

Univerzita Karlova v Praze

Přírodovědecká fakulta

Studijní program: Mikrobiologie



Mgr. Jaroslav Nunvář

**Bakteriální REP elementy:
původ, variabilita a využití**

Bacterial REP elements: origins, variability and application

Dizertační práce

Školitel: RNDr. Irena Lichá, CSc.

Praha, 2013

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracoval samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne

Podpis:

Poděkování

V první řadě bych velice rád poděkoval své školitelce Dr. Ireně Liché za vstřícný, tolerantní a lidský přístup, mnoho věnovaného času a nespočetné odborné konzultace a korektury během naší šestileté společné anabáze. Dále děkuji konzultantovi mé dizertační práce Dr. Bohdanu Schneiderovi za vedení grantového projektu a možnost na něm participovat. Rovněž jsem zavázán Doc. Alexandru Nemcovi ze Státního zdravotního ústavu a Doc. Pavlu Dřevínkovi z Ústavu lékařské mikrobiologie 2. lékařské fakulty UK za poskytnutí bakteriálních kmenů a expertní pomoc při makrorestrikci a interpretaci dat.

Můj dík také patří všem studentům a technikům Laboratoře bakteriální genetiky a fyziologie Katedry genetiky a mikrobiologie PřF UK a Laboratoře biomolekulárního rozpoznávání Biotechnologického ústavu AV za vytvoření přátelského pracovního prostředí a výbornou technickou asistenci. Zvláště příjemnou a přínosnou pro mě byla interakce se studentkami, na jejichž projektech jsem se podílel, jmenovitě s Mgr. Terezou Bauerovou, Mgr. Terezou Poláčkovou a Mgr. Ivou Zuskovou, jimž tímto děkuji za milou společnost.

Na závěr bych rád vyjádřil hlubokou vděčnost své rodině a příteli Petrovi za všestrannou a neutuchající podporu v někdy nesnadných situacích, jež doktorské studium obnáší. Poděkování si zaslouží i náš leguán Yoko, jenž je pro mě kontinuálním zdrojem zábavy a rozptýlení.

ABSTRACT (English)

This thesis is based on three published research papers studying bacterial REP (repeat extragenic palindrome) elements. REP elements are one of the best-characterized groups of bacterial DNA repeats, distributed mostly in gammaproteobacteria, including enterobacteria. They are present in noncoding parts of host genomes, usually occurring in hundreds of copies. REPs are typically aggregated in higher order repeats. In the Gram-negative model *Escherichia coli*, interactions of several proteins important for cell's physiology with REPs were described, indicating significant role for these elements for host cells.

The first work (Nunvar et al. 2010) presents the discovery of a protein class, related to IS200/IS605 transposases. These proteins, termed RAYTs (**R**EP-**a**ssociated **t**yrosine **t**ransposases), contain characteristic motifs in their amino acid sequences, which are absent in canonical IS200/IS605 transposases. Another attribute of RAYTs is the arrangement of their encoding genes. These are single copy genes, always flanked at both termini by at least two REPs in inverted orientation. Based on the similarity between the REP-*rayt*-REP unit and insertion sequences of the IS200/IS605 family, between RAYTs and tyrosine transposases and between REPs and subterminal sequences of the IS200/IS605 family, the hypothesis about RAYTs being the mobilizers of REP elements was proposed.

The second work (Nunvar et al. 2013) explores the variability of REP copy numbers with respect to evolution of host bacteria. The analysis covers large dataset of genomic sequences from two bacterial lineages – fluorescent pseudomonads (63 strains) and stenotrophomonads (10 strains). Tens of unique classes of REP sequences and their cognate RAYTs were identified. The copy numbers of particular REP classes varied significantly among phylogenetic clades, as well as within the clades. High REP copy numbers were typically conditioned by the presence of

cognate RAYT genes. The results imply that long term presence of RAYTs is likely to lead to the proliferation of cognate REP elements.

The third work (Nunvar et al. 2012) examines the usefulness of species-specific REP sequences and their dynamics for easy genotyping of *Stenotrophomonas maltophilia* isolates. The developed method, termed SmrepPCR, employs polymerase chain reaction that uses single primer, complementary to the most abundant REP class in *S. maltophilia*. The grouping of 34 isolates, based on the similarity of their SmrepPCR profiles (banding patterns), correlated well with the branching of phylogram constructed from sequences of an essential gene (*gyrB*). The novel SmrepPCR is therefore suitable for estimation of clonal or phylogenetic relationships of environmental and clinical strains of *S. maltophilia*.

Keywords

Pseudomonas fluorescens, *Stenotrophomonas maltophilia*, REP element, IS200/IS605, RAYT, SmrepPCR, *gyrB*, molecular phylogenetics, comparative genomics

ABSTRAKT (česky)

Tato práce shrnuje tři publikace, jejichž jednotícím tématem jsou bakteriální REP (angl. repetitive extragenic palindrome) elementy. REP elementy jsou jednou z nejznámějších a nejlépe charakterizovaných skupin repetitivních DNA sekvencí u bakterií. Jsou známe především u gamaproteobakterií, včetně enterobakterií. Vyskytují se v nekódujících oblastech hostitelských genomů, zpravidla ve stovkách kopií. REP elementy jsou typicky agregovány do repetitivních útvarů vyšších řádů. U gramnegativní modelové bakterie *Escherichia coli* byly popsány interakce REP elementů s několika důležitými proteiny buněčné fyziologie, svědčící o významné roli těchto sekvencí pro hostitelskou buňku.

První studie (Nunvar et al. 2010) prezentuje objev třídy proteinů příbuzných tyrosinovým transponázám ze skupiny IS200/IS605 inzerčních sekvencí. Tyto proteiny, pojmenované RAYT (angl. REP-assoiated tyrosine transposase), jsou charakteristické přítomností specifických motivů v jejich aminokyselinové sekvenci, nepřítomných u kanonických IS200/IS605 transponáz. Dalším společným znakem je uspořádání genů kódujících RAYT proteiny. Tyto geny se v hostitelském chromozomu vyskytují vždy v jedné kopii a jsou z obou stran ohraničeny alespoň dvěma REP elementy v inverzní orientaci. Na základě podobnosti mezi systémem REP-*rayt*-REP a inzerčními sekvencemi rodiny IS200/IS605, mezi RAYT proteiny a tyrosinovými transponázami a mezi REP elementy a subterminálními sekvencemi inzerčních sekvencí rodiny IS200/IS605 byla předložena hypotéza o mobilizaci REP elementů RAYT transponázami.

Druhá studie (Nunvar et al. 2013) zkoumá variabilitu počtu kopií REP elementů v závislosti na evoluční příslušnosti hostitelských bakterií. Podkladem analýzy byly rozsáhlé soubory genomových sekvencí dvou bakteriálních fylogenetických skupin – fluorescentních pseudomonád (63 kmenů) a stenotrofomonád (10 kmenů). Byly identifikovány desítky unikátních tříd REP

sekvencí a s nimi asociovaných RAYT proteinů. Počty kopií jednotlivých třídy REP elementů vykazovaly významnou variabilitu mezi fylogenetickými liniemi hostitelů i mezi kmeny v rámci jedné linie. Vysoké četnosti daných REP sekvencí byly zpravidla podmíněny přítomností asociovaného RAYT genu. Z výsledků lze vyvodit, že dlouhodobá přítomnost RAYT proteinů vede k proliferaci příslušných REP elementů.

Třetí studie (Nunvar et al. 2012) pojednává o možnostech využití dynamiky druhově specifických REP sekvencí pro jednoduchou genotypizaci izolátů druhu *Stenotrophomonas maltophilia*. Vyvinutá metoda, nazvaná SmrepPCR, využívá polymerázovou řetězovou reakci s jediným primerem, komplementárním k nejčetnější třídě REP elementů u *S. maltophilia*. Rozdělení 34 zkoumaných izolátů do skupin na základě příbuznosti jejich SmrepPCR profilů (soubory amplifikovaných proužků různých molekulových hmotností) velmi dobře korelovalo s fylogenetickým stromem sekvencí esenciálního genu *gyrB*. Novou metodu SmrepPCR lze tedy použít pro odhad klonální, resp. fylogenetické příbuznosti environmentálních a klinických kmenů *S. maltophilia*.

Klíčová slova

Pseudomonas fluorescens, *Stenotrophomonas maltophilia*, REP element, IS200/IS605, RAYT, SmrepPCR, *gyrB*, molekulární fylogenetika, srovnávací genomika

OBSAH

| | |
|--|----|
| Abstract (English)..... | 4 |
| Abstrakt (česky)..... | 6 |
| 1. ÚVOD..... | 9 |
| 2. LITERÁRNÍ PŘEHLED..... | 11 |
| 2.1. Repetitivní DNA a hustota kódující informace..... | 11 |
| 2.2. Dlouhé repetitivní sekvence DNA u prokaryot..... | 13 |
| 2.2.1 Strukturní RNA..... | 13 |
| 2.2.2. Inzerční sekvence, transpozony a jejich transponázy..... | 14 |
| 2.2.2.1. DDE transponázy..... | 17 |
| 2.2.2.2. Y1 transponázy..... | 18 |
| 2.3. Krátké nekódující repetitivní sekvence DNA u prokaryot..... | 26 |
| 2.3.1. Rozptýlené repetitivní oligonukleotidové motivy..... | 26 |
| 2.3.2. CRISPR elementy..... | 28 |
| 2.3.3. Neautonomní transponibilní elementy..... | 30 |
| 2.3.4. REP elementy..... | 34 |
| 2.3.4.1. Výskyt..... | 34 |
| 2.3.4.2. Struktura..... | 36 |
| 2.3.4.3. Funkce..... | 38 |
| 2.3.4.4. Využití..... | 40 |
| 3. PUBLIKACE AUTORA..... | 42 |
| 4. DISKUZE..... | 45 |
| REFERENCE..... | 56 |

1. ÚVOD

Typickým znakem genomů prokaryotních buněk, který je zásadně odlišuje od vyšších eukaryot, je vysoká hustota kódujících sekvencí. Tato vlastnost patrně souvisí s bakteriálním životním stylem, pro nějž jsou charakteristické úspornost a efektivita. Přesto se i zde vyskytují repetitivní, nekódující sekvence DNA. Jednou z několika tříd těchto repetitivních sekvencí jsou REP (angl. repetitive extragenic palindrome) elementy. REP elementy se vyznačují vysokým počtem kopií, značnou variabilitou a dynamikou a predikovanými funkcemi ve fyziologii hostitelských buněk. V této práci popisuji variabilitu REP elementů, zkoumám příčiny této variability a navrhuji její využití v genotypizaci gramnegativních bakterií druhu *Stenotrophomonas maltophilia*.

Dizertační práce vznikla na Katedře genetiky a mikrobiologie PŘF UK v Praze pod vedením Dr. Ireny Liché a na Biotechnologickém ústavu AV ČR v.v.i. v laboratoři Dr. Bohdana Schneidera. Některé experimenty jsem prováděl v Laboratoři molekulární genetiky Pediatrické kliniky Fakultní nemocnice Motol a 2. lékařské fakulty UK pod vedením Doc. Pavla Dřevínka.

Díličními cíli mého výzkumu bylo:

1. Bioinformaticky charakterizovat novou rodinu tyrosinových transponáz (RAYT – angl. REP-assoiated tyrrosine transposase), již jsem objevil v rámci analýzy sekvence DNA *S. maltophilia*, z hlediska asociace jejích kódujících genů s REP elementy, unikátních motivů v jejich aminokyselinových sekvencích a taxonomického rozšíření. Najít definující znaky REP elementů asociovaných s geny RAYT transponáz. Na základě podobnosti mezi inzerčními sekvencemi rodiny IS200/IS605 a RAYT geny asociovanými s REP elementy navrhnout hypotézu o jejich funkčním propojení.

2. Podrobně analyzovat diverzitu REP elementů a RAYT proteinů v rámci rozsáhlé skupiny genomových sekvencí bakterií z komplexu *Pseudomonas fluorescens* a *S. maltophilia*. Bioinformaticky prozkoumat závislost dynamiky REP elementů na evoluci hostitelských bakterií a přítomnosti RAYT proteinů. Odvodit roli RAYT proteinů v proliferaci a dlouhodobé perzistenci REP elementů v bakteriálních genomech.
3. Vyvinout a otestovat jednoduchou metodu (SmrepPCR) pro genotypizaci bakteriálních kmenů *S. maltophilia* založenou na dynamice REP elementů. Porovnat informaci o identitě nebo příbuznosti izolátů získanou pomocí SmrepPCR s výsledky metod makrorestrikčního štěpení genomové DNA a sekvenací esenciálního genu.

Tento výzkum byl finančně podporován následujícími granty:

MŠMT: LC06066 - Centrum environmentální mikrobiologie

GA ČR: GAP305/12/1801

Výzkumný záměr AV ČR: AV0Z50520701

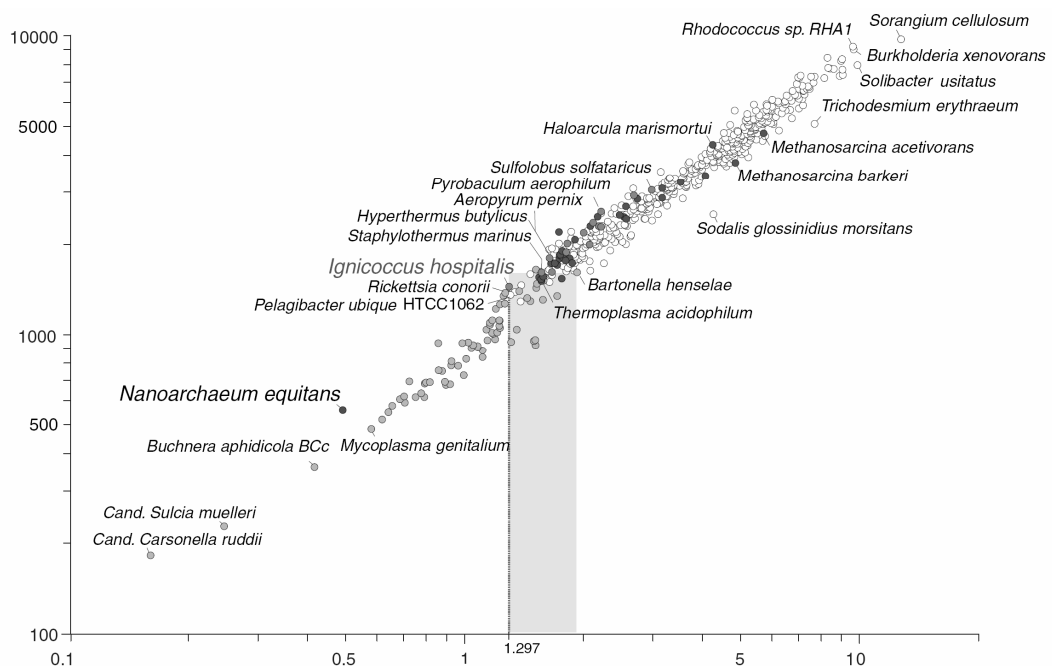
Specifický vysokoškolský výzkum: 12/261 214, 13/267 205

2. LITERÁRNÍ PŘEHLED

2.1. Repetitivní DNA a hustota kódující informace

Bakterie jsou nejjednodušší buněčnou formou života, která přetrvala do současnosti. Mnoho bakteriálních vlastností vykazuje podstatně jednodušší (primitivnější) úroveň než u eukaryotních organismů. Jedná se zejména o velikost a vnitřní členění buňky, rozmnožování, komplexitu mechanismů účastnících se genové exprese, celkovou velikost genomu a absolutní počet genů.

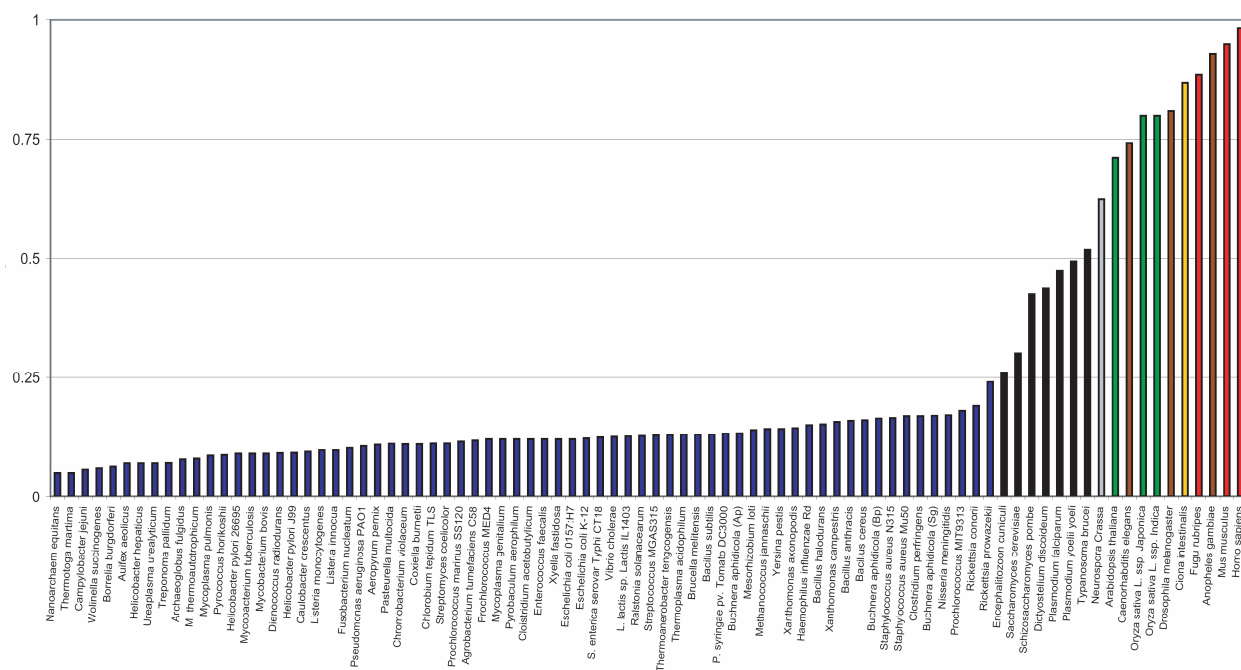
Pro bakteriální genomy je charakteristická vysoká kódující denzita, tj. ta část genetické informace (podíl z celkového počtu bází), která je přítomna v genech (Koonin a Wolf 2008). Hodnoty této veličiny jsou v rámci prokaryotní fylogenetické diverzity poměrně stálé (**Obr. 1**).



Obrázek 1 – Vztah mezi celkovou velikostí genomu v Mbp (horizontální osa) a celkovým počtem genů (vertikální osa) v genomech 623 prokaryot. Každý bod odpovídá jednomu prokaryotnímu druhu, vybrané druhy bakterií a archeí jsou označeny.

Podle Podar et al. 2008

Kódující denzita eukaryotních organismů je oproti prokaryotům násobně nižší (Ahnert et al. 2008). Evolučním trendem je zde zvyšování podílu nekódující DNA v genomech mnohobuněčných, vývojově pokročilých linií eukaryot (typicky zelených rostlin a živočichů – Obr. 2).



Obrázek 2 – Podíl nekódující DNA k celkovému množství haploidní DNA (vertikální osa) u prokaryot (modře), jednobuněčných eukaryot (černě) a mnohobuněčných eukaryot (ostatní). Hodnoty jsou nepřímo úměrné kódující denzitě.

Podle Mattick 2004

Nekódující DNA v eukaryotních genomech se skládá převážně z repetitivních sekvencí. Během evoluce eukaryot docházelo k expanzi repetitivní DNA. V důsledku toho narůstala celková velikost genomů a snižovala se kódující denzita, neboť množství genů potřebných pro funkce organismů se zvyšovalo nepatrně nebo mnohem pomaleji než četnost repetitivních sekvencí (Lopez-Flores a Garrido-Ramos 2012). Zatímco například lidský genom je tisícinásobně větší než genom typické bakterie, obsahuje jen asi desetinásobek genů. Jeho kódující denzita je pak přibližně 3 % (Obr. 2).

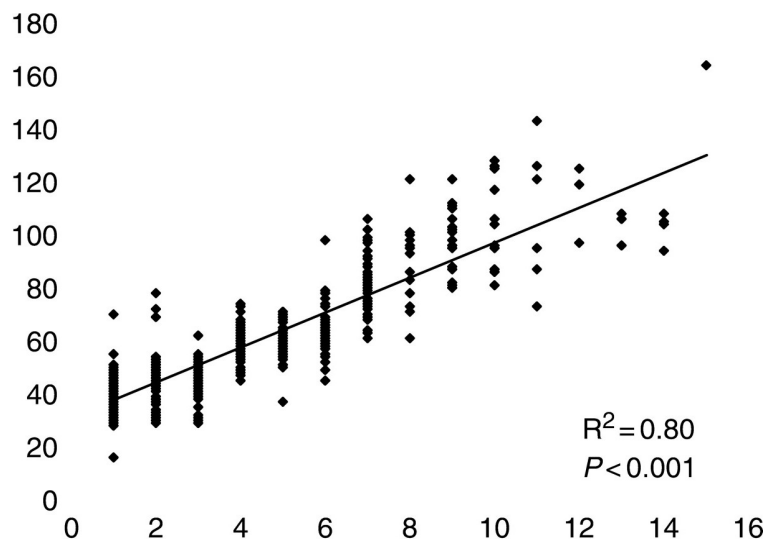
Prokaryotní kódující denzita se typicky pohybuje na úrovni 90 % (**Obr. 2**) – na jeden průměrně velký gen (1 kbp, **Obr. 1**) pak připadá přibližně 100 bp nekódujících oblastí, kam spadají převážně operátorové oblasti, promotory a transkripční terminátory. Je tedy zřejmé, že bakterie ve svých genomech netolerují vyšší množství nekódující repetitivní DNA. O povaze selekčního tlaku na zachování vysoké kódující denzity není, vzhledem k nemožnosti rekonstruovat výsledky takto dlouhých evolučních procesů v laboratoři, mnoho známo. Všeobecně se soudí, že se jedná o důsledek intenzivní selekce na rychlé množení ve vysoce kompetitivním přirozeném prostředí bakterií, neboť replikace nadbytečné DNA představuje pro buňku nezanedbatelnou biosyntetickou zátěž. Nižší kódující denzita je v rámci prokaryotní diverzity typická jen pro bakterie obligátně asociované s eukaryotními hostiteli – intracelulární patogeny a endosymbionty, což je vysvětlováno nižší intenzitou selekce (Gil a Latorre 2012).

2.2. Dlouhé repetitivní sekvence DNA u prokaryot

Repetitivní elementy této skupiny typicky kódují funkční produkt – ribonukleovou kyselinu (tRNA, rRNA) (**Kap. 2.2.1.**) nebo enzym transponázu (**Kap. 2.2.2.**). Bývají méně početné než kratší, nekódující repetece (**Kap. 2.3.**).

2.2.1. Strukturní RNA

Geny, jejichž konečnými produkty exprese jsou ribozomální a transferové RNA, jsou někdy nazývány strukturními RNA, vzhledem ke složitým sekundárním strukturám, jimiž se odlišují od mRNA. Operony kódující rRNA i tRNA se v genomech bakterií zpravidla vyskytují zmnožené v jednotkách až desítkách kopií, přičemž multiplicita genů obou typů těchto RNA navzájem koreluje (**Obr. 3**).



Obrázek 3 – Korelace mezi celkovým počtem genů pro 16S rRNA (horizontální osa) a celkovým počtem genů tRNA (vertikální osa) v genomech 590 bakterií.

Podle Lee et al. 2009

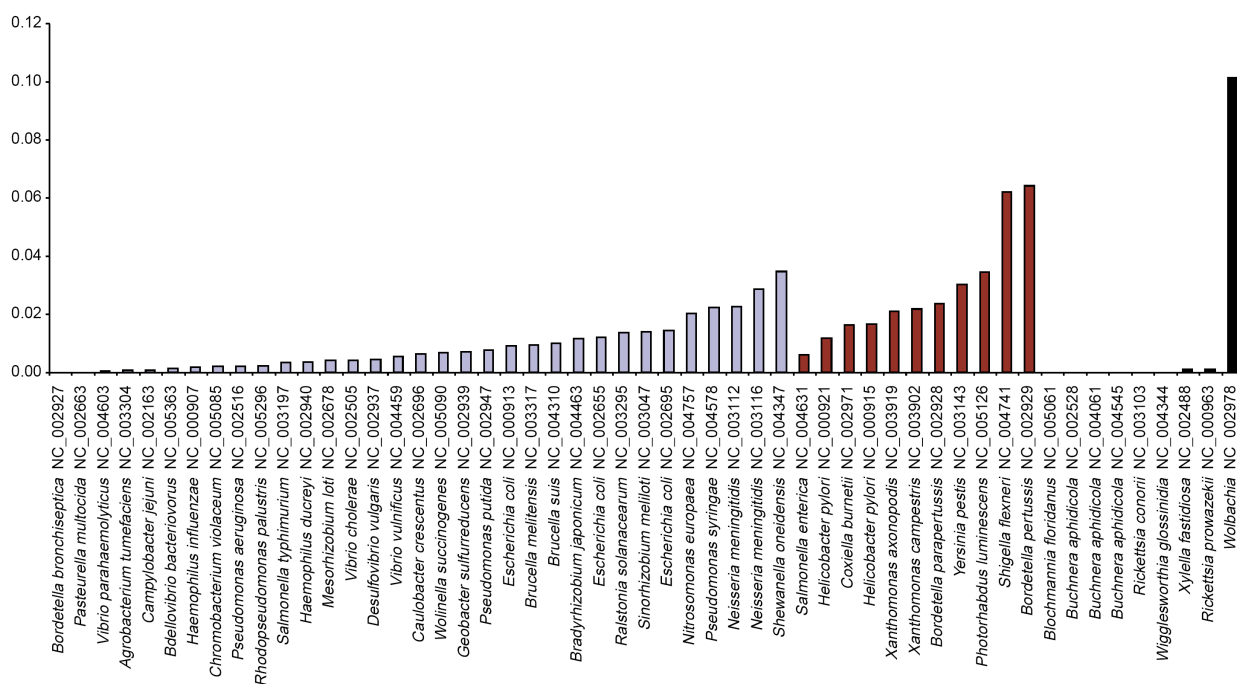
Počet kopií genů rRNA pozitivně koreluje se schopností buňky rychle růst a efektivně využívat živiny (Klappenbach et al. 2000). Jako důvod multiplicity genů rRNA se obecně uvažuje absence translační amplifikace při syntéze jejich konečných produktů. rRNA se vyskytuje v ribozomech v ekvimolárním poměru s ribozomálními proteiny. Ribozomální proteiny jsou při expresi z DNA kromě transkripce amplifikovány i translací, přítomnost několika kopií genu rRNA tak kompenzuje nemožnost translační amplifikace (Kobayashi 2011).

2.2.2. Inzerční sekvence, transpozony a jejich transponázy

Prokaryotní inzerční sekvence jsou jednoduché mobilní DNA elementy, schopné přemisťovat se v rámci hostitelského genomu z místa na místo (Mahillon a Chandler 1998). Lze na ně nahlížet jako na sobecké elementy, které využívají funkce hostitelské buňky (replikace, transkripce, translace) k vlastním účelům. Inzerční sekvence reprezentují nejjednodušší formu autonomních jednotek. Gen kódující enzym zajišťující mobilitu, tj. transponázu, je ohraničen z obou stran terminálními sekvencemi, sloužícími jako místo rozpoznávané transponázou. Vzácněji se

u prokaryot setkáváme s transpozony, což jsou deriváty inserčních sekvencí vzniklé buď jejich sloučením, nebo získáním přídatných genů.

Inserční sekvence jsou typické značnou variabilitou výskytu u jednotlivých druhů prokaryot. Známe jak kmeny zcela bez inserčních sekvencí (např. *Bacillus subtilis* 168), tak kmeny hostící několik typů inserčních sekvencí, s různým počtem kopií každé z nich (např. *Escherichia coli* K-12). Počty kopií v rámci jednoho druhu, např. *Helicobacter pylori*, mohou značně kolísat mezi jednotlivými kmeny (Mahillon et al. 1999). V průměru se u volně žijících bakterií vyskytuje jedna inserční sekvence na 100 kbp genomové DNA (**Obr. 4**).



Obrázek 4 – Počty predikovaných inserčních sekvencí příslušejících 1 kbp genomové DNA (vertikální osa) v genomech osekvenovaných proteobakterií. Šedé sloupce – volně žijící druhy; červené a černé sloupce – druhy obligátně asociované s eukaryotními organismy.

Podle Moran a Plague 2004

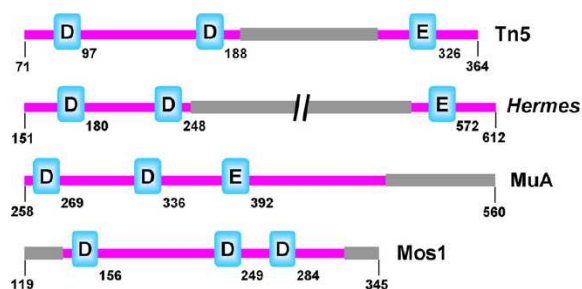
Díky své mobilitě jsou inserční sekvence důležitým zdrojem evoluční variability u prokaryot. Jejich inserce do esenciálních nebo neesenciálních genů jsou letální nebo mění fitness hostitele. Exprese transponáz je proto přísně regulovaná (Nagy a Chandler 2004). Na druhou stranu, variabilita generovaná transpozicí inserčních sekvencí může být zdrojem prospěšných evolučních novinek. V laboratorních experimentech simulujících dlouhodobou evoluci (Schneider et al. 2000; Schneider a Lenski 2004) byly mezi typy evolučně prospěšných mutací nalezeny inserce do promotorových oblastí a rekombinační události zprostředkované inserčními sekvencemi (delece, inverze). Pro potenciální výhodnost pohybu inserčních sekvencí svědčí i skutečnost, že mnoho stresorů (vysoká teplota, UV záření, nutriční limitace a oxidativní stres) má stimulační vliv na frekvenci transpozice (Eichenbaum a Livneh 1998; Ilves et al. 2001; Ohtsubo et al. 2005; Twiss et al. 2005; Drevinek et al. 2010). Tento fenomén je interpretován jako mechanismus adaptivní mutagenese, jejímž cílem je stresové podmínky překonat ziskem výhodné mutace.

Inserční sekvence se člení do několika desítek rodin na základě homologie jimi kódovaných transponáz a terminálních sekvencí. Protože transponázy a terminální sekvence spolu během transpoziční reakce interagují (podléhají tedy koevoluci), jsou jejich vlastnosti pro danou rodinu inserčních sekvencí charakteristické.

Následující text pojednává o DDE transponázách, jež jsou u inserčních sekvencí zdaleka nejrozšířenější, a dále o Y1 transponázách, jež jsou důležité ve vztahu k tématu dizertační práce. Zbylé dvě skupiny prokaryotních transponáz, tj. typ Y2 a S (Curcio a Derbyshire 2003), v tomto přehledu nejsou zahrnuty.

2.2.2.1. DDE transponázy

Velká většina rodin inzerčních sekvencí kóduje tzv. DDE transponázy (Mahillon et al. 1999). Název pochází z jednopísmenných zkratk trojice kyselých katalytických aminokyselin, z nichž se skládá aktivní místo proteinu. Pozice těchto zbytků v primární struktuře transponáz je značně variabilní (**Obr. 5**). Pro strukturu DDE transponáz je charakteristický tzv. RNase H fold (Nesmelova a Hackett 2010).



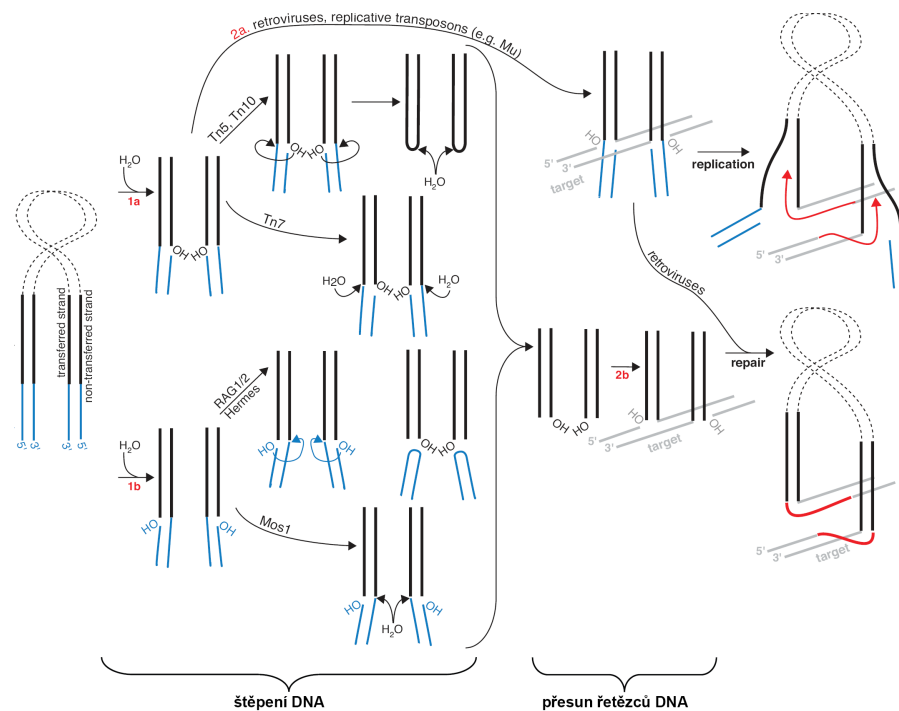
Obrázek 5: Schematické znázornění umístění triády katalytických aminokyselin (modře) u čtyř reprezentantů DDE transponáz. Čísla odpovídají pořadí aminokyselin v sekvenci proteinů.

Podle Nesmelova a Hackett 2010

Primární funkcí triády kyselých aminokyselin je vazba a koordinace dvou dvojmocných kovových iontů. Tyto ionty orientují molekuly vody a terminální hydroxylové skupiny DNA v průběhu transpoziční reakce a propůjčují jim nukleofilní vlastnosti. Transpozice je pak souborem hydrolýzy esterových vazeb DNA a následných transesterifikačních reakcí. Jednotlivé mezireakce jsou poměrně variabilní, jak ilustruje **Obr. 6**.

Inzerční sekvence využívající DDE transponázy se vyznačují společným uspořádáním terminálních oblastí, tj. koncových sekvencí. Jsou lemovány krátkými přímými repeticemi, vzniklými při integraci do daného místa genomu duplikací krátkých jednořetězcových komplementárních oblastí, k nimž se DNA inzerční sekvence připojuje (**Obr. 6**). Vlastní terminální sekvence mají podobu inverzních repetit různých délek a sekvence. Sekvence terminálních inverzních repetit je zodpovědná za rozpoznání mobilního elementu transponázou (Ichikawa et al. 1987).

Inverzní uspořádání terminálních sekvencí umožňuje, aby homodimer transponázy štěpil oba konce mobilního elementu, s nutností rozpoznávat pouze jednu specifickou sekvenci DNA.



Obrázek 6: Alternativní průběh reakcí katalyzovaných DDE transponázami. Černě plně – terminální sekvence inzerčních sekvencí/transpozonů, černě tečkovaně – vnitřní sekvence mobilního elementu (někde vynecháno), modře – hostitelská DNA přiléhající k mobilnímu elementu, šedě – hostitelská DNA přiléhající k místu inzerce, červeně – nově syntetizovaná DNA.

Podle Montano a Rice 2011

2.2.2.2. Y1 transponázy

Skupinou inzerčních sekvencí zcela odlišných od většiny ostatních, jež jsou mobilizovány DDE transponázami (**Kap. 2.2.2.1**), jsou inzerční sekvence patřící do rodiny IS200/IS605. Kromě základního společného rysu, tj. schopnosti přemísťovat se v genomu hostitelů, se odlišují v téměř všech ostatních aspektech. Společné a rozdílné znaky jsou shrnuty v **Tab. 1**. Inzerční sekvence rodiny IS200/IS605 jsou

mobilizovány tzv. Y1 transponázami, které obsahují jeden katalytický tyrosin (Ronning et al. 2005), čímž se odlišují od inzerčních sekvencí rodiny IS91, jejichž transponázy využívají dva katalytické tyrosiny (Curcio a Derbyshire 2003).

| | Třída inzerčních sekvencí podle kódované transponázy | |
|--|--|------------------------------|
| | DDE | Y1 |
| transponázy tvoří dimer | ano | ano |
| dvojmocný iont kovu jako kofaktor | ano | ano |
| tvorba přímých repetic při inzerci | ano | ne |
| typy terminálních sekvencí | inverzní repetice | palindromy + tetranukleotidy |
| forma transponované inzerční sekvence | dvouřetězcová DNA | jednořetězcová DNA |
| tvorba kovalentního komplexu DNA/transponáza | ne | ano |
| katalytické aminokyseliny | DDE | Y |

Tabulka 1: Společné (nahore) a rozdílné (dole) vlastnosti inzerčních sekvencí využívajících DDE transponázy nebo Y1 transponázy.

Inzerční sekvence rodiny IS200/IS605 se dělí na dvě podskupiny. Podskupina IS200 kóduje jediný protein, Y1 transponázu (TnpA). Členové podskupiny IS605 kódují navíc druhý protein, TnpB, jenž nemá transponázovou aktivitu (Kersulyte et al. 2002). U *Deinococcus radiodurans* byl nedávno prokázán negativní vliv TnpB na frekvenci transpozice (Pasternak et al. 2013). V subterminálních oblastech všech inzerčních sekvencí této rodiny se vyskytují palindromické sekvence (Hickman et al. 2010). V podmínkách vysoké negativní superhelicity, nebo pokud se objeví ve formě jednořetězcové DNA, dávají palindromické sekvence vzniknout vlásenkovým

(angl. hairpin, stem-loop) strukturám (Bikard et al. 2010). Tyto subterminální vlásenky jsou rozpoznávány Y1 transponázami.

Y1 transponázy inserčních sekvencí rodiny IS200/IS605 z různých bakteriálních a archeálních druhů jsou nejmenšími transponázami vůbec. Vykazují značnou sekvenční variabilitu (**Obr. 7**). Mezi nejkonzervovanější aminokyseliny patří invariantní katalytický tyrosin a dvojice histidinů, oddělených hydrofobní aminokyselinou, tzv. HUH motiv ("U" značí hydrofobní zbytek). Funkcí HUH motivu je vazba hořčnatého iontu, který se spolupodílí na katalýze.



Obrázek 7: Alignment aminokyselinových sekvencí členů třídy Y1 transponáz IS200/IS605 inserčních sekvencí. Pozice dvou histidinů HUH motivu a katalytického tyrosinu jsou znázorněny svislými šipkami.

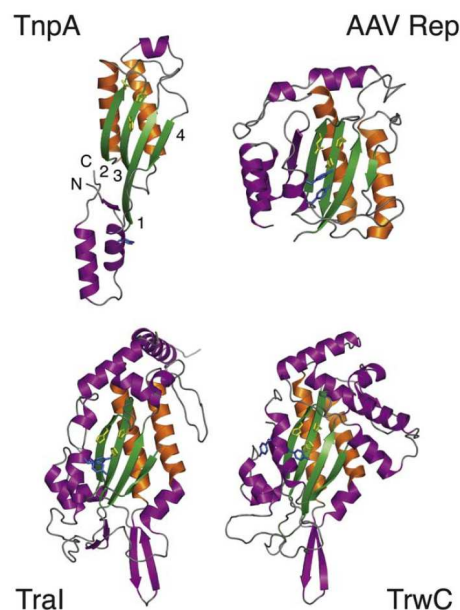
Podle Ton-Hoang et al. 2005

Metodami rentgenové krystalografie byly dosud vyřešeny struktury čtyř Y1 transponáz (**Tab. 2**). Všechny tyto proteiny vytvářejí homodimery s podobnou terciární strukturou.

| Organismus | Název inzerční sekvence | Identifikační číslo v databázi PDB | Literární zdroj |
|--------------------------------|-------------------------|------------------------------------|---------------------|
| <i>Helicobacter pylori</i> | IS608 | 2a6m | Ronning et al. 2005 |
| <i>Deinococcus radiodurans</i> | ISDra2 | 2xqc | Hickman et al. 2010 |
| <i>Sulfolobus solfataricus</i> | nepojmenována | 2f4f | Lee et al. 2006 |
| <i>Sulfolobus tokodaii</i> | ISSto1 | 2ec2 | nepublikováno |

Tabulka 2: Souhrnné údaje k vyřešeným krystalovým strukturám Y1 transponáz.

Proteiny strukturně i funkčně nejpříbuznějšími s Y1 transponázami se ukázaly být nukleázy spojené s replikací valivou kružnicí – konjugativní relaxázy a virové Rep proteiny (Ronning et al. 2005). Ačkoliv jejich vzájemná sekvenční homologie je zanedbatelná, obě skupiny mají podobnou konfiguraci aktivního místa, které obsahuje katalytický tyrosin a HUH motiv. Všechny prostorové struktury se vyznačují podobným uspořádáním sekundárních struktur kolem aktivního místa, což svědčí o sdíleném, byť dávném, evolučním původu (**Obr. 8**). Dalším společným znakem s výše uvedenými nukleázami je schopnost vázat vlásenkové struktury, které se vyskytují v místech replikačních počátků plasmidů, resp. virů (Ronning et al. 2005).



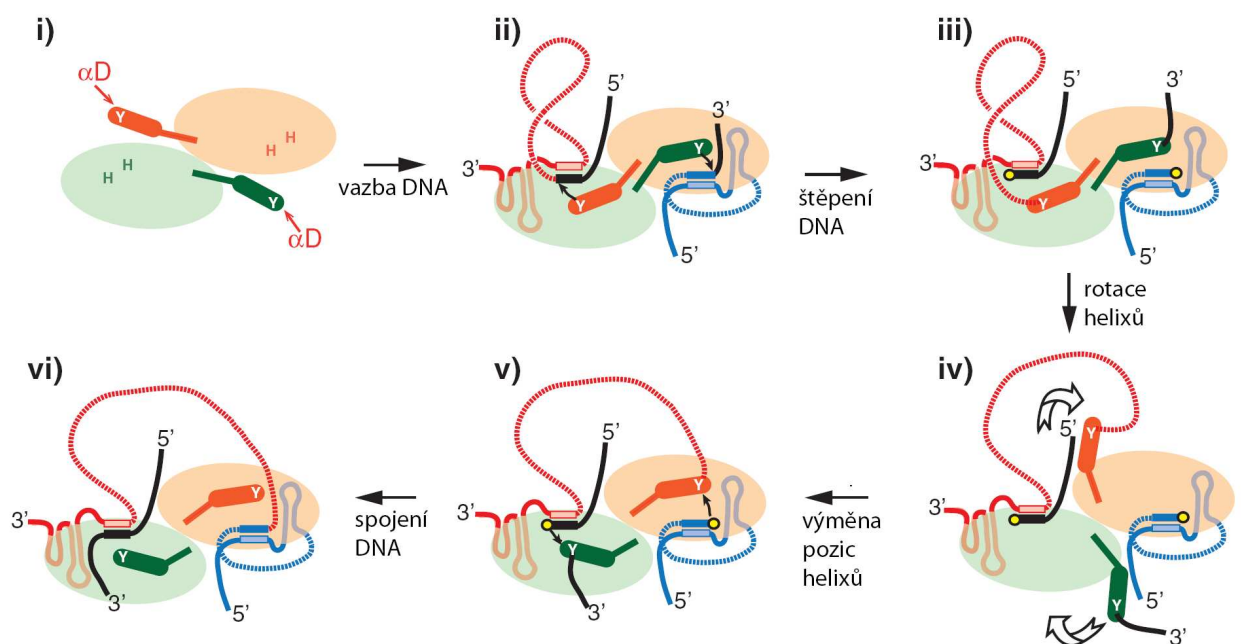
Obrázek 8: Porovnání prostorových struktur Y1 transponázy inzerční sekvence IS608 (TnpA), konjugativních relaxáz TraI a TrwC a Rep proteinu adeno-asociovaného viru AAV5. Sekundární struktury s konzervovaným uspořádáním jsou vyznačeny žlutě (α -helixy) a zeleně (β -listy).

Podle Ronning et al. 2005

Mechanismy transpozice prostřednictvím Y1 transponáz byly detailně studovány pouze u inzerčních sekvencí IS608 (Barabas et al. 2008; Guynet et al. 2008; He et al. 2011; He et al. 2013) a IS*Dra2* (Hickman et al. 2010; Pasternak et al. 2010). Ačkoliv tyto transponázy vykazují poměrně nízkou vzájemnou sekvenční identitu (36 % shodných aminokyselin) a liší se především nukleotidovými sekvencemi terminálních oblastí inzerčních sekvencí, transpoziční mechanismus je velmi podobný (Hickman et al. 2010).

Jednotlivé kroky transpoziční reakce jsou znázorněny na **Obr. 9**. Po navázání dimeru transponázy na DNA vlásenky dochází ke konformační změně proteinu a přesunu α -helixů s katalytickými tyrosiny do blízkosti HUH motivu s hořčnatým kationtem (i-ii). Tímto způsobem je zformováno aktivní místo a dochází ke štěpení DNA a její kovalentní vazbě na tyrosin (iii). Následně si oba helixy s navázanou DNA vymění pozice v rámci dimeru (iv-v) a mechanismem reverzním v porovnání s bodem iii) spojí fragmenty DNA (vi). Spojením 5'-konce inzerční sekvence

(navázaného na tyrosin) a uvolněné hydroxylové skupiny jejího 3'-konce dojde k vyštěpení jednořetězcové cirkulární formy inserční sekvence (Guyenet et al. 2008). Zároveň jsou spojeny oba konce hostitelské DNA přiléhající k inserční sekvenci. Cirkulární forma je dále integrovaná do nového místa hostitelského genomu procesem opačným k jejímu vyštěpení (He et al. 2013).

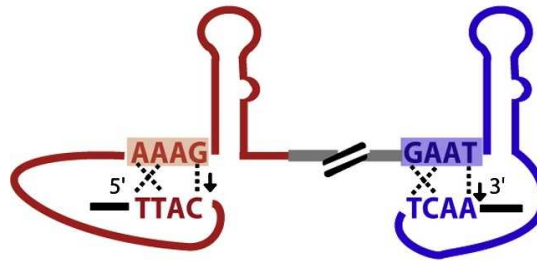


Obrázek 9: Schéma transpozičního mechanismu inserční sekvence IS608. Znáznorněny jsou pouze kroky vedoucí k vyštěpení inserční sekvence. Jednotlivé monomery transponázy, histidiny HUH motivu (H), katalytické tyrosiny (Y) a mobilní α -helixy (α D) jsou vyznačeny stejnou (oranžovou nebo zelenou) barvou. Terminální úseky DNA inserční sekvence jsou vyznačené červeně, resp. modře. Sekvence sousedící hostitelské DNA jsou vyznačeny černě. Párující tetranukleotidy jsou znázorněné jako obdélníky příslušných barev. Volné hydroxylové skupiny 3'-konců jsou znázorněné žlutými kroužky. Bližší vysvětlení v textu výše.

Podle He et al. 2013

Jedinečným rysem transpozice inzerčních sekvencí rodiny IS200/IS605 je způsob determinace štěpného místa. Tuto funkční úlohu plní dvojice tetranukleotidových motivů lokalizované na obou koncích inzerční sekvence (**Obr. 10**). Jeden tetranukleotid z této dvojice se vždy vyskytuje těsně u paty vlásenky, druhý je vzdálenější. Toto prostorové oddělení je sterickou podmínkou umožňující rotaci α -helixů při transpozici (**Obr. 9**). Dvojice tetranukleotidů spolu párují kombinací Watson-Crickovského a Hoogstenova typu párování, což je v DNA velmi vzácným jevem. S tetranukleotidem u paty vlásenky je následkem párování bází fyzicky asociován vzdálenější tetranukleotid, čímž dochází k jeho definovanému ukotvení v aktivním místě. Je tak umožněno přesné, sekvenčně-specifické štěpení substrátové DNA, přestože transponáza sama tetranukleotidové motivy přímo neváže (Barabas et al. 2008).

Pro determinaci štěpného místa je tedy využívána fyzická blízkost paty vlásenky a aktivního místa enzymu a přítomnost dvou párujících tetranukleotidů. Při inzerci IS608 do chromozomu je jako cílové místo rozpoznáván právě jeden z čtveřice tetranukleotidů. Cílenou mutagenézí tetranukleotidových motivů přiléhajících k DNA vlásenky tak lze, díky párování komplementárních bází tetranukleotidů, programově měnit sekvenční specifitu inzercí tohoto elementu, což otevírá možnosti bioinženýrského využití těchto enzymů pro cílenou modifikaci DNA (Guynet et al. 2009).



Obrázek 10: Určení štěpných míst inserční sekvence IS608. Terminální sekvence jsou znázorněné červeně, resp. modře. Párování bází je vyznačeno přerušovanými čarami. Místa štěpení jsou znázorněna šipkami.

Podle Guynet et al. 2009

Regulace transpozice inzerčních sekvencí rodiny IS200/IS605 souvisí s dostupností jednořetězcové DNA, která je substrátem pro transpozici. Jednořetězcová DNA se v buňce vyskytuje jen ve zvláštních případech. Obligátně se objevuje např. při replikaci zpoždujícího se řetězce. Bylo prokázáno, že transponované vlákno IS608 se preferenčně integruje do zpoždujícího se řetězce, resp. je z něj vyštěpováno. Experimentálně indukované zvýšení množství jednořetězcové DNA v replikační vidličce vedlo k razantnímu zvýšení frekvence transpozice IS608 (Ton-Hoang et al. 2010). U IS*Dra2* z radiorezistentní bakterie *D. radiodurans* byla zjištěna až stonásobná stimulace transpoziční frekvence po vystavení účinkům UV a γ záření (Mennecier et al. 2006). Vlivem záření rovněž docházelo ve zvýšené míře k excizi cirkulárních jednořetězcových forem IS*Dra2*. Tato pozorování souvisí s masivním výskytem jednořetězcové DNA během opravy radiací fragmentovaných chromozomů (Pasternak et al. 2010).

2.3. Krátké nekódující repetitivní sekvence DNA u prokaryot

Repetitivní sekvence tohoto typu jsou výrazně kratší než formy repetitivní DNA popsané v **Kap. 2.2.**, neboť nekódují žádné proteiny ani strukturní RNA. V hostitelských genomech je zpravidla nalézáme v řádově vyšších počtech (typicky stovkách) kopií. Většinou mají významnou roli pro hostitelské buňky. Následující kapitoly přehledně shrnují informace o nejlépe prozkoumaných typech těchto repetitivních sekvencí.

2.3.1. Rozptýlené repetitivní oligonukleotidové motivy

Některé sekvenční motivy, kratší než 10 bází a rozmístěné po celé délce genomu (včetně kódujících oblastí, tj. genů), plní v hostitelských buňkách významné funkce. Mezi nejlépe studované repetice tohoto typu patří GATC sekvence, χ motivy a USS, resp. DUS sekvence.

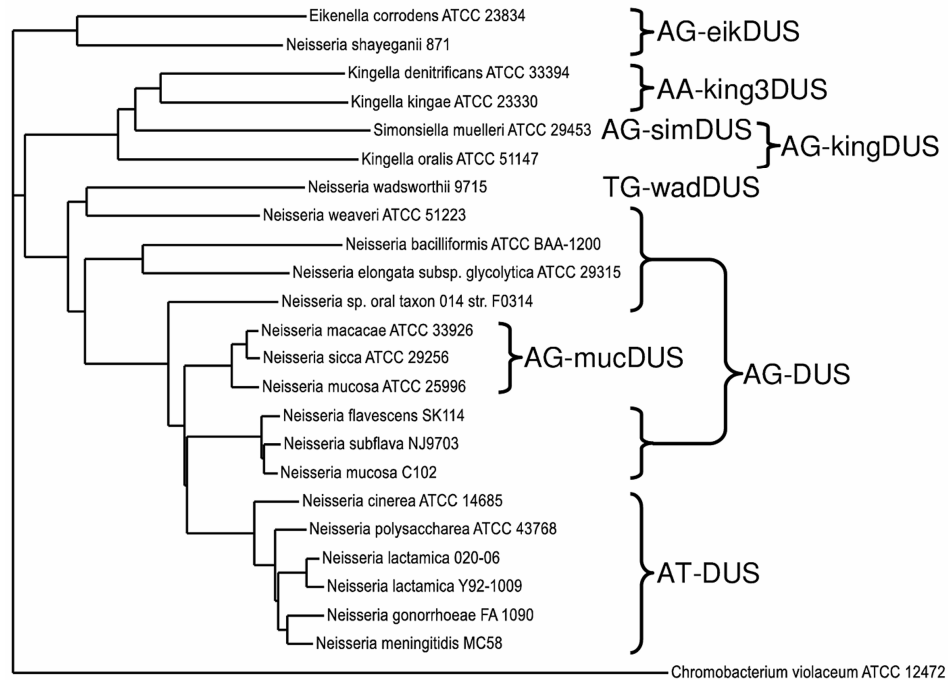
Palindromické sekvence 5'-GATC-3' jsou u některých gamaproteobakterií, mezi nimi i *E. coli*, rozpoznávány a methylovány na adeninu sekvenčně-specifickou metylázou Dam (Lobner-Olesen et al. 2005). Jelikož při replikaci dochází k inkorporaci nemodifikovaného adeninu, jsou po replikaci GATC palindromy v buněčné dvouřetězcové DNA dočasně hemimetylované. Hemimetylovaná DNA je důležitým typem signálu. Zaprvé se na ni váže protein SeqA, jenž následně inhibuje předčasnou reiniciaci replikace chromozomu (Molina a Skarstad 2004). Zadruhé, hemimetylace dceřiného vlákna umožňuje opravu replikačních chyb (mutací), jež se v něm vyskytují (Li 2008). Rozhodující roli v tomto procesu hraje endonukleáza MutH, která specificky zavádí utajené přerušení do nemetylovaného řetězce hemimetylovaného GATC palindromu (Welsh et al. 1987).

χ motivy jsou známy z širokého spektra bakterií. Tyto sekvence fungují jako tzv. horká místa (angl. hot spots) homologní rekombinace. Při opravě dvojřetězcových zlomů v chromozomální DNA dochází k jednosměrnému

odbourávání obou řetězců enzymovým komplexem RecBCD. V okamžiku, kdy komplex dorazí k χ motivu, vyvolá vazba sekvence motivu konformační změnu proteinů. Důsledkem této reorganizace enzymového komplexu je pokračující exonukleázové odbourávání již jen jednoho vlákna DNA, což vede k obnažení rekombinogenní jednořetězcové DNA (Dillingham a Kowalczykowski 2008; Smith 2012). U *E. coli* je frekvence oktanukleotidových χ sekvencí mnohonásobně vyšší než četnost ostatních možných oktamerů (Halpern et al. 2007). Vysoká abundance χ motivů je tedy patrně evolučním mechanismem zefektivnění rekombinační opravy DNA.

Repetice třídy USS (angl. uptake signal sequences), resp. DUS (angl. DNA uptake sequences), reprezentují extrémně početnou skupinu oligonukleotidových repetit, jejichž počet v genomu typicky přesahuje tisíc kopií, a dohromady zabírají přibližně jedno procento celkové velikosti genomové DNA (Frye et al. 2013). Vyskytují se v genomech bakterií z čeledi *Pasteurellaceae* (Redfield et al. 2006) a *Neisseriaceae* (Frye et al. 2013). Tyto repetice se uplatňují při přirozené transformaci, tj. aktivnímu příjmu nahé, lineární DNA z vnějšího prostředí skrz cytoplazmatickou membránu dovnitř buňky, kde je posléze rekombinačně včleněna do homologních úseků hostitelského chromozomu. Specifická vazba USS (DUS), zprostředkována proteinem ComP (Cehovin et al. 2013), je nezbytná pro transformaci. Sekvence USS (DUS) motivů jsou druhově specifické (**Obr. 11**) a zodpovídají tak za selektivní transformaci DNA pocházející z příslušníků vlastního druhu. Za podmínek intenzivní frekvence přirozené transformace hostitelských bakterií (Feil a Spratt 2001) je vysoká abundance USS (DUS) motivů v chromozomu předpokladem pro efektivní rozpoznání exogenní DNA vlastního druhu. Příčiny této abundance, resp. jejího udržení, jsou komplexní (Treangen et al. 2008) a zásadní roli v nich hraje šíření repetit horizontálním (transformačním) přenosem, jež nadto podmiňují (přenášejí se spolu s transformovanou DNA) (Smith et al. 1995).

| | | |
|-------------|---|---------------|
| AT-DUS | : | ATGCCGCTCTGAA |
| TG-wadDUS | : | TGCCGCTCTGAA |
| AG-DUS | : | AGGCCGCTCTGAA |
| AG-mucDUS | : | AGGTCGCTCTGAA |
| AG-simDUS | : | AGGCTCCCTGAA |
| AG-kingDUS | : | AGGCAACCTGAA |
| AA-king3DUS | : | AAGCAACCTGCA |
| AG-eikDUS | : | AGGCTACCTGAA |



Obrázek 11: Specificita DUS sekvencí u příslušníků čeledi *Neisseriaceae*. Nahoře – sekvence jednotlivých variant DUS repetit. Dole – fylogram hostitelských druhů a distribuce jednotlivých variant repetit.

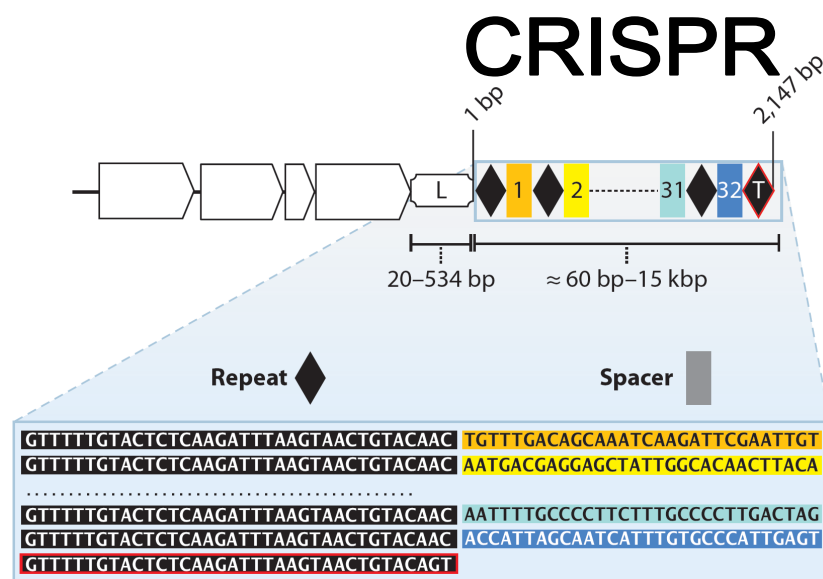
Podle Frye et al. 2013

2.3.2. CRISPR elementy

Velká a dobře definovaná třída CRISPR (angl. clustered regularly interspaced short palindromic repeats) elementů byla popsána relativně nedávno (Jansen et al. 2002). Tyto elementy se vyskytují v mikroorganismech napříč prokaryotní fylogenetickou diverzitou (Sorek et al. 2013). Soudě dle počtu souvisejících publikací, jedná se o jedno z nejzajímavějších témat současné bakteriální genetiky a evoluční mikrobiologie.

Funkcí CRISPR elementů je obrana buňky před infekcí mimochromozomálními mobilními genetickými elementy – konjugativními

plazmidy a fágy. K tomuto účelu je využívána sada enzymů, převážně nukleáz (Cas proteiny) a krátké molekuly RNA. Všechny tyto komponenty jsou kódované jedním genovým shlukem (**Obr. 12**). Kardinální součástí CRISPR elementů jsou mnohonásobně zmnožené krátké, přímé, slabě palindromické repetice o délce menší než 50 bp. Mezi repeticemi jsou vmezeřeny rovněž krátké úseky DNA, tzv. spacers, které mají shodnou délku, ale unikátní sekvence. Sekvence spacerů jsou homologní téměř výhradně se sekvencemi exogenních mobilních genetických elementů (Bolotin et al. 2005; Pourcel et al. 2005). Nové spacers vznikají integrací fragmentů mimochromozomální DNA vždy na jeden konec CRISPR. Pořadí spacerů v rámci CRISPR tedy odpovídá časové posloupnosti evolučních událostí, při nichž docházelo k infekcím daného mikrobiálního kmene (Barrangou et al. 2007).



Obrázek 12: Obecné schéma typického CRISPR lokusu a asociovaných genů. Součástí CRISPR lokusu jsou přímé repetice (černý kosočtverec) a unikátní spacers (každé sekvenci odpovídá jiná barva a číslo). Geny Cas proteinů jsou znázorněny bílými šipkami.

Podle Deveau et al. 2010

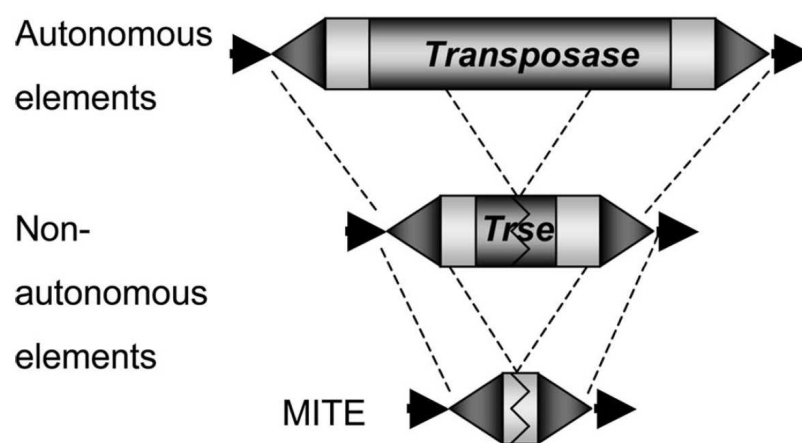
Přítomnost homologního spaceru vede k rezistenci (imunitě) vůči danému mobilnímu elementu. Rezistence je způsobená interferencí malých molekul RNA komplementárních k sekvenci spaceru, které asistují Cas proteinům ve štěpení exogenních nukleových kyselin. Interferující RNA vznikají štěpením transkriptu CRISPR lokusu. V tomto procesu se rovněž uplatňují Cas proteiny (Horvath a Barrangou 2010). Ačkoliv přítomnost různých homologních tříd *cas* genů se liší mezi jednotlivými typy CRISPR a pouze dva z těchto proteinů jsou zcela konzervované (Makarova et al. 2011), struktura vlastních CRISPR lokusů (přímé repetice + unikátní spacery) je neměnná.

Diverzita CRISPR lokusů je podivuhodná. Často se vyskytuje několik typů CRISPR elementů v genomech téhož mikrobiálního druhu. Například u bakterie *Streptococcus thermophilus* jsou známy tři typy CRISPR a každý byl zjištěn v nadpoloviční většině testovaných kmenů. Další úroveň variability je daná proměnlivým počtem spacerů v CRISPR. CRISPR lokusy *S. thermophilus* průměrně obsahují 23 spacerů, resp. přímých repetic (Horvath et al. 2008). Přibývání spacerů závisí na setkáních s mimochromozomálními mobilními elementy, jež jsou unikátní pro danou buňku, resp. linii jejích potomků. Snižování počtu spacerů se realizuje obecnou homologní rekombinací mezi přímými, slabě palindromickými repeticemi (Deveau et al. 2010). K jejich delecím může dále docházet při tzv. sklouznutí templátu DNA polymerázy během replikace (Levinson a Gutman 1987).

2.3.3. Neautonomní transponibilní elementy

Repetitivní elementy, jež budou popsány v této kapitole, se významně liší od krátkých oligonukleotidových motivů (**Kap. 2.3.1.**). Protože jsou delší (několik desítek bp), jejich přítomnost v hostitelských genech by významně ovlivňovala aminokyselinové složení jejich proteinových produktů. Z toho důvodu se vyskytují téměř výlučně v mezigenových (nekódujících) oblastech (Delihias 2011).

Charakteristickou vlastností neautonomních transponibilních elementů je jejich mobilita, tj. schopnost měnit svoji polohu v hostitelském genomu mechanismem transpozice (srv. **Kap. 2.2.2.**) – odtud část názvu. Neautonomními jsou nazývány kvůli neschopnosti samostatného přemísťování v genomech, neboť nekódují vlastní transponázu, na rozdíl od autonomních elementů – inzerčních sekvencí (**Kap. 2.2.2.**). Pro svou mobilizaci využívají transponázy kódované autonomními elementy. Neautonomní elementy vznikaly v evoluci delecemi části nebo celého genu transponázy. Vždy však mají zachované terminální sekvence, jejichž rozpoznání příslušnou transponázou je nutné pro transpozici (**Obr. 13**). Zatímco autonomní elementy lze z funkčního hlediska považovat za parazity, sobecky využívající funkce hostitelské buňky, neautonomní elementy jsou parazité druhého řádu – hyperparazité (parazité parazitů), kteří sobecky využívají transponázy svých "parentálních" autonomních elementů (Gonzalez a Petrov 2009). Pokud jsou terminální sekvence (inverzní repetice) lokalizovány blízko sebe, nazýváme tyto elementy MITEs (angl. **miniature inverted-repeat transposable elements**) (Fattash et al. 2013) (**Obr. 13**).



Obrázek 13: Vztah mezi neautonomními transponibilními elementy a autonomními parentálními elementy. Černý trojúhelník – přímé repetice, šedý trojúhelník – inverzní repetice, tmavší obdélník – gen transponázy (typu DDE), světlejší obdélníky – regulační oblasti.

Podle Wessler 2006

Bylo popsáno mnoho skupin prokaryotních MITEs, ty nejvýznamnější (nejčetnější a nejvíce prostudované) jsou shrnuty v **Tab. 3**. Největší diverzita byla zatím zjištěna v genomech cyanobakterií (Elhai et al. 2008; Lin et al. 2011) a archeí (Filee et al. 2007), kde se často vyskytuje několik typů MITEs u jednoho druhu. U rodu *Synechococcus* byla analýzou metagenomických sekvencí prokázána mobilita dvou forem MITEs a jejich parentálních inzerčních sekvencí v hostitelské populaci (Nelson et al. 2012).

Zatím nebylo zcela objasněno, jakou roli hrají MITEs pro svého hostitele a jaké jsou evoluční důsledky jejich přítomnosti v hostitelských genomech. Díky svým terminálním inverzním repetitívám se mohou v transkribované formě sbalovat do stabilních vlásenkových struktur a ovlivňovat tak terminaci transkripce sousedících genů, což může přispívat k optimalizaci regulace exprese hostitelských genů (Knutsen et al. 2006; Delihias 2011). Podobně mohou MITEs ve svých sekvencích obsahovat promotory (pozůstatky původních elementů regulace transkripce transponázy v paternálním autonomním elementu, **Obr. 13**) (Buisine et al. 2002) a aktivovat tak transkripci přilehlých genů (Matus-Garcia et al. 2012). Přítomnost vyššího počtu MITEs v genomu tak může představovat snadno využitelný "rezervoár variability".

| Hostitelský organismus | Název elementu | Název parentální inzerční sekvence (rodina inzerčních sekvencí) | Délka v bp | Počet kopií v genomu hostitele | Literární zdroj |
|--|--------------------|---|------------|--------------------------------|-------------------------|
| <i>Streptococcus pneumoniae</i> | RUP | IS630- <i>Spn1</i> (IS630) | 107 | 108 | Oggioni a Claverys 1999 |
| | BOX | ISSpn2/ISStso1 (IS630) | 67 - 637 | 127 | Knutsen et al. 2006 |
| <i>Neisseria meningitidis</i> | CREE | neidentifikována | 155 | 127 | Liu et al. 2002 |
| <i>Enterobacteriaceae</i> (více druhů) | ERIC | neidentifikována | 127 | 20-247 | De Gregorio et al. 2005 |
| <i>Synechococcus</i> sp. | ISSoc2 δ -1 | ISSoc2 (IS200/IS605) | 300 | 21-44 | Nelson et al. 2012 |
| | ISSoc2 δ -2 | | 265 | 9-10 | |
| <i>Cyanobacteria</i> (více druhů) | mnoho | parentální IS definovány (více rodin) | <500 | <600 | Lin et al. 2011 |
| <i>Archaea</i> (více druhů) | mnoho | parentální IS definovány (více rodin) | <368 | <69 | Filee et al. 2007 |

Tabulka 3: Souhrnné informace o vybraných skupinách prokaryotních MITE elementů.

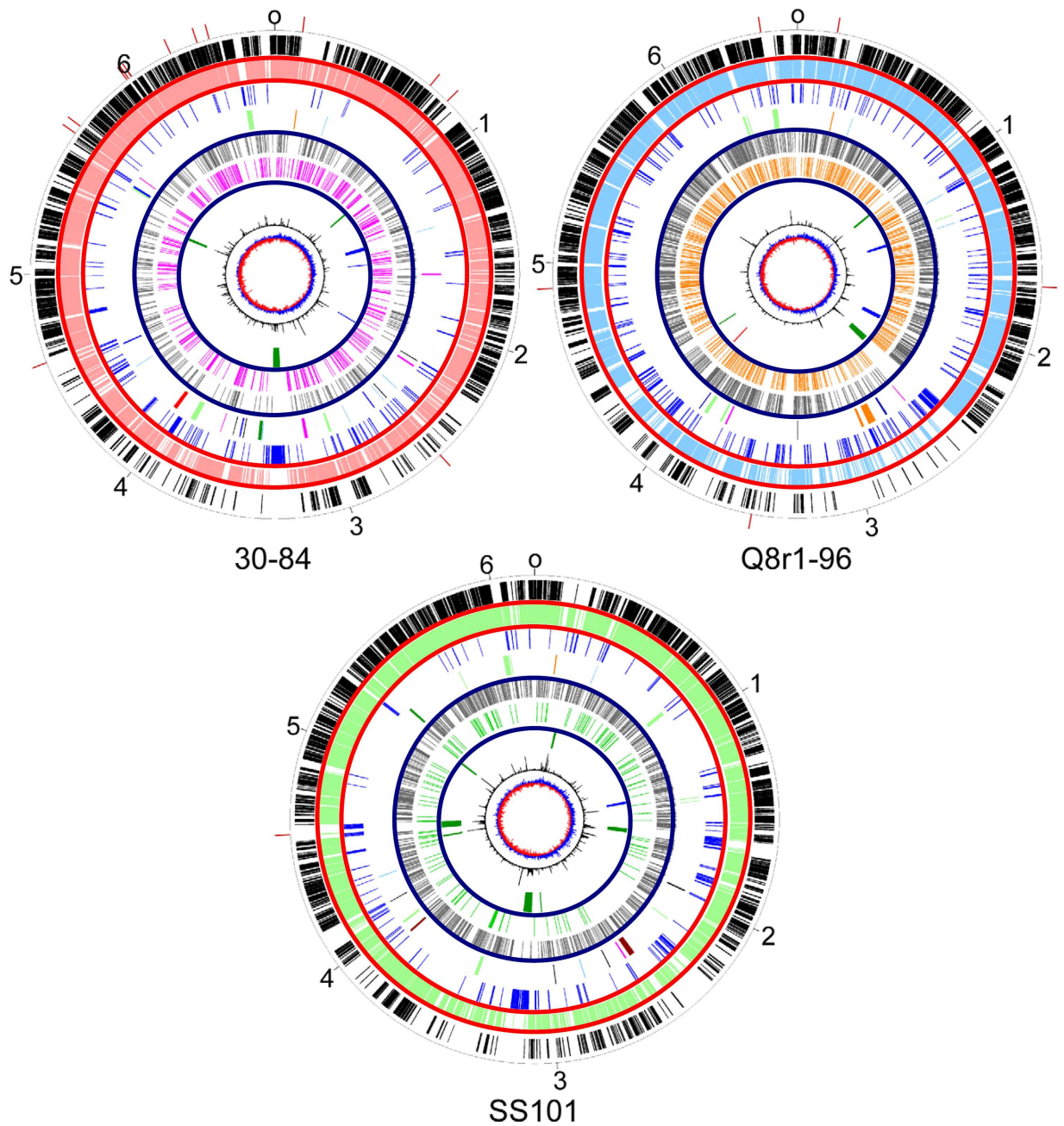
2.3.4. REP elementy

2.3.4.1. Výskyt

Bakteriální REP (angl. repetitive extragenic palindrome) elementy byly poprvé objeveny v genomu *E. coli* v roce 1982 (Higgins et al. 1982). Jak název napovídá, jde o repetitivní elementy (typicky stovky kopií v hostitelském genomu) vyskytující se v nekódujících oblastech DNA, které obsahují ve své sekvenci palindrom (vzájemně komplementární báze).

REP elementy byly kromě *E. coli* identifikovány v genomech příbuzných enterobakterií, *Salmonella typhimurium* (Gilson et al. 1987) a *Klebsiella pneumoniae* (Bachelier et al. 1993). Dále se vyskytují u evolučně vzdálenějších gamaproteobakterií, pseudomonád *Pseudomonas putida* (Aranda-Olmedo et al. 2002; Weinel et al. 2002), *P. syringae* (Tobes a Pareja 2005), *P. entomophila* (Vodovar et al. 2006), *P. fluorescens* (Silby et al. 2009; Bertels a Rainey 2011) a *P. aeruginosa* (Weinel et al. 2002), druhu *Stenotrophomonas maltophilia* (Roschetto et al. 2008; Rocco et al. 2010) a dalších (Tobes a Ramos 2005). Sekvence REP elementů jsou vždy unikátní pro daný druh. Mezi blízkými příbuznými druhy se mohou lišit jen málo (Gilson et al. 1987).

REP elementy nejsou v hostitelských genomech rozloženy zcela rovnoměrně. U *P. fluorescens* (Silby et al. 2009; Loper et al. 2012) a *S. maltophilia* (Rocco et al. 2010) se vyskytují téměř výlučně v tzv. core genomu, což jsou chromozomální segmenty sdílené všemi kmeny daného bakteriálního druhu (Segerman 2012). Core genomy jsou typicky fragmentovány inzercemi horizontálně se šířících mobilních elementů, v nichž REP sekvence chybí (**Obr. 14**). K horizontálnímu šíření REP elementů tedy ve větší míře nedochází.



Obrázek 14: Rozložení REP elementů v genomech tří kmenů *P. fluorescens* (30-84, Q8r1-96 a SS101). Pozice REP sekvencí je znázorněna vždy ve dvou kruzích v oblasti mezi tmavomodrými kružnicemi (1 čára = 1 REP). Různé barvy reprezentují různé typy REP sekvencí. Core genom fylogenetické linie každého kmene je barevně vyznačen v oblasti mezi červenými kružnicemi.

Podle Loper et al. 2012

2.3.4.2. Struktura

Kromě přítomnosti komplementárních (palindromických) sekvencí se většina popsaných REP elementů vyznačuje přítomností terminálního tetranukleotidu 5'-GTAG-3'. Délky palindromu a jejich sekvence jsou v rámci diverzity REP elementů značně variabilní, někdy mohou být palindromy přerušené krátkými navzájem nekompatibilními úseky. REP elementy jednoho druhu mohou být polymorfní z hlediska délky a sekvence (**Tab. 4**).

| Hostitelský druh bakterie | Sekvence REP elementů | Počet kopií v genomu | Literární zdroj |
|-----------------------------|---|----------------------|---------------------------|
| <i>E. coli</i> K-12 | <u>GTAGGCCGGATAAGGCGT</u> <u>TYACGCCGCATCCGGC</u> | 79 | Bachelier et al. 1999 |
| | <u>GTAGGCCTGATAAGCGTAGCGCATCAGGC</u> | 45 | |
| | <u>GTAGGCCTGATAAGACGCGCMAGCGT</u> <u>CGCATCAGGC</u> | 21 | |
| <i>P. aeruginosa</i> PAO1 | <u>GTAGGGCGAATAACCGC</u> <u>NNNGCGGTTATCCGCC</u> | 17 | Weinel et al. 2002 |
| | <u>GTAGGGCGAATAACGCC</u> <u>NNNGGCGTTATCCGCC</u> | 26 | |
| <i>P. putida</i> KT2440 | <u>GTAGGAGCGGGTT</u> <u>ACCCGCGAA</u> | 205 | Aranda-Olmedo et al. 2002 |
| <i>S. maltophilia</i> K279a | <u>GTAGTGCCGGCCGCT</u> <u>GGCCGGCA</u> | 437 | Roschetto et al. 2008 |

Tabulka 4: Sekvence a četnosti vybraných publikovaných REP elementů. Palindromické části sekvencí jsou tučné a podtržené. Konzervovaný tetranukleotid je zvýrazněn šedě. Zkratky variabilních bází jsou v souladu s nomenklaturou IUPAC. Počty kopií jednotlivých elementů byly určeny v programu pDRAW32 (www.acaclone.com).

Většina REP elementů se v hostitelských genomech nevyskytuje samostatně, nýbrž jsou agregovány do shluků. Častým typem shluků je dimer dvou blízkých inverzně orientovaných REP sekvencí, REPIN (angl. **REP** doublet forming hairpin), původně identifikovaný u *P. fluorescens* (Bertels a Rainey 2011). Elementy typu

REPIN dále reprezentují predominantní formu REP sekvencí v genomech *P. putida* (Aranda-Olmedo et al. 2002) a *S. maltophilia* (Rocco et al. 2010). REP sekvence mohou být nadto uspořádány do repetitivních struktur vyšších řádů, BIME (angl. **b**acterial **i**nterspersed **m**osaic **e**lement) (Gilson et al. 1991). BIME vykazují unikátní architekturu. Opakují se v nich dimery inverzních REP sekvencí typu REPIN. Tyto jednotky jsou spolu s mezilehlými sekvencemi tandemově uspořádány. Celková struktura BIME tak sestává jak z přímých, tak z inverzních repetitivních sekvencí (**Obr. 15**). BIME jsou častou formou výskytu REP v genomu *E. coli* (Bachellier et al. 1999). U této bakterie se vyskytuje několik sekvenčních variant REP elementů (**Tab. 4**). V tandemově se opakujících dimerech v rámci BIME vždy alternují dvě různé REP varianty.



Obrázek 15: Schématické znázornění struktury BIME elementu *E. coli*. Bílé obdélníky reprezentují REP sekvence, poloha vnitřních černých trojúhelníků vyznačuje jejich orientaci. Žluté, resp. modré obdélníky značí dvě různé mezilehlé sekvence. Z² a Y značí dva typy REP sekvencí. Závorky značí celou jednotku, jež je v rámci BIME tandemově opakována (*n* je typicky menší než 10).

Převzato z <http://www.pasteur.fr/recherche/unites/pmtg/repet/intro.BIME.html>

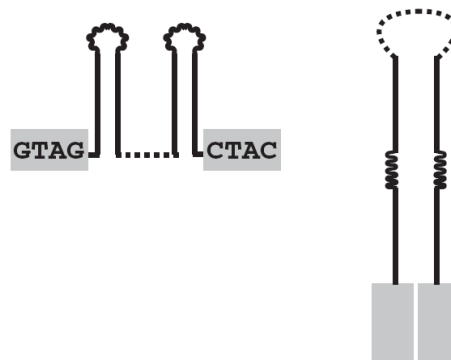
Celkový počet tandemově opakovaných jednotek se liší mezi jednotlivými kmeny stejného druhu, což vede k délkové variabilitě BIME (Bachellier et al. 1997; Roschetto et al. 2008). Délka (počet jednotek) BIME je podobná nebo shodná u příbuzných kmenů. BIME tedy pravděpodobně podléhají relativně pomalé, leč konstantní evoluci.

2.3.4.3. Funkce

Bylo zjištěno, že REP elementy, resp. jejich agregáty (REPIN, BIME – **Kap. 2.3.4.2.**), slouží jako vazebné sekvence několika bakteriálních proteinů. Dimer dvou blízkých, inverzně orientovaných REP elementů (typ REPIN) váže u *E. coli* integrační hostitelský faktor (IHF) (Boccard a Prentki 1993; Oppenheim et al. 1993), neesenciální protein, jenž reguluje širokou škálu buněčných procesů, včetně genové exprese (Freundlich et al. 1992). Na REP elementy *E. coli* se rovněž váže DNA polymeráza I (Gilson et al. 1990), zásadní pro replikaci zpoždujícího se vlákna DNA a opravné mechanismy DNA (Savic et al. 1990). REP elementy, které jsou součástí BIME, jsou rozpoznávány a štěpeny DNA gyrázou (Yang a Ames 1988; Espeli a Boccard 1997), topoizomerázou katalyzující tvorbu negativních nadobrátek v DNA. BIME-dependentní modulace superhelicity tak může mít vliv na kompaktaci chromozomu do topologických domén (Yang a Ames 1988).

Je obtížné přesně určit rozsah modulace hostitelských funkcí přítomností REP elementů, vzhledem k jejich vysokému počtu kopií, který znesnadňuje potenciální mutační studie. Ač bylo vytvořeno několik kmenů *E. coli* s výrazně redukovanými genomy, zbavenými mobilních a horizontálně přenášených elementů (Kolisnychenko et al. 2002; Hashimoto et al. 2005), REP elementy jsou v těchto pracích opomíjeny. Mutantní kmen, zcela prostý REP elementů, zatím nebyl sestrojen.

Několik studií ukázalo, že REP elementy hrají roli v regulaci transkripce. Pro tuto funkci je důležitá přítomnost dvou inverzních REP sekvencí typu REPIN, jež mohou v tomto uspořádání tvořit velké vlásenkové sekundární struktury (díky své palindromicitě mohou tvořit vlásenky i samostatné REP elementy – **Obr. 16**). V tomto ohledu se struktury REPIN podobají MITEs (**Kap. 2.3.3.**).



Obrázek 16: Tvorba alternativních vlásenkových struktur u dimeru REP elementů v inverzní orientaci (typ REPIN). Vlevo - dvojice malých vlásenek (každá odpovídá jednomu REP elementu), vpravo – velká vlásenka z REP dimeru. Konzervované tetranukleotidy jsou vyznačeny šedým rámečkem.

Podle Rocco et al. 2010

U *E. coli* vede přítomnost transkribovaných BIME sekvencí v mRNA k terminaci transkripce, závislé na transkripčním terminátoru Rho (Espeli et al. 2001). Terminačně fungují též inverzní REP dimery u *P. putida*, kde se však nejedná o Rho-závislý proces, nýbrž o nespecifickou terminaci závislou na přítomnosti obecné vlásenkové struktury (Aranda-Olmedo et al. 2002). Přítomnost sekundárních vlásenkových struktur v BIME zvyšuje stabilitu transkriptu u *E. coli* (Espeli et al. 2001). Obdobná stabilizace mRNA byla popsána i u REP elementů *S. maltophilia* (Rocco et al. 2010).

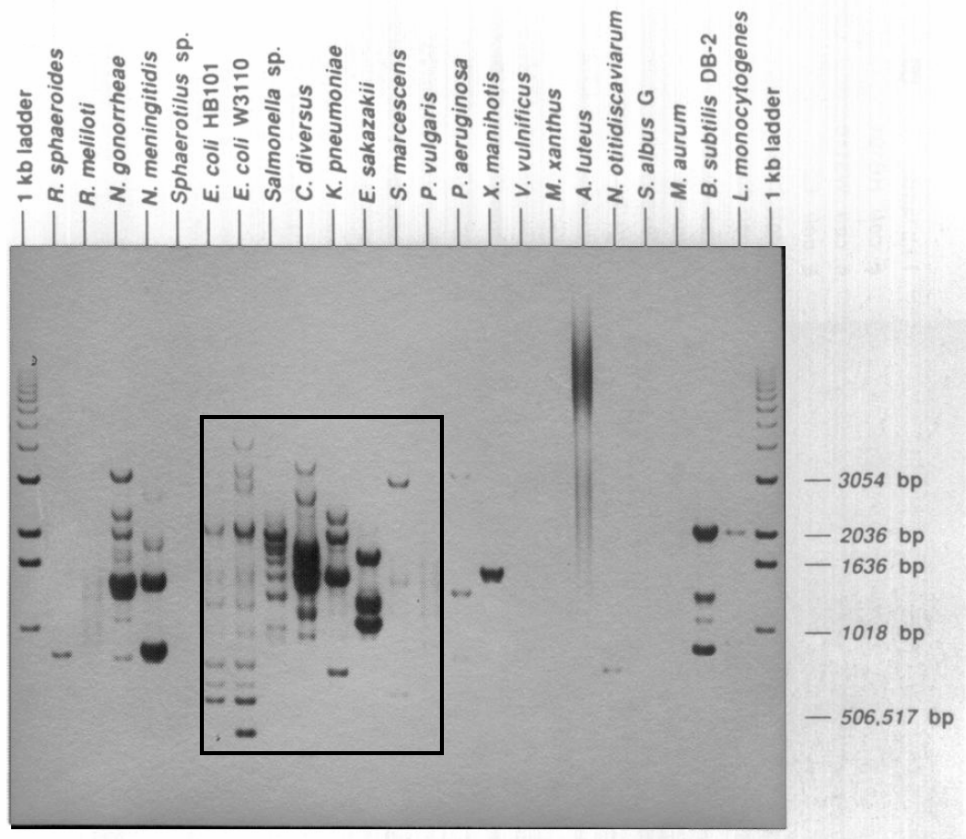
Některé skupiny inzerčních sekvencí, patřící do DDE skupiny podle typu jimi kódovaných transponáz (**Kap. 2.2.2.1.**), využívají REP elementy jako cílová místa pro svou integraci. Jedná se zaprvé o rodinu inzerčních sekvencí IS110/IS492. Zde byla sekvenční specifita inzerce do REP elementů zjištěna u ISPpu10 z *P. putida* (Ramos-Gonzalez et al. 2006), IS621 z *E. coli* a IS1594 z *Anabaena* sp. (Choi et al. 2003) a dalších (Tobes a Pareja 2006). Tyto inzerční sekvence se integrují do místa v REP elementu, které se nachází u paty vlásenky. Dvě inzerční sekvence z rodiny IS3, IS1397 z *E. coli* (Clement et al. 1999) a ISKpn1 z *Klebsiella pneumoniae* (Wilde et al. 2001), se naopak integrují do jednořetězcové sekvence na vrcholu vlásenky. Ačkoliv

hostitelské bakterie a sekvence cílových REP elementů dvou posledně zmiňovaných inzerčních sekvencí jsou si vzájemně velmi podobné, transponázy dokáží mezi nimi specificky rozlišovat (Wilde et al. 2003).

2.3.4.4. Využití

Výskytu REP elementů v bakteriálních genomech lze využít k účelům molekulární taxonomie. V dnes již klasické publikaci (Versalovic et al. 1991) byl vyvinut typ PCR reakce (angl. **polymerase chain reaction**), tzv. REP-PCR, která využívala sekvence REP elementů z *E. coli* jako primery pro amplifikaci segmentů bakteriální genomové DNA. Vzhledem k vysoké četnosti REP elementů dochází k amplifikaci fragmentů ležících mezi dvěma blízkými REP kopiemi. Jelikož sekvence REP elementů jsou druhově specifické (**Kap. 2.3.4.1.**), PCR využívající tyto primery dávala, až na výjimky, vzniknout souboru amplifikovaných fragmentů u *E. coli* a příbuzných druhů z čeledi *Enterobacteriaceae*, u nichž se vyskytují sekvenčně podobné REP elementy (Gilson et al. 1990) (**Obr. 17**). Přesto se REP-PCR velice často využívá pro genotypizaci zcela nepříbuzných a evolučně vzdálených bakterií. U těchto druhů při absenci REP elementů dochází k nesespecifické amplifikaci fragmentů ohraničených sekvencemi se slabou náhodnou homologií k REP-PCR primerům (Herman a Heyndrickx 2000). Výstupy REP-PCR pak spíše připomínají tzv. arbitrární PCR (McClelland a Welsh 1994).

Jiný typ genotypizace využívá variability počtu jednotek v BIME elementech (**Kap. 2.3.4.2.**). Při vhodně navržených primerech, těsně hraničících s BIME, vypovídá velikost amplifikovaných fragmentů o délce BIME. Analýza velikostí několika BIME lokusů současně byla využita pro rychlou genotypizaci izolátů *S. maltophilia* (Roschetto et al. 2008).



Obrázek 17: Výsledky REP-PCR u různých bakterií. Druhy patřící mezi enterobakterie jsou vyznačeny obdélníkem.

Podle Versalovic et al. 1991

3. PUBLIKACE AUTORA

1.

Nunvář J, Hucková T, Lichá I:

Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes.

BMC Genomics 2010, **11**:44

Elektronická verze publikace: <http://www.biomedcentral.com/1471-2164/11/44>

Celý text publikace v tištěné podobě práce: **Příloha č. 1**

IF₂₀₁₀=4,206; 18 citačních ohlasů (Google Scholar)

Bakteriální REP elementy jsou repetitivní sekvence DNA, charakteristické segmenty komplementárních bází (palindromy), značnou četností v hostitelských genomech a druhovou specificitou. Mohou se sdružovat do repetitivních sekvencí vyšších řádů, tzv. BIME. BIME jsou tvořeny tandemově multiplikovanými dimery REP elementů v inverzní orientaci. Jsou známy interakce REP a BIME elementů s proteiny důležitými pro fyziologii hostitelské buňky a jejich role v transkripčních procesech. Mechanismy multiplikace REP elementů a vzniku BIME, stejně jako příčiny druhové specificity REP sekvencí, nebyly v době publikace této studie známy.

Práce prezentuje odborné veřejnosti objev nové třídy proteinů pojmenované RAYT. RAYT proteiny jsou vzdáleně příbuzné s Y1 typem tyrosinových transponáz inserčních sekvencí rodiny IS200/IS605, v jejich aminokyselinových sekvencích se vyskytují specifické motivy, nepřítomné u kanonických IS200/IS605 transponáz. Inverzní repeticity asociované s geny RAYT proteinů jsou značně abundanční v hostitelských genomech a odpovídají již charakterizovaným REP elementům. Na základě podobnosti mezi jednotkou REP-*rayt*-REP a inserčními sekvencemi rodiny IS200/IS605 jsem vyslovil hypotézu o RAYT proteinech jakožto nukleázách štěpících a diseminujících REP sekvence. Mobilní původ a transponibilní povaha REP elementů nabízí relativně jednoduché vysvětlení jejich typických vlastností, např. vnitro- a mezidruhové variability, a vzniku unikátních struktur BIME elementů.

Nunvář J, Lichá I, Schneider B:

Evolution of REP diversity: a comparative study.

BMC Genomics 2013, **14**:385

Elektronická verze publikace: <http://www.biomedcentral.com/1471-2164/14/385>

Celý text publikace v tištěné podobě práce: **Příloha č. 2**

IF₂₀₁₂= 4.397; bez citačních ohlasů

V této práci analyzuji diverzitu REP elementů s ohledem na evoluci hostitelských bakterií a zkoumám vliv RAYT proteinů, pravděpodobně zodpovědných za mobilizaci REP elementů, na evoluční distribuci a celkovou početnost těchto repetitivních sekvencí.

Analyzoval jsem sadu genomových sekvencí dvou fylogenetických linií environmentálních bakterií – fluorescentních pseudomonád a stenotrofomonád. Tyto skupiny byly zvoleny z důvodů dostatečně vysokého množství přečtených kompletních genomových sekvencí, vysokého počtu přítomných sekvenčních tříd REP elementů a asociovaných RAYT proteinů a také kvůli jejich biotechnologickému, biokontrolnímu a klinickému významu.

Celková diverzita REP elementů sestává z 22, resp. 13 unikátních sekvenčních tříd a s nimi asociovaných genů RAYT proteinů. Často se vyskytuje několik tříd REP elementů společně v jednom bakteriálním kmeni. Distribuce jednotlivých sekvenčních tříd vykazuje značnou variabilitu v rámci fylogenetických linií hostitelů i mezi nimi a jejich četnost pozitivně koreluje s přítomností příslušných RAYT proteinů. Syntenní analýza detekovala opakovaná nahrazení REP elementů asociovaných s RAYT geny jinou sekvenční třídou, aniž by došlo ke změně chromozomální lokalizace genu. REP elementy obou bakteriálních skupin vykazují charakteristické sekvenční rysy (molekulární apomorfie) a asociované RAYT proteiny tvoří monofyletické skupiny.

Výsledky implikují dlouhodobou přítomnost systémů RAYT/REP v genomech obou fylogenetických linií a naznačují jejich možnou výhodnost pro hostitelské bakterie.

Nunvář J, Dřevínek P, Lichá I:

DNA profiling of *Stenotrophomonas maltophilia* by PCR targeted to its species-specific repetitive palindromic sequences.

Letters in Applied Microbiology 2012, **54**(1):59-66

Elektronická verze publikace: <http://onlinelibrary.wiley.com/doi/10.1111/j.1472-765X.2011.03172.x/abstract>

Celý text publikace v tištěné podobě práce: **Příloha č. 3**

IF₂₀₁₂= 1,629; bez citačních ohlasů

Tato práce se zabývá využitím druhově specifických REP elementů v genomech bakterií *Stenotrophomonas maltophilia* pro genotypizační účely. Genotypizace je nutná pro určení příbuzenských vztahů izolátů z klinických nebo environmentálních vzorků, ev. determinaci jejich klonální spřízněnosti. Je s výhodou, pokud je použita metoda rychlá a jednoduchá na provedení. *S. maltophilia* je velmi variabilní na genetické úrovni a genetická struktura její metapopulace je stále předmětem bádání.

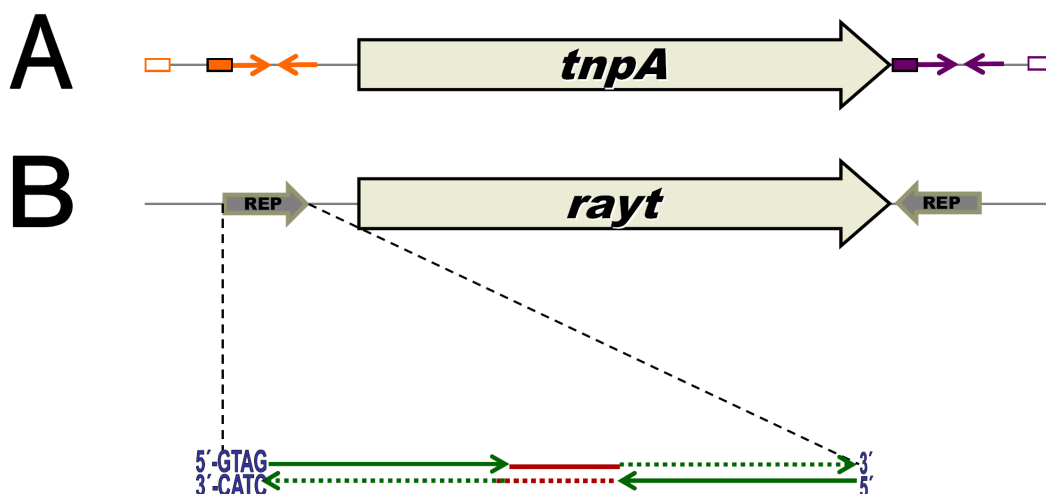
Díky vysokým četnostem REP elementů se mohou vyskytovat dvě kopie ve vzájemné blízkosti a inverzní orientaci, čehož lze využít pro amplifikaci mezilehlého úseku pomocí PCR reakce, využívající jediný primer komplementární k REP elementům (SmrepPCR). Porovnání amplikonů získaných pomocí SmrepPCR poskytlo odhad příbuznosti izolátů, který velmi dobře koreloval s jejich fylogenetickou příbuzností (vyvozenou z nukleotidových sekvencí esenciálního genu *gyrB*). Klonalita izolátů s identickými SmrepPCR amplikony byla potvrzena porovnáním jejich makrorestrikčních profilů. SmrepPCR je tak časově a finančně nenáročnou metodou využitelnou v populační genetice *S. maltophilia*.

Veškeré analýzy a experimenty publikované v uvedených třech publikacích jsem navrhnul a provedl samostatně, rovněž jsem sám sepisoval původní rukopisy. Tereza Hucková, druhá autorka první publikace, poskytla inspiraci pro analýzu genomu *S. maltophilia*, jež vedla k objevu RAYT proteinů. Role senior autorů spočívala v odborném vedení a připomínkování rukopisů.

4. DISKUZE

REP-asociované tyrosinové transponázy

Stěžejním tématem této práce jsou RAYT proteiny (angl. **REP-associated tyrosine transposases**) (Nunvar et al. 2010). Tyto proteiny jsou vzdáleně příbuzné Y1 transponázám inzerčních sekvencí rodiny IS200/IS605, s nimiž sdílejí podobnou délku a konzervované aminokyseliny v jejich primární struktuře – katalytický tyrosin a HUH motiv (**Kap. 2.2.2.2.**). RAYT proteiny jsou charakteristické uspořádáním svých genů, které jsou z obou konců obklopeny jedním nebo několika REP elementy (**Kap. 2.3.4.2.**) v inverzní orientaci (**Obr. 18**).



Obrázek 18: Uspořádání genů kódujících IS200/IS605 transponázy (A) a RAYT proteiny (B).

Barevnými plnými šipkami jsou znázorněné palindromické sekvence. Plné a prázdné obdélníky značí párující tetranukleotidy u IS200/IS605 transponáz (vždy stejnou barvou). Obecná struktura REP elementu je rozvedena na úroveň jednotlivých vláken DNA.

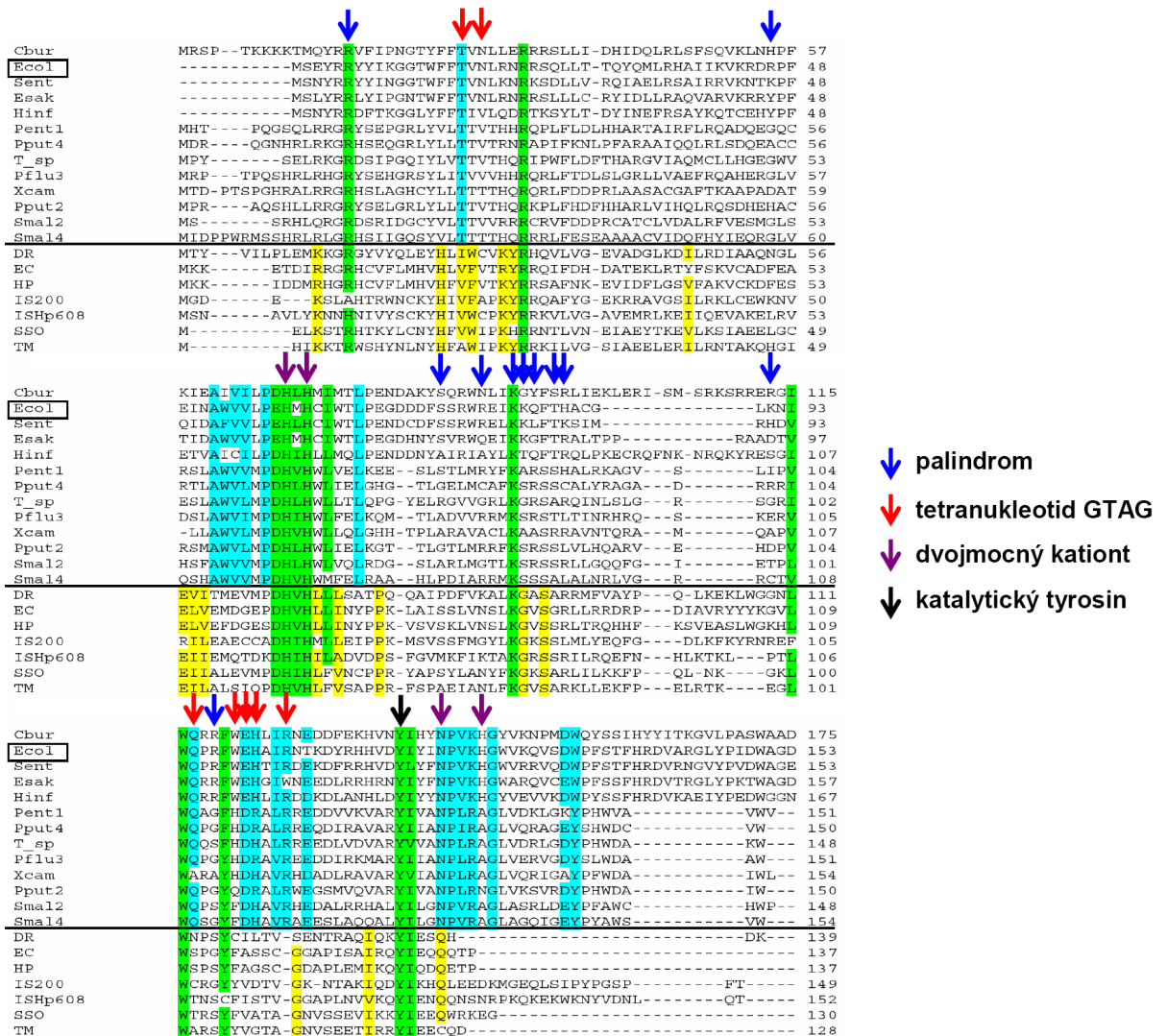
U bakterií je fyzická asociace nukleotidových sekvencí (nejen genů) znakem jejich funkčního propojení. Typickými příklady jsou operony jakožto shluky genů stejné dráhy nebo funkce, ale i např. CRISPR lokusy (**Kap. 2.3.2.**). Transponázy inzerčních sekvencí rovněž rozpoznávají a štěpí sekvence přítomné v blízkosti jejich genů, čímž si tyto sobecké elementy zajišťují svoji mobilitu. Typem rozpoznávané DNA u inzerčních sekvencí rodiny IS200/IS605 jsou palindromy v jednořetězcovém

stavu, kdy zaujímají konformaci vlásenky. Přítomnost palindromu v sekvenci je definující vlastností REP elementů (**Tab. 4**). Na základě výše uvedených faktů jsem vyvodil, že RAYT proteiny jsou nukleázy specificky štěpící REP elementy. Tento předpoklad byl následně experimentálně potvrzen týmem Dr. Michaela Chandlera při pokusech *in vitro*, jež využívaly purifikovaný RAYT z *E. coli* a jednořetězcové sekvence sousedící s *rayt* genem, jejichž součástmi byly REP elementy. Přítomnost vlásenky a konzervovaného tetranukleotidu GTAG, jenž je obligátní součástí REP elementů sousedících s *rayt* geny (Nunvar et al. 2010), se ukázaly jako nezbytné podmínky pro štěpení (Ton-Hoang et al. 2012).

Funkční a evoluční vazba mezi RAYT proteiny a REP elementy

Kromě asociace jejich genů s REP elementy jsou RAYT proteiny definovány přítomností vysoce konzervovaných sekvenčních motivů, které chybí u IS200/IS605 transponáz (Nunvar et al. 2010). Dalo se předpokládat, že tyto motivy budou souviset s vazbou REP elementů. Nejkonzervovanějším v rámci sekvencí REP elementů je tetranukleotid GTAG (**Tab. 4**). Nedávno publikovaná struktura RAYT proteinu z *E. coli* v komplexu se substrátovou REP vlásenkou (Messing et al. 2012) ukázala, že motivy specifické pro RAYT proteiny se z velké části účastní právě vazby tetranukleotidu GTAG. Jedná se o invariantní N-koncový threonin a C-koncový segment o sekvenci W,Q,x,x,F/Y,x,W/H/F,D/E,H/R,x,I/L/V,R,x,D/E (**Obr. 19**).

Tyto GTAG-vazebné aminokyseliny jsem našel v sekvencích všech RAYT proteinů, jež jsem dosud identifikoval (stovky proteinů, převážně neanotovaných), bez výjimky (vlastní nepublikovaná data).



Obrázek 19: Interakce aminokyseliny RAYT proteinu z *E. coli* (vyznačeno obdélníkem) se substrátovou DNA a jejich konzervovanost. V alignmentu reprezentativního vzorku RAYT proteinů (nad čarou) a kanonických IS200/IS605 transponáz (pod čarou) jsou pozice aminokyselin interagujících s palindromem, GTAG tetranukleotidem a dvojmocným kationtem vyznačeny šipkami příslušných barev. Konzervované aminokyseliny jsou zvýrazněné: modře (specifické pro RAYT proteiny), žlutě (specifické pro IS200/IS605 transponázy), zeleně (sdílené oběma skupinami).

Podle Nunvar et al. 2010; Messing et al. 2012

Charakteristickým znakem funkčně propojených entit je koevoluce, tj. obdobný průběh evoluce obou partnerů, v níž se odráží jejich vzájemný vztah. Typickým případem koevoluce na organismální úrovni jsou systémy parazit-hostitel (Schulte et al. 2010). Na suborganismální, molekulární úrovni dochází ke koevoluci

mezi interagujícími proteiny (Clark et al. 2011) nebo mezi proteiny a nekódujícími sekvencemi DNA, což je například příčinou skupinově-specifických charakteristik terminálních repetitivních sekvencí (Kap. 2.2.2.). Obdobný typ koevoluce patrně probíhá mezi RAYT proteiny a jim příslušejícími REP elementy. REP sekvence tří bakteriálních evolučních linií (enterobakterií, stenotrofomonád a fluorescentních pseudomonád) vykazují skupinově-specifické charakteristiky, jako je celková délka palindromu, přítomnost přídatných sekvencí vedle palindromu, přerušení palindromu nebo dimorfismus REP sekvencí (Tab. 5). RAYT proteiny, asociované s REP elementy každé skupiny, tvoří vždy monofyletické linie (Nunvar et al. 2010; Nunvar et al. 2013). Příbuznost REP sekvencí tak odráží příbuznost asociovaných RAYT proteinů a patrně vypovídá o specifických požadavcích na rozpoznávání či štěpení substrátové DNA REP elementů RAYT proteiny.

| enterobakterie | stenotrofomonády | fluorescentní pseudomonády |
|--|------------------------------------|----------------------------------|
| <u>GTAGGCCCGGTAAGCGAAGCGCCACCGGGC</u> <u>GTAGGCCCGGATAAGGCGCTTGCGCCGCCATCCGGC</u> | <u>GGTGGGTGCCGACCGCTTGGTCGGCAC</u> | <u>GTGGGAGGGGGCTTGCCCCGAT</u> |
| <u>GTAGGGCCGGTAAGCGGAGCGCACCCGCC</u> <u>GTAGGGTGGTAAGCGCAGCGCACCCACC</u> | <u>GGTGGGTGCCAACCTTGGTTGGCAC</u> | <u>GTGGGAGGGGGCTTGCTCCCGAT</u> |
| <u>GTAGGACGGATAAGGCGTTCAGCCGCATCCGGCA</u> <u>GTAGGCATGATAAGACGCGCCAGCGTCCATCAGGCA</u> | <u>GTAGWTGCCAACCTTGGTTGGCA</u> | <u>GTAGGAGCYGGCTTGCCRCGAA</u> |
| <u>GTAGGCCCGGATAAGGCGTAGCCGCATCCGGC</u> <u>GTAGGCCCGGATAAGCAACGCCATCCGGC</u> | <u>GTRGATCCACGCATGCGTTGAT</u> | <u>GTAGGAGCCGGCTTGCTGGCGAT</u> |
| <u>GTAGGCCCGGCAAGCGCAGCGCCGCCGGGC</u> <u>GTAGGCCCGGATAAGGCGAAGCCGCATCCGGC</u> | <u>GTAGAGCCACCCATGGGTGGCT</u> | <u>GTGGYGAGGGGATTTATCCCCG</u> |
| | <u>GGTAGAGTCGACTGTTAGTCGACT</u> | <u>GTGGCAGGGGGCTTGCCCCCG</u> |
| | <u>GTAGMGCCGGYTCTRCCCGGCK</u> | <u>GTGGCAGGGGGCTTGCCCCCG</u> |
| | <u>GGTAGTGCCGGCCCTGGCCGGCA</u> | <u>GTGGGAGCGAGCTTGCTCGCGAT</u> |
| | <u>TGTAGAGCCGAGCCATGCTCGGCT</u> | <u>GTGGGAGCGGGCTTGCTCGCGAA</u> |
| | <u>GGTAGCGCCGGCCATGCCGGCG</u> | <u>GTGGGAGCGAGCTTGCTCGCGAA</u> |
| | <u>TGTAGAGTCGAGCCATGCTCGACT</u> | <u>GTAGGAGTGAGCTTGCTCGCGAT</u> |
| | <u>GTAGAGTCGAGCTTGCTCGACT</u> | <u>GTGGGAGCTGGCTTGCTCGCGAT</u> |
| | <u>GTAGAGCCGACCGCTTGGTCGGCT</u> | <u>GTGGGAGCGGGCTTGCCCCGAT</u> |
| | | <u>GTGGGAGCTGGCTTGCCAGCGAT</u> |
| | | <u>GTGGGAGCCAGCTTGCTGGCGAT</u> |
| | | <u>GTGGGAGCTAGCTGCTAGCGAT</u> |
| | | <u>GTGGCAGGGAGCTTGCTCCCGCT</u> |
| | | <u>GTAGGAGCGAGCYGCTCGCGA</u> |
| | | <u>GTRGGAGCAAGGCTTGCCCCGCGAT</u> |
| | | <u>GTAGGAGCTGCCGAAGGCTGCGAT</u> |
| | | <u>GTAGGAGCGAGGCTTGCCCCGCGA</u> |
| | | <u>GTRGTAGCGGGCTTGCCCCCGCT</u> |

Tabulka 5: Porovnání sekvencí REP elementů u různých fylogenetických skupin bakterií.

Palindromické části sekvencí jsou tučné a podtržené. Konzervovaný tetranukleotid je zvýrazněn šedě.

Zkratky variabilních bází jsou v souladu s nomenklaturou IUPAC.

Podle Nunvar et al. 2010; Nunvar et al. 2013

Cis-mobilizace REP elementů

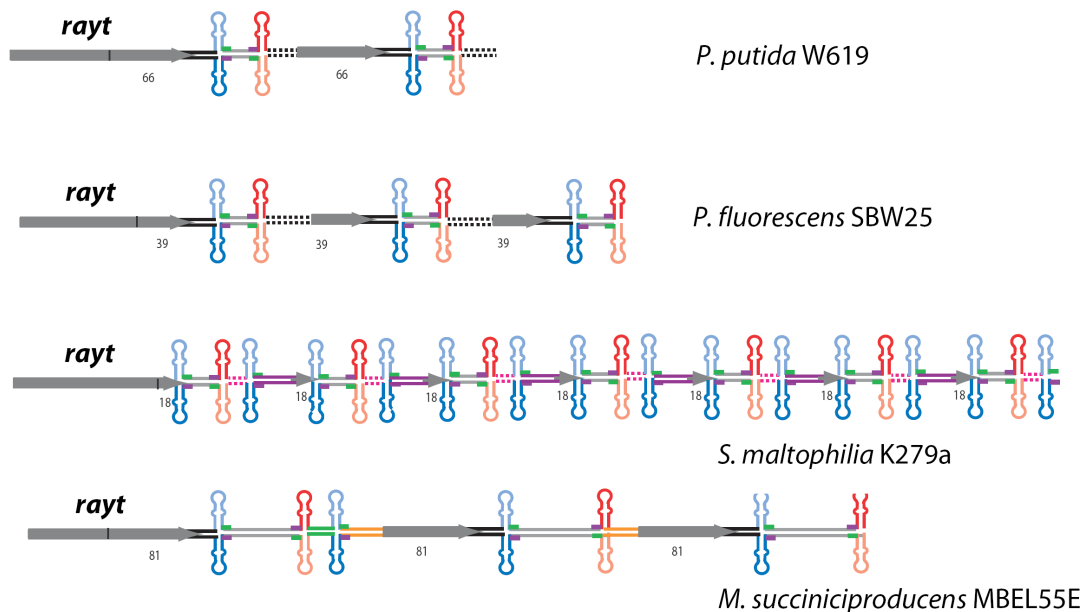
Několik pozorování, jež budou diskutována níže, naznačuje, že REP elementy, sousedící s *rayt* geny, mohou být preferenčně multiplikovány a diseminovány po hostitelském chromozomu. Tato tzv. *cis*-mobilizace je jev typický pro mnoho transponibilních elementů (Adams et al. 2006) a souvisí s prostorovým spřažením transkripce a translace v prokaryotické buňce, jež postrádá oddělené kompartmenty. Nově syntetizovaná transponáza má proto vysokou šanci, že se dostane do kontaktu s terminálními sekvencemi její vlastní inzerční sekvence. Vzhledem k asociaci mezi *rayt* geny a REP elementy (**Obr. 18**) tak může podobně docházet k jejich *cis*-mobilizaci. Stejně jako u transponáz inzerčních sekvencí rodiny IS200/IS605 je substrátem štěpící aktivity RAYT proteinu jednořetězcová DNA (Ton-Hoang et al. 2012). Zvláštností jednořetězcové transpozice/mobilizace je přenos "poloviny", tj. jednoho vlákna DNA daného elementu (komplementární řetězec není vyštěpován a zůstává v původní pozici). Tento proces tedy může být replikativní a vést k multiplikaci REP elementů.

Pro *cis*-mobilizaci svědčí výrazně kolokalizovaná distribuce REP elementů v genomech některých bakterií. Zde se vyskytuje značná část z celkového počtu REP sekvencí právě v oblasti chromozomu, kde se nachází *rayt* gen (Nunvar et al. 2010 a další nepublikované případy). U osekvenovaných příbuzných těchto bakterií se RAYT proteiny nevyskytují, z čehož lze soudit, že přítomnost těchto *rayt* genů je důsledkem horizontálního přenosu do hostitelských bakterií. Pokud by se po přenosu preferenčně multiplikovaly REP elementy v blízkosti *rayt* genu, vedlo by to k pozorované výrazné kolokalizaci.

V publikaci Nunvar et al. 2013 jsem analyzoval závislost abundance různých typů REP sekvencí na přítomnosti RAYT proteinů s nimi asociovaných. K tomu účelu jsem zvolil dvě fylogenetické skupiny gamaproteobakterií – stenotrofomonády a fluorescentní pseudomonády, jež vykazují nejvyšší diversitu REP elementů (často

se vyskytuje několik sekvenčních typů v genomu jednoho hostitelského kmene). Nejčtenější sekvenční třídy byly zpravidla asociovány s příslušným *rayt* genem, což pravděpodobně svědčí o jejich specifické *cis*-mobilizaci. Zajímavé bylo zjištění, že poměrně často docházelo u obou bakteriálních skupin k nahrazení REP elementů asociovaných s daným *rayt* genem jinými REP elementy odlišných sekvencí, neboť ortologní a syntenní (tj. s totožnou chromozomální lokací) *rayt* geny byly u různých kmenů asociovány s různými, byť blízkými příbuznými, REP sekvencemi. Pro tento jev jsem zavedl pojem orthoswitch (angl. **orthologous REP switch**). Ač o přesném mechanismu, jenž by orthoswitch způsoboval, lze jen spekulovat, je jisté, že k tomuto fenoménu docházelo opakovaně a že se jedná o způsob zvyšování variability REP sekvencí.

RAYT proteiny se rovněž mohou uplatňovat při vzniku a šíření BIME, agregátů REP elementů s enigmatickou strukturou (**Kap. 2.3.4.2.**). Opakující se jednotka v BIME vždy obsahuje dvě REP sekvence v navzájem inverzním uspořádání (**Obr. 15**), což připomíná jejich pozici na obou koncích *rayt* genu, kde se nadto často vyskytují přídatné inverzní REP elementy. Katalytické působení RAYT proteinů nabízí relativně představitelnou možnost vzniku BIME (jeden z možných scénářů je prezentován v Nunvar et al. 2010 , další v Ton-Hoang et al. 2012). Dále se v BIME, v rámci sekvencí vmezeřených mezi REP elementy, mohou vyskytovat 3'-terminální fragmenty *rayt* genů (**Obr. 20**). To dokazuje, že tytéž procesy, cílené na REP elementy v blízkosti *rayt* genů, mohou vést k formaci BIME. Vzhledem k tomu, že BIME elementy se vyskytují u mnoha bakterií, u nichž byly detekovány REP sekvence (Nunvar et al. 2010), mechanismus jejich vzniku je patrně konzervovaný.



Obrázek 20: Inkorporace 3'-konce *rayt* genů (šedé šipky, resp. jejich části) do BIME u různých bakteriálních druhů. REP elementy jsou znázorněny vlásenkami.

Podle Ton-Hoang et al. 2012 (tam upraveno z Nunvar et al. 2010)

Pokud přijmeme teorii *cis*-mobilizace REP elementů RAYT proteiny, můžeme relativně snadno vysvětlit dva typické rysy těchto repetitivních sekvencí: druhovou specificitu a vysokou četnost v hostitelských genomech. Za předpokladu, že "stávající" REP sekvence v průběhu evoluce degenerují v důsledku hromadění mutací, je nutné katalytické působení RAYT proteinů, vedoucí k diseminaci REP variant, jež sousedí s jejich příslušnými geny. Diverzifikace sekvencí těchto REP elementů (mechanismem typu orthoswitch), fixovaná při speciaci hostitelských bakterií, by pak vedla k jejich druhově-specifickému výskytu.

RAYT proteiny jako domestikované transponázy

U eukaryotních organismů, pro něž je typická vysoká abundance a diversita transponibilních elementů, bylo popsáno mnoho případů, kdy došlo k přizpůsobení transponáz jiným buněčným funkcím než mobilizace příslušných elementů. Tento jev se nazývá domestikace nebo neofunkcionalizace (Alzohairy et al. 2013), ev. exaptace (Feschotte a Pritham 2007) transponáz.

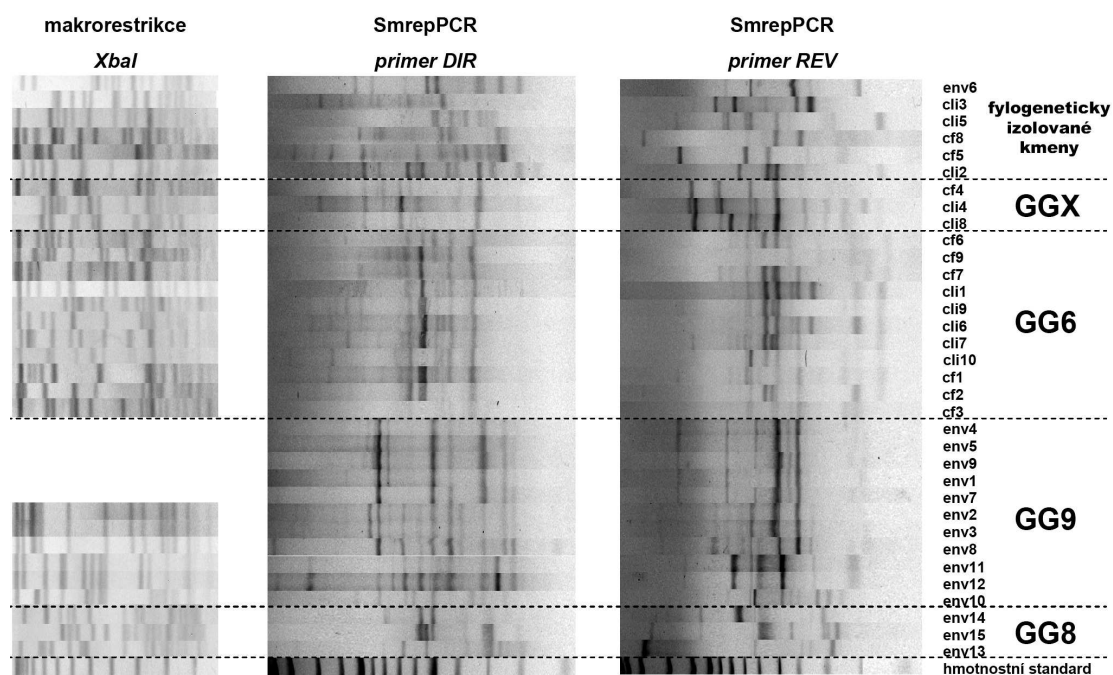
Na RAYT proteiny lze, díky sekvenčním homologiím (**Obr. 19**), nahlížet jako na vzdálené příbuzné Y1 transponáz inzerčních sekvencí rodiny IS200/IS605. RAYT proteiny však nevykazují chování typických transponáz, neboť se v hostitelských genomech vyskytují výhradně v jedné kopii (Nunvar et al. 2010). Evoluční předek RAYT proteinů pravděpodobně ztratil funkci transpozice, namísto toho se u něj vyvinula schopnost diseminovat REP elementy. Je zajímavé, že v evoluci fluorescentních pseudomonád a stenotrofomonád očividně došlo nezávisle na sobě k radiaci RAYT proteinů, spojené s rapidní diverzifikací REP elementů a jejich proliferací až na úroveň stovek nebo dokonce tisíců kopií v jednom hostitelském kmeni (Nunvar et al. 2013). Je zřejmé, že vyšší počty REP elementů jsou bakteriemi nejen tolerované, ale patrně i dochází k jejich pozitivní selekci. Tato radiace REP elementů a RAYT proteinů se mohla podílet na evoluční úspěšnosti bakteriálních hostitelů, vzhledem k popsaným funkcím REP elementů (jejich blízkého inverzního dimeru typu REPIN, což je predominantní forma u obou bakteriálních linií – **Kap. 2.3.4.2.**), nejčastěji v regulaci transkripce (**Kap. 2.3.4.3.**). RAYT proteiny tak pravděpodobně představují jeden z prvních a zároveň nejvýznamějších popsaných případů domestikace transponáz u prokaryot (Chandler et al. 2013).

Využití REP elementů v genotypizaci *S. maltophilia*

Již dvě dekády se v genotypizaci bakterií využívá PCR cílená na REP elementy *E. coli* (**Kap. 2.3.4.4.**). V rámci třetí práce (Nunvar et al. 2012) jsem vyvinul protokol pro genotypizaci izolátů druhu *S. maltophilia*, jež je založená na polymorfismu druhově specifických REP elementů této bakterie. Analogicky s původní REP-typizací (Versalovic et al. 1991) využívá tato nová metoda, pojmenovaná SmrepPCR, primery komplementární k REP sekvencím. Reakce vždy probíhaly s využitím jediného primeru, čímž docházelo k amplifikaci DNA mezi inverzně orientovanými REP elementy. Metoda byla testována na souboru 34 izolátů klinického a environmentálního původu.

Přestože se jedná o metodu nenáročnou na čas a materiál, poskytovala SmrepPCR hodnotné výsledky. Podle podobnosti profilů amplifikovaných fragmentů bylo možné přesně určit, do jaké evoluční linie v rámci druhu *S. maltophilia* daný izolát patří. V tomto ohledu bylo použití SmrepPCR výhodnější než makrorestrikční analýza, jež vykazovala přílišný rozptyl profilů restrikčních fragmentů (**Obr. 21**). Je možné, že tato překvapivě vysoká variabilita restrikčních profilů souvisí s velmi variabilním rozložením genomových ostrovů u *S. maltophilia* (Rocco et al. 2009). REP elementy se u této bakterie vyskytují jen v core genomu (**Kap. 2.3.4.1.**), což může vysvětlovat vyšší konzervovanost SmrepPCR profilů.

Dále bylo možné pomocí SmrepPCR spolehlivě určit klonální přízvěnost environmentálních kmenů, izolovaných z téže lokality. V jednom případě byla SmrepPCR dokonce citlivější než makrorestrikce a odlišila kmen env8 od klonální dvojice env2 a env3, zatímco jejich makrorestrikční profily byly stejné. Tento výsledek patrně svědčí o recentním pohybu REP elementů.



Obrázek 21: Porovnání fragmentů DNA získaných makrorestrikcí genomové DNA a amplifikací pomocí SmrepPCR (s využitím primerů DIR a REV, jež jsou vzájemně komplementární). Profily fragmentů byly rozdělány agarózovou elektroforézou a seřazeny podle podobnosti v programu Bionumerics. Skupiny izolátů podle příbuznosti nukleotidových sekvencí genu *gyrB* (popis vpravo) jsou odděleny přerušovanými čarami. DNA kmenů env1,env4,env5,env7 a env9 se v průběhu elektroforézy makrorestrikčních fragmentů nespecificky degradovala, jejich makrorestrikční profily proto nejsou v obrázku zahrnuty.

Podle Nunvar et al. 2012

Perpektivy studia RAYT proteinů a REP elementů

Dosud nebyla zcela experimentálně prokázána role RAYT proteinů v mobilizaci a multiplikaci REP elementů. V tomto směru probíhá experiment v laboratoři P. B. Raineyho, který porovnává genomové sekvence populací divokého kmene *P. fluorescens* a jeho delečních mutant v *rayt* genech po roční simulaci evoluce v laboratoři, s cílem detekovat inserce a delece REP elementů a dokázat podmíněnost těchto událostí přítomností RAYT proteinů (Bertels 2012). Další důležitou nevyřešenou otázkou je vznik unikátních struktur BIME, na tomto tématu se pracuje v laboratoři Micka Chandlera (Ton-Hoang et al. 2012). Prostor k zajímavým zjištěním nabízí též porovnání shodných a rozdílných rysů v mechanismech vazby a štěpení substrátové REP DNA mezi různými, evolučně odlišnými RAYT proteiny.

REFERENCE

- Adams, C. D., B. Schnurr, et al. (2006). "Tn5 transposase loops DNA in the absence of Tn5 transposon end sequences." *Mol Microbiol* **62**(6): 1558-68.
- Ahnert, S. E., T. M. A. Fink, et al. (2008). "How much non-coding DNA do eukaryotes require?" *Journal of Theoretical Biology* **252**(4): 587-592.
- Alzohairy, A. M., G. Gyulai, et al. (2013). "Transposable elements domesticated and neofunctionalized by eukaryotic genomes." *Plasmid* **69**(1): 1-15.
- Aranda-Olmedo, I., R. Tobes, et al. (2002). "Species-specific repetitive extragenic palindromic (REP) sequences in *Pseudomonas putida*." *Nucleic Acids Res* **30**(8): 1826-33.
- Bachellier, S., J. M. Clement, et al. (1999). "Short palindromic repetitive DNA elements in enterobacteria: a survey." *Res Microbiol* **150**(9-10): 627-39.
- Bachellier, S., J. M. Clement, et al. (1997). "Bacterial interspersed mosaic elements (BIMEs) are a major source of sequence polymorphism in *Escherichia coli* intergenic regions including specific associations with a new insertion sequence." *Genetics* **145**(3): 551-62.
- Bachellier, S., D. Perrin, et al. (1993). "Bacterial interspersed mosaic elements (BIMEs) are present in the genome of *Klebsiella*." *Mol Microbiol* **7**(4): 537-44.
- Barabas, O., D. R. Ronning, et al. (2008). "Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection." *Cell* **132**(2): 208-20.
- Barrangou, R., C. Fremaux, et al. (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." *Science* **315**(5819): 1709-12.
- Bertels, F. (2012). The evolution of selfish genetic elements within bacterial genomes: a thesis submitted in partial fulfilment of the requirements for the degree of Ph.D. in molecular evolution at Massey University, Auckland, New Zealand, Massey University, Auckland.
- Bertels, F. and P. B. Rainey (2011). "Within-genome evolution of REPINs: a new family of miniature mobile DNA in bacteria." *PLoS Genet* **7**(6): e1002132.
- Bikard, D., C. Loot, et al. (2010). "Folded DNA in action: hairpin formation and biological functions in prokaryotes." *Microbiol Mol Biol Rev* **74**(4): 570-88.
- Boccard, F. and P. Prentki (1993). "Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3' end of transcription units." *Embo J* **12**(13): 5019-27.
- Bolotin, A., B. Quinquis, et al. (2005). "Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin." *Microbiology* **151**(Pt 8): 2551-61.
- Buisine, N., C. M. Tang, et al. (2002). "Transposon-like Correia elements: structure, distribution and genetic exchange between pathogenic *Neisseria* sp." *FEBS Lett* **522**(1-3): 52-8.
- Cehovin, A., P. J. Simpson, et al. (2013). "Specific DNA recognition mediated by a type IV pilin." *Proc Natl Acad Sci U S A* **110**(8): 3065-70.
- Clark, G. W., V. U. Dar, et al. (2011). "Using coevolution to predict protein-protein interactions." *Methods Mol Biol* **781**: 237-56.
- Clement, J. M., C. Wilde, et al. (1999). "IS1397 is active for transposition into the chromosome of *Escherichia coli* K-12 and inserts specifically into palindromic units of bacterial interspersed mosaic elements." *J Bacteriol* **181**(22): 6929-36.
- Curcio, M. J. and K. M. Derbyshire (2003). "The outs and ins of transposition: from Mu to Kangaroo." *Nat Rev Mol Cell Biol* **4**(11): 865-77.

- De Gregorio, E., G. Silvestro, et al. (2005). "Enterobacterial repetitive intergenic consensus sequence repeats in yersiniae: genomic organization and functional properties." J Bacteriol **187**(23): 7945-54.
- Delihias, N. (2011). "Impact of small repeat sequences on bacterial genome evolution." Genome Biol Evol **3**: 959-73.
- Deveau, H., J. E. Garneau, et al. (2010). "CRISPR/Cas system and its role in phage-bacteria interactions." Annu Rev Microbiol **64**: 475-93.
- Dillingham, M. S. and S. C. Kowalczykowski (2008). "RecBCD enzyme and the repair of double-stranded DNA breaks." Microbiol Mol Biol Rev **72**(4): 642-71, Table of Contents.
- Drevinek, P., A. Baldwin, et al. (2010). "Oxidative stress of *Burkholderia cenocepacia* induces insertion sequence-mediated genomic rearrangements that interfere with macrorestriction-based genotyping." J Clin Microbiol **48**(1): 34-40.
- Eichenbaum, Z. and Z. Livneh (1998). "UV light induces IS10 transposition in *Escherichia coli*." Genetics **149**(3): 1173-81.
- Elhai, J., M. Kato, et al. (2008). "Very small mobile repeated elements in cyanobacterial genomes." Genome Res **18**(9): 1484-99.
- Espeli, O. and F. Boccard (1997). "In vivo cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterization of the target and identification of the cut site." Mol Microbiol **26**(4): 767-77.
- Espeli, O., L. Moulin, et al. (2001). "Transcription attenuation associated with bacterial repetitive extragenic BIME elements." J Mol Biol **314**(3): 375-86.
- Fattash, I., R. Rooke, et al. (2013). "Miniature inverted-repeat transposable elements: discovery, distribution, and activity." Genome: 1-12.
- Feil, E. J. and B. G. Spratt (2001). "Recombination and the population structures of bacterial pathogens." Annu Rev Microbiol **55**: 561-90.
- Feschotte, C. and E. J. Pritham (2007). "DNA transposons and the evolution of eukaryotic genomes." Annu Rev Genet **41**: 331-68.
- Filee, J., P. Siguier, et al. (2007). "Insertion sequence diversity in archaea." Microbiol Mol Biol Rev **71**(1): 121-57.
- Freundlich, M., N. Ramani, et al. (1992). "The role of integration host factor in gene expression in *Escherichia coli*." Mol Microbiol **6**(18): 2557-63.
- Frye, S. A., M. Nilsen, et al. (2013). "Dialects of the DNA uptake sequence in *Neisseriaceae*." PLoS Genet **9**(4): e1003458.
- Gil, R. and A. Latorre (2012). "Factors behind junk DNA in bacteria." Genes **3**(4): 634-650.
- Gilson, E., S. Bachellier, et al. (1990). "Palindromic unit highly repetitive DNA sequences exhibit species specificity within *Enterobacteriaceae*." Res Microbiol **141**(9): 1103-16.
- Gilson, E., D. Perrin, et al. (1990). "DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organization." Nucleic Acids Res **18**(13): 3941-52.
- Gilson, E., D. Perrin, et al. (1987). "Species specificity of bacterial palindromic units." J Mol Evol **25**(4): 371-3.
- Gilson, E., W. Saurin, et al. (1991). "Palindromic units are part of a new bacterial interspersed mosaic element (BIME)." Nucleic Acids Res **19**(7): 1375-83.
- Gonzalez, J. and D. Petrov (2009). "MITEs - the ultimate parasites." Science **325**(5946): 1352-3.
- Guynet, C., A. Achard, et al. (2009). "Resetting the site: redirecting integration of an insertion sequence in a predictable way." Mol Cell **34**(5): 612-9.
- Guynet, C., A. B. Hickman, et al. (2008). "In vitro reconstitution of a single-stranded transposition mechanism of IS608." Mol Cell **29**(3): 302-12.
- Halpern, D., H. Chiapello, et al. (2007). "Identification of DNA motifs implicated in maintenance of bacterial core genomes by predictive modeling." PLoS Genet **3**(9): 1614-21.
- Hashimoto, M., T. Ichimura, et al. (2005). "Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome." Mol Microbiol **55**(1): 137-49.

- He, S., C. Guynet, et al. (2013). "IS200/IS605 family single-strand transposition: mechanism of IS608 strand transfer." Nucleic Acids Res **41**(5): 3302-13.
- He, S., A. B. Hickman, et al. (2011). "Reconstitution of a functional IS608 single-strand transpososome: role of non-canonical base pairing." Nucleic Acids Res **39**(19): 8503-12.
- Herman, L. and M. Heyndrickx (2000). "The presence of intragenically located REP-like elements in *Bacillus sporothermodurans* is sufficient for REP-PCR typing." Res Microbiol **151**(4): 255-61.
- Hickman, A. B., J. A. James, et al. (2010). "DNA recognition and the precleavage state during single-stranded DNA transposition in *D. radiodurans*." Embo J **29**(22): 3840-52.
- Higgins, C. F., G. F. Ames, et al. (1982). "A novel intergenic regulatory element of prokaryotic operons." Nature **298**(5876): 760-2.
- Horvath, P. and R. Barrangou (2010). "CRISPR/Cas, the immune system of bacteria and archaea." Science **327**(5962): 167-70.
- Horvath, P., D. A. Romero, et al. (2008). "Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*." J Bacteriol **190**(4): 1401-12.
- Chandler, M., F. de la Cruz, et al. (2013). "Breaking and joining single-stranded DNA: the HUH endonuclease superfamily." Nat Rev Microbiol **11**(8): 525-38.
- Choi, S., S. Ohta, et al. (2003). "A novel IS element, IS621, of the IS110/IS492 family transposes to a specific site in repetitive extragenic palindromic sequences in *Escherichia coli*." J Bacteriol **185**(16): 4891-900.
- Ichikawa, H., K. Ikeda, et al. (1987). "Specific binding of transposase to terminal inverted repeats of transposable element Tn3." Proc Natl Acad Sci U S A **84**(23): 8220-4.
- Ilves, H., R. Horak, et al. (2001). "Involvement of sigma(S) in starvation-induced transposition of *Pseudomonas putida* transposon Tn4652." J Bacteriol **183**(18): 5445-8.
- Jansen, R., J. D. Embden, et al. (2002). "Identification of genes that are associated with DNA repeats in prokaryotes." Mol Microbiol **43**(6): 1565-75.
- Kersulyte, D., B. Velapatino, et al. (2002). "Transposable element ISHp608 of *Helicobacter pylori*: nonrandom geographic distribution, functional organization, and insertion specificity." J Bacteriol **184**(4): 992-1002.
- Klappenbach, J. A., J. M. Dunbar, et al. (2000). "rRNA operon copy number reflects ecological strategies of bacteria." Appl Environ Microbiol **66**(4): 1328-33.
- Knutsen, E., O. Johnsborg, et al. (2006). "BOX elements modulate gene expression in *Streptococcus pneumoniae*: impact on the fine-tuning of competence development." J Bacteriol **188**(23): 8307-12.
- Kobayashi, T. (2011). "Regulation of ribosomal RNA gene copy number and its role in modulating genome integrity and evolutionary adaptability in yeast." Cell Mol Life Sci **68**(8): 1395-403.
- Kolisnychenko, V., G. Plunkett, 3rd, et al. (2002). "Engineering a reduced *Escherichia coli* genome." Genome Res **12**(4): 640-7.
- Koonin, E. V. and Y. I. Wolf (2008). "Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world." Nucleic Acids Res **36**(21): 6688-719.
- Lee, H. H., J. Y. Yoon, et al. (2006). "Crystal structure of a metal ion-bound IS200 transposase." J Biol Chem **281**(7): 4261-6.
- Lee, Z. M., C. Bussema, 3rd, et al. (2009). "rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea." Nucleic Acids Res **37**(Database issue): D489-93.
- Levinson, G. and G. A. Gutman (1987). "Slipped-strand mispairing: a major mechanism for DNA sequence evolution." Mol Biol Evol **4**(3): 203-21.
- Li, G. M. (2008). "Mechanisms and functions of DNA mismatch repair." Cell Res **18**(1): 85-98.
- Lin, S., S. Haas, et al. (2011). "Genome-wide comparison of cyanobacterial transposable elements, potential genetic diversity indicators." Gene **473**(2): 139-49.

- Liu, S. V., N. J. Saunders, et al. (2002). "Genome analysis and strain comparison of *correia* repeats and *correia* repeat-enclosed elements in pathogenic *Neisseria*." J Bacteriol **184**(22): 6163-73.
- Lobner-Olesen, A., O. Skovgaard, et al. (2005). "Dam methylation: coordinating cellular processes." Curr Opin Microbiol **8**(2): 154-60.
- Loper, J. E., K. A. Hassan, et al. (2012). "Comparative genomics of plant-associated *Pseudomonas* spp.: insights into diversity and inheritance of traits involved in multitrophic interactions." PLoS Genet **8**(7): e1002784.
- Lopez-Flores, I. and M. A. Garrido-Ramos (2012). "The repetitive DNA content of eukaryotic genomes." Genome Dyn **7**: 1-28.
- Mahillon, J. and M. Chandler (1998). "Insertion sequences." Microbiol Mol Biol Rev **62**(3): 725-74.
- Mahillon, J., C. Leonard, et al. (1999). "IS elements as constituents of bacterial genomes." Res Microbiol **150**(9-10): 675-87.
- Makarova, K. S., D. H. Haft, et al. (2011). "Evolution and classification of the CRISPR-Cas systems." Nat Rev Microbiol **9**(6): 467-77.
- Mattick, J. S. (2004). "RNA regulation: a new genetics?" Nat Rev Genet **5**(4): 316-23.
- Matus-Garcia, M., H. Nijveen, et al. (2012). "Promoter propagation in prokaryotes." Nucleic Acids Res **40**(20): 10032-40.
- McClelland, M. and J. Welsh (1994). "DNA fingerprinting by arbitrarily primed PCR." PCR Methods Appl **4**(1): S59-65.
- Menecier, S., P. Servant, et al. (2006). "Mutagenesis via IS transposition in *Deinococcus radiodurans*." Mol Microbiol **59**(1): 317-25.
- Messing, S. A., B. Ton-Hoang, et al. (2012). "The processing of repetitive extragenic palindromes: the structure of a repetitive extragenic palindrome bound to its associated nuclease." Nucleic Acids Res **40**(19): 9964-79.
- Molina, F. and K. Skarstad (2004). "Replication fork and SeqA focus distributions in *Escherichia coli* suggest a replication hyperstructure dependent on nucleotide metabolism." Mol Microbiol **52**(6): 1597-612.
- Montano, S. P. and P. A. Rice (2011). "Moving DNA around: DNA transposition and retroviral integration." Curr Opin Struct Biol **21**(3): 370-8.
- Moran, N. A. and G. R. Plague (2004). "Genomic changes following host restriction in bacteria." Curr Opin Genet Dev **14**(6): 627-33.
- Nagy, Z. and M. Chandler (2004). "Regulation of transposition in bacteria." Res Microbiol **155**(5): 387-98.
- Nelson, W. C., D. Bhaya, et al. (2012). "Novel miniature transposable elements in thermophilic *Synechococcus* strains and their impact on an environmental population." J Bacteriol **194**(14): 3636-42.
- Nesmelova, I. V. and P. B. Hackett (2010). "DDE transposases: Structural similarity and diversity." Adv Drug Deliv Rev **62**(12): 1187-95.
- Nunvar, J., P. Drevinek, et al. (2012). "DNA profiling of *Stenotrophomonas maltophilia* by PCR targeted to its species-specific repetitive palindromic sequences." Lett Appl Microbiol **54**(1): 59-66.
- Nunvar, J., T. Huckova, et al. (2010). "Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: implications for REP evolution and dynamics in bacterial genomes." BMC Genomics **11**: 44.
- Nunvar, J., I. Licha, et al. (2013). "Evolution of REP diversity: a comparative study." BMC Genomics **14**: 385.
- Oggioni, M. R. and J. P. Claverys (1999). "Repeated extragenic sequences in prokaryotic genomes: a proposal for the origin and dynamics of the RUP element in *Streptococcus pneumoniae*." Microbiology **145 (Pt 10)**: 2647-53.
- Ohtsubo, Y., H. Genka, et al. (2005). "High-temperature-induced transposition of insertion elements in *Burkholderia multivorans* ATCC 17616." Appl Environ Microbiol **71**(4): 1822-8.

- Oppenheim, A. B., K. E. Rudd, et al. (1993). "Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli*." Mol Microbiol **10**(1): 113-22.
- Pasternak, C., R. Dulermo, et al. (2013). "ISDra2 transposition in *Deinococcus radiodurans* is downregulated by TnpB." Mol Microbiol **88**(2): 443-55.
- Pasternak, C., B. Ton-Hoang, et al. (2010). "Irradiation-induced *Deinococcus radiodurans* genome fragmentation triggers transposition of a single resident insertion sequence." PLoS Genet **6**(1): e1000799.
- Podar, M., I. Anderson, et al. (2008). "A genomic analysis of the archaeal system *Ignicoccus hospitalis*-*Nanoarchaeum equitans*." Genome Biol **9**(11): R158.
- Pourcel, C., G. Salvignol, et al. (2005). "CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies." Microbiology **151**(Pt 3): 653-63.
- Ramos-Gonzalez, M. I., M. J. Campos, et al. (2006). "Characterization of the *Pseudomonas putida* mobile genetic element ISPPu10: an occupant of repetitive extragenic palindromic sequences." J Bacteriol **188**(1): 37-44.
- Redfield, R. J., W. A. Findlay, et al. (2006). "Evolution of competence and DNA uptake specificity in the *Pasteurellaceae*." BMC Evol Biol **6**: 82.
- Rocco, F., E. De Gregorio, et al. (2009). "*Stenotrophomonas maltophilia* genomes: a start-up comparison." Int J Med Microbiol **299**(8): 535-46.
- Rocco, F., E. De Gregorio, et al. (2010). "A giant family of short palindromic sequences in *Stenotrophomonas maltophilia*." FEMS Microbiol Lett **308**(2): 185-92.
- Ronning, D. R., C. Guynet, et al. (2005). "Active site sharing and subterminal hairpin recognition in a new class of DNA transposases." Mol Cell **20**(1): 143-54.
- Roschetto, E., F. Rocco, et al. (2008). "PCR-based rapid genotyping of *Stenotrophomonas maltophilia* isolates." BMC Microbiol **8**: 202.
- Savic, D. J., M. Jankovic, et al. (1990). "Cellular role of DNA polymerase I." J Basic Microbiol **30**(10): 769-84.
- Segerman, B. (2012). "The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories." Front Cell Infect Microbiol **2**: 116.
- Schneider, D., E. Duperchy, et al. (2000). "Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements." Genetics **156**(2): 477-88.
- Schneider, D. and R. E. Lenski (2004). "Dynamics of insertion sequence elements during experimental evolution of bacteria." Res Microbiol **155**(5): 319-27.
- Schulte, R. D., C. Makus, et al. (2010). "Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite." Proc Natl Acad Sci U S A **107**(16): 7359-64.
- Silby, M. W., A. M. Cerdeno-Tarraga, et al. (2009). "Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*." Genome Biol **10**(5): R51.
- Smith, G. R. (2012). "How RecBCD enzyme and Chi promote DNA break repair and recombination: a molecular biologist's view." Microbiol Mol Biol Rev **76**(2): 217-28.
- Smith, H. O., J. F. Tomb, et al. (1995). "Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome." Science **269**(5223): 538-40.
- Sorek, R., C. M. Lawrence, et al. (2013). "CRISPR-mediated adaptive immune systems in bacteria and archaea." Annu Rev Biochem **82**: 237-66.
- Tobes, R. and E. Pareja (2005). "Repetitive extragenic palindromic sequences in the *Pseudomonas syringae* pv. *tomato* DC3000 genome: extragenic signals for genome reannotation." Res Microbiol **156**(3): 424-33.
- Tobes, R. and E. Pareja (2006). "Bacterial repetitive extragenic palindromic sequences are DNA targets for Insertion Sequence elements." BMC Genomics **7**: 62.
- Tobes, R. and J. L. Ramos (2005). "REP code: defining bacterial identity in extragenic space." Environ Microbiol **7**(2): 225-8.

- Ton-Hoang, B., C. Guynet, et al. (2005). "Transposition of ISHp608, member of an unusual family of bacterial insertion sequences." Embo J **24**(18): 3325-38.
- Ton-Hoang, B., C. Pasternak, et al. (2010). "Single-stranded DNA transposition is coupled to host replication." Cell **142**(3): 398-408.
- Ton-Hoang, B., P. Siguier, et al. (2012). "Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences." Nucleic Acids Res **40**(8): 3596-609.
- Treangen, T. J., O. H. Ambur, et al. (2008). "The impact of the neisserial DNA uptake sequences on genome evolution and stability." Genome Biol **9**(3): R60.
- Twiss, E., A. M. Coros, et al. (2005). "Transposition is modulated by a diverse set of host factors in *Escherichia coli* and is stimulated by nutritional stress." Mol Microbiol **57**(6): 1593-607.
- Versalovic, J., T. Koeth, et al. (1991). "Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes." Nucleic Acids Res **19**(24): 6823-31.
- Vodovar, N., D. Vallenet, et al. (2006). "Complete genome sequence of the entomopathogenic and metabolically versatile soil bacterium *Pseudomonas entomophila*." Nat Biotechnol **24**(6): 673-9.
- Weinel, C., D. W. Ussery, et al. (2002). "Comparative genomics of *Pseudomonas aeruginosa* PAO1 and *Pseudomonas putida* KT2440: orthologs, codon usage, repetitive extragenic palindromic elements, and oligonucleotide motif signatures." Genome Lett **1**(4): 175-187.
- Welsh, K. M., A. L. Lu, et al. (1987). "Isolation and characterization of the *Escherichia coli* *mutH* gene product." J Biol Chem **262**(32): 15624-9.
- Wessler, S. R. (2006). "Transposable elements and the evolution of eukaryotic genomes." Proc Natl Acad Sci U S A **103**(47): 17600-1.
- Wilde, C., S. Bachellier, et al. (2001). "Transposition of IS1397 in the family *Enterobacteriaceae* and first characterization of ISKpn1, a new insertion sequence associated with *Klebsiella pneumoniae* palindromic units." J Bacteriol **183**(15): 4395-404.
- Wilde, C., F. Escartin, et al. (2003). "Transposases are responsible for the target specificity of IS1397 and ISKpn1 for two different types of palindromic units (PUs)." Nucleic Acids Res **31**(15): 4345-53.
- Yang, Y. and G. F. Ames (1988). "DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences." Proc Natl Acad Sci U S A **85**(23): 8850-4.