

**Univerzita Karlova**

**Přírodovědecká fakulta**

Studijní program: Speciální chemicko-biologické obory

Studijní obor: Molekulární biologie a biochemie organismů



**Tadeáš Staněk**

Bakteriální REP elementy

Bacterial REP elements

Bakalářská práce

Školitel: RNDr. Jaroslav Nunvář, Ph.D.

Praha, 2021

**Prohlášení:**

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

V Praze, 5. 8. 2021

Podpis

### **Poděkování:**

Zde bych chtěl poděkovat mému školiteli RNDr. Jaroslavu Nunváři, Ph.D. za trpělivost a ochotu při vedení mé práce. Za rychlé a podrobné odpovědi na email a možnost osobní konzultace.

## Abstrakt

Inserční sekvence jsou nejmenším typem mobilních genetických elementů. Jsou tvořeny specifickými terminálními sekvencemi, mezi kterými je umístěn gen pro transpozázu, enzym, s jehož pomocí jsou schopny se propagovat genomem.

REP elementy jsou krátké repetitivní sekvence hojně zastoupené v řadě bakterií. Vyskytují se v nekódujících oblastech genomu. Mají strukturu palindromu, díky tomu ve formě ssRNA či ssDNA vytváří vlásenku. V hostitelském genomu lze nalézt stovky těchto elementů. Ty jsou často ve shlucích vytvářejících vyšší strukturní jednotky: REPIN, mající strukturu REP-mezera-REP, nebo BIME se strukturou iREP-mezera-iREP. Tyto struktury jsou asociované s RAYT, tyrozinovou nukleázou příbuznou rodině IS200/IS605 transpozáz. Jedná se o domestikovaný enzym propagující REP elementy skrze genom. Sama přítomnost REP elementů má vliv na jak na translaci, tak transkripci přilehlých genů. Krom toho REP interagují s proteiny důležitými pro fyziologii buňky.

Práce se zabývá rozšířením REP elementů, jejich vlastnostmi, funkcí v buňce a jejich asociací s RAYT nukleázou. Dále práce shrnuje poznatky o RAYT a jejich podobnostech a rozdílech v porovnání s ostatními rodinami transpozáz.

Klíčová slova:

REP elementy, BIME, REPIN, RAYT, IS200/IS605, Y1 transpozázy, mobilní genetické elementy

## Abstract

Insertion sequences are the smallest type of mobile genetic element. They are formed from specific repetitive elements, between which gene for transposition is placed. That is an enzyme used for their genome propagation.

REP elements are short repetitive sequences abundantly represented in many bacterial classes. They are placed in noncoding regions of genome. They have a palindromic structure, thanks to that their sides are complementary. ssRNA and ssDNA forms of REP can form a stem-loop structure. In the host genome, there are hundreds of these elements. REPs are usually grouped in higher structural units: REPIN, which has structure REP-spacer-REP, or BIME with iREP-spacer-iREP structure. These structures are associated with RAYT, tyrosine nuclease related to IS200/IS605 transposase family. The RAYT is a domesticated enzyme propagating REP elements through the genome. The presence of REP elements has an influence on transcription and translation of neighbouring genes. In addition, REPs interact with proteins important for cell physiology.

This thesis is focused on the dissemination of REP elements, their characteristics, function in cells, and cooperation with RAYT nucleases. This thesis also summarizes the findings of RAYT and their similarities and differences from other families of transposases.

Key words:

REP elements, BIME, REPIN, RAYT, IS200/IS605, Y1 transposases, mobile genetic elements

# Obsah

1. Úvod .....	1
2. Inzerční sekvence a transpozázy .....	2
2.1. DDE transpozázy .....	2
2.2. Tyrozinové transpozázy s HUH motivem.....	3
3. REP elementy .....	5
3.1. Struktura REP sekvence .....	6
3.2. Porovnání REP mezi druhy .....	7
3.2.1. <i>Pseudomonas fluorescens</i> .....	7
3.2.2. <i>Pseudomonas putida</i> .....	8
3.2.3. <i>Stenotrophomonas maltophilia</i> .....	8
3.2.4. <i>Escherichia coli</i> .....	9
3.3. Funkce REP elementů.....	9
3.4. Lokalizace REP elementů.....	10
4. BIME .....	12
4.1. Vznik BIME variability .....	13
4.2. Struktura BIME a jednotlivé motivy.....	14
4.3. BIME rodiny.....	15
4.3.1. BIME-1 .....	15
4.3.2. BIME-2 .....	16
5. RAYT a REPIN .....	17
5.1. Porovnání RAYT a IS200/IS605 transpozáz.....	19
5.1.1. Mechanismus štěpení ssDNA.....	19
5.1.2. Funkčně-specifické motivy RAYT transpozáz .....	20
5.1.3. Genomová distribuce genů transpozáz a s nimi asociovaných repetitivních sekvencí.....	21
5.2. Evoluce RAYT z IS200/IS605 transpozáz .....	23
5.3. Populace REPIN elementů.....	24
5.4. Rozdělení RAYT proteinů.....	26
5.4.1. Podtřída 2.1 (G1).....	27
5.4.2. Podtřída 2.2 (G2).....	28
5.4.3. Podtřída 2.5 (G3).....	28
5.4.4. Podtřída G6.....	28
6. Závěr.....	29
7. Seznam použité literatury .....	30

## Přílohy

1. Srovnání sekvencí REP elementů u vybraných druhů.....	1
2. Srovnání RAYT jednotlivých podtříd.....	2

## Seznam použitých zkratek

<b>AK</b>	Aminokyselina
<b>BIME</b>	Roztroušené bakteriální mozaikové elementy, anglicky: „Bacterial interspersed mosaic element“
<b>bp</b>	Párů bází, anglicky: „base pair“
<b>CRISPR</b>	Pravidelně rozmístěné segmenty krátkých repetitivních, anglicky: „Clustered regularly interspaced short palindromic repeats“
<b>DDE</b>	Katalytická triáda aminokyselin, kyselina asparagová (D) kyselina asparagová (D) kyselina glutamová (E)
<b>dsDNA</b>	Dvouřetězcová DNA, anglicky: „Double strand DNA“
<b>HUH</b>	Aminokyselinový motiv, histidin (H) hydrofobní aminokyselina (U) histidiny (H)
<b>IHF</b>	Integrační faktor hostitele, anglicky: „Integration host factor“
<b>iREP</b>	Inverzní REP element
<b>IS</b>	Inzerční sekvence, anglicky: „Insertion sequence“
<b>nt</b>	Nukleotid/y
<b>Orthoswitch</b>	Záměna REP sekvence za novou, anglicky: „orthologous REP switch“
<b>RAYT (či TnpA<sub>rep</sub>)</b>	Tyrozínová transpozáza asociovaná s REP, anglicky: „REP associated tyrosine transposase“
<b>REP</b>	Repetitivní mimogenové palindromické sekvence, anglicky: „Repetitive extragenic palindrome“
<b>REPIN</b>	Pár REP elementů schopných vytvářet vlásenkovou strukturu, anglicky: „REP doublet forming hairpin“
<b>REPtron</b>	<i>rayt</i> gen a přilehlé REP elementy, které jeho sekvenci obklopují
<b>RIB</b>	Opakující se IHF BIME, anglicky: „reiterative IHF BIME“
<b>RIP</b>	Repetitivní IHF vazebný palindrom, anglicky: „repetitive IHF binding palindromic element“
<b>SMAG</b>	<i>Stenotrophomonas maltophilia</i> REP elementy s GTAG tetranukleotidem
<b>ssDNA</b>	Jednořetězcová DNA, anglicky: „Single strand DNA“
<b>ssRNA</b>	Jednořetězcová RNA, anglicky: „Single strand RNA“
<b>TE</b>	Transponující elementy, anglicky: „Transposable elements“
<b>TnpA</b>	Čtecí rámec transpozázy A, anglicky: „Transposase A“
<b>TnpB</b>	Čtecí rámec transpozázy B, anglicky: „Transposase B“
<b>TIRYT</b>	Tyrozínová transpozáza terminálních inverzních opakování, anglicky: „Terminal inverted repeat tyrosine transposase“

# 1. Úvod

Bakterie jsou nejstarší a také největší skupinou organismů, které najdeme takřka v jakémkoliv prostředí. Náš svět je světem bakterií, počtem druhů zdaleka převyšují všechny ostatní skupiny. Obývají všechna prostředí vhodná pro život, včetně těch nejextrémnějších. Musí neustále bojovat s limitací prostředí a konkurencí ostatních organismů. S tím souvisí vysoký selekční tlak, kdy o každou volnou niku probíhá intenzivní boj. Bakterie jsou tak nuceny být vysoce efektivní. To je jedním z důvodů, proč kódující denzita bakteriálního genomu (tj. frakce genomu, která kóduje proteiny) dosahuje typicky hodnot okolo 90% (Silby *et al.*, 2009; McCutcheon and Moran, 2012). DNA nekódující přímo proteiny má především funkci regulační, je tvořena promotory, represory, terminátory transkripce a dále také geny pro rRNA a tRNA (Rogozin *et al.*, 2002).

Nekódující DNA dále zahrnuje řadu sobeckých genetických elementů, včetně transpozibilních elementů (TE). Základním mechanismem, který pro svou životní strategii využívají, je transpozice, tj. schopnost přemísťovat se skrze genom. Sekvence sobecké DNA se vyštěpí, přenesa a vloží na nové místo. To je zprostředkováno pomocí enzymu transpozázy, kterou si element kóduje. TE jsou považovány za součást nekódující DNA, přestože kódují proteiny, ty jsou však pro fyziologii buňky redundantní a slouží jen potřebám TE. Nejjednodušší TE (inzerční sekvence, IS elementy) jsou tvořeny pouze genem pro transpozázu a specifickou sekvencí okrajů, kterou je transpozáza schopna rozpoznat, navázat se na ni a provést transpozici. Díky výskytu ve více identických kopiích jsou cílem rekombinačních procesů, které pak mohou vést k amplifikaci, deleci či přeskupení genetického materiálu (Doolittle and Sapienza, 1980). Složitější mobilní elementy krom genu pro transpozici mohou nést geny výhodné pro hostitele. Tím si mohou zajistit pozitivní selekci v rámci hostitelského druhu. Vztah mezi mobilním genetickým elementem a hostitelem je velice dynamický, plyně přechází od parazitismu po mutualismus. Z pohledu čistě praktického, představují IS a příbuzné elementy nástroj pro genetické inženýrství. Mechanismus transpozice je možné upravit a využít pro nesespecifické či specifické vkládání či delece genů. Každý bakteriální kmen má jinak rozmístěny mobilní genetické elementy, jejich analýza je dobrým nástrojem pro evoluční a diagnostické účely (De Bruijn, 1992; Woodford, Johnson and Ridley, 2003; Treangen *et al.*, 2009).

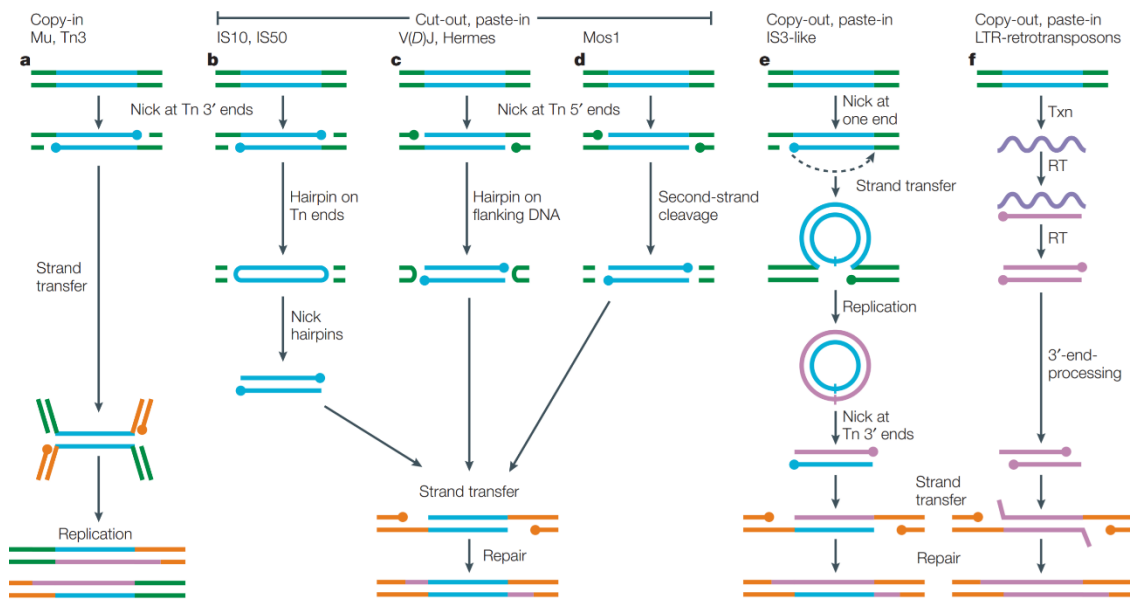
Práce si klade za cíl uvést do kontextu bakteriálních TE elementů REP sekvence a shrnout současné znalosti o těchto repetitivních sekvencích: popsat jejich strukturu, výskyt v bakteriích a funkci. Práce je dále zaměřena na vyšší struktury, které tyto elementy tvoří (BIME a REPIN). V poslední kapitole bude řešena spojitost REPIN se skupinou RAYT nukleáz, které způsobují jejich diseminaci, příbuznost RAYT s rodinou IS200/IS605 transpozáz a jejich podobnosti a rozdíly.

## 2. Inzerční sekvence a transpozázy

Prokaryotní inzerční sekvence (IS) jsou krátké sekvence DNA schopné samostatného pohybu po genomu, případně i (zprostředkovaného) pohybu mezi genomy. Pro mobilizaci využívají vlastní specializované rekombinační systémy. IS jsou nejjednodušší typ transpozičních elementů, kódující pouze vlastní transpozázu ohraničenou specifickými konci. Transpozáza je schopna rozeznat konce IS, navázat se na ně, element vyštěpit a přesunout jinam v genomu. Je známo více mechanismů, které IS pro svůj přenos využívají. Jelikož IS kódují pouze transpozázu, nemají přímý vliv na fenotyp (kódovaný produkt nemá v buňce roli, krom vlastní transpozice). Nicméně mohou fenotyp ovlivnit nepřímo svou polohou. Vložení do genu způsobí vznik nefunkčního produktu. DNA bude navíc ovlivněna i po vystřížení IS, neboť po ní typicky zůstávají zbytky DNA (krátké přímé repetice). V současnosti jsou známy tisíce typů inzerčních sekvencí, dají se rozdělit do skupin podle řady kritérií (mechanismu transpozice, délky sekvence, typu kódované transpozázy a dalších). V rámci práce bude nastíněno rozdělení IS podle transpozázy, kterou využívají.

### 2.1. DDE transpozázy

DDE transpozázy jsou využívány největším počtem inzerčních sekvencí (Mahillon, Léonard and Chandler, 1999). Název je odvozen od konzervované triády aminokyselin, kterými jsou dvě molekuly kyseliny asparagové (D) a kyselina glutamová (E). Vzdálenost těchto AK se mezi DDE transpozázami liší. Tyto aminokyseliny koordinují dva dvojmocné ionty kovů, které se podílejí na transesterifikační reakci (orientují během ní molekuly vody). Dochází k štěpení jednoho vlákna DNA, čímž se generuje 3' OH konec. Ten bude následně použit jako nukleofil v dalších krocích, které mezi jednotlivými rodinami liší mezikroky (Hickman, Chandler and Dyda, 2010), jak je znázorněno na obrázku 1. Při transpozici se přenáší DNA ve formě dvouřetězce. Pro inzerční sekvence využívající tyto transpozázy jsou typické inverzní repetice, které sekvenci ohraničují. Jejich účelem je pomoci transpozáze rozeznat inzerční sekvenci (Ichikawa *et al.*, 1987). Transpozáza se naváže na inverzní repetice na obou koncích inzerční sekvence a vyštěpí ji. Z toho důvodu DDE transpozázy pracují jako homodimery, aby byly schopny vázat oba konce sekvence zároveň.



Obrázek 1: Mechanismy transpozice několika rodin DDE transpozáz. Znáznorněna je řada meziproductů, které mohou mít například kruhovou či vlásenkovou formu. Převzato z (Curcio and Derbyshire, 2003).

## 2.2. Tyrozinové transpozázy s HUH motivem

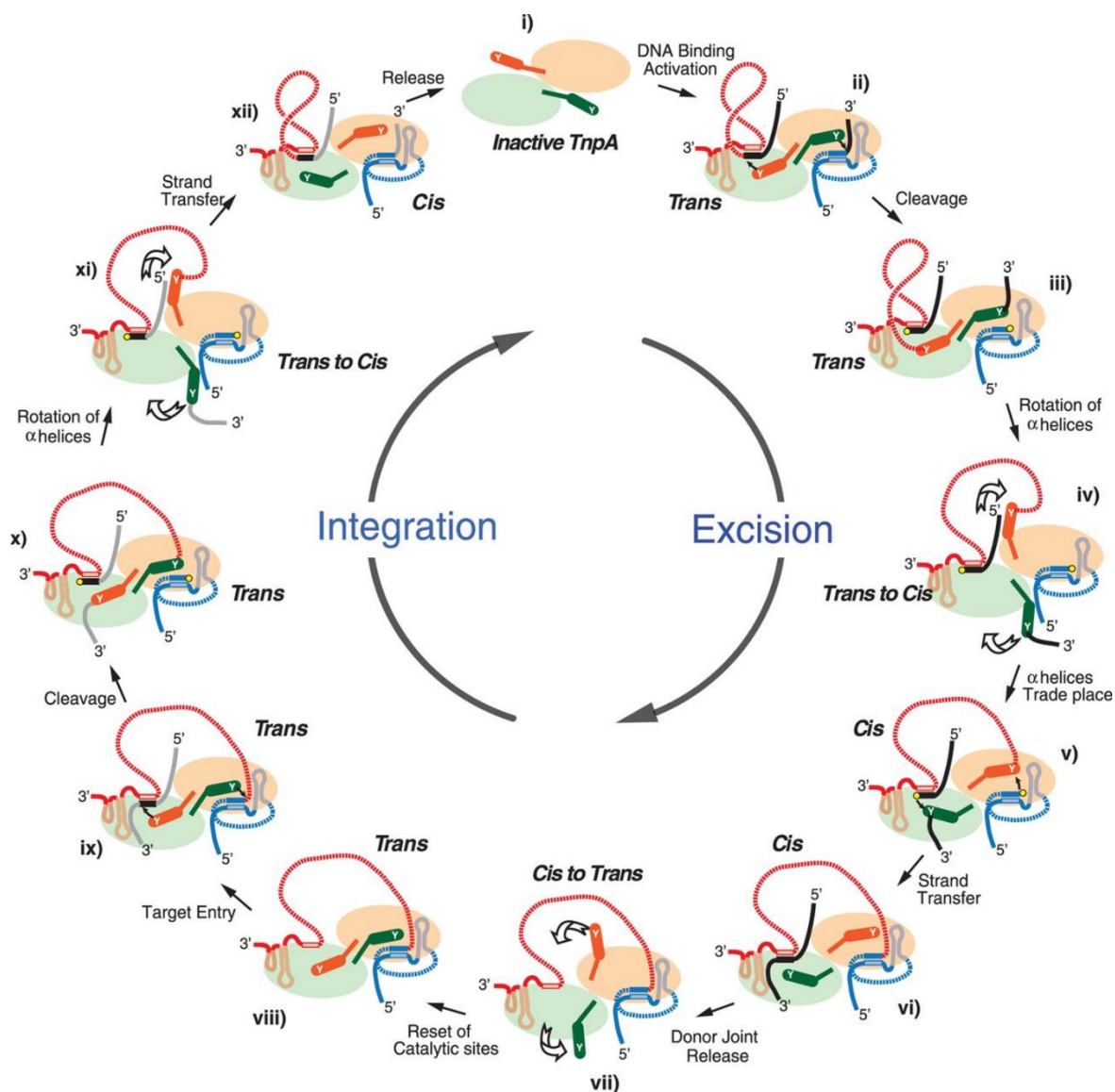
Jak již název napovídá, katalytickou skupinou těchto transpozáz je aminokyselina tyrozin. Oproti DDE transpozázám jde o podstatně menší skupinu. HUH motiv je tvořen konzervovaným párem histidinů oddělených hydrofobní aminokyselinou. Tento motiv váže hořčičnaté ionty, které se podílí na katalýze. Inzerční sekvence je přenesena ve formě jednovláknové DNA, čímž se tyrozinové transpozázy odlišují od DDE. Také na rozdíl od DDE nevznikají při inzerci inverzní repetice. Konce IS obsahují palindromy a konzervované oligonukleotidové motivy. HUH motiv se krom transpozáz vyskytuje u dalších nukleáz sobeckých genetických elementů (virů a plazmidů), které jako substrát také využívají jednovláknovou DNA. Proteiny s tímto motivem iniciují replikaci metodou valivého kruhu, která hraje roli např. při konjugaci plazmidů (Garcillán-Barcia and De La Cruz, 2002; Hickman *et al.*, 2002). Tyto transpozázy se rozlišují podle toho, zda mají jednu či dvě molekuly tyrozinu podílejících se na transpozici.

Y2 transpozázy jsou skupinou obsahující první identifikovanou tyrozinovou transpozázu s HUH motivem. Jsou asociovány s rodinou IS91 (Diaz-Aroca *et al.*, 1984; Garcillán-Barcia and De La Cruz, 2002; Curcio and Derbyshire, 2003). Transpozázy IS91 jsou schopny integrace do konzervovaného tetranukleotidu CTTG či GTTC. Při inzerci nedojde k duplikaci cílové sekvence.

Y1 transpozázy jsou unikátní skupinou transpozáz využívající pouze jeden katalytický tyrozin (Ronning *et al.*, 2005). Y1 transpozázy jsou využívány rodinou IS200/IS605 (Lam and Roth, 1983). Terminální oblasti sekvencí této rodiny jsou tvořené nedokonalými palindromickými sekvencemi, které jsou schopny ve formě jednovláknové DNA či RNA vytvářet vlásenky. Tyto vlásenky jsou

rozeznány transpozázou a IS následně vyštěpeny (obrázek 2). Výsledkem štěpení je 5' fosfo-tyrosin kovalentně navázaný na jednovláknový DNA meziproduct na jedné straně. Strana druhá je zakončena volným 3' OH zbytkem (He *et al.*, 2015). Krom vlásenky je nezbytná také přítomnost vedoucí (anglicky guide) sekvence. Jde o konzervovaný oligonukleotid umístěný blízko vlásenky. Díky tomu lze štěpit velice specificky vždy konkrétní nukleotidy, bez nutnosti dodatečné DNA vazebné domény v TnpA (Ton-Hoang *et al.*, 2012). Inzerce vyštěpeného elementu probíhá na replikační vidličce preferenčně do lagging vlákna. Také se zdá, že tyto transpozázy cílí na replikační vidličku při zastavení replikace (He *et al.*, 2013).

Rodina IS200/IS605 se dělí na podskupiny odlišné počtem kódovaných proteinů. IS200 kódují pouze jeden protein, TnpA, což je samotná transpozáza. IS605 kódují krom transpozázy ještě protein TnpB, ten pro transpozici nutný není (Kersulyte *et al.*, 2002). Ještě existuje třetí podskupina, IS1341. Ta kóduje pouze TnpB, z toho důvodu není jasné, zda jde o skupinu schopnou samostatné transpozice (Murai *et al.*, 1995). TnpB má patrně regulační funkci a snižuje transpoziční aktivitu IS (Pasternak *et al.*, 2013). Homology TnpB byly krom prokaryotních IS200/IS605 transpozáz také nalezeny u transpozonů eukaryot (Bao and Jurka, 2013).



Obrázek 2: Model transpozice IS608 z rodiny IS200/IS605. V pravé polovině je znázorněno vyštěpení IS z genomu hostitele, v levé polovině inserce do nového místa (proces je funkčně reverzní ve vztahu k vyštěpení). Transpozázá (elipsa) pracuje jako homodimer a díky tomu může naráz vázat oba konce IS elementu (oranžová, resp. modrá čára). DNA hostelské buňky je znázorněna černě (excize), resp. šedě (integrace). Převzato z (He et al., 2013).

### 3. REP elementy

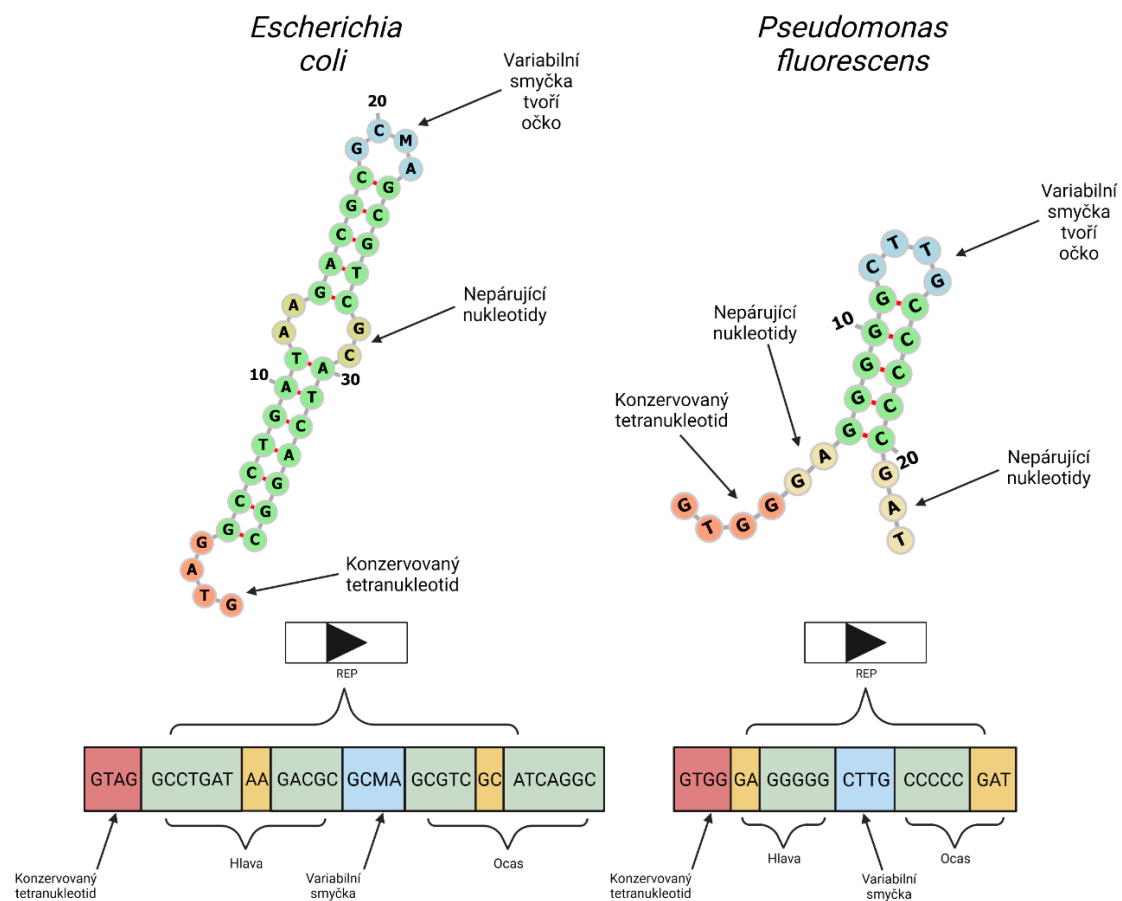
REP elementy byly objeveny u *Escherichia coli* a *Salmonella typhimurium*, již v roce 1982 (Higgins et al., 1982). V roce 1984 autoři původní práce element definovali a odhadli, že v genomu *E. coli* je více jak 500 takových sekvencí (Gilson et al., 1984; Stern et al., 1984). Element pojmenovali jako REP (repetitive extragenic palindromic sequence, česky repetitivní mimogenová palindromická sekvence) (Stern et al., 1984). Název shrnuje klíčové vlastnosti REP. Vyskytují se v hojném počtu, mimo čtecí rámce genů a jejich sekvence vytváří palindromy. To znamená, že sekvenci lze rozdělit na

dvě (téměř) identické, antiparalelní poloviny (viz kapitola 3.1.). Většina REP sekvencí je transkribovaná (Qian *et al.*, 2015).

Do konce tisíciletí byly REP nalezeny u několika druhů enterobakterií. Poté byly nalezeny i mimo enterobakterie, první u *Pseudomonas putida* a *P. aeruginosa* (Weinel *et al.*, 2002). Dnes se ukazuje, že REP elementy jsou rozšířeny napříč bakteriální diverzitou.

### 3.1. Struktura REP sekvence

Pro REP je charakteristické, že mají v bezprostřední blízkosti samotného palindromu konzervovaný 5'-GT(A/G)G-3' tetranukleotid (Nunvar, Huckova and Licha, 2010; Messing *et al.*, 2012). Podle pozice GTAG můžeme orientovat REP element, který se dá rozdělit na 4 části, jak je znázorněno na obrázku 3. Kromě samotného GTAG rozlišujeme hlavu (anglicky head) (první část palindromu), variabilní smyčku a ocas (anglicky tail), který je druhou částí palindromu. REP mají délku mezi 20 až 65 nukleotidy.



Obrázek 3: Schéma typického REP elementu s předpokládanou sekundární strukturou RNA. Žlutá barva vyznačuje nepárující nukleotidy. REP sekvence je z genomu *E. coli* (tato konkrétní se vyskytuje v genomu ve 45 kopiích) a *P. fluorescens* SBW25 (ta je v genomu ve stovkách kopií). Data převzata z (Bertels and Rainey, 2011b; Bertels, Gokhale and Traulsen, 2017). Schéma vytvořeno v biorender.com. RNA model vytvořen ve forna.

Palindrom v rámci REP elementu nemusí být dokonalý, u některých elementů se vyskytují úseky dlouhé 1-2 nukleotidy, které nejsou komplementární. Také nukleotidy variabilní smyčky nejsou komplementární, díky tomu jsou schopny vytvářet ve vlásence očko (anglicky loop). Komplementarita hlavy a ocasu umožňuje vytvářet vlásenkové struktury (anglicky stem-loop). V případě mutace jedné strany palindromu pozorujeme selekci na obnovení komplementarity (viz kapitola 5.3.). Také se často vyskytuje krátká mezera mezi konzervovaným tetranukleotidem a začátkem palindromu (obrázek 3, REP element *P. fluorescens*).

## 3.2. Porovnání REP mezi druhy

REP elementy se mezi druhy významně liší (příloha 1). Konsensus sekvence může být zcela odlišná, zachovává se pouze schopnost vytvářet palindrom a přítomnost GT(A/G)G tetranukleotidu. Celková délka se pohybuje v rozmezí 20 až 65 nukleotidů. Počet kopií v genomu, i v rámci příbuzných druhů, nebo dokonce v rámci různých linií jednoho druhu, se významně liší (Nunvar, Licha and Schneider, 2013). Některé bakteriální druhy mají méně než sto REP elementů, jiné jich mohou mít přes tisíc. Zatím nebyla nalezena spojitost mezi počtem REP elementů a charakteristikou daného druhu.

Je možné, že REP krom obecných mechanismů, které fungují univerzálně napříč diverzitou hostitelských bakterií, mohou získat v rámci druhu unikátní využití. Proto je vhodné považovat REP elementy každého druhu za unikátní skupinu. Mnoho mechanismů bude pravděpodobně společných, například funkční asociace s RAYT transpozázou (kapitola 5). Nicméně některé vlastnosti mohou být unikátní pro konkrétní bakteriální druh. Každý druh má REP elementy ve specifických pozicích, dokonce i jednotlivé kmeny daného druhu mají značně rozdílné umístění REP. Analýza REP se proto dá využít jako efektivní metoda pro identifikaci bakteriálního druhu i kmene (Dimri *et al.*, 1992; Versalovic *et al.*, 1994; Sheehan *et al.*, 1995).

### 3.2.1. *Pseudomonas fluorescens*

*P. fluorescens* je druh, u něž jsou REP extrémně rozšířeny. Jednotlivé kmeny mají jeden až dva tisíce REP na genom (obrázek 5) (Loper *et al.*, 2012). Řada kmenů obsahuje několik skupin REP elementů (viz kapitola 5.3.), které mají každá specifickou sekvenci. Obecná struktura REP elementu je v rámci druhu stejná (obrázek 4). GT(A/G)G je oddělen mezerou od palindromu (dva až tři nukleotidy). Variabilní smyčka je velmi krátká a po konci palindromu následuje konzervovaný konec (dva až tři nukleotidy). Nejčastější sekvence REP elementů je v příloze 1. Množství elementů je seskupeno ve vyšších strukturách, nejčastěji REPIN (Di Nocera, De Gregorio and Rocco, 2013). Vysoká konzervovanost REP a fakt, že nevykazují známky degradace, vedou k hypotéze, že se tyto elementy

rozšířily v *P. fluorescens* relativně nedávno (Nunvar, Licha and Schneider, 2013). *S. maltophilia* a *P. fluorescens* jsou druhy, u nichž je evoluce REP/RAYT systémů nejvíce zkoumána.

### 3.2.2. *Pseudomonas putida*

Genom obsahuje 751 kopií REP. Většina z nich tvoří páry, kde jsou REP v inverzní orientaci a lze je tak označit za REPIN (viz kapitola 5.). Obecné strukturní rysy REP *P. putida* jsou identické s *P. fluorescens* (obrázek 4). GT(A/G) je oddělen od palindromu mezerou (GA). Hlava, variabilní smyčka i ocas mají u obou druhů stejnou délku, liší se jejich sekvence (Di Nocera, De Gregorio and Rocco, 2013). Po konci palindromu také následují tři konzervované nepárující nukleotidy (GAA). Samotná symetrická struktura je tvořena 14 nukleotidy, právě jejich symetrie je evidentně pod vysokým selekčním tlakem a při mutaci je často udržena komplementární mutací (Aranda-Olmedo *et al.*, 2002).

Velký rozdíl od například *E. coli* je v zastoupení BIME. V rámci genomu *E. coli* existuje 582 REP elementů rozložených do 264 BIME (Espéli, Moulin and Boccard, 2001). *P. putida* má téměř 800 kopií REP, ale přitom bylo nalezeno pouze 8 BIME. REPIN v rámci *P. putida* nejsou schopny atenuovat transkripci (Espéli, Moulin and Boccard, 2001). Nejčastěji jsou umístěny mezi dvěma konvergentními geny, ve vzdálenosti do 30 bází od STOP kodonu přiléhajícího genu (Aranda-Olmedo *et al.*, 2002).

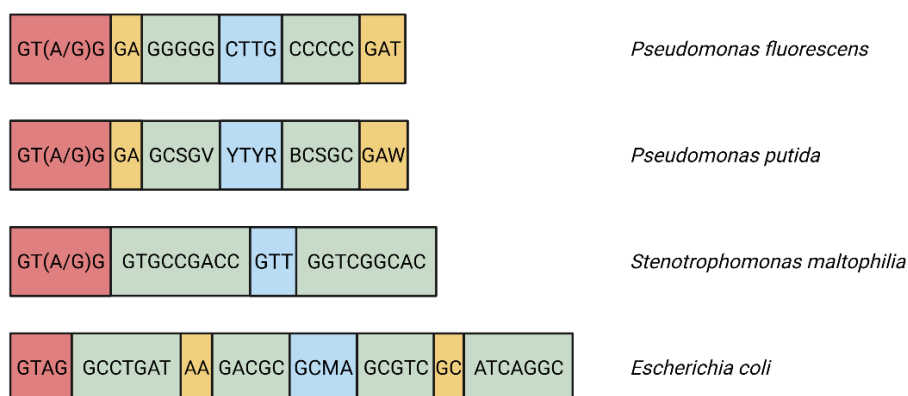
### 3.2.3. *Stenotrophomonas maltophilia*

Jde o druh s velkou populací REP elementů. Kmen *S. maltophilia* K279a má 1650 elementů, kmeny SKA14 a R551-3 pak obdobně (Di Nocera, De Gregorio and Rocco, 2013). GT(A/G)G je u *S. maltophilia* striktně konzervován. Z toho důvodu se REP elementy *S. maltophilia* také označují jako SMAG (*Stenotrophomonas maltophilia* GTAG). GTAG je nejčastější tetranukleotid, dále se vyskytují GTGG a (vzácně) GTCG tetranukleotidy. Na rozdíl od předchozích druhů chybí konzervované nekomplementární nukleotidy, GTAG je tedy umístěn hned vedle palindromu (obrázek 4). Velká populace GTAG-obsahujících REP elementů naznačuje, že bakterie mají funkční RAYT (viz kapitola 5.). Podle sekvence tvořící palindrom a variabilní smyčku jsou SMAG rozděleny do pěti rodin. Jednotlivé kmeny *S. maltophilia* se liší zastoupením rodin SMAG elementů. SMAG tvořící dvojice orientované TT (angl. tail to tail) a dvojice orientované HH (angl. head to head), korespondují k BIME-1, respektive BIME-2 elementům (Rocco, De Gregorio and Di Nocera, 2010).

SMAG elementy jsou velice konzervované, hlava a ocas jsou téměř vždy plně komplementární, což naznačuje, že zachování jejich komplementarity je pod vysokým selekčním tlakem.

### 3.2.4. *Escherichia coli*

*E. coli* je druh, u kterého byly REP elementy objeveny. Jejich počet se odhadoval mezi 500 až 1000, dnes genomy *E. coli* (podle použitého algoritmu a analyzovaného kmene) obsahují mezi 500 až 600 REP. Velká část elementů, kolem tří stovek, vytváří organizované shluky zvané BIME (viz kapitola 2.). BIME jsou specifické elementy, velice hojně zastoupeny právě u *E. coli*. Většina ostatních druhů tyto vyšší REP struktury neobsahuje. REP elementy *E. coli* mají v porovnání s ostatními druhy delší palindrom (obrázek 4). Ten je navíc přerušen páry nekomplementárních nukleotidů (jak hlava, tak ocas).



Obrázek 4: Srovnání struktury vybraných REP elementů u jednotlivých druhů. Konzervovaný tetranukleotid (červená), nepárující nukleotidy (žlutá), palindrom (zelená), variabilní smyčka (modrá). Data převzata z (Di Nocera, De Gregorio and Rocco, 2013; Nunvar, Licha and Schneider, 2013). Vytvořeno v biorender.com.

### 3.3. Funkce REP elementů

Ohledně REP elementů zůstává mnoho otázek nezodpovězených. Dosud bylo do větší či menší hloubky charakterizováno několik funkcí, jaké v buňce mohou zastávat. Vložení REP sekvence do 3' nepřepisovaných oblastí může efektivně zvýšit expresi daného genu o desítky procent. Transkribovaná mRNA je díky REP elementu lépe chráněná proti 3'-5' exonukleázám (Becerril *et al.*, 1985; McLaren *et al.*, 1991; Khemici and Carpousis, 2004), v důsledku toho je stabilnější a může být déle translatována (Merino *et al.*, 1987). Ochrana před degradací je patrně důsledkem přítomnosti vlásenek v mRNA, které jsou pro REP elementy tak typické. Tyto vlásenky zamezí nasednutí většiny 3'-5' exonukleáz. Tato mRNA může být rozštěpena pouze speciálními exonukleázami schopnými štěpit mRNA tvořící výrazné sekundární struktury, např. RNasa R a polynukleotid fosforyláza (PNPasa) (Newbury *et al.*, 1987). V případě jejich absence dochází ke kumulaci mRNA fragmentů obsahujících REP (Newbury *et al.*, 1987; Cheng and Deutscher, 2005). Vložení REP do 3' nepřepisované oblasti genu může také způsobit atenuaci transkripce (viz kapitola 4.2.). K tomu nicméně dochází při inzerci jen u některých genů a REP sníží exprese maximálně na polovinu (Espéli, Moulin and Boccard, 2001).

Velký vliv má vzdálenost, ve které je REP od terminačního kodonu umístěn. Ve vzdálenosti do 15 nukleotidů dochází, i přes zvýšení stability mRNA, k snížení exprese genu (Deng *et al.*, 2019). Důvodem je nejspíše opět vlásenková struktura. Umístění REP příliš blízko STOP kodonu pravděpodobně způsobí zastavení pohybu ribozomu a ukončení translace této mRNA (Liang, Rudd and Deutscher, 2015). Vlášenska na mRNA fyzicky znemožní ribozomu dokončit translaci. Výsledkem je nefunkční protein, který je poté degradován. Takový gen vykazuje až 3x nižší expresi, není tedy vypnut zcela. Na vlásenku mRNA jsou schopny se vázat RNA helikázy (Iggo *et al.*, 1990; Liang, Rudd and Deutscher, 2015). Jejich interakce způsobí dočasné rozvolnění vlásenky, díky čemuž je ribozom schopen dokončit translaci. Rozvolňování vlásenky je periodický děj, který závisí na množství RNA helikázy v buňce. Při poklesu hladiny RNA helikázy dojde k dalšímu poklesu množství funkčního proteinu. Zvýšení hladiny RNA helikázy naopak způsobí zvýšení. Ke zvýšení hladiny RNA helikázy dochází při vystavení buňky stresu, například UV záření. Lze tak předpokládat, že u některých druhů mají REP elementy funkci jako velice rychlé regulátory odpovědi na stres (Liang and Deutscher, 2016). U kmenů R551-3 a K279a druhu *S. maltophilia* bylo pozorováno, že REP elementy jsou často vzhledem k ORF posunuty o 10 až 15 nukleotidů (Rocco, De Gregorio and Di Nocera, 2010), což může být důsledkem selekčních procesů - buňky mohou tímto umístěním REP elementů upravit, které geny budou mít při stresu zvýšenou expresi.

Vliv na REP-závislou modulaci genové exprese má mnoho faktorů, např. délka REP vlásenky. Jak bylo výše zmíněno, roli hraje též vzdálenost REP od STOP kodonu, a dále samotná sekvence mezi REP a STOP kodonem. Výhodou přítomnosti REP jako nástroje pro regulaci hladiny exprese genů jsou malé účinky na ostatní geny a jimi kódované funkce (Deng *et al.*, 2019). Jedná se tak o potenciální nástroj pro genetické inženýrství bakterií.

Mobilní genetické elementy jsou schopny reorganizace genomu (Nevers and Saedler, 1977). REP jsou časté inzerční cíle pro některé rodiny těchto elementů. Konkrétně IS1397 u *E. coli* (Bachellier *et al.*, 1997; Clément *et al.*, 1999), ISKpn1 u *Klebsiella pneumoniae* (Wilde *et al.*, 2001, 2003) či IS621 u *E. coli* (Choi, Ohta and Ohtsubo, 2003). Díky pozici REP mimo sekvence kódující proteiny, jejich charakteristické stavbě a vysokému množství kopií jsou ideálními cíli pro inzerce (Tobes and Pareja, 2006). Díky transpozici do REP jsou tyto elementy místy hrajícími významnou roli při reorganizaci bakteriálního genomu (Shyamala, Schneider and Ames, 1990; Lee *et al.*, 2016).

### 3.4. Lokalizace REP elementů

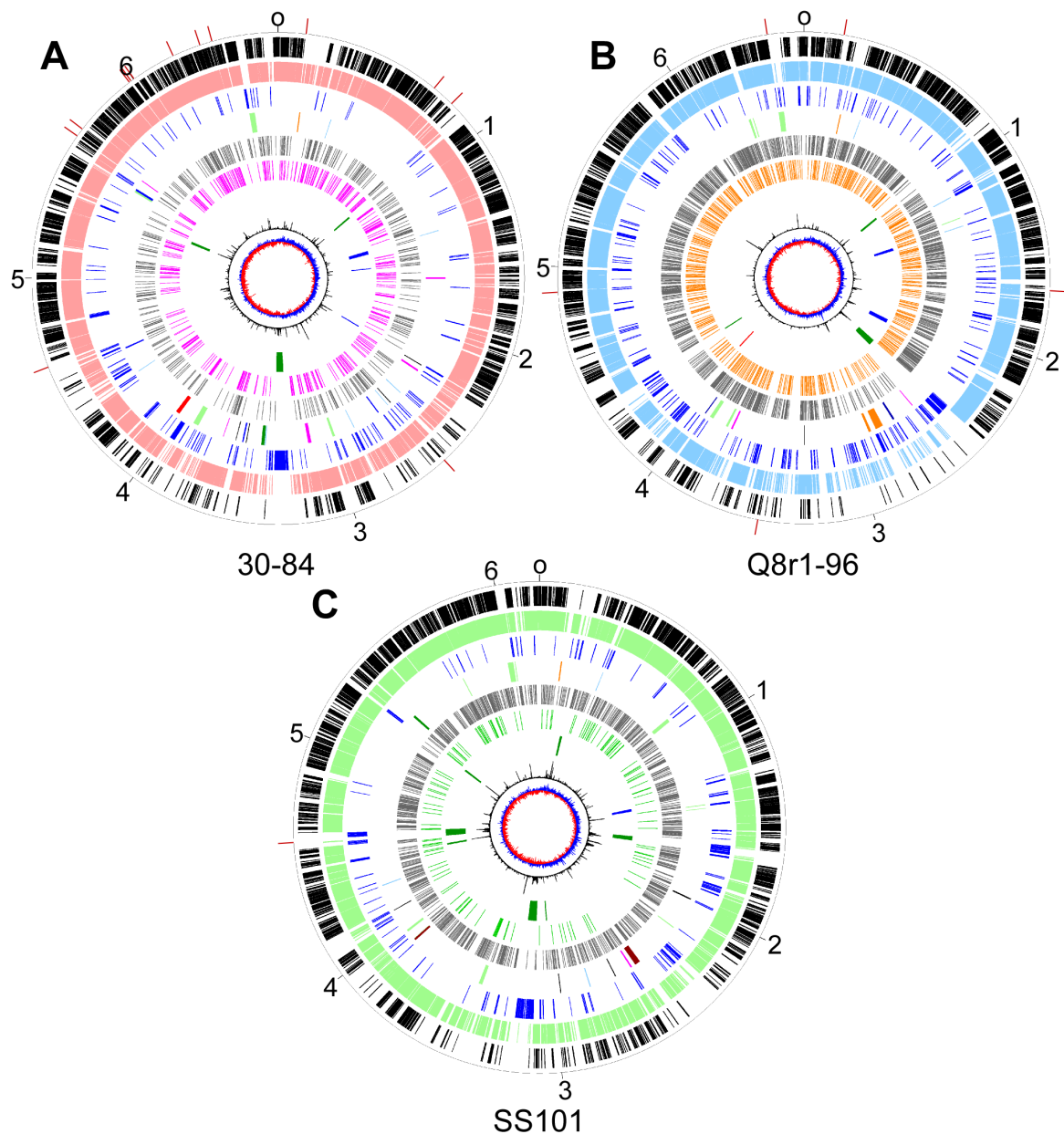
Nejčastější umístění REP elementů je mezi konvergentními operony (tabulka 1), tedy výlučně po proudu transkripce obou okolních genů. Při transkripci konvergentních genů dochází ke vzniku

pozitivních nadobrátek v genomu (tvoří se při pohybu RNA polymerázy). Je možné, že toto rozmístění REP elementů tak slouží ke snižování nadobrátek DNA při transkripci pomocí cílení aktivity DNA gyrázy (viz kapitola 4.3.2.). REP se téměř nikdy nevyskytují uvnitř sekvence, která kóduje proteiny. Vzdálenost od přiléhajícího genu se pohybuje do 100 bází od terminačního kodonu (Aranda-Olmedo *et al.*, 2002). První nukleotidy REP sekvence mohou být součástí terminačního kodonu (často poslední tři báze tetranukleotidu GTAG slouží jako STOP kodon).

Druh	RF/EF →→	RF/EF →←	RF/EF ←→	RF/EF ←←
<i>Escherichia coli</i> K12	0,64	2,93	0,05	0,78
<i>Pseudomonas aeruginosa</i>	0,58	3,35	0,14	0,59
<i>Pseudomonas putida</i> KT2440	0,82	2,06	0,39	0,88
<i>Salmonella enterica typhimurium</i>	0,77	2,25	0,14	0,95

Tabulka 1: Porovnání výskytu REP v extragenových prostorech genomu. Uvedeny jsou poměry mezi reálným počtem REP elementů (real frequency, RF) v oblasti a počtem očekávaným při náhodném rozložení REP (expected frequency, EF). Symboly u sloupců značí umístění REP mezi stejnosměrnými geny na pozitivním vláknu (→→), konvergentními geny (→←), divergentními geny (←→) a stejnosměrnými geny na negativním vláknu (←←). Data převzata z (Tobes and Ramos, 2005).

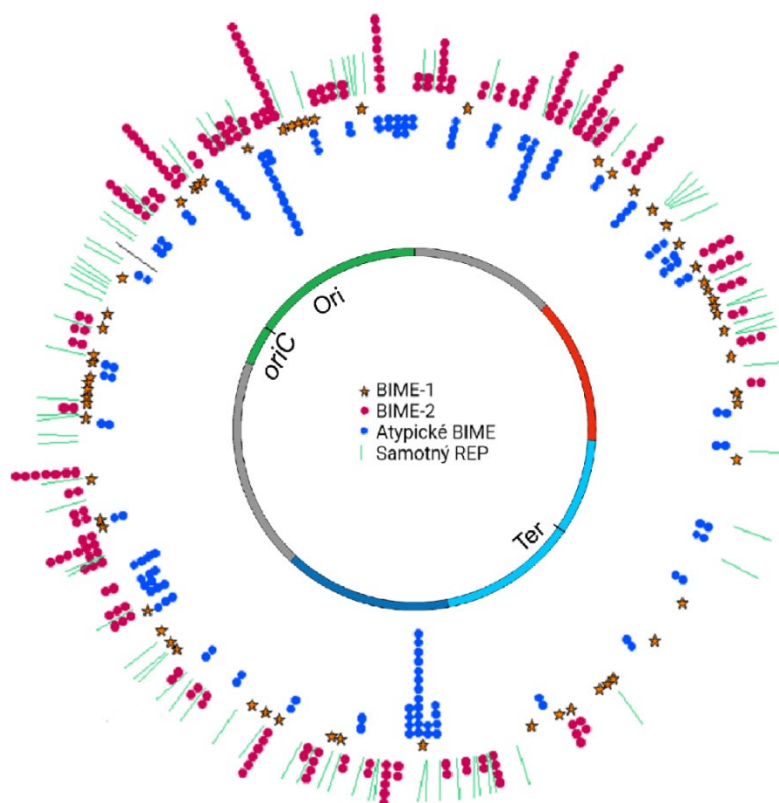
Předpokládá se, že většina REP sekvencí je transkribována a má poté vliv na strukturu mRNA a průběh translace. Nebyla nalezena žádná spojitost mezi funkční kategorií genů a přítomností REP elementů v jejich downstream oblastech. Jak ukazuje obrázek 5, REP se vyskytují napříč genomem hostitelské bakterie. Nicméně jejich výskyt je soustředěn především v core genomu (Loper *et al.*, 2012). Core genom je soubor homologních genů, které jsou společné pro všechny kmeny daného druhu. Typicky tvoří na chromozomu mozaikovou strukturu, kdy jsou stabilní oblasti s core geny přerušovány oblastmi, které jsou mezi kmeny variabilní. Dále REP chybí v oblastech vzniklých šířením mobilních elementů. To koresponduje s jejich malou tendencí šířit se horizontálním přenosem a nízkou tendencí kolonizovat DNA získanou (více či méně) recentním horizontálním přenosem.



Obrázek 5: Lokalizace REP u tří genomů *P. fluorescens*. Druhý vnější kruh vyznačuje core genom. REP elementy jsou vyznačeny v pátém a šestém kruhu od okraje. U 30-84 purpurovou, Q8r1-96 oranžovou a u SS101 zelenou barvou. Převzato z (Loper *et al.*, 2012).

## 4. BIME

V roce 1982 byla popsána první REP sekvence u *E. coli* (Higgins *et al.*, 1982). V roce 1991 bylo zjištěno, že REP jsou velice často součástí vyšších repetitivních elementů, takzvaných BIME (anglicky bacterial interspersed mosaic element) (Gilson *et al.*, 1991b, 1991a). BIME jsou sekvence DNA dlouhé 40 až 500 nukleotidů. V genomu *E. coli* je přibližně 300 BIME elementů, které jsou relativně rovnoměrně rozesety po genomu, s výjimkou oblasti *ori*, kde jsou REP zastoupeny více, a oblasti terminace replikace, kde elementy téměř zcela chybí (obrázek 6) (Espéli, Moulin and Boccard, 2001).

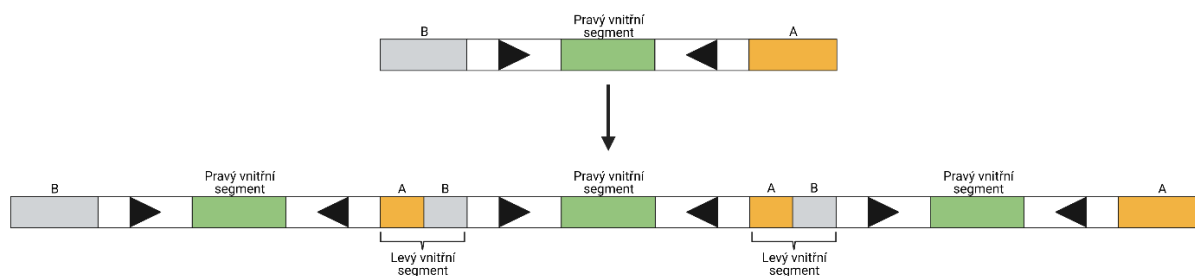


Obrázek 6: Mapa REP a BIME elementů v genomu *E. coli*. Převzato a upraveno z (Bachelier, Clément and Hofnung, 1999; Nolivos *et al.*, 2016).

BIME jsou mozaiky z 8 typů menších elementů (10, pokud rozlišujeme i jednotlivé typy REP) (viz kapitola 4.2.). Tyto menší elementy jsou poskládány v různých kombinacích. Několik elementů může být poskládáno do široké škály mozaik (Gilson *et al.*, 1991b). BIME mohou obsahovat pouze jeden, tak i více než 10 REP. Jednotlivé REP jsou od sebe vždy odděleny jiným motivem, nikdy tak dvě REP nejsou bezprostředně vedle sebe. Také se pravidelně střídá orientace REP elementů.

#### 4.1. Vznik BIME variability

Zdá se, že odlišné BIME mozaiky vznikly z elementu typu REPIN (Nunvar, Huckova and Licha, 2010). Původní element podobný REPIN (rozdílem je, že REPIN má strukturu REP-mezera-REP, zatímco BIME-1 je iREP-mezera-iREP) se rozšířil genomem v mnoha kopiích. Poté u některých kopií došlo k amplifikaci a tandemovému zmnožení původního elementu. Některé části byly časem smazány, či u nich došlo ke konverzi. Tímto mechanismem se dá vysvětlit jednak variabilita BIME sekvencí, dále skutečnost, že jsou vlastně jedním základním vzorem, který je tandemově opakován (obrázek 7). Jednotlivá opakování se poté liší třeba jen jedním elementem. Již první práce předpokládají, že BIME variabilita vznikala tímto mechanismem. Původní dvojice REP elementů nazývají jako „generating dublet“ (Gilson *et al.*, 1991b)



Obrázek 7: Hypotetický model BIME amplifikace. Původní BIME (typ BIME-1) je amplifikován a několikrát za sebou se opakuje. Původně vnější motivy A a B jsou místem překryvu. Přibližně polovina délky motivů je ztracena, dochází k rekombinaci a vytvářejí společně motiv levého vnitřního segmentu. Tento motiv se poté označuje jako r. iREP elementy (REP na komplementárním vlákně, tj. 5'-stem-loop-stem-CTAC-3') jsou znázorněny černými šipkami. Vypracováno podle (Gilson et al., 1991b). Vytvořeno v biorender.com.

## 4.2. Struktura BIME a jednotlivé motivy

Jednotlivé elementy tvořící BIME lze rozdělit do dvou skupin. První skupinou jsou samotné REP, ty lze rozdělit podle páru nukleotidů na 7. a 32. pozici sekvence na REP typu Y a typu Z. Nukleotidy na 7. a 32. pozici jsou komplementární, podílí se tak na vzniku vlásenek v sekundární struktuře. REP typu Y mají GC pár nukleotidů na 7. a 32. pozici, REP typu Z pak AT pár. Typ Z se dále ještě dá rozlišit podle délky na Z<sup>1</sup> a Z<sup>2</sup>. Typ Z<sup>1</sup> má délku 33 nukleotidů, Z<sup>2</sup> 40 nukleotidů (tabulka 2). Rovněž variabilní smyčka uprostřed REP palindromu je specifická pro jednotlivé skupiny.

REP typu Y jsou schopny atenuace transkripce DNA. Přesný mechanismus není znám, nicméně se zdá, že tento REP element je schopen zastavit pohyb RNA polymerázy. Poté je nutná přítomnost Rho faktoru, jenž uvolní RNA polymerázu a dochází k terminaci transkripce. Nejde o příliš účinný způsob atenuace, transkripce upstream DNA je snížena o maximálně 50%, obvykle méně (Gilson et al., 1986; Espéli, Moulin and Boccard, 2001).

Typ REP elementu	Nukleotidy na 7 a 32 pozici	Délka sekvence	Homogenita sekvence	Variabilní smyčka	Rodina BIME, ve které se motiv vyskytuje
Typ Y	GC	38 nt	92%	GCGT(G/A)AA	BIME-1 i -2
Typ Z1	AT	33 nt	94%	CTA	BIME-1
Typ Z2	AT	40 nt	93%	ACGCT(G/T)GCG	BIME-2

Tabulka 2: Rozdělení REP elementů v rámci genomu *E. coli*. Vypracováno podle (Gilson et al., 1991b).

Druhou skupinou elementů tvořících BIME jsou takzvané extra-REP (extra-PU) motivy (tabulka 3). Jde o 7 sekvencí, které na rozdíl od REP nemají v sekvenci palindrom. Krom L elementu jsou podstatně méně homogenní (jejich sekvence je méně konzervovaná) v porovnání s REP. Motivы A a B jsou na vnějších okrajích BIME, mohou se vyskytovat i mezi REP uvnitř BIME, k tomu dochází, pokud je původní element typu REPIN amplifikován.

Vnitřní motivy jsou mezi jednotlivými dvojicemi REP. Jsou rozlišeny na pravé a levé podle toho, jak jsou REP vzhledem k nim orientované. Motiv je pravý, pokud jsou okolní REP elementy orientované směrem k němu. Levý motiv je obklopen REP, které jsou orientovány směrem od něj, znázorněno na obrázku 7.

Pravé vnitřní motivy jsou L a S motiv. L motiv je nejdelší (32 až 34 nukleotidů) a nejvíce konzervovaný extra-REP motiv. Levé vnitřní motivy jsou l, s, r motiv. s motiv se vymyká od ostatních, neboť je dlouhý pouze 1 nukleotid, přičemž v 60% případů jde o molekulu guanosinu. R motiv, na rozdíl od ostatních je spíše skupinou 5 segmentů délky od 18 po 31 nukleotidů.

Extra-PU motiv		Homogenita sekvence	Délka sekvence	Rodina BIME, ve které se motiv vyskytuje	
Vnější	A	55%	---	BIME-1	
	B	60%	---	BIME-1	
Vnitřní	Pravé	L	80%	22-34 nt	BIME-1
		S	50%	11-14 nt	BIME-2
	Levé	l	68%	8-9 nt	BIME-2
		s	---	1 nt	BIME-2
		r	---	18-31 nt	BIME-2

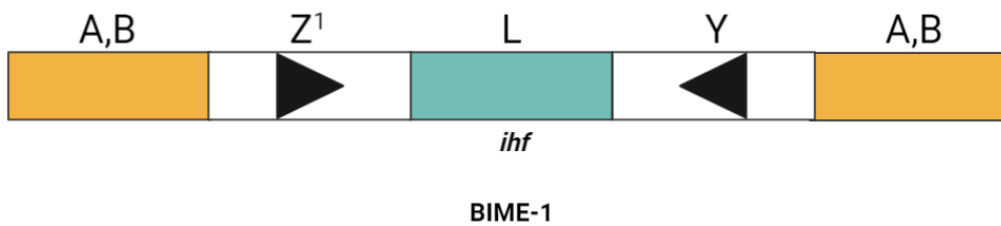
Tabulka 3: Přehled extra-REP motivů tvořících BIME mozaiky v rámci genomu *E. coli*. Vypracováno podle (Bachelier et al., 1994).

### 4.3. BIME rodiny

BIME se svou strukturou dají rozdělit do dvou rodin. Ty se liší tím, zda obsahují REP typu Z<sup>1</sup> či Z<sup>2</sup>, celkovým počtem REP motivů, zastoupením ostatních motivů a posloupností motivů. V genomu *E. coli* byly klasifikovány dvě velké rodiny, BIME-1 a BIME-2. Nicméně, ne všechny BIME se dají takto rozdělit. Sekvence, které nevyhovují kritériím ani jedné rodiny, se označují jako atypické BIME.

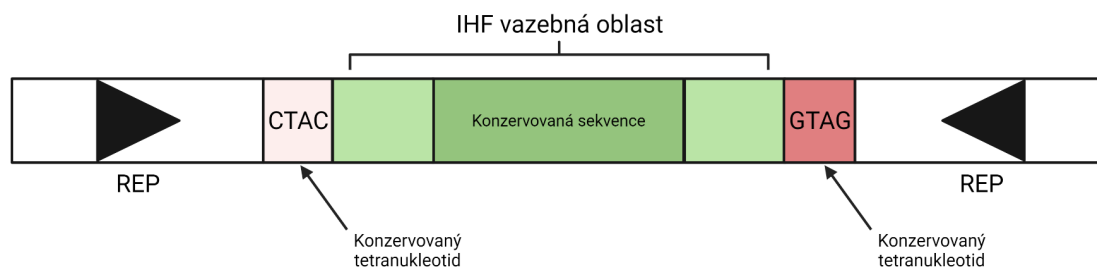
#### 4.3.1. BIME-1

BIME-1 rodina obsahuje většinu REP typu Z<sup>1</sup>. BIME-1 nevytváří tandemové kopie, na rozdíl od BIME-2. BIME-1 se skládá ze dvou konvergentně orientovaných iREP, jeden typu Z<sup>1</sup>, druhý Y (obrázek 8). L motiv tyto dva REP odděluje. V rámci BIME-1 rodiny je L motiv vysoce konzervován. Sekvence L motivu je schopna vázat IHF protein (Friedman, 1988; Tsui and Freundlich, 1988). Tyto BIME se označují jako RIB (anglicky reiterative IHF BIME) či RIP (anglicky repetitive IHF binding palindromic element) elementy. Jejich schématickou strukturu znázorňuje obrázek 9. RIB jsou přítomné pouze v genomu *E. coli*, nebyly nalezeny ani v blízkce příbuzných druzích jako je *S. typhimurium*.



Obrázek 8: Schéma motivu BIME-1 rodiny. Bílé obdélníky označují REP element, trojúhelník uvnitř pak naznačuje orientaci elementu. Vnější motiv je buď A či B. Vypracováno podle (Bachelier, Clément and Hofnung, 1999). Vytvořeno v biorender.com.

V genomu *E. coli* se nachází kolem 100 BIME-1, z nich asi 70 lze označit za RIB elementy. IHF vazebná oblast RIB motivu, v porovnání s IHF vazebnými oblastmi obecně, je extrémně dobře konzervovaná (Boccard and Prentki, 1993). Je tak možné, že k rozšíření RIB elementů v genomu *E. coli* došlo díky selekci na jejich schopnost vázat IHF. BIME-1 jsou typicky umístěny na 3' koncích operonů.



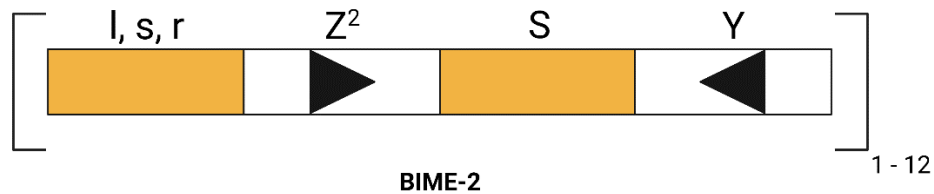
Obrázek 9: Z 28 RIB elementů *E. coli*, které v práci zkoumali, 25 odpovídá svou strukturou tomuto schéma. Za zmínku stojí vysoká konzervovanost centrální oblasti pro vazbu IHF proteinu. Dále přítomnost konzervovaných tetranukleotidů na obou koncích IHF-vazebné sekvence. Vypracováno podle dat (Bachelier *et al.*, 1993). Vytvořeno v biorender.com.

IHF je DNA-vazebný protein vyskytující se u *E. coli*, *S. enterica* a příbuzných bakterií (Oppenheim *et al.*, 1993; Dorman and Deighan, 2003). Byl objeven při rekombinaci fága  $\lambda$ , krom rekombinace ho bakteriofág využívá i ke sbalení své DNA (Kikuchi and Nash, 1978). IHF je schopen se v DNA vázat na specifickou sekvenci a ohnout poté dvoušroubovici DNA o přibližně 180 stupňů (Rice *et al.*, 1996; Rice, 1997). Tento ohyb pomáhá DNA kompaktně se složit v buňce, také DNA dovoluje vytvářet nadobrátky. IHF má vliv na řadu dalších procesů probíhajících na chromozomu. Dokáže ovlivnit replikaci (Ryan *et al.*, 2002), upravit genovou expresi (Charlier *et al.*, 1995) a zajistit interakci vzdálených oblastí genomu (Goosen and van de Putte, 1995).

#### 4.3.2. BIME-2

Tyto elementy mohou obsahovat až 12 REP sekvencí. BIME-2 se skládají ze základní struktury podle obrázku 10, tato struktura je poté tandemově opakována až dvanáctkrát. V genomu *E. coli* je přibližně 200 BIME-2 elementů. BIME-2 se predominantně vyskytují v oblastech mezi dvěma operony. Toto rozmístění může souviset s jejich schopností vázat DNA gyrázu (Gellert *et al.*, 1976; Yang and Ferro-Luzzi Ames, 1988; Reece, Maxwell and Wang, 1991). DNA gyrázu váží především konzervované sekvence Y motivu specificky nepárující nukleotidy uvnitř Y REP vlásenky a očko

(Sutormin *et al.*, 2019). Přesný reakční mechanismus není znám (Espéli and Boccard, 1997). Jedním z důsledků působení DNA gyrázy je snižování torzního stresu na vláknu DNA. Tento stres může vznikat při transkripci DNA a způsobit její zpomalení. Díky své pozici na 3' koncích operonů mohou BIME-2 cílením aktivity DNA gyrázy zabránit zvyšování nadobrátek DNA vznikajícím při transkripci okolních genů (Yang and Ferro-Luzzi Ames, 1988).



Obrázek 10: Schéma BIME-2, původní element je tandemově opakován 1-12x za sebou. Občas se vyskytnou výjimky, kdy v jednom opakování chybí motiv, či je změněný. Vypracováno podle (Bachelier, Clément and Hofnung, 1999). Vytvořeno v biorender.com.

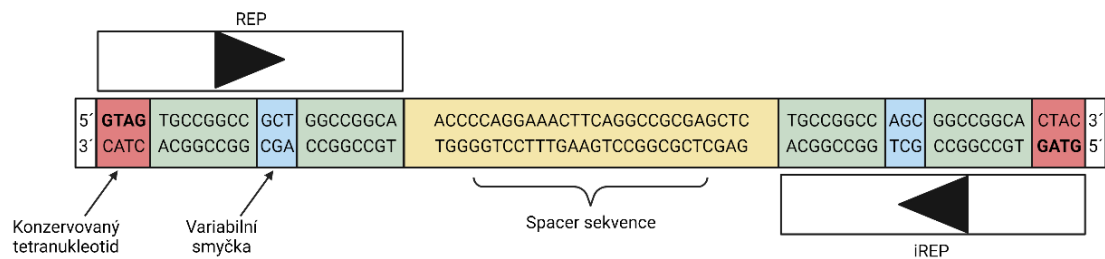
IHF protein se neváže na BIME-2 rodinu. DNA gyráza neinteraguje s BIME-1 rodinou. Obě rodiny se liší svým lokálním rozmístěním. Můžeme hypotetizovat, že obě rodiny mají rozdílnou funkci uvnitř buňky a působí na ně odlišné selekční tlaky.

## 5. RAYT a REPIN

RAYT (angl. REP-associated tyrosine transposase) jsou relativně nedávno objevenou třídou nukleáz. RAYT, jak plyne z názvu, byly původně pojmenovány jako transpozázy, vzhledem k jejich sekvenční příbuznosti k IS200/IS605 transpozázám (viz dále). RAYT však nejsou schopny vlastní mobilizace, označovat je za transpozázy není tedy přesné, nicméně jde o zavedený pojem, který bude v práci respektován.

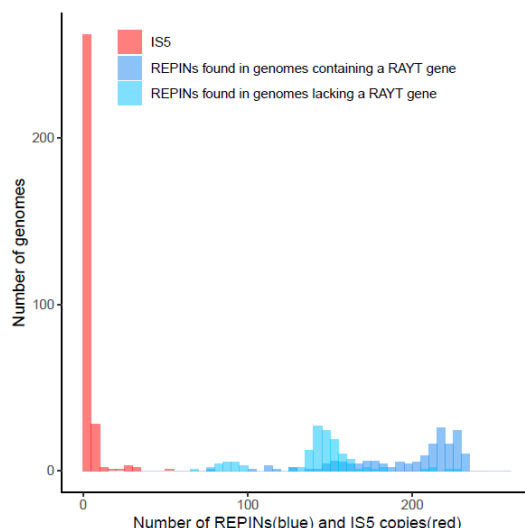
RAYT jsou příbuzné rodiny IS200/IS605 transpozáz (Nunvar, Huckova and Licha, 2010), v sekvenci aminokyselin mají konzervován katalytický tyrosin a HUH motiv. (Ilyina and Koonin, 1992; Koonin and Ilyina, 1993). Krom těchto dvou konzervovaných oblastí je zbytek sekvence variabilní. Zdá se, že na *rayt* geny působí diverzifikující selekce (Bertels and Rainey, 2011a). Tyto geny nejsou pro buňku nezbytné, při jejich odstranění se fyziologie buňky nemění. (Bertels and Rainey, 2011a). Dodnes není známo, jakou roli RAYT v genomu může hrát, přesto se *rayt* geny nalézají v asi 24 až 26% všech bakteriálních druhů, u kterých je sekvenován alespoň jeden celý genom (Bertels, Gallie and Rainey, 2017; Quentin *et al.*, 2018). Jak již bylo zmíněno, pro geny kódující RAYT je typické, že jsou z obou stran ohraničeny REP elementy. Oblast *rayt* genu a okolních REP/REPIN se nazývá REPtron. Častým ohraničením *rayt* jsou samotné REP v inverzní orientaci. Další možností je ohraničení BIME mozaikami či REPIN elementy. To je důvodem, proč již při objevení byly RAYT považovány za proteiny interagující s REP elementy (Nunvar, Huckova and Licha, 2010). REPIN je tvořen ze dvou REP

elementů oddělených spacer sekvencí (obrázek 11). Spacer sekvence má v rámci jedné třídy REPIN (REPIN asociované se společným RAYT) podobnou délku.



Obrázek 11: Příklad sekvence REPIN elementu z genomu *Stenotrophomonas maltophilia* K279a. Data z (Rocco, De Gregorio and Di Nocera, 2010). Sekvence z databáze *biocyc.org*. Vytvořeno v *biorender.com*.

Prozatím nebyl nalezen přímý (experimentální) důkaz kauzálního spojení RAYT transpozáz se šířením REP/REPIN elementů. Existuje však řada nepřímých důkazů, díky kterým se lze domnívat, že REPIN jsou genomem šířeny právě pomocí RAYT transpozáz (Bertels and Rainey, 2011b). Bylo dokázáno, že RAYT jsou schopny se vázat na REP sekvenci (Messing *et al.*, 2012) a také štěpit DNA v blízkosti REP. Dále, REP v rámci REPIN vykazují vyšší konzervovanost než REP vyskytující se samotně, což implikuje REPIN jakožto jednotku podléhající specifické propagaci genomem. Je tak možné, že všechny REP nalezené samostatně jsou degenerovanými pozůstatky REPIN. Důvod, proč se jeden samostatný REP mohl zachovat, může být například jeho užitečnost v daném místě genomu pro hostitelskou buňku (Bertels and Rainey, 2011a). Významným nepřímým důkazem je analýza genomů několika stovek kmenů *E. coli*, které mají či nemají *rayt* geny (Park, Gokhale and Bertels, 2021). Obrázek 12 zachycuje zřetelnou korelaci, kdy genomy s funkčním *rayt* genem mají více REPIN, oproti genomům bez *rayt* genu. Jakožto kvantitativní analýza toto pozorování nenabízí odpověď, jak spolu RAYT a REPIN interagují. Z dat se dá nicméně vyvodit, že REPIN jsou buď pomocí RAYT propagovány nebo alespoň udržovány v genomu.



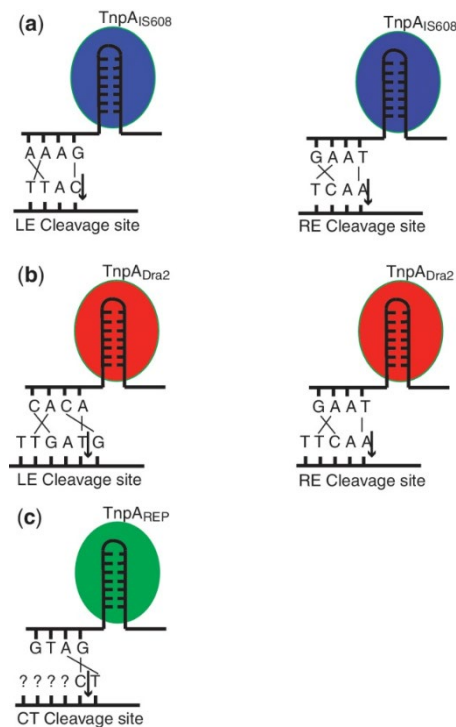
Obrázek 12: Distribuce IS5 a REPIN elementů v rámci 300 de-replikovaných *E. coli* genomů. Světle modrou barvou jsou zvýrazněny genomy bez *rayt* genu, tmavě modrou genomy s *rayt* genem. Červená označuje IS5. Převzato z (Park, Gokhale and Bertels, 2021).

## 5.1. Porovnání RAYT a IS200/IS605 transpozáz

### 5.1.1. Mechanismus štěpení ssDNA

RAYT jsou příbuznými transpozázami rodiny IS200/IS605. Obě rodiny patří do HUH transpozáz, v jejich rámci pak do Y1 podskupiny (Kapitola 2.2). IS200/IS605 transpozázy typicky tvoří dimery. Dimerizaci využívají při excizi ssDNA vlákna, kdy se vážou na oba konce vlákna. RAYT dimery nevytváří, funkční jednotkou je monomer (Messing *et al.*, 2012). Schopnost vystřihnout a přenést vlákno je jedním z hypotetických kroků potřebných k tomu, aby byl RAYT schopen vytvářet BIME elementy (viz kapitola 3.). Alternativně by vznik BIME elementů mohl být řízen zcela jinou transpozázou. Všechny Y1 transpozázy využívají dvojmocný iont kovu jako kofaktor v aktivní oblasti. IS200/IS605 transpozázy využívají  $Mg^{2+}$  zatímco RAYT transpozáza preferuje  $Mn^{2+}$ , nicméně při snížené efektivitě pracuje *in vitro* i s  $Mg^{2+}$  (Ton-Hoang *et al.*, 2012). Důvodem nižší afinity k  $Mg^{2+}$  může být záměrná selekce, kdy RAYT protein hůře váže  $Mg^{2+}$ , který je v buňce výrazně častější než  $Mn^{2+}$ , a tím je snížena jeho aktivita. Pro nalezení místa štěpení využívají IS200/IS605 dva specifické oligonukleotidové motivy, které jsou umístěny na koncích inzerční sekvence. RAYT transpozáza pro determinaci místa štěpení také využívá oligonukleotid, konkrétně motiv GTAG (obrázek 13). GTAG motiv je umístěn typicky do vzdálenosti 2 nukleotidů od začátku palindromu REP. Na 3' konci, na rozdíl od IS200/IS605, využívá RAYT transpozáza pouze konzervovaný dinukleotid 5'-CT-3', ten může být od REP elementu umístěn i ve větší vzdálenosti. GTAG je kritický pro rozpoznání REP elementu RAYT transpozázou, mezi nukleotidy CT pak dochází ke štěpení vlákna DNA. V případě posunutí CT motivu bude transpozáza stále fungovat a štěpné místo se přesune s motivem. IS200/IS605

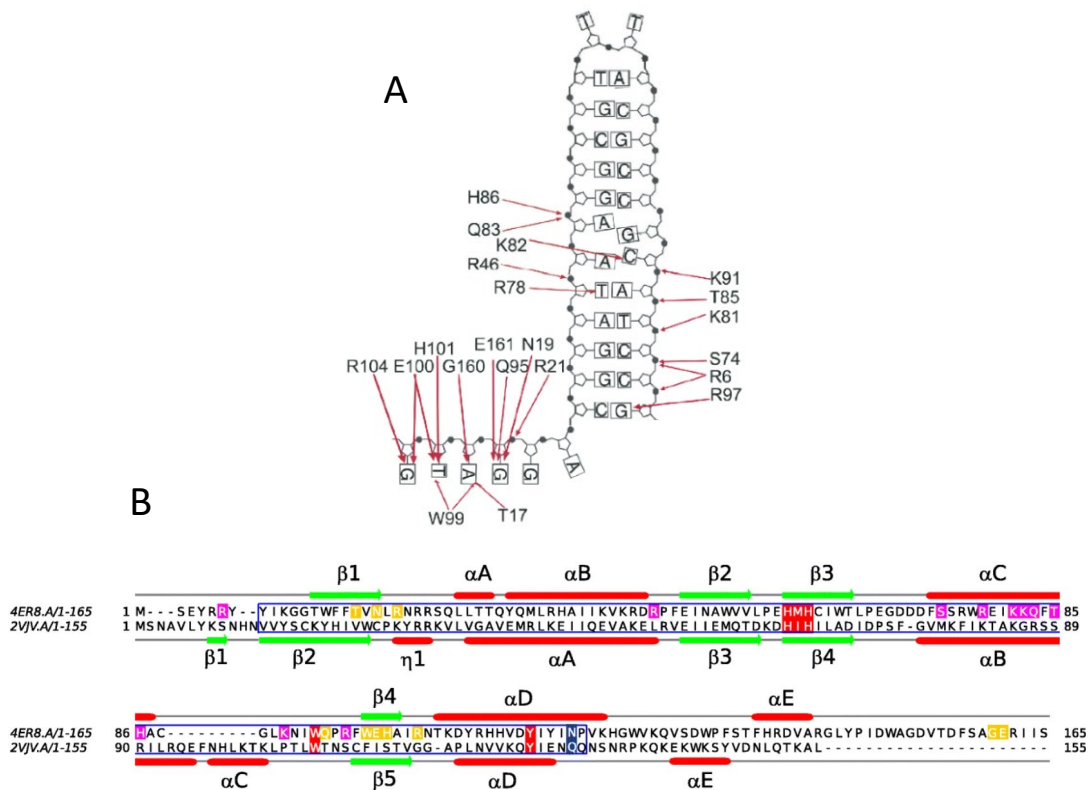
transpozázy využívají specifitější sekvence pro nalezení štěpného místa (obrázek 13). Stejně jako u IS200/IS605 transpozázy dochází při determinaci štěpného místa k interakci báze-báze. Z GTAG tetranukleotidu A3 komplementárně interaguje s T2, G4 interaguje s C1.



Obrázek 13: Porovnání štěpných míst Y1 transpozáz. Pro IS200/IS605 transpozázy IS608 a Dra2 jsou známy vždy dva oligonukleotidy, které transpozáza specificky štěpí. RAYT štěpí pouze CT dinukleotid, ten se musí nacházet v blízkosti REP sekvence. Převzato z (Messing et al., 2012).

### 5.1.2. Funkčně-specifické motivy RAYT transpozáz

RAYT interaguje s REP především prostřednictvím bází GTAG tetranukleotidu. U *E. coli* jsou vazebnými aminokyselinami 17T, 19N, 21R, 95Q, 99W, 100E, 101H a 104R. Dále se interakce účastní ještě 160G a 161E (obrázek 14B), které jsou ovšem součástí auto-inhibiční C-terminální domény, jež není konzervována v rámci diverzity RAYT (obrázek 14A). Vazba AK na REP vlásenku v drtivé většině probíhá prostřednictvím cukr-fosfátové kostry. S výjimkou C-terminálních AK jsou všechny AK interagující s GTAG konzervované u RAYT asociovaných s REP elementy (Kapitola 5.4).



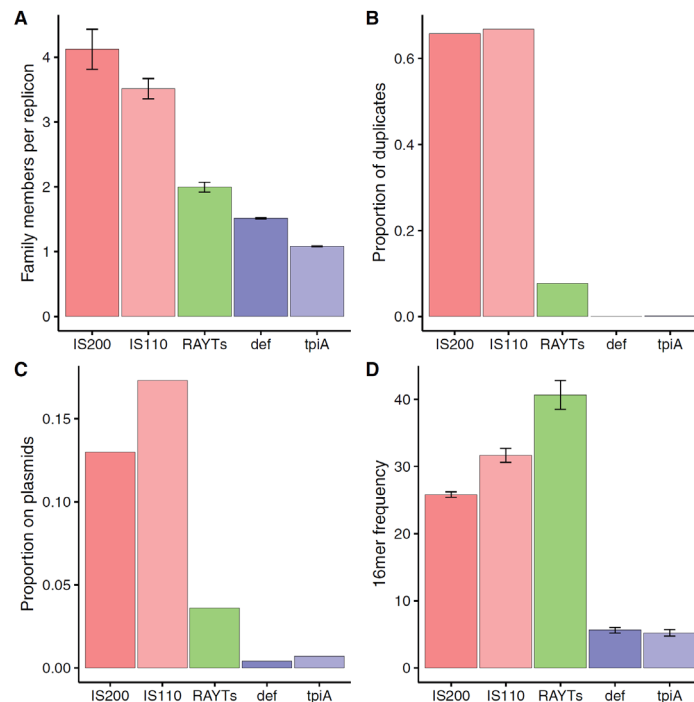
Obrázek 14: A) Schéma sekvence REP elementu s GTAG tetranukleotidem. Konzervované aminokyseliny RAYT jsou přiřazeny k nukleotidům, na které se vážou. Převzato z (Messing *et al.*, 2012). B) Porovnání AK sekvence RAYT (PDB 4ER8) a IS608 (PDB 2VJV) transpozázy. Vnější čáry značí sekundární struktury proteinu,  $\alpha$  helix (červená) a  $\beta$  list (zelená). Konzervované aminokyseliny jsou zvýrazněny červenou (HUH motiv, katalytický tyrosin), modrou (AK nezbytné pro transpozici), žlutou (AK interagující s GTAG), purpurovou (AK interagující s REP vláseňkou). V modrém čtverci je vyznačeno konzervované jádro proteinu. Převzato z (Quentin *et al.*, 2018).

Část C-terminální domény RAYT z *E. coli* slouží jako regulátor štěpné aktivity. Jde o první známý mechanismus autoregulace DNA transpozázy, kdy sama část transpozázy pracuje jako kompetitivní auto-inhibitor (Messing *et al.*, 2012). C-terminální doména kompetuje s ssDNA o vazbu do aktivní oblasti. Jedná se pravděpodobně o mechanismus pro snížení aktivity transpozázy. Tato hypotetická schopnost regulace RAYT nebyla zatím do hloubky prozkoumána. Navíc není známo, nakolik je inhibice aktivity RAYT C-terminální doménou rozšířená v rámci diversity RAYT proteinů, z nichž významná část C-terminální doménu postrádá (viz kapitola 5.4.).

### 5.1.3. Genomová distribuce genů transpozáz a s nimi asociovaných repetitivních sekvencí

Počet (téměř) identických kopií je jeden z hlavních rozdílů mezi RAYT a transpozázami inzerčních sekvencí (např. IS200 a IS110). Pro IS200 a IS110 je typický vyšší počet kopií. Geny pro *rayt* se v genomu téměř nevyskytují ve více kopiích (obrázek 15B). Další rozdíl mezi transpozázami je v počtu genů stejné skupiny přítomných v genomu, kdy *rayt* je dvakrát méně početný oproti IS200 (obrázek 15A). Horizontální přenos genů probíhá pomocí plazmidů, jde tak o ukazatel frekvence, s jakou se daná rodina transpozáz mezi genomy šíří (Bichsel, Barbour and Wagner, 2013). Minimální

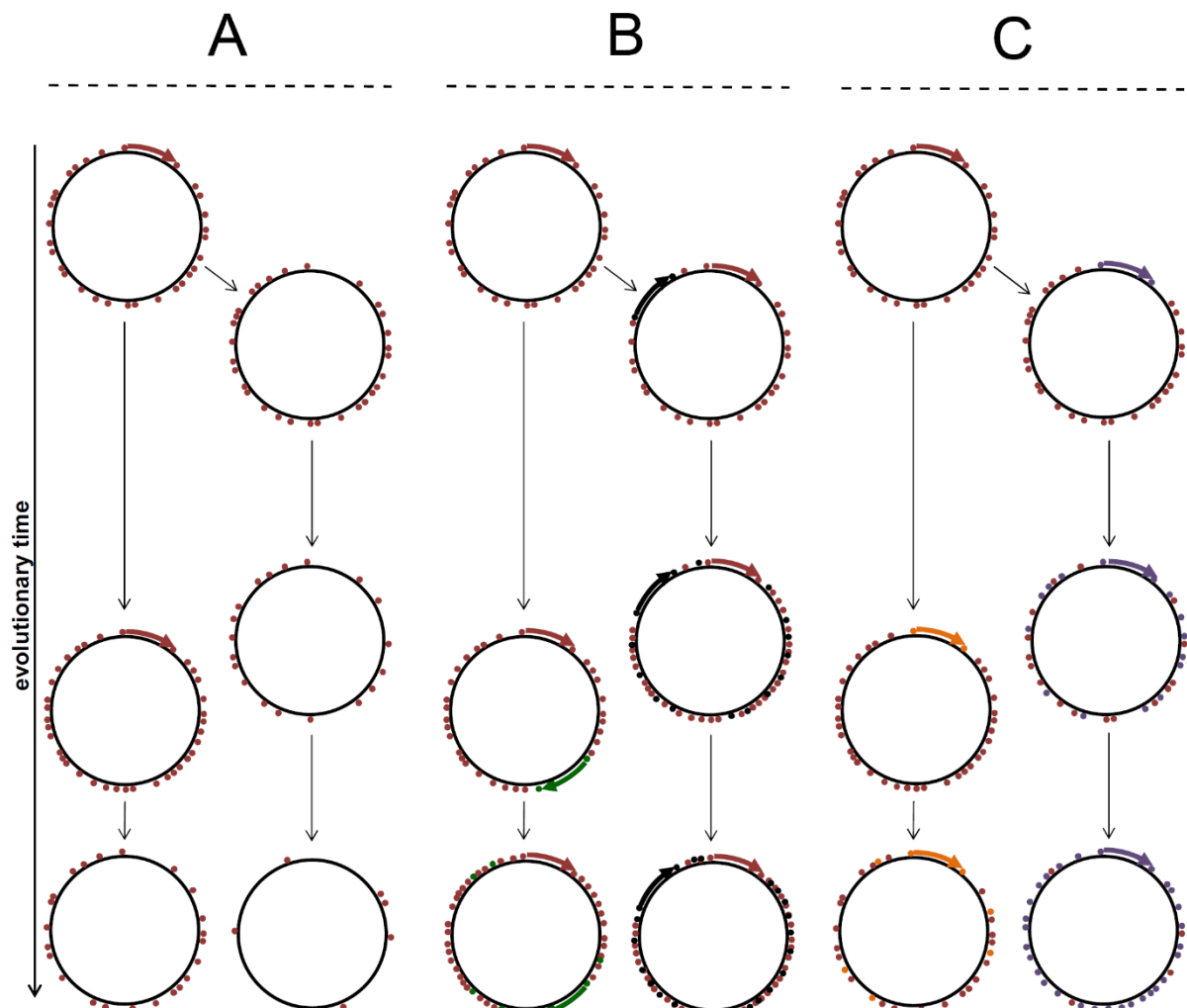
množství *rayt* genů na plazmidech také ukazuje na nízkou frekvenci horizontálního přenosu (obrázek 15C). RAYT, IS200 a IS110 mají geny obklopeny konzervovanými sekvencemi. V případě RAYT jsou těmito konzervovanými sekvencemi REP elementy (nejčastěji REPIN), které jsou četnější než terminální sekvence IS elementů, což ukazuje na jejich šíření bez nutnosti mobilizace samotného *rayt* genu (obrázek 15D).



Obrázek 15: Porovnání vlastností genů transpozáz IS elementů (červeně), RAYT (zeleně) a housekeeping genů (modře). A) Počet genů dané rodiny na jeden replikon. B) Frekvence intragenomových dupli/multiplikátů (více kopií). C) Frekvence výskytu na plazmidech. D) Frekvence nejčastějších 16 bází dlouhých sekvencí v okolí genu v hostitelských genomech. Převzato z (Bertels, Gallie and Rainey, 2017).

V genomu hostitelské bakterie může být přítomno více *rayt* genů, nejvýše byly nalezeny čtyři. V takovém případě se typicky dají REP elementy přítomné na okrajích *rayt* genů porovnat s REP elementy rozšířenými po genomu. Lze tak ke každému *rayt* genu přiřadit jednu třídu REP elementů, kterou pravděpodobně rozšířil genomem. Vyskytují se případy, kdy se třídu REP elementů přiřadit k *rayt* genu nepodaří. Tyto případy se většinou dají vysvětlit ztrátou *rayt* genu (obrázek 16A), kdy u příbuzných druhů *rayt* gen zachován je. Existují ovšem případy, kdy v genomu je přítomen jeden *rayt* gen a více tříd REP elementů, přičemž z analýzy příbuzných druhů se nezdá, že by došlo ke ztrátě *rayt* genu. V této situaci pravděpodobně došlo k takzvanému orthoswitchi (anglicky „Orthologous REP switch“) (Nunvar, Licha and Schneider, 2013). Jde o mechanismus, kdy dojde k mutaci REP v okolí *rayt* genu, čímž vznikne nová třída REP elementů, kterou tento RAYT začne propagovat genomem (obrázek 16C). Jde o jeden z mechanismů, který zvyšuje variabilitu REP elementů, dalším je duplikace

a přesun *rayt* genu do nové genomické lokace, následované diverzifikací a proliferací s ním asociovaných REP elementů (obrázek 16B).



Obrázek 16: Model dlouhodobé evoluce REP elementů u pseudomonád a stentrophomonád. Černý kruh znázorňuje genom, barevné šipky představují *rayt* gen, REP elementy asociované s daným *rayt* genem jsou znázorněny tečkami stejné barvy. Levá část každého sloupce (A, B, C) je model, kdy k evoluční události došlo nedávno. Pravá část modeluje, když ke změně došlo dávno v evoluci. A) Ztráta RAYT následována postupnou ztrátou REP elementů vlivem mutací a/nebo absencí proliferace REP. B) Duplikace RAYT a vznik dvou rozdílných RAYT. Časem dojde k rozšíření nové skupiny REP elementů, které propaguje nově vzniklý RAYT. Dojde k zvýšení celkové populace REP elementů v genomu. C) REP orthoswitch následovaný propagací nové skupiny REP a degradací té původní. Převzato z (Nunvar, Licha and Schneider, 2013)

## 5.2. Evoluce RAYT z IS200/IS605 transponáz

Nabízí se otázka, jak v evoluci vznikl vztah mezi REPIN a RAYT. Jedním z nejjednodušších vysvětlení je, že v evolučně dávne historii na původním elementu podobném IS200 začaly parazitovat jeho terminální sekvence (REP/REPIN), jako elementy neschopné vlastní transpozice začaly využívat IS200 transpozázu. Tím také způsobily selekční tlak, který nakonec vedl v evoluci k divergenci původní transpozázy od ostatních IS200, a následnému vzniku RAYT (Bertels and Rainey, 2011b) a jeho charakteristických vlastností (viz kapitola 5.1.3.)

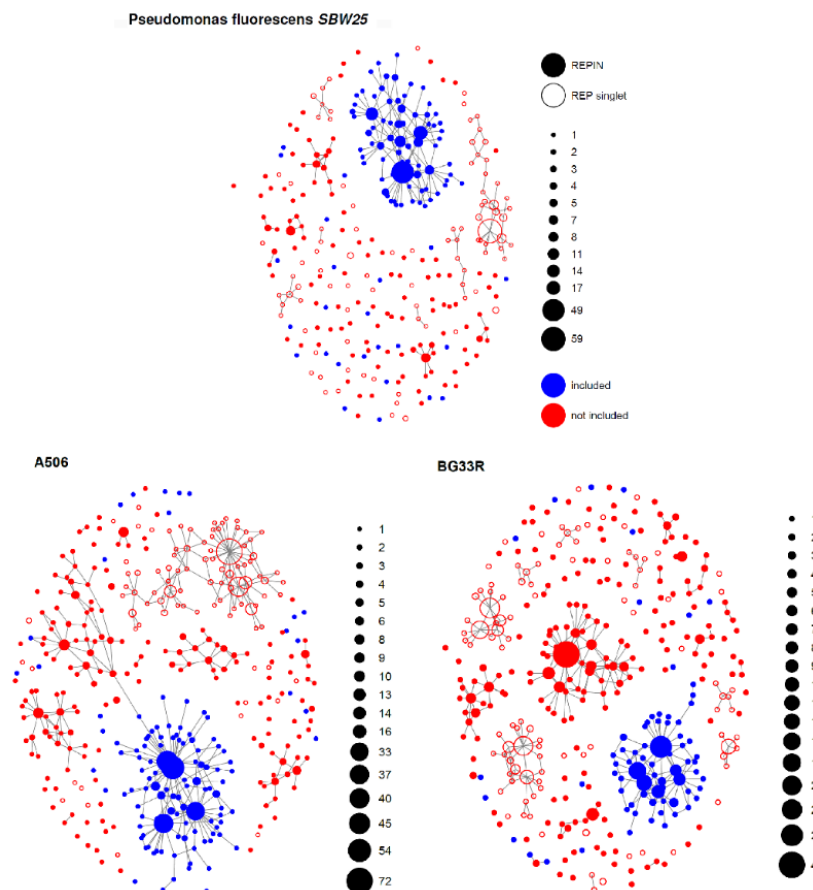
Geny kódující RAYT se vyskytují téměř výlučně v jedné kopii na genom (v případě, že jich je v genomu více, je každý zřetelně odlišný a je asociován s jinou třídou REP/REPIN). Tím se RAYT odlišují od transpozáz klasických IS elementů. Nepřítomnost většího počtu kopií znamená, že pokud dojde k poškození *rayt* genu, RAYT proteiny se již v buňce nebudou moci vyrábět. Taková „křehkost“ je pro transpozázu velice nezvyklá. Nicméně, přítomnost RAYT v genomu hostitelské bakterie je v evolučním čase pozoruhodně stabilní (Nunvar, Licha and Schneider, 2013; Park, Gokhale and Bertels, 2021). Nezdá se, že by taková transpozáza sama o sobě měla velkou šanci dlouhodobě v genomu prosperovat, pokud by její přítomnost nebyla pro hostitelskou buňku prospěšná. Propagace REPIN genomem je pravděpodobným selekčním faktorem zodpovědný za evoluční perzistenci RAYT. Vzniká vztah dlouhodobé koexistence, kdy se buňkám vyplatí do jisté míry tolerovat propagaci těchto elementů skrze svoje genomy. Aktivita RAYT by pak mohla být evolučně harmonizována na optimální úroveň (viz kapitola 5.1.1. a 5.1.2.). Buňka by se tak chránila před příliš masivní propagací REPIN elementů, ale zároveň využívala benefity, které jejich přítomnost přináší.

### 5.3. Populace REPIN elementů

REP elementy jsou šířeny genomem jako součást REPIN. Ty jsou pravděpodobně propagovány pomocí interakce s RAYT. *rayt* gen je v genomu obklopen z obou stran REPIN elementy, právě ty pak pravděpodobně šíří. Analýzy distribuce REP elementů v genomech bakterií tuto teorii podporují. REP, které jsou součástí REPIN, bývají svou sekvencí podstatně více konzervovány než REP nacházející se samostatně. Navíc, párové REP elementy typu REPIN jsou predominantní formou výskytu REP v genomu, zatímco podle náhodného modelu by se REP elementy měly vyskytovat převážně samostatně (Bertels and Rainey, 2011b). Tento jev může mít dvě příčiny. Pravděpodobnou možností je, že se genomem šíří REPIN, ty se občas rozpadnou a zůstanou jako samotné REP elementy. Druhou možností je, že se REP a REPIN liší svou schopností udržet se v genomu, tedy že samotný REP element rychleji podléhá mutačním procesům. Zdá se, že dochází ke kombinaci obou příčin. REPIN se vyznačují schopností odolat mutačnímu tahu, jsou pod purifikující selekcí (z genomových analýz se dá vysledovat tendence REPIN, udržet si konsensus sekvenci, jak znázorňuje obrázek 17).

V sekvenci REP elementu může dojít vlivem mutace ke změně jednoho nukleotidu, tím dojde ke ztrátě komplementarity části sekvence ve vlásenkové struktuře. Časem by tak mělo docházet k postupné degradaci REP, nicméně se zdá, že k tomu příliš nedochází. Mutace v jedné polovině palindromu je často následována mutací na druhé straně palindromu, aby báze ve vlásence opět párovaly. Bylo zjištěno, že frekvence kompenzačních mutací u REP elementů je řádově vyšší, než by mělo vznikat bez selekčního tlaku. Je tak zřejmý selekční tlak na udržení vnitřní symetrie elementů (Aranda-Olmedo *et al.*, 2002). Tlak na sekvenci REPIN elementů, udržovat své REP v rozmezí jisté

odchylky od konsensus sekvence se dá graficky znázornit (obrázek 17). V tomto případě se jedná o populační mapu REPIN v genomech tří kmenů *Pseudomonas fluorescens* (SBW25, BG33R a A50). Čarou jsou spojeny pouze ty populace REPIN, které se od sebe liší právě jednou pozicí. SBW25 má v genomu velkou populaci REP elementů, která čítá asi tisíc elementů. SBW25 má také tři *rayt* geny (a tedy tři třídy REP elementů), dá se tak předpokládat, že v genomu by se měly šířit tři populace rozdílných REPIN. Struktura REP populace ukazuje, že minimálně jedna skupina elementů je udržována kolem „master sekvence“. Vzhledem k přítomnosti tří aktivních RAYT transpozáz lze spekulovat, že část transpozáz má výrazně sníženou aktivitu. Zbývající dva kmeny *P. fluorescens* mají každý jeden *rayt* gen, to odpovídá jedné skupině REPIN elementů udržujících se kolem master sekvence. A506 má jednu a BG33R dvě skupiny REPIN. Menší skupina u BG33R (obrázek 17, červeně) není asociovaná s *rayt*, je tedy nejasné, jakým mechanismem se udržuje v genomu. Z možných evolučních scénářů se jako nejpravděpodobnější jeví recentní ztráta příslušného *rayt* genu (obrázek 16A). Znalost evolučních vztahů bakteriálních hostitelů a evoluční „pattern“ přítomnosti/absence jednotlivých RAYT je nezbytná pro interpretaci dat z analýzy vnitrobuněčných populací REPIN.

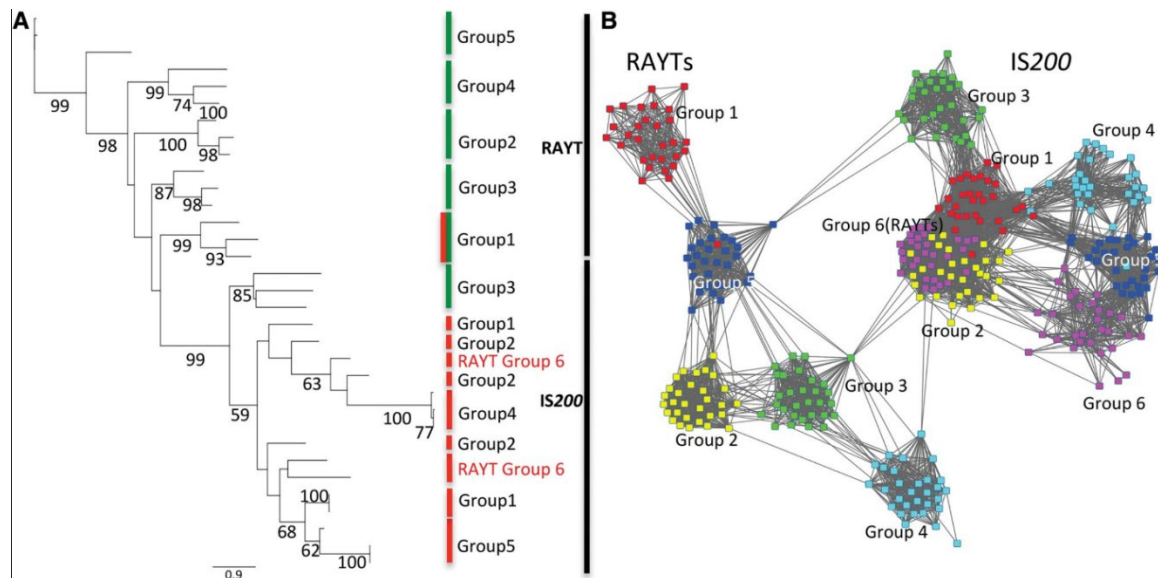


Obrázek 17: Struktura populace REPIN elementů v genomu *Pseudomonas fluorescens*, kmenech SBW25, A506 a BG33R. Každá tečka značí jednu sekvenci REPIN elementu. Průměr tečky značí počet kopií této sekvence v genomu. Mapy se liší měřítkem velikosti skupiny stejných REPIN. Plné tečky jsou REP elementy tvořící REPIN, prázdné tečky jsou vyskytující se samostatně. Čarou jsou spojeny populace elementů lišící se jednou pozicí v sekvenci. Modrou jsou zvýrazněny REP, které svou sekvenci udržují kolem master sekvence, s maximálně třemi mutovanými pozicemi. Červenou jsou zvýrazněny REP více rozdílné od master sekvence. Převzato z (Bertels, Gokhale and Traulsen, 2017).

## 5.4. Rozdělení RAYT proteinů

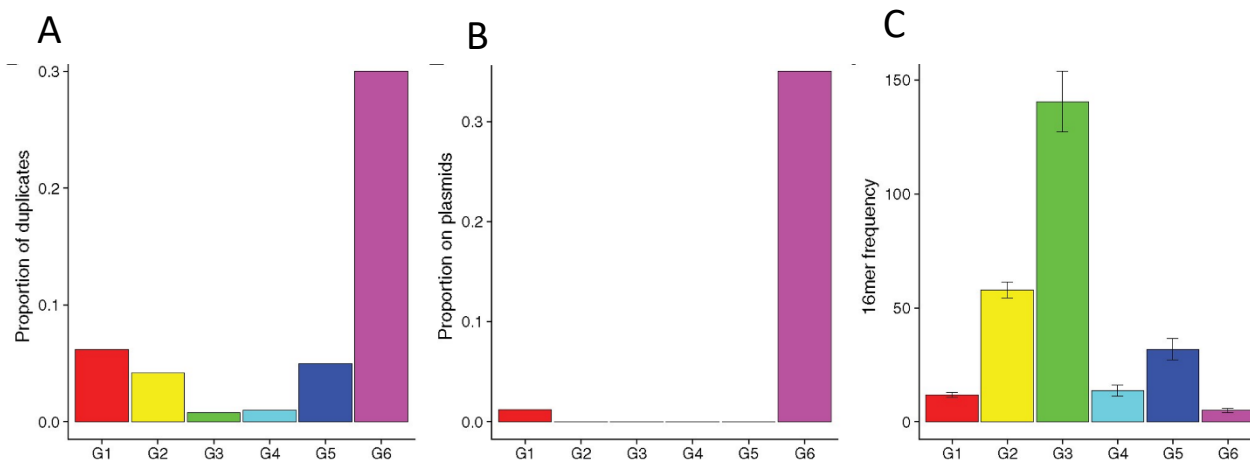
RAYT proteiny jsou velmi diverzifikované a vyskytují se v řadě rozdílných bakterií, z nichž každý druh má charakteristické RAYT. Sekvence přidružených REP se liší, stejně jako jejich počty. Samotný *rayt* gen bývá obklopen více REP, které tvoří REPIN či BIME mozaiky. Některé genomy jako *P. fluorescens* SBW25 obsahují přes tisíc REP elementů a stovky REPIN (Bertels and Rainey, 2011b). *E. coli* MG1655 má oproti tomu několik stovek REP elementů a řadu z nich ve velice konzervovaných BIME elementech. Snaha rozdělit RAYT do skupin a ty použít jako detailnější optiku pro další výzkumu může pomoci odhalit dosud neznámé vlastnosti těchto proteinů. "

RAYT proteiny, jak ukazuje obrázek 18, byly rozděleny do charakteristických skupin (podtříd) na základě podobnosti aminokyselinové sekvence, nejprve do šesti (G1 – G6) (Bertels, Gallie and Rainey, 2017), posléze do osmi skupin, respektive podtříd (anglicky subclass) (Quentin *et al.*, 2018).



Obrázek 18: A) Fylogenetický strom RAYT (zelená svislá čára) a IS200/IS605 transpozáz. B) Klastrovací analýza příbuznosti zástupců jednotlivých skupin analyzovaných RAYT a IS200/IS605 transpozáz. Převzato z (Bertels, Gallie and Rainey, 2017).

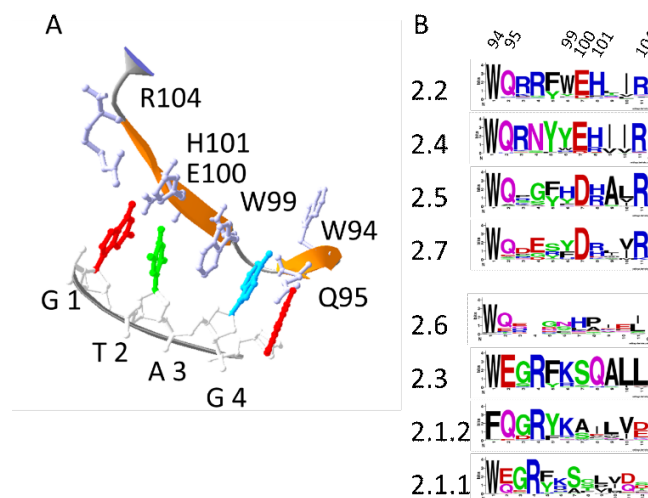
Skupiny jsou číslovány podle velikosti. Skupina G1 (podtřída 2.1) má přes 400 členů, nejmenší skupina G6 pak méně jak 30 (Bertels, Gallie and Rainey, 2017). Jak ukazuje obrázek 19, skupina G6 má krom malého počtu členů značný počet genových duplikátů a nízkou asociaci s REP. V kontextu dřívějších výzkumů, do skupiny G2 (podtřída 2.2) patří RAYT z *E. coli*, který byl použit pro experimentální důkaz REP-RAYT interakce (příloha 2) (Messing *et al.*, 2012; Ton-Hoang *et al.*, 2012). Většina skupin je rozšířených skrze více bakteriální tříd, to platí zvláště pro podtřídu G1, která byla nalezena v 21 různých bakteriálních třídách a ve 150 druzích, což vzhledem k její velikosti není tak překvapivé. V kontrastu, skupina G3 byla nalezena jen v 61 druzích, tyto druhy ale pokrývají 18 bakteriálních tříd.



Obrázek 19: Podobná data, jako byla na obrázku 15. Zde se hodnotí rozdíly mezi jednotlivými skupinami RAYT. A) Množství RAYT genů, které jsou replikované (více identických kopií) v hostitelských genomech B) Tento graf zobrazuje, jak velká část skupiny byla nalezena na plasmidech. C) Frekvence nejčastějších 16 nukleotidů-dlouhých sekvencí z okolí RAYT genu v hostitelském genomu. Převzato z (Bertels, Gallie and Rainey, 2017).

#### 5.4.1. Podtřída 2.1 (G1)

Jde o největší skupinu RAYT, původně s 292 členy (Bertels, Gokhale and Traulsen, 2017). Později byla rozšířena a nyní je do ní řazeno 493 RAYT (Quentin *et al.*, 2018), což tvoří 61% všech identifikovaných *rayt* genů. Podtřída 2.1 má v průměru delší sekvenci než většina ostatních podtříd. Důvodem je C-terminální doména. Podlé té se také podtřída 2.1 dá dále rozdělit na dvě skupiny, 2.1.1 a 2.1.2. Podtřída 2.1.1 má C-terminální doménu kratší, ale více homogenní v porovnání s 2.1.2. Zdá se, že u obou skupin tato doména vytváří sekundární strukturu helix-turn-helix. Není jasné, jakou má funkci. Jako jediné se RAYT podtřídě 2.1 vyskytují na plasmidech, jde o asi 1% nalezených členů (Bertels, Gallie and Rainey, 2017). Lze se domnívat, že i v současnosti může docházet k horizontálnímu přenosu této skupiny a jejímu šíření do dalších druhů. Také jsou do této podtřídě zahrnuty TIRYT transpozázy (Di Nocera, De Gregorio and Rocco, 2013). *rayt* geny nejsou ohraničeny REP elementy a jimi kódované proteiny postrádají konzervované AK zbytky, které vážou GTAG (Obrázek 20).



Obrázek 20: A) interakce RAYT s GTAG tetranukleotidem u *E. coli* (viz kapitola 5.1.2.). B) AK na pozicích klíčových pro interakci s GTAG u jednotlivých podtříd RAYT. Je patrná různá konzervovanost triády 99/100/101 (aromatická/kyselá/zásaditá AK) a R104. Převzato z (Quentin *et al.*, 2018).

#### 5.4.2. Podtřída 2.2 (G2)

Tato podtřída se vyskytuje v *E. coli* a řadě dalších proteobakteriích. *rayt* geny jsou ohraničeny REP elementy a GTAG-vazebné AK jsou přítomny. Obsahuje první objevené REP (Higgins *et al.*, 1982) a zároveň i první identifikovaný *rayt* gen (Nunvar, Huckova and Licha, 2010). Struktura REP elementů této skupiny je ukázkou typického elementu, vyobrazeno v kapitole 3 (obrázek 3). Pro tuto podtřidu REP je typická přítomnost nepárujících nukleotidů uvnitř vlásenky. V souvislosti s tím má RAYT této podtřídy dvě krátké inserce vedle oblasti, která rozeznává REP vlásenku (Quentin *et al.*, 2018).

#### 5.4.3. Podtřída 2.5 (G3)

Do této podtřídy jsou řazeny RAYT nalezené v prvních studiích na pseudomonádách (Feil *et al.*, 2005; Tobes and Pareja, 2005; Rocco, De Gregorio and Di Nocera, 2010; Bertels and Rainey, 2011b). Kromě pseudomonád jsou tyto RAYT rozšířené i v xanthomonádách a stenotrofomonádách. Jeden genom může obsahovat více genů této podtřídy. *rayt* geny jsou ohraničeny REP elementy a GTAG-vazebné AK jsou přítomny. V současnosti pravděpodobně stále dochází k amplifikaci a diverzifikaci těchto elementů, což se projevuje vysokým počtem REPIN elementů v genomu. Samotné REP jsou kratší v porovnání s podtřídou 2.2 (ukazuje kapitola 3, obrázek 3). Ramena palindromu párují v celé své délce.

#### 5.4.4. Podtřída G6

Nejmenší skupina RAYT je G6. Od ostatních se liší (obrázek 19) velmi výrazně. Jde o skupinu se zdaleka nejvyšším intragenových replikátů (30%), G6 má také 35% členů na plasmidech. Pro srovnání, jediná další skupina vyskytující se na plasmidech je G1, a to pouze 1,2% členů. Skupina G6 není obklopena REP či REPIN elementy. V těchto statistikách se G6 značně odlišuje od ostatních RAYT a je více podobná transpozázám IS110 a IS200 než housekeeping genům. K těmto tvrzením přispívá i proteinová sekvenční síť (obrázek 18), která jasně zařazuje G6 RAYT k IS200 transpozázám. Vzhledem k předpokládané evoluci RAYT z IS200/IS605 rodiny transpozáz je možné, že RAYT skupiny G6 jsou v porovnání s ostatními RAYT mladší skupina, které ještě není buňkami plně domestikována. Funkčně se tak nemusí jednat o RAYT a při analýzách RAYT může být prospěšné tuto skupinu vyřazovat a zkoumat samostatně.

## 6. Závěr

Bakterie jsou lidským okem neviditelné organismy, a přesto jedna z největších biologických sil formujících svět. Prakticky všechny vyšší organismy, včetně člověka, sdílejí svá těla s těmito mikroorganismy. Znalost principů fungování bakteriální buňky, kromě porozumění jejich životnímu stylu, má velký potenciál pro cílené modifikace bakteriálních procesů.

Jednou ze složek bakteriálního genomu jsou repetitivní sekvence DNA, k nimž se řadí mobilní genetické elementy. Složka genomu do jisté míry samostatná, schopná vlastního přenosu a případně množení. Jednou ze tříd těchto elementů jsou REP. Jde o konzervované palindromické sekvence přítomné v řadě bakteriálních tříd. Díky vlastnostem palindromu mohou vytvářet sekundární vlásečkové struktury (jako ssDNA či ssRNA). V genomu mohou být stovky i tisíce REP elementů. Často jsou umístěny společně a vytváří vyšší strukturní elementy, jako REPIN či BIME. Právě ty jsou skrze genom pravděpodobně rozšiřovány. Klíčovým enzymem je RAYT. Jedná se o enzym příbuzný IS200/IS605 transpozázám, obsahující HUH motiv a jeden katalytický tyrosin. *rayt* je v genomu obklopen shlukem REP elementů (vytvářející BIME či REPIN struktury). Tato oblast, včetně *rayt* genu, se označuje souhrnně jako REPtron. Právě REP elementy, které jsou součástí REPtron, jsou pravděpodobně cílem RAYT. Přesný mechanismus zatím není znám, nicméně tyto REPIN/BIME jsou rozeznány RAYT, pomocí konzervovaného tetranukleotidu (podobně jako u IS200/IS605). Následně jsou rozšířeny skrze genom. Na rozdíl od většiny transpozáz, se RAYT jeví jako buňkou zcela domestikované. Nejsou schopny se samy genomem propagovat a jen minimálně využívají horizontálního přenosu mezi genomy. RAYT a s nimi asociované REP jsou značně variabilní. Dají se rozdělit do skupin v závislosti na struktuře RAYT proteinu, délce REP a vlastnostech vlásečky.

RAYT-REP interakce je systém prospěšný pro buňku. REP elementy umístěné na 3' koncích operonů mohou efektivně snížit genovou expresi blokáci translace. REP elementy jsou také často využity jako cíle pro specifickou transpozici jiných inserčních sekvencí. Díky vysokému počtu kopií a umístění mimo čtecí rámce jsou ideálními oblastmi pro inzerci. Při takové transpozici může dojít k reorganizaci genomu. Krom mechanismů, pro které jsou REP využívány ve všech hostitelských genomech, se vyvinuly mechanismy využívány jen u některých druhů. Příkladem je vazba IHF proteinu na RIB elementy (BIME se specifickou sekvencí uprostřed) u *E. coli* nebo interakce s DNA gyrázou, pozorovaná u BIME-2 elementů *E. coli*. Nevíme, nakolik se RAYT a REP podílejí na dalších procesech hostitelských buněk – v tomto směru lze očekávat nové objevy. REP elementy lze využít k rychlé identifikaci bakteriálních kmenů a jako nástroj genetického inženýrství, například pro úpravy genové exprese. I vzhledem k jejich hojnému rozšíření věřím, že jejich studium je perspektivní oblastí dalšího výzkumu.

## 7. Seznam použité literatury

- Aranda-Olmedo, I. *et al.* (2002) 'Species-specific repetitive extragenic palindromic (REP) sequences in *Pseudomonas putida*', *Nucleic Acids Research*, 30(8). doi: 10.1093/nar/30.8.1826.
- Bachellier, S. *et al.* (1993) 'Bacterial interspersed mosaic elements (BIMEs) are present in the genome of *Klebsiella*', *Molecular Microbiology*, 7(4). doi: 10.1111/j.1365-2958.1993.tb01144.x.
- Bachellier, S. *et al.* (1994) 'Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs)', *Molecular Microbiology*. doi: 10.1111/j.1365-2958.1994.tb00995.x.
- Bachellier, S. *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).
- Bachellier, S., Clément, J. M. and Hofnung, M. (1999) 'Short palindromic repetitive DNA elements in enterobacteria: A survey', *Research in Microbiology*, 150(9–10), pp. 627–639. doi: 10.1016/S0923-2508(99)00128-X.
- Bao, W. and Jurka, J. (2013) 'Homologues of bacterial TnpB-IS605 are widespread in diverse eukaryotic transposable elements', *Mobile DNA*, 4(1). doi: 10.1186/1759-8753-4-12.
- Becerril, B. *et al.* (1985) 'Repetitive extragenic palindromic (REP) sequences in the *Escherichia coli* *gdhA* gene', *Gene*, 37(1–3), pp. 53–62. doi: 10.1016/0378-1119(85)90257-4.
- Bertels, F., Gallie, J. and Rainey, P. B. P. B. (2017) 'Identification and characterization of domesticated bacterial transposases', *Genome Biology and Evolution*, 9(8), pp. 2110–2121. doi: 10.1093/gbe/evx146.
- Bertels, F., Gokhale, C. S. and Traulsen, A. (2017) 'Discovering complete quasispecies in bacterial genomes', *Genetics*. doi: 10.1534/genetics.117.201160.
- Bertels, F. and Rainey, P. B. (2011a) 'Curiosities of REPINs and RAYTs', *Mobile Genetic Elements*, 1(4). doi: 10.4161/mge.18610.
- Bertels, F. and Rainey, P. B. (2011b) 'Within-Genome evolution of REPINs: A new family of miniature mobile DNA in bacteria', *PLoS Genetics*, 7(6). doi: 10.1371/journal.pgen.1002132.
- Bichsel, M., Barbour, A. D. and Wagner, A. (2013) 'Estimating the fitness effect of an insertion sequence', *Journal of Mathematical Biology*, 66(1–2). doi: 10.1007/s00285-012-0504-2.
- Boccard, F. and Prentki, P. (1993) 'Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3' end of transcription units', *EMBO Journal*, 12(13), pp. 5019–5027. doi: 10.1002/j.1460-2075.1993.tb06195.x.
- De Bruijn, F. J. (1992) 'Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria', *Applied and Environmental Microbiology*, 58(7). doi: 10.1128/aem.58.7.2180-2187.1992.
- Charlier, D. *et al.* (1995) 'Pyrimidine regulation of the *Escherichia coli* and *Salmonella typhimurium* *carP* operons: *CarP* and integration host factor (IHF) modulate the methylation status of a GATC site present in the control region', *Journal of Molecular Biology*. doi: 10.1006/jmbi.1995.0384.
- Cheng, Z. F. and Deutscher, M. P. (2005) 'An important role for RNase R in mRNA decay', *Molecular Cell*, 17(2). doi: 10.1016/j.molcel.2004.11.048.

- Choi, S., Ohta, S. and Ohtsubo, E. (2003) 'A novel is element, IS621, of the IS110/IS492 family transposes to a specific site in repetitive extragenic palindromic sequences in *Escherichia coli*', *Journal of Bacteriology*, 185(16). doi: 10.1128/JB.185.16.4891-4900.2003.
- Clément, J.-M. *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', *Journal of Bacteriology*, 181(22), pp. 6929–6936. doi: 10.1128/jb.181.22.6929-6936.1999.
- Curcio, M. J. and Derbyshire, K. M. (2003) 'The outs and ins of transposition: From MU to kangaroo', *Nature Reviews Molecular Cell Biology*. doi: 10.1038/nrm1241.
- Deng, C. *et al.* (2019) 'Synthetic repetitive extragenic palindromic (REP) sequence as an efficient mRNA stabilizer for protein production and metabolic engineering in prokaryotic cells', *Biotechnology and Bioengineering*, 116(1). doi: 10.1002/bit.26841.
- Diaz-Aroca, E. *et al.* (1984) 'Characterization of the new insertion sequence IS91 from an alpha-hemolysin plasmid of *Escherichia coli*.', *Molecular & general genetics : MGG*, 193(3), pp. 493–499. doi: 10.1007/BF00382089.
- Dimri, G. P. *et al.* (1992) 'Physical mapping of repetitive extragenic palindromic sequences in *Escherichia coli* and phylogenetic distribution among *Escherichia coli* strains and other enteric bacteria', *Journal of Bacteriology*, 174(14). doi: 10.1128/jb.174.14.4583-4593.1992.
- Doolittle, W. F. and Sapienza, C. (1980) 'Selfish genes, the phenotype paradigm and genome evolution', *Nature*, 284(5757). doi: 10.1038/284601a0.
- Dorman, C. J. and Deighan, P. (2003) 'Regulation of gene expression by histone-like proteins in bacteria', *Current Opinion in Genetics and Development*. doi: 10.1016/S0959-437X(03)00025-X.
- Espéli, O. and Boccard, F. (1997) 'In vivo cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: Genetic characterization of the target and identification of the cut site', *Molecular Microbiology*. doi: 10.1046/j.1365-2958.1997.6121983.x.
- Espéli, O., Moulin, L. and Boccard, F. (2001) 'Transcription attenuation associated with bacterial repetitive extragenic BIME elements', *Journal of Molecular Biology*, 314(3). doi: 10.1006/jmbi.2001.5150.
- Feil, H. *et al.* (2005) 'Comparison of the complete genome sequences of *Pseudomonas syringae* pv. *syringae* B728a and pv. *tomato* DC3000', *Proceedings of the National Academy of Sciences of the United States of America*, 102(31). doi: 10.1073/pnas.0504930102.
- Friedman, D. I. (1988) 'Integration host factor: A protein for all reasons', *Cell*. doi: 10.1016/0092-8674(88)90213-9.
- Garcillán-Barcia, M. P. and De La Cruz, F. (2002) 'Distribution of IS91 family insertion sequences in bacterial genomes: Evolutionary implications', *FEMS Microbiology Ecology*, 42(2). doi: 10.1016/S0168-6496(02)00340-9.
- Gellert, M. *et al.* (1976) 'DNA gyrase: an enzyme that introduces superhelical turns into DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 73(11). doi: 10.1073/pnas.73.11.3872.
- Gilson, E. *et al.* (1984) 'A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*.', *The EMBO journal*, 3(6). doi: 10.1002/j.1460-2075.1984.tb01986.x.
- Gilson, E. *et al.* (1986) 'A subfamily of *E. coli* palindromic units implicated in transcription termination?', *Annales de l'Institut Pasteur Microbiology*, 137(1). doi: 10.1016/S0769-2609(86)80116-8.

- Gilson, E. *et al.* (1991a) 'Palindromic units are part of a new bacterial interspersed mosaic element (BIME)', *Nucleic Acids Research*, 19(7), pp. 1375–1383. doi: 10.1093/nar/19.7.1375.
- Gilson, E. *et al.* (1991b) 'The BIME family of bacterial highly repetitive sequences', *Research in Microbiology*, 142(2–3). doi: 10.1016/0923-2508(91)90033-7.
- Goosen, N. and van de Putte, P. (1995) 'The regulation of transcription initiation by integration host factor', *Molecular Microbiology*. doi: 10.1111/j.1365-2958.1995.tb02386.x.
- He, S. *et al.* (2013) 'IS200/IS605 family single-strand transposition: Mechanism of IS608 strand transfer', *Nucleic Acids Research*, 41(5), pp. 3302–3313. doi: 10.1093/nar/gkt014.
- He, S. *et al.* (2015) 'The IS200/IS605 family and "peel and paste" single-strand transposition mechanism', *Microbiology Spectrum*, 3(4). doi: 10.1128/microbiolspec.MDNA3-0039-2014.
- Hickman, A. B. *et al.* (2002) 'Structural unity among viral origin binding proteins: Crystal structure of the nuclease domain of adeno-associated virus Rep', *Molecular Cell*, 10(2). doi: 10.1016/S1097-2765(02)00592-0.
- Hickman, A. B., Chandler, M. and Dyda, F. (2010) 'Integrating prokaryotes and eukaryotes: DNA transposases in light of structure', *Critical Reviews in Biochemistry and Molecular Biology*. doi: 10.3109/10409230903505596.
- Higgins, C. F. *et al.* (1982) 'A novel intercistronic regulatory element of prokaryotic operons', *Nature*. doi: 10.1038/298760a0.
- Ichikawa, H. *et al.* (1987) 'Specific binding of transposase to terminal inverted repeats of transposable element Tn3.', *Proceedings of the National Academy of Sciences of the United States of America*, 84(23). doi: 10.1073/pnas.84.23.8220.
- Iggo, R. *et al.* (1990) 'Identification of a putative RNA helicase in *E. coli*', *Nucleic Acids Research*, 18(18). doi: 10.1093/nar/18.18.5413.
- Ilyina, T. V. and Koonin, E. V. (1992) 'Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaeobacteria', *Nucleic Acids Research*, 20(13). doi: 10.1093/nar/20.13.3279.
- Kersulyte, D. *et al.* (2002) 'Transposable element ISHp608 of *Helicobacter priori*: Nonrandom geographic distribution, functional organization, and insertion specificity', *Journal of Bacteriology*, 184(4). doi: 10.1128/jb.184.4.992-1002.2002.
- Khemici, V. and Carpousis, A. J. (2004) 'The RNA degradosome and poly(A) polymerase of *Escherichia coli* are required in vivo for the degradation of small mRNA decay intermediates containing REP-stabilizers', *Molecular Microbiology*, 51(3). doi: 10.1046/j.1365-2958.2003.03862.x.
- Kikuchi, Y. and Nash, H. A. (1978) 'The bacteriophage lambda *int* gene product. A filter assay for genetic recombination, purification of *int*, and specific binding to DNA.', *Journal of Biological Chemistry*, 253(20), pp. 7149–7157. doi: [https://doi.org/10.1016/S0021-9258\(17\)34477-0](https://doi.org/10.1016/S0021-9258(17)34477-0).
- Koonin, E. V. and Ilyina, T. V. (1993) 'Computer-assisted dissection of rolling circle DNA replication', *BioSystems*, 30(1–3). doi: 10.1016/0303-2647(93)90074-M.
- Lam, S. and Roth, J. R. (1983) 'IS200: A *salmonella*-specific insertion sequence', *Cell*, 34(3). doi: 10.1016/0092-8674(83)90552-4.
- Lee, H. *et al.* (2016) 'Insertion sequence-caused large-scale rearrangements in the genome of *Escherichia coli*', *Nucleic Acids Research*, 44(15). doi: 10.1093/nar/gkw647.
- Liang, W. and Deutscher, M. P. (2016) 'REP sequences: Mediators of the environmental stress

response?', *RNA Biology*, 13(2). doi: 10.1080/15476286.2015.1112489.

Liang, W., Rudd, K. E. K. E. and Deutscher, M. P. M. P. (2015) 'A Role for REP sequences in regulating translation', *Molecular Cell*, 58(3), pp. 431–439. doi: 10.1016/j.molcel.2015.03.019.

Loper, J. E. *et al.* (2012) 'Comparative genomics of plant-associated *pseudomonas spp.*: Insights into diversity and inheritance of traits involved in multitrophic interactions', *PLoS Genetics*, 8(7). doi: 10.1371/journal.pgen.1002784.

Mahillon, J., Léonard, C. and Chandler, M. (1999) 'IS elements as constituents of bacterial genomes', *Research in Microbiology*, 150(9–10). doi: 10.1016/S0923-2508(99)00124-2.

McCutcheon, J. P. and Moran, N. A. (2012) 'Extreme genome reduction in symbiotic bacteria', *Nature Reviews Microbiology*. doi: 10.1038/nrmicro2670.

McLaren, R. S. *et al.* (1991) 'mRNA degradation by processive 3'-5' exoribonucleases in Vitro and the implications for prokaryotic mRNA decay in Vivo', *Journal of Molecular Biology*, 221(1). doi: 10.1016/0022-2836(91)80206-A.

Merino, E. *et al.* (1987) 'Deletion of a repetitive extragenic palindromic (REP) sequence downstream from the structural gene of *Escherichia coli* glutamate dehydrogenase affects the stability of its mRNA', *Gene*, 58(2–3), pp. 305–309. doi: 10.1016/0378-1119(87)90386-6.

Messing, S. A. J. S. A. J. *et al.* (2012) 'The processing of repetitive extragenic palindromes: The structure of a repetitive extragenic palindrome bound to its associated nuclease', *Nucleic Acids Research*, 40(19), pp. 9964–9979. doi: 10.1093/nar/gks741.

Murai, N. *et al.* (1995) 'A novel insertion sequence (IS)-like element of the thermophilic bacterium PS3 promotes expression of the alanine carrier protein-encoding gene', *Gene*, 163(1). doi: 10.1016/0378-1119(95)00384-I.

Nevers, P. and Saedler, H. (1977) 'Transposable genetic elements as agents of gene instability and chromosomal rearrangements', *Nature*. doi: 10.1038/268109a0.

Newbury, S. F. S. F. *et al.* (1987) 'Stabilization of translationally active mRNA by prokaryotic REP sequences', *Cell*, 48(2), pp. 297–310. doi: 10.1016/0092-8674(87)90433-8.

Di Nocera, P. P., De Gregorio, E. and Rocco, F. (2013) 'GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes', *BMC Genomics*, 14(1). doi: 10.1186/1471-2164-14-522.

Nolivos, S. *et al.* (2016) 'MatP regulates the coordinated action of topoisomerase IV and MukBEF in chromosome segregation', *Nature Communications*, 7. doi: 10.1038/ncomms10466.

Nunvar, J., Huckova, T. and Licha, I. (2010) 'Identification and characterization of repetitive extragenic palindromes (REP)-associated tyrosine transposases: Implications for REP evolution and dynamics in bacterial genomes', *BMC Genomics*, 11(1). doi: 10.1186/1471-2164-11-44.

Nunvar, J., Licha, I. and Schneider, B. (2013) 'Evolution of REP diversity: A comparative study', *BMC Genomics*, 14(1). doi: 10.1186/1471-2164-14-385.

Oppenheim, A. B. *et al.* (1993) 'Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli*', *Molecular Microbiology*, 10(1). doi: 10.1111/j.1365-2958.1993.tb00908.x.

Park, H. J., Gokhale, C. S. and Bertels, F. (2021) 'How sequence populations persist inside bacterial genomes', *Genetics*. doi: 10.1093/genetics/iyab027.

Pasternak, C. *et al.* (2013) 'ISDra2 transposition in *Deinococcus radiodurans* is downregulated by TnpB', *Molecular Microbiology*, 88(2). doi: 10.1111/mmi.12194.

- Qian, Z. *et al.* (2015) 'A new noncoding RNA arranges bacterial chromosome organization', *mBio*, 6(4). doi: 10.1128/mBio.00998-15.
- Quentin, Y. *et al.* (2018) 'Single-strand DNA processing: Phylogenomics and sequence diversity of a superfamily of potential prokaryotic HuH endonucleases', *BMC Genomics*, 19(1). doi: 10.1186/s12864-018-4836-1.
- Reece, R. J., Maxwell, A. and Wang, J. C. (1991) 'DNA gyrase: Structure and function', *Critical Reviews in Biochemistry and Molecular Biology*, 26(3–4). doi: 10.3109/10409239109114072.
- Rice, P. A. *et al.* (1996) 'Crystal structure of an IHF-DNA complex: A protein-induced DNA U-turn', *Cell*, 87(7). doi: 10.1016/S0092-8674(00)81824-3.
- Rice, P. A. (1997) 'Making DNA do a U-turn: IHF and related proteins', *Current Opinion in Structural Biology*, 7(1). doi: 10.1016/S0959-440X(97)80011-5.
- Rocco, F., De Gregorio, E. and Di Nocera, P. P. (2010) 'A giant family of short palindromic sequences in *Stenotrophomonas maltophilia*', *FEMS Microbiology Letters*, 308(2). doi: 10.1111/j.1574-6968.2010.02010.x.
- Rogozin, I. B. *et al.* (2002) 'Congruent evolution of different classes of non-coding DNA in prokaryotic genomes', *Nucleic Acids Research*. doi: 10.1093/nar/gkf549.
- Ronning, D. R. *et al.* (2005) 'Active site sharing and subterminal hairpin recognition in a new class of DNA transposases', *Molecular Cell*, 20(1), pp. 143–154. doi: 10.1016/j.molcel.2005.07.026.
- Ryan, V. T. *et al.* (2002) 'IHF and HU stimulate assembly of pre-replication complexes at *Escherichia coli* oriC by two different mechanisms', *Molecular Microbiology*, 46(1). doi: 10.1046/j.1365-2958.2002.03129.x.
- Sheehan, C. *et al.* (1995) 'Genomic fingerprinting *Acinetobacter baumannii*: amplification of multiple inter-repetitive extragenic palindromic sequences', *Journal of Hospital Infection*, 31(1). doi: 10.1016/0195-6701(95)90081-0.
- Shyamala, V., Schneider, E. and Ames, G. F.-L. F. L. (1990) 'Tandem chromosomal duplications: Role of REP sequences in the recombination event at the join-point', *EMBO Journal*, 9(3), pp. 939–946. doi: 10.1002/j.1460-2075.1990.tb08192.x.
- Silby, M. W. *et al.* (2009) 'Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*', *Genome Biology*, 10(5). doi: 10.1186/gb-2009-10-5-r51.
- Stern, M. J. *et al.* (1984) 'Repetitive extragenic palindromic sequences: A major component of the bacterial genome', *Cell*, 37(3). doi: 10.1016/0092-8674(84)90436-7.
- Sutormin, D. *et al.* (2019) 'Single-nucleotide-resolution mapping of DNA gyrase cleavage sites across the *Escherichia coli* genome', *Nucleic Acids Research*, 47(3). doi: 10.1093/nar/gky1222.
- Tobes, R. and Pareja, E. (2005) 'Repetitive extragenic palindromic sequences in the *Pseudomonas syringae* pv. *tomato* DC3000 genome: Extragenic signals for genome reannotation', *Research in Microbiology*, 156(3). doi: 10.1016/j.resmic.2004.10.014.
- Tobes, R. and Pareja, E. (2006) 'Bacterial repetitive extragenic palindromic sequences are DNA targets for Insertion Sequence elements', *BMC Genomics*, 7. doi: 10.1186/1471-2164-7-62.
- Tobes, R. and Ramos, J. L. (2005) 'REP code: Defining bacterial identity in extragenic space', *Environmental Microbiology*. doi: 10.1111/j.1462-2920.2004.00704.x.
- Ton-Hoang, B. *et al.* (2012) 'Structuring the bacterial genome: Y1-transposases associated with REP-BIME sequences', *Nucleic Acids Research*, 40(8), pp. 3596–3609. doi: 10.1093/nar/gkr1198.

- Treangen, T. J. T. J. *et al.* (2009) *Genesis, effects and fates of repeats in prokaryotic genomes*, *FEMS Microbiology Reviews*. doi: 10.1111/j.1574-6976.2009.00169.x.
- Tsui, P. and Freundlich, M. (1988) 'Integration host factor binds specifically to sites in the *ilvGMEDA* operon in *Escherichia coli*', *Journal of Molecular Biology*, 203(3). doi: 10.1016/0022-2836(88)90212-4.
- Versalovic, J. *et al.* (1994) 'Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction', *Methods in Molecular and Cellular Biology*, 5(1).
- Weinel, C. *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 Letters*, 1(4). doi: 10.1166/gl.2002.021.
- Wilde, C. *et al.* (2001) 'Transposition of IS1397 in the family *Enterobacteriaceae* and first characterization of IS*Kpn1*, a new insertion sequence associated with *Klebsiella pneumoniae* palindromic units', *Journal of Bacteriology*, 183(15). doi: 10.1128/JB.183.15.4395-4404.2001.
- Wilde, C. *et al.* (2003) 'Transposases are responsible for the target specificity of IS1397 and IS*Kpn1* for two different types of palindromic units (PUs)', *Nucleic Acids Research*, 31(15). doi: 10.1093/nar/gkg494.
- Woodford, N., Johnson, A. and Ridley, A. M. (2003) 'Genomic Fingerprinting by Application of rep-PCR', in *Molecular Bacteriology*. Humana Press, pp. 103–116. doi: 10.1385/0-89603-498-4:103.
- Yang, Y. and Ferro-Luzzi Ames, G. (1988) 'DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences', *Proceedings of the National Academy of Sciences of the United States of America*, 85(23), pp. 8850–8854. doi: 10.1073/pnas.85.23.8850.

# Přílohy

## 1. Srovnání sekvencí REP elementů u vybraných druhů

Hostitelský druh	SEKVENCE		Zastoupení	Zdroj
<i>P. putida</i>	<b>GTRG</b> GA GCGGGY KY RCCCGC GAA			GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes
<i>P. putida</i> KT2440	<b>GTAG</b> GAGCGGGT TT ACCCGCGAA		205 kopií	Aranda-Olmedo et al. 2002
<i>P. aeruginosa</i> PAO1	2002	<b>GTAG</b> GGCGAATAACCGCNNNGCGGTTATCCGCC	17 kopií	Weinell et al. 2002
	2002	<b>GTAG</b> GGCGAATAACCGCNNNGCGGTTATCCGCC	26 kopií	Weinell et al. 2002
	2006	<b>GGTAG</b> GGCGGATAACCGCTCGCGGTTATCCGCCG		(Tobes and Pareja, 2006)
<i>Salmonella typhimurium</i>	<b>GTAG</b> GCCBGR T AA GGC N3-4 CGCC GCC AYCSMGGS			GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes
<i>Klebsiella pneumoniae</i>	<b>GTAG</b> SCCSGRT AA GGCG N3-4 CGCC GCC AYCSGGS			GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes
<i>E. coli</i>	<b>GTAG</b> GYCKGAT AA GRCGY N2-6 CGCC GCC AYCSGGS			GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes
<i>E. coli</i> K-12	<b>GTAG</b> GCCGATAAGGCGT TY ACGCCGCATCCGGC		79 kopií	Bachelier et al. 1999
	<b>GTAG</b> GCCTGATAAGCGTAGCGCATCAGGC		45 kopií	Bachelier et al. 1999
	<b>GTAG</b> GCCTGATAAGACGCGCMAGCGTCGCATCAGGC		21 kopií	Bachelier et al. 1999
<i>Neisseria meningitidis</i>	<b>GTAG</b> GTCGGATWC TY GWATCCGAC			GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes
<i>Stenotrophomonas maltophilia</i> K279a	Obecný (2008)	<b>GTAG</b> TGCCGGCC GCT GGCCGGCA	437 kopií	Rosetto et al. 2008
	Obecný (2013)	<b>GTAG</b> MGYCGASY N2-4 RSTCGRCK		GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes
	SMAG-1	<b>GTAG</b> HGCCGGSC GCN GSCCGGCD	56%	A gene family of non-palindromic sequences in <i>Stenotrophomonas maltophilia</i>
	SMAG-2	<b>GTAG</b> MGYCGAGC (TT/CRY/CYRY) GCTCGRCK	14%	A gene family of non-palindromic sequences in <i>Stenotrophomonas maltophilia</i>
	SMAG-3	<b>GTAG</b> AGYCGACY GYY RGTCGRCT	11%	A gene family of non-palindromic sequences in <i>Stenotrophomonas maltophilia</i>
	SMAG-4	<b>GTAG</b> AKCCACGC CAY GCGTGGMT	19%	A gene family of non-palindromic sequences in <i>Stenotrophomonas maltophilia</i>
	SMAG-5	<b>GTAG</b> GTGYSRACC (TT/GTT) GGTYSRCAC	21%	A gene family of non-palindromic sequences in <i>Stenotrophomonas maltophilia</i>
	Asociované s RAYT (SM1)	<b>GGTGG</b> GTGCCGACC GTTGGTCGGCAC		(Nurvar, Licha and Schneider, 2013)
Asociované s RAYT (SM8)	<b>GGTAG</b> TGCCGGCCGCTGGCCGGCA		(Nurvar, Licha and Schneider, 2013)	
<i>Pseudomonas aeruginosa</i>	<b>GTAG</b> GGCGAATAACSSC (TY/ANN) GSSGTTATCCGCC			GTAG- and CGTC-tagged palindromic DNA repeats in prokaryotes
<i>P. fluorescens</i> SBW25 (3 rayt)	Nejčastější REP	<b>ATGTGG</b> GAGGGGGCTTGCCCCGAT	265 kopií	(Bertels, Gokhale and Traulsen, 2017)
	Asociované s RAYT (PF1)	<b>GTGG</b> GAGGGGGCTTGCCCCGAT		
	Asociované s RAYT (PF17)	<b>GTGG</b> CGAGGGAGCTTGCTCCCGCT		(Nurvar, Licha and Schneider, 2013)
	Asociované s RAYT (PF22)	<b>GTRG</b> TGAGCGGGCTTGCCCCGCGCT		
<i>P. fluorescens</i> GB1 (3 rayt)	<b>TGTAG</b> GAGCGGCCTTGTCGCGAT		109 kopií	(Bertels, Gokhale and Traulsen, 2017)
<i>P. fluorescens</i> A506 (1 rayt)	Nejčastější REP	<b>TGTGG</b> GAGCTGGCTTGCTGCGATA	314 kopií	(Bertels, Gokhale and Traulsen, 2017)
	Asociované s RAYT (PF12)	<b>GTGG</b> GAGCTGGCTTGCTGCGAT		(Nurvar, Licha and Schneider, 2013)
<i>P. fluorescens</i> BG33R (1 rayt)	<b>ATGTGG</b> GAGCTGGCTTGCTGCGAT		160 kopií	(Bertels, Gokhale and Traulsen, 2017)

## 2. Srovnání RAYT jednotlivých podtříd

Podtřída	Odpovídající skupina z (Bertels, Gallie and Rainey, 2017)	Strukturní odchylky	Počet členů	Taxonomické rozšíření	Významní bakteriální hostitelé s detailně analyzovanými REP elementy/RAYT proteiny	Přítomnost/absence REP elementů u RAYT genů, charakteristika REP elementů	Poznámka
2.1	G1	Přidatná C-terminální doména – vykazuje helix-turn-helix	493 (292)	Proteobakterie		Chybí, asociují s dlouhými inverzními repeticemi (TIR)	Patří sem transposázy TIRYT (Di Nocera, De Gregorio and Rocco, 2013)
2.2	G2	Nejdelší RAYT proteiny – průměr 325 AK	270	Proteobakterie	<i>Escherichia coli</i>	Přítomny, dlouhé, často s mismatch, vysoce repetitivní	Obsahuje první objevené REP, původní clade II
2.3	Žádná		224	Chromatiales Alteromonadales		Chybí	
2.4	G4		124	Cyanobakterie Bacteroidetes		Přítomny, středně dlouhé, minimálně repetitivní	Inzerce REP do samotného <i>rayt</i> genu
2.5	G3		87	xanthomonády pseudomonády stenotrofomonády	<i>Pseudomonas fluorescens</i> <i>Stenotrophomonas sp.</i>	Přítomny, krátké, vysoce repetitivní	Až 5 typů v jednom genomu, původní clade I
2.6	G5		63	Bacteroidetes		Chybí, atypické sekundární struktury	
2.7	Žádná		36	Cyanobakterie		Přítomny, dlouhé	
2.8	Žádná		8	Cyanobakterie		Přítomny, dlouhé	
3	Žádná		14	Planctomycetes Acidobacteria Proteobakterie		Chybí, atypické sekundární struktury	
Žádná	G6		30			Chybí	Charakteristiky transposáz IS