zatížen chybou. Některé chyby mohou být detekovány a odstraněny nebo minimalizovány změnami parametrů v jednotlivých krocích. K nejčastějším chybám dochází během PCR reakce. Při amplifikaci nemusí primery a DNA polymeráza identifikovat správnou cílovou sekvenci, ale mohou nasedat na alternativní místa podobná cílové DNA. Taková chyba se nejčastěji vyskytuje s používáním nových primerů, které byly navrženy podle jiných sekvencí z příbuzných druhů. Tyto primery nenacházejí homologní místa pro nasednutí na cílovou sekvenci. S tím souvisí i stupeň degenerace primeru (viz úvod, kapitola 1.4.2). Další komplikaci způsobuje tzv. kompetice

o primery. K tomuto jevu nejčastěji dochází, pokud jsme zacílili DNA s nízkým počtem kopií nebo DNA získána z izolace symbiontů byla příliš naředěná. Výsledkem je pak snadnější amplifikace kratších fragmentů oproti delším. K chybě mohlo dojít i během manipulace s DNA, která mohla být poškozena častým používáním a dlouhodobým skladováním.

5.3 Informační hodnota purB

Pro fylogenetické analýzy byla získaná sekvence purB izolována z bakterie rodu *Arsenophonus*. Pomocí BLASTu podle nukleotidových pozic byl potvrzen blízký vztah s enterobakteriemi, kde nejlepší hodnota vykazovala 75% podobnost se sekvencí purB u bakterií rodu *Photorhabdus*. Všeobecně se předpokládá, že bakterie rodu *Photorhabdus* tvoří s bakteriemi rodu *Arsenophonus* monofyletickou skupinu (Trowbridge et al., 2005). Provedenou fylogenetickou analýzou byl sice prokázán blízký fylogenetický vztah mezi těmito rody, ale v kladogramu tvořily parafyletickou skupinu (štěpily se za sebou na bázi skupiny volně žijících bakterií). Analýza je zatížena především nedostatkem reprezentativních vzorků genu purB v databázi GenBank pro bakterie skupiny *Enterobacteriaceae* (tento gen je v podstatě k dispozici jen u bakterií s kompletně osekvenovaným genomem). Dalším faktorem ovlivňujícím tuto analýzu byla délka získané sekvence, která se ukázala být nedostačující. Analýza pomocí genu purB nebyla schopna potvrdit monofylii skupiny tvořené rody *Arsenophonus* a *Photorhabdus*. Současné výsledky naznačují, že další sekvence genu purB, které byly k dispozici, vytváří dostatečnou strukturu a odráží fylogenetické vztahy (viz obr.č1). Lze předpokládat, že přidáním dalších vzorků by mohly mít sekvence genu purB informační hodnotu spolu s 16S rRNA při fylogenetických analýzách enterobakterií. Svými vlastnostmi by mohly kompenzovat nedostatky 16S rRNA analýz. V předchozích studiích bylo prokázáno, že 16S rRNA nedokáže správně odhalit vztahy mezi symbiotickými bakteriemi, u kterých došlo k výrazným změnám v genomu různými evolučními událostmi. Tyto změny mají velký vliv na strukturu 16S rRNA a její rozlišovací schopnost. Přidáním dalších genů do srovnávacích analýz by se vliv těchto faktorů eliminoval a získali bychom jasnější představu o fylogenetických vztazích mezi enterobakteriemi.

6. Závěr

Práce ukázala, že zvoleným postupem lze identifikovat geny vhodné pro doplnění fylogenetických souborů v rámci bakterií *Enterobacteriaceae* a sestavit primery schopné amplifikovat požadované geny. Vyhledávání vhodných genových sekvencí je omezeno několika faktory (přítomnost genu ve zkoumaných organismech, délka sekvence, míra variability v obou směrech). Z těchto sekvencí se podařilo získat sekvenci genu *purB*. Bylo zjištěno, že gen *purB* je sice schopen odrážet příbuzenské vztahy, ale má zřejmě nízký informační obsah pro přesné odvození topologických vztahů. Nicméně tento nízký informační obsah by bylo v budoucnu možné kompenzovat zvětšením vzorků (tj. získáním sekvence *purB* z dalších enterobakterií) a přidáním dalších genů do fylogenetických analýz.

Literatura

- Aksoy S., (1995) *Wigglesworthia* gen-nov and *wigglesworthia-glossinidia* sp-nov, taxa consisting of the mycetocyte-associated, primary endosymbionts of tsetse-flies. *Internat J System Bacteriol* 45 (4): 848-851
- Aksoy S., (2000) Tsetse—A haven for microorganisms. *Parasitol. Today* 16: 114–118
- Andersson S.G., Kurland C.G., (1998) Reductive evolution of resident genomes *Trends Microbiol.* pp. 263–268
- Baptiste E., Susko E., Leigh J., MacLeod D., Charlebois R.L., Doolittle W.F., (2005) Do orthologous gene phylogenies really support tree-thinking? *Evol Biol.* 5(1):33.
- Bergsten J., (2005) A review of long-branch attraction *Cladistics* 21 (2), 163–193.
- Chavali S., Mahajan A., Tabassum R., Maiti S., Bharadwaj D., (2005) Oligonucleotide properties determination and primer designing: a critical examination of predictions. *Bioinformatics* 21(20): 3918-3925
- Chen D. Q., Montllor C. B., Purcell A. H., (2000) Fitness effects of two facultative endosymbiotic bacteria on the pea aphid, *Acyrtosiphon pisum*, and the blue alfalfa aphid *A. kondoi*. *Entomol. Exp. Appl.* 95:315–323
- Chen L.L., Zhang C.T., (2003) Seven GC-rich microbial genomes adopt similar codon usage patterns regardless of their phylogenetic lineages. *Biochem. Biophys. Res. Commun.* 306 pp. 310–317.
- Chen S.H., Lin C.Y., Cho C.S., Lo C.Z., Hsiung C.A., (2003) Primer Design Assistant (PDA) a web-based primer design tool. *Nucleic Acids Res.*, 31:, 3751–3754
- Coenye T., Gevers D., Van de Peer Y., Vandamme P., Swings J., (2005) Towards a prokaryotic genomic taxonomy *FEMS Microbiol Rev* 29, 147–167
- Coenye T., Vandamme P., (2003) Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microbiol Lett* 228:45-49.
- Devulder G., Pérouse de Montclos M., Flandrois J.P., (2005) A multigene approach to phylogenetic analysis using the genus *Mycobacterium* as a model. *International Journal of Systematic and Evolutionary Microbiology*, 55 :293–302
- Fitz-Gibbon S.T., House C.H., (1999) Whole genome-based phylogenetic analysis of free-living microorganisms. *Nucleic Acids Res.* 27 pp. 4218–4222

- Gadberry M. D., Simon T. M., Doust A. N., Kellogg E.A., (2005) Primaclade - a flexible tool to find conserved PCR primers across multiple species. *Bioinformatics* 21(7): 1263-1264
- Galtier N., Gouy M., (1995) Inferring phylogenies from DNA sequences of unequal base compositions. *Proc. Natl. Acad. Sci. USA* 92:11317–11321
- Galtier N., Gouy M., Gautier C., (1997) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput. Appl. Biosci.* 12:543–548.
- Garcia-Vallve S., Guzman E., Montero M. A., A. Romeu A., (2003) HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Research.* pp. 187-189(3)
- Gevers D., Vandepoele K., Simillion C., Van de Peer Y., (2004) Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends Microbiol.* vol 12 p. 148-54.
- Gogarten P., Doolittle W. F., Lawrence J. G., (2002) Prokaryotic Evolution in Light of Gene Transfer. *Mol. Biol. Evol.*, 19: 2226 - 2238.
- Gontcharov A. A., Marin B., and Melkonian M., (2004) Are Combined Analyses Better Than Single Gene Phylogenies? A Case Study Using SSU rDNA and *rbcL* Sequence Comparisons in the *Zygnematophyceae* (*Streptophyta*). *Mol. Biol. Evol.* 21(3):612-624
- Herbeck J.T., Degnan P.H., Wernegreen J.J., (2005) Nonhomogeneous model of sequence evolution indicates independent origins of primary endosymbionts within the enterobacteriales (gamma-*Proteobacteria*). *Mol Biol Evol.* 22(3): 520-32.
- Hoffmeister M., Martin W., (2003) Interspecific evolution: microbial symbiosis, endosymbiosis and gene transfer. *Environmental Microbiology* 5: 641-649
- Hooper D., Berg O.G., (2003) On the nature of gene innovation: duplication patterns in microbial genomes. *Mol. Biol. Evol.* 20, pp. 945–954
- Hypša V., Dale. C., (1997) In vitro culture and phylogenetic analysis of “*Candidatus Arsenophonus triatominarum*,” an intracellular bacterium from the triatomine bug, *Triatoma infestans*. *Int. J. Syst. Bacteriol.* 47:1140–1144
- Huynen M.A. and Bork P., (1998) Measuring genome evolution. *Proc. Natl. Acad. Sci. USA* 95 pp. 5849–5856
- Jabado O., Palacios G., Kapoor V., Hui J., Renwick N., Zhai J., Briese T., Lipkin W. I., (2006) "Greene SCPrimer: a rapid comprehensive tool for designing degenerate primers from multiple sequence alignments" *Nucl Acids Res* 34 (22) 6605-6611
- Klasson L., Andersson S.G., (2003) Evolution of minimal-gene-sets in host-dependent bacteria. *Trends Microbiol.* 12:37-43

- Koga R., Tsuchida T., Fukatsu T., (2003) Changing partners in an obligate symbiosis: a facultative endosymbiont can compensate for loss of the essential endosymbiont *Buchnera* in an aphid. *Proc. R. Soc. Lond. B* 270: 2543–2550
- Lawrence J.G., (2002) Gene transfer in bacteria: Speciation without species? *Theor. Pop. Biol.* 61:449-460
- Lerat E., Daubin V., Moran N. A., (2003) From gene trees to organismal phylogeny in the Prokaryotes: the case of the gamma-Proteobacteria. *PLoS Biol. Oct.* 1:E19
- Ludwig W., Schleifer K. H., (1994) Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. *FEMS Microbiol. Rev* 15:155-173.
- Maddison D.R., Swofford D.L., Maddison W.P., (1997) NEXUS: an extensible file format for systematic information. *Syst Biol* 46(4) 590-621. pmid:11975335.
- Mira A., Moran N. A., (2002) Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb. Ecol.* 44: 137–143.
- Mira A., Ochman H., Moran N.A., (2001) Deletional bias and the evolution of bacterial genomes, *Trends Genet.* 17, pp. 589–596
- Moran N.A., (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria *Proc. Natl. Acad. Sci. (7)*: 2873-2878
- Moran N.A., (2002) Microbial minimalism: genome reduction in bacterial pathogens. *Cell* 108:583-586.
- Moran N.A., Russell J.A., Koga R., Fukatsu T., (2004) Evolutionary relationships of three new species of *Enterobacteriaceae* living as symbionts of aphids and other insects. *Applied and Environmental Microbiology* 71: 3302-3310.
- Moran N.A., Telang A., 1998. Bacteriocyte-associated symbionts of insects. *Bioscience* 48: 295–304
- Novichkov P.S., Omelchenko M.V., Gelfand M.S., Mironov A.A., Wolf Y.I., (2004) Genome-wide molecular clock and horizontal gene transfer in bacterial evolution. *J Bacteriol* 186:6575–6585
- Ochman H., Elwyn S., Moran N. A., (1999) Calibrating bacterial evolution. *Proc Natl Acad Sci* 96 (22): 12638-12634
- Olivek K. M., Moran N. A., Hunter M. S. (2003) Variation in resistance to parasitism in aphids is due to symbionts not host genotype. *Proc. Natl. Acad. Sci. USA* 102:12795–12800
- Olivier A., Lee H.-Y., Côté J.-C., (2005) Study of the heterogeneity of 16S rRNA genes in γ -proteobacteria. *J. Gen. Appl. Microbiol.*; 51(6): 395-405

- Palacios C., Wernegreen J.J., (2002) A strong effect of AT mutational bias on amino acid usage in *Buchnera* is mitigated at high-expression genes. *Mol. Biol. Evol.* 19: 1575-1584
- Philippe H., Laurent J., (1998) How good are deep phylogenetic trees? *Curr. Opin. Genet. Dev.* (6):616–623.
- Pride D.T., Meinersmann R.J., Wassenaar T.M., and Blaser M.J., (2003) Evolutionary implications of microbial genome tetranucleotide frequency biases. *Genome Res.* 13 pp. 145–158
- Reis M., Savva R., Wernisch L., (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res* 32:5036–5044 .
- Rivera M., Jain R., Moore J., Lake J., (1998) Genomic evidence of two functionally distinct gene classes. *Proc. Natl. Acad. Sci. USA* 95: 6239–6244
- Rose T.M., Schultz ER, Henikoff JG, (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Research* 26 (7): 1628-1635 APR 1
- Sharp P., Bailes E., Grocock R., Peden J., Sockett R., (2005) Variation in the strength of selected codon usage bias among bacteria *Nucleic Acids Res.*, vol. 33, pp. 1141–1153,
- Snel B., Bork P., Huynen M.A., (2002) Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome Res.* 12 pp. 17–25
- Steinert M., Hentschel U., Hacker J., (2000) Symbiosis and pathogenesis: evolution of the microbe-host interaction. *Naturwissenschaften* 87:1-11
- Suyama M., Bork P., (2001) Evolution of prokaryotic gene order: genome rearrangements in closely related species *Trends Genet.* 17, pp. 10–13
- Swofford D., (1998) PAUP (Phylogenetic Analysis Using Parsimony) [computer program]. Version 4*. In: Wisconsin Package Version 9.1, Genetics Computer Group, Madison, Wisconsin
- Tatusov R. L., Galperin M. Y., Natale D. A., Koonin E. V., (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.*; 28:33–36.
- Tatusov R. L., Koonin E. V., Lipman D. J., (1997) A genomic perspective on protein families. *Science* 278:631–637.
- Thao M. L., Baumann P. (2004) Evidence for Multiple Acquisition of *Arsenophonus* by Whitefly Species (Sternorrhyncha: Aleyrodidae) *Curr. Microbiology* 48: 140-144
- Thompson J.D., Higgins D.G., Gibson T.J., (1994) *Comput. Applic. Biosci.*,10, 19–29.

- Trowbridge D., Richard E., Dittmar, K., Whiting M. F., (2005) Identification and phylogenetic analysis of Arsenophonus- and Photorhabdus-type bacteria from adult *Hippoboscidae* and *Streblidae* (Hippoboscoidea). *Journal of Invertebrate Pathology* 91 64–68
- Yamamoto S., Kasai H., Arnold D.L., Jackson R.W., Vivian A., Harayama S., (2000) Phylogeny of the genus pseudomonas: intrageneric structure reconstructed from the nucleotide sequences of gyrB and rpoD genes. *Microbiology-Sgm* 146, pp. 2385–2394
- Yanai I., Derti A., DeLisi C., (2001) Genes linked by fusion events are generally of the same functional category: a systematic analysis of 30 microbial genomes *Proc. Natl Acad. Sci.* 98, 7940–7945
- Vandamme P., Pot B., Gillis M., De Vos P., Kersters, K., Swings, J. (1996). Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol Rev* 60, 407-438.
- Werren, J.H., Zhang, W. and Guo, L.R. (1995) Evolution and phylogeny of *Wolbachia*: reproductive parasites of arthropods. *Proc. R. Soc. Lond. B.* 261: 55-71
- Wernegreen J. J., Genome evolution in bacterial endosymbionts of insects, *Nat Rev Genet*, 3:850-61 (2002)
- Wernegreen J. J., Moran N., (1999) A Evidence for genetic drift in endosymbionts (*Buchnera*): analyses of protein-coding genes. *Mol. Biol. Evol.* 16: 83–97
- Wernegreen J.J., Degnan P.H., Lazarus A.B., Palacios C., Bordenstein S.R. (2003), Genome evolution in an insect cell: Distinct features of an ant-bacterial partnership, *Biol. Bull.*, 204:221-232
- Wolf Y.I., Rogozin I.B., Grishin N.V., Koonin E.V., (2002) Genome trees and the tree of life, *Trends Genet.* 18 pp. 472–479
- Zientz, E., Silva F. J., Gross R., (2001) Genome interdependence in insect-bacterium symbioses. *Genome Biol.* 2:Reviews 1032.

