



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA ELEKTROTECHNIKY A KOMUNIKAČNÍCH TECHNOLOGIÍ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

ÚSTAV BIOMEDICÍNSKÉHO INŽENÝRSTVÍ

DEPARTMENT OF BIOMEDICAL ENGINEERING

DETAILNÍ ANALÝZA GENŮ REZISTENCE NA ANTIBIOTIKA A TĚŽKÉ KOVY A ZNAKŮ HORIZONTÁLNÍHO PŘESOSU GENŮ U ZÁSTUPCŮ ANAEROBNÍCH BAKTERIÍ

DETAIL ANALYSIS ON ANTIBIOTIC AND HEAVY-METAL RESISTANCE GENES AND HORIZONTAL GENE
TRANSFER TRAITS IN ANAEROBIC BACTERIA

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Kateřina Vancová

VEDOUCÍ PRÁCE

SUPERVISOR

Mgr. Bc. Darina Čejková, Ph.D.

BRNO 2023

Bakalářská práce

bakalářský studijní program **Biomedicínská technika a bioinformatika**

Ústav biomedicínského inženýrství

Studentka: Kateřina Vancová

ID: 231165

Ročník: 3

Akademický rok: 2022/23

NÁZEV TÉMATU:

Detailní analýza genů rezistence na antibiotika a těžké kovy a znaků horizontálního přesosu genů u zástupců anaerobních bakterií

POKYNY PRO VYPRACOVÁNÍ:

1) Vypracujte literární rešerši zabývající se detekcí genů rezistence (na antibiotika a těžké kovy) u bakterií. 2) Seznamte se s veřejnými databázemi pro detekci genů rezistence a horizontálního přenosu - např. MEGARes, ISEScan, ICEberg a INTEGRALL. 3) Detekujte hledané geny pomocí různých algoritmů a/nebo databází. Navrhněte metodu pro detekci a detailní charakterizaci těchto genů z genomických dat bakteriálních anaerobů. Dílčí části metod realizujte a otestujte. 4) Implementujte navrženou metodu ve zvoleném programovacím jazyce. 5) Algoritmus otestujte na souboru 300 genomů. 6) Výsledky vyhodnoťte a diskutujte.

DOPORUČENÁ LITERATURA:

[1] JURICOVA H., J. MATIASOVICOVA, T. KUBASOVA, et al. The distribution of antibiotic resistance genes in chicken gut microbiota commensals. Scientific Reports. 2021, 11(1):3290.

[2] DOSTER E., S. M. LAKIN, C. J. DEAN, et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence

Termín zadání: 6.2.2023

Termín odevzdání: 14.8.2023

Vedoucí práce: Mgr. Bc. Darina Čejková, Ph.D.

doc. Ing. Jana Kolářová, Ph.D.
předseda rady studijního programu

UPOZORNĚNÍ:

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

ABSTRAKT

Tato práce se zabývá detailní analýzou genů rezistence na antibiotika a těžké kovy a znaků horizontálního přesosu genů u zástupců anaerobních bakterií. Antimikrobiální rezistence (AMR) u bakterií je celosvětově rostoucí hrozbou pro veřejné zdraví. U těchto bakterií jsou geny AMR často spojeny s mobilními genetickými elementy (MGE), které podporují jejich mobilitu a umožňují jim rychlé šíření mezi různými bakteriemi. V práci je popsán horizontální přenos genetické informace a problematika antimikrobiální rezistence. Bylo otestováno 20 různých databází pro detekci genů AMR a MGE, databáze byly popsány a zhodnoceny. Poté bylo 6 z těchto databází vybráno pro analýzu 452 bakteriálních anaerobů. Byla navržena metoda pro analýzu a charakterizaci genů rezistence na antibiotika a těžké kovy a horizontálního přenosu, navržená metoda byla implementována v Pythonu.

KLÍČOVÁ SLOVA

horizontální přenos genů, mobilní genetické elementy, bakterie, antimikrobiální rezistence, geny rezistence, databáze pro detekci genů rezistence, antibiotika, těžké kovy

ABSTRACT

This thesis deals with a detail analysis on antibiotic and heavy-metal resistance genes and horizontal gene transfer traits in anaerobic bacteria. Antimicrobial resistance (AMR) in bacteria is a growing threat to public health globally. In these bacteria, AMR genes are often associated with mobile genetic elements (MGEs), which promote their mobility, enabling them to rapidly spread throughout a bacterial community. The work describes the horizontal gene transfer and the issue of antimicrobial resistance. 20 different databases for the detection of AMR genes and MGEs were tested, the databases were described and evaluated. Then 6 of these databases were selected for analysis of 452 bacterial anaerobes. A method for the analysis and characterization of antibiotic and heavy metal resistance genes and horizontal transfer was proposed, the proposed method was implemented in Python.

KEYWORDS

horizontal gene transfer, mobile genetic elements, bacteria, antimicrobial resistance, resistance genes, database for resistance gene detection, antibiotics, heavy-metals

VANCOVÁ, Kateřina. *Detailní analýza genů rezistence na antibiotika a těžké kovy a znaků horizontálního přesosu genů u zástupců anaerobních bakterií*. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství, 2023, 54 s. Bakalářská práce. Vedoucí práce: Mgr. Bc. Darina Čejková, Ph.D.

Prohlášení autora o původnosti díla

Jméno a příjmení autora:	Kateřina Vancová
VUT ID autora:	231165
Typ práce:	Bakalářská práce
Akademický rok:	2022/23
Téma závěrečné práce:	Detailní analýza genů rezistence na antibiotika a těžké kovy a znaků horizontálního přesosu genů u zástupců anaerobních bakterií

Prohlašuji, že svou závěrečnou práci jsem vypracovala samostatně pod vedením vedoucí/ho závěrečné práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autorka uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušila autorská práva třetích osob, zejména jsem nezasáhla nedovoleným způsobem do cizích autorských práv osobnostních a/nebo majetkových a jsem si plně vědoma následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

Brno

.....

podpis autorky*

*Autor podepisuje pouze v tištěné verzi.

PODĚKOVÁNÍ

Ráda bych poděkovala vedoucí bakalářské práce paní Mgr. Bc. Darině Čejkové, Ph.D., za odborné vedení, konzultace, trpělivost a podnětné návrhy k práci.

Obsah

Úvod	11
1 Horizontální přenos genetické informace	12
1.1 Konjugace	13
1.2 Transformace	14
1.3 Transdukce	14
1.3.1 Generalizovaná transdukce	15
1.3.2 Specializovaná transdukce	15
1.4 Mobilní genetické elementy	15
1.4.1 Inzerční sekvence	15
1.4.2 Transpozony	17
1.4.3 Integrony	19
1.4.4 Plazmidy	21
2 Rezistence na antibiotika a těžké kovy	22
2.1 Zdroje genů rezistence na antibiotika	22
2.2 Dopady antimikrobiální rezistence	23
2.3 Detekce genů rezistence a horizontálního přenosu	23
2.3.1 Databáze pro detekci genů rezistence a horizontálního přenosu	24
3 Detekce a charakterizace genů rezistence	33
3.1 Analyzovaný dataset	33
3.2 Testování databází	33
3.2.1 Stažené databáze	35
3.2.2 Online databáze	36
3.3 Metody detekce a charakterizace genů rezistence	41
3.3.1 Vybrané databáze	41
3.3.2 Unifikace výsledků	41
3.3.3 Přenos genů rezistence	42
3.3.4 Přehled získaných výsledků pro vybranou bakterii	42
3.4 Výsledky detekce a charakterizace genů rezistence	43
3.4.1 Geny rezistence	43
3.4.2 Antibiotika a těžké kovy	44
3.4.3 Mobilní genetické elementy	45
3.4.4 Přenos genů rezistence	46
4 Diskuze	47

Závěr	49
Literatura	50

Seznam obrázků

1.1	Mechanismy horizontálního přenosu genetické informace	12
1.2	Struktura inzerční sekvence	16
1.3	Mechanismy přenosu inzerční sekvence mezi molekulami DNA	16
1.4	Struktura transpozonů	18
1.5	Struktura integronu a integrace genových kazet	20
3.1	Rozdělení genů rezistence na antibiotika	44
3.2	Zastoupení mobilních genetických elementů	45

Seznam tabulek

3.1	Databáze pro detekci genů rezistence a mobilních genetických elementů	34
-----	---	----

Úvod

V současné době je problém rezistence bakterií na antibiotika a těžké kovy jednou z největších výzev v oblasti veřejného zdraví. Kvůli rostoucí rezistenci na antibiotika u patogenů se onemocnění způsobená těmito patogeny stávají vážnou hrozbou. Antimikrobiální rezistence proto představuje vážnou hrozbu pro velkou část zdravotnictví, jak ji známe. Bakterie mohou také obsahovat geny, které kódují rezistenci na biocidy, které jsou důležité jako dezinfekční prostředky, a kovy, které mají antimikrobiální účinky.

Tato bakalářská práce se zabývá detailní analýzou genů rezistence na antibiotika a těžké kovy a znaků horizontálního přenosu genetické informace u zástupců anaerobních bakterií. Práce se zaměřuje na využití moderních bioinformatických metod a veřejných databází pro detekci a charakterizaci těchto genů z genomických dat.

Prvním cílem této práce je provést literární rešerši, která bude zaměřena na stávající poznatky o genech rezistence na antibiotika a těžké kovy u bakterií. Budou zkoumány různé mechanismy horizontálního přenosu genů mezi bakteriemi. Druhým cílem je seznámení s veřejnými databázemi a bioinformatickými nástroji určenými pro detekci genů rezistence a horizontálního přenosu.

Třetím cílem je provést detekci genů rezistence na antibiotika a těžké kovy u zástupců anaerobních bakterií. Bude navržena metoda pro detekci a charakterizaci těchto genů z genomických dat bakteriálních anaerobů a dílčí části metod budou realizovány a otestovány. To bude zahrnovat využití různých databází a bioinformatických algoritmů.

Čtvrtým cílem je implementace navržené metody ve zvoleném programovacím jazyce. Pátým cílem je otestování algoritmu na souboru 453 genomů. Posledním cílem je získané výsledky vyhodnotit a diskutovat.

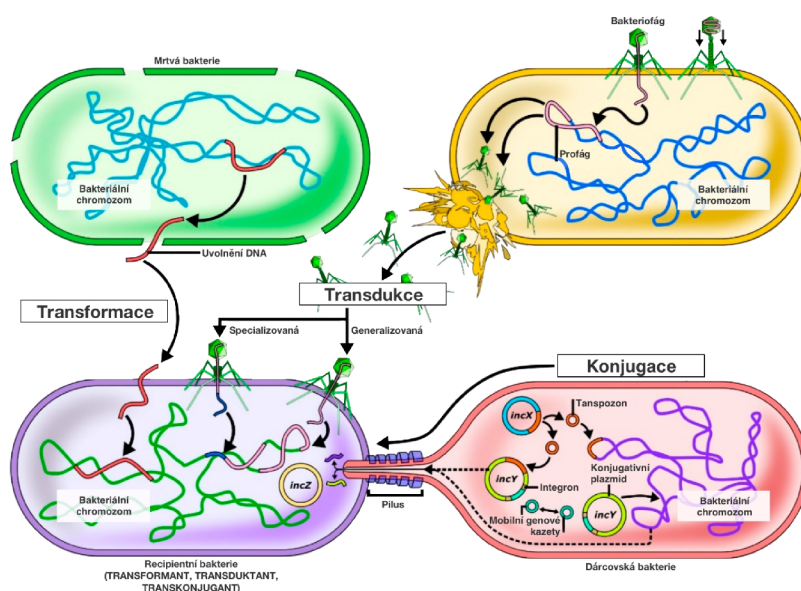
Tato bakalářská práce může přispět k porozumění různých přístupů k detekci a charakterizaci genů rezistence. Výsledky této bakalářské práce poskytnou důležité informace o rozšíření genů rezistence na antibiotika a těžké kovy mezi anaerobními bakteriemi a mechanismech jejich přenosu. Mohou poskytnout informace pro budoucí diagnostiku a prevenci.

1 Horizontální přenos genetické informace

Genetická informace se mezi organizmy přenáší dvěma cestami - vertikálně a horizontálně (laterálně). Vertikální přenos je z generace na generaci (z rodičů na potomky). Horizontální přenos je sdílení genetického materiálu mezi nepříbuznými organizmy (mezi organizmy, které nejsou ve vztahu rodič-potomek).[1]

Horizontální přenos genetické informace (HGT) je známý mechanismus adaptace u bakterií a archea. U prokaryot probíhá bez jakéhokoliv imunitního rozpoznávání. Je pro ně univerzálním prostředkem, jak reorganizovat své genomy, aby získaly nové metabolické reakce a nové vlastnosti jako patogenitu, virulenci a rezistenci na různé chemické látky vyskytující se v prostředí včetně léků a antibiotik.[2] Mutace je pomalý proces a většina mutací je pro bakterie škodlivá nebo neutrální, není pro ně tak přínosná. Oproti tomu je HGT rychlým způsobem, jak najednou získat velké kusy DNA z jiné bakterie. Mikrobiální antibiotická rezistence a patogenita jsou nejčastěji spojeny s HGT. HGT ale nesouvisí jen s organizmy způsobující nemoci, vyskytuje se mezi prokaryoty, prokaryoty a eukaryoty a i mezi mnohobuněčnými eukaryoty.[3]

Nejvíce známé mechanismy horizontálního přenosu genetické informace u prokaryot jsou konjugace, transformace a transdukce (obrázek 1.1).[3] Konjugace je přenos genů, který je zprostředkován určitými plazmidy nebo konjugativními transpozony s relevantními přenosovými geny. Na rozdíl od transformace nebo transdukce je pro konjugaci nutný kontakt mezi dárcovskou a recipientní buňkou. Transformace je přenos genů, který je zprostředkován vychytáváním volné DNA z okolního prostředí. Transdukce je přenos genů, který je zprostředkován určitými typy bakteriofágů.[4]



Obr. 1.1: Mechanizmy horizontálního přenosu genetické informace.[5]

1.1 Konjugace

Bakteriální konjugace je promiskuitní mechanismus transportu DNA, kdy dochází k přenosu genetické informace mezi dvěma bakteriemi prostřednictvím přímého kontaktu a výměny plasmidů, což jsou kruhové molekuly DNA. Konjugativní plasmidy se přenášejí mezi většinou bakterií, jsou tedy jednou z hlavních příčin šíření antibiotické rezistence mezi patogenními bakteriemi. Konjugace poskytuje bakteriím jedinečný prostředek pro genetickou výměnu a je to silný zdroj genetické variability.[6]

Ačkoli je konjugace definována jako přenos DNA z bakterie na bakterii, je možné, že jako příjemce může sloužit jakýkoli typ buňky, protože všechny události, ke kterým dochází během konjugace, jsou řízeny dárcovskou buňkou a jsou typicky kódovány plasmidem.[7] Schopnost dárce je určena specifickými konjugativními plasmidy nazývanými plasmidy plodnosti (F-plasmidy).[8]

Konjugace vyžaduje fyzický kontakt mezi dárcovskou a recipientní buňkou prostřednictvím konjugačního pilu (vlasovitý útvar na povrchu mnoha bakterií sloužící ke konjugaci), přes který se přenáší genetický materiál.[3]

Replikace DNA a transport makromolekul přes membrány jsou základními procesy života. Bakteriální konjugační systémy vykazují podobnosti s replikací DNA a makromolekulárními transportními systémy.[6] Mechanismus konjugace v F-plasmidech zahrnuje relaxozom a transferozom. Relaxozom je proteinový komplex, který hraje důležitou roli při přenosu plasmidů v bakteriích. Shromažďuje se na *oriT* (počátek přenosu na plasmidu). Skládá se z několika proteinů, z nichž každý plní specifickou funkci.[9] Transferozom je komplex konjugativních proteinů, který sestavuje membránový transportér. Patří do sekrečního systému typu IV (T4SS). Sousedí s konjugačním pílím.[6] Konjugativní systémy navíc zahrnují vazebný protein, který působí jako spojovací článek mezi těmito dvěma komplexy (relaxozomem a transferozomem).[9]

Pro konjugaci je nutný plasmid zvaný F-faktor. Bakterie obsahující F-faktor se nazývají F+, ty bez něj se nazývají F-. Buňka F+, tedy dárce, vytváří pilus, kterým se spojuje s recipientní buňkou. Pro zahájení konjugace je F-faktor přerušen na *oriT* relaxozomem, který se spojuje s řetězcem, který má být přenesen (T-DNA). Dojde k uvolnění pomocných proteinů relaxozomu, ale část relaxozomu zvaná relaxáza zůstává připojena k T-DNA. Tento komplex T-DNA a relaxázy je rozpoznán vazebným proteinem a přenesen na membránový transportér. Transportér tlačí komplex T-DNA a relaxázy přes pilus do buňky příjemce. Přenesený řetězec T-DNA se převede na kruhovou dvouřetězcovou F-plazmidovou DNA v recipientní buňce a v dárcovské buňce se syntetizuje nový řetězec, který nahradí přenesený řetězec. Spojení přes pilus je uvolněno. Obě exkonjugantní bakterie jsou F+ a F-plasmid se proto může dále šířit infekcí mezi geneticky kompatibilní populací bakterií.[8]

1.2 Transformace

Transformace je proces, při kterém bakterie přijímá cizí DNA z okolního prostředí a začleňuje ji do svého genomu.[3]

Schopnost bakterií přijímat extracelulární DNA a transformovat se, nazývaná kompetence, se liší podle fyziologického stavu bakterií. Bakterie přijímá cizí DNA dovnitř buňky pomocí speciálních proteinů na povrchu buněčné membrány, které se nazývají kompetenční faktory. Tyto faktory umožňují vstup cizí DNA do buněčného cytoplazmatického prostoru.[8]

V cytoplazmatickém prostoru bakterie se cizí DNA zpracovává a rekombinací může být začleněna do genomu bakterie. Rekombinace je proces, při kterém dochází k přeskupení genetické informace mezi molekulami DNA. Tento proces je katalyzován speciálními enzymy nazývanými rekombinantní proteiny, které umožňují přesun cizí DNA do genomu recipientní bakterie. DNA z dárcovské bakterie se obvykle nemůže replikovat v recipientní bakterii, pokud se rekombinací nestane součástí replikonu, což jsou sekvence DNA, které obsahují informaci nutnou pro replikaci.[8]

1.3 Transdukce

Transdukce je proces, při kterém se genetický materiál přenáší mezi bakteriemi pomocí bakteriofágů. To jsou viry, které infikují bakterie. Bakteriofág vloží některé bakteriální geny do svého genomu, které se následně mohou přenést na další bakterie, které jsou tímto bakteriofágem infikovány.[3]

Transdukce začíná infekcí bakterie. Bakteriofág se váže na specifické receptory na povrchu bakteriální buňky a vstupuje dovnitř. Po vstupu do bakterie využívá bakteriofág replikační systém bakterie k replikaci svého vlastního genomu a k tvorbě nových virových částic. Tento proces může probíhat v cyklu lytickém nebo v cyklu lysogenním.[10]

V lytickém cyklu se virus po infekci hostitelské buňky replikuje. Po replikaci viru dochází k lýze (rozpadu) hostitelské buňky, při které se uvolňují nově vytvořené virové částice. Ty mohou infikovat další buňky. V lysogenním cyklu se virus po infekci hostitelské buňky integruje do genomu buňky a je přítomen ve formě profágu. Profág je neaktivní a replikuje se společně s genomem hostitelské buňky, aniž by způsoboval její poškození. Za určitých podmínek, jako je poškození DNA (např. UV zářením, ionizujícím zářením) nebo stresová situace (např. nedostatek živin, změny v prostředí), může dojít k indukci profágu. Indukce profágu je proces, při kterém je profág uvolněn z genomu buňky a dochází k přechodu viru z cyklu lysogenního do cyklu lytického.[11]

Při transdukcii je genetická informace dárcovské bakterie během lytické replikace začleněna do bakteriofágu a je přenesena do recipientní buňky infekcí[10]. Existují dva typy transdukce, generalizovaná a specializovaná[3].

1.3.1 Generalizovaná transdukce

Při generalizované transdukcii je náhodný kousek hostitelské DNA začleněn během buněčné lýzy.[3] Generalizované transdukující fágy produkované během lytického cyklu obsahují místo části fágové DNA náhodný segment bakteriálního genomu. Každý jednotlivý generalizovaný transdukující fág nese odlišnou sadu genů, které představují tento malý segment bakteriálního genomu. Každá část bakteriálního genomu má přibližně stejnou pravděpodobnost začlenění do fágu a přenosu z dárcovské na recipientní bakterii.[8]

1.3.2 Specializovaná transdukce

Při specializované transdukcii může být do recipientních bakterií přeneseno pouze několik specifických genů, které se nacházejí v místě integrovaného profágu. Specializovaná transdukce je tedy výsledkem lysogenizace.[10] Specializované transdukující fágy se tvoří tehdy, když dárcovská bakterie z cyklu lysogenního vstoupí do cyklu lytického a dojde k indukci integrovaných profágů. Při indukci mohou tyto fágy do svého genomu začlenit sousední geny z genomu hostitelské buňky a přenést je na recipientní bakterii. Mnoho specializovaných transdukujících fágů je defektních a nemůže dokončit lytický cyklus v infikovaných buňkách, pokud zde nejsou přítomny pomocné fágy, které poskytují chybějící fágové funkce.[8]

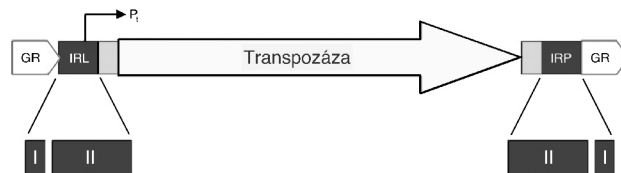
1.4 Mobilní genetické elementy

Mobilní genetické elementy (MGE) jsou segmenty DNA, které kódují enzymy a další proteiny, které zprostředkovávají pohyb DNA v rámci genomů (intracelulární mobilita) nebo mezi bakteriálními buňkami (intercelulární mobilita). Antibiotická rezistence bakterií se šíří získáváním genů rezistence, které existují v transpozonech, integronech a plazmidech.[4]

1.4.1 Inzerční sekvence

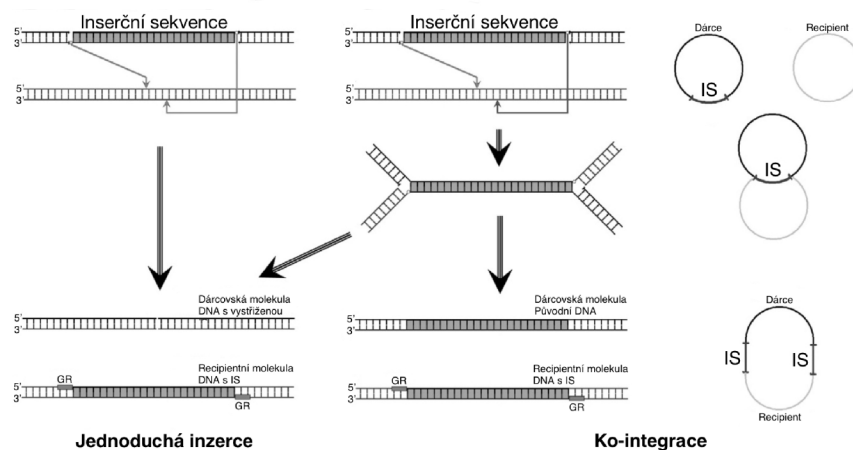
Inzerční sekvence (IS) jsou krátké úseky DNA, které jsou schopné nezávislé transpozice v mikrobiálním genomu. Nekódují žádné jiné funkce než ty, které se podílejí na jejich mobilitě.[12] Mimo bakterií se vyskytují také u eukaryot. Způsobují inzerční mutace, rekombinace, případně mobilizují geny virulence nebo rezistence.[13]

Základní složkou inserční sekvence je gen transpozázy, který je ohraničen invertovanými repeticemi (IRL, IRP), které rozpoznává transpozáza. Invertované repetice se skládají ze dvou podjednotek, první podjednotka (I) je úsek rozpoznávaný transpozázou jako štěpné místo, druhá podjednotka (II) je vazebným místem transpozázy. Poslední součástí u většiny inserčních sekvencí jsou generované repetice (GR), které jsou generovány při inserci inserční sekvence do cílové molekuly DNA (obrázek 1.4.1).[13]



Obr. 1.2: Struktura inserční sekvence. IRL - invertovaná repetice levá, IRP - invertovaná repetice pravá, GR - generované repetice, P_t - promotor genu transpozázy, I - úsek rozpoznávaný transpozázou jako štěpné místo, II - vazebné místo transpozázy.[13]

K transpozici inserční sekvence dochází přímou vazbou transpozázy na invertované repetice s následným rozpojením vazby a přenesením na cílové místo recipientní molekuly DNA. Rozlišují se dva mechanismy transpozice inserční sekvence, jednoduchá inserce, kdy je inserční sekvence přenesena kompletním vystřížením a volné konce na dárcovské molekule DNA jsou následně spojeny, a ko-integrace, kdy po replikaci dojde ke spojení dárcovské a recipientní molekuly DNA (obrázek 1.4.1).[13]



Obr. 1.3: Mechanizmy přenosu inserční sekvence mezi molekulami DNA. Jednoduchá inserce a ko-integrace. IS - inserční sekvence, GR - generované repetice.[13]

Inzerční sekvence jsou klasifikovány podle složení transpozázy, invertovaných a generovaných repetit, atd. U prokaryot bylo popsáno kolem 25 různých rodin, které obsahují několik set různých inzerčních sekvencí.[13]

Významnou úlohou inzerčních sekvencí v šíření antibiotické rezistence je zvýšení exprese genů rezistence. Invertované repetice inzerčních sekvencí jsou silné promotory a lokalizace takového promotoru vede ke zvýšené expresi genů nacházejících se za inzerční sekvencí.[13]

1.4.2 Transpozony

Transpozony jsou mobilní genetické elementy, definovány jako sekvence DNA schopné transpozice. Mohou přeskakovat do různých míst genomu, proto se jim říká skokové geny.[14] Na rozdíl od inzerčních sekvencí kódují další geny, které přímo nesouvisí s jejich funkcí[13]. Transpozici transpozonů katalyzuje enzym transpozáza, který rozeznává koncové obrácené repetice transpozonu, štěpí cílové sekvence, do kterých se vkládá transpozon, a po transpozici spojuje cílové sekvence s vloženým transpozonem[14].

Transpozony obsahují tři funkční složky, gen transpozázy, která umožňuje transpozici, genetickou informaci nepotřebnou k transpozici transpozonu (např. geny rezistence) a invertované repetice na obou koncích transpozonu, které jsou stejně jako v případě inzerční sekvence rozpoznávány transpozázou (obrázek 1.4.2 A).[13]

Rozlišují se dva druhy transpozice, intermolekulární transpozice, která probíhá mezi dárcovskou a recipientní molekulou DNA (např. mezi plazmidem a chromosomem), a intramolekulární transpozice, která probíhá mezi donorovým a cílovým místem na jedné molekule DNA (např. na molekule plazmidu)[15]. Dále se rozlišují dva způsoby transpozice, konzervativní transpozice, kdy dojde k excisi transpozonu z donorového místa a inserci do cílového místa, a replikativní transpozice, kdy se transpozon v donorovém místě zreplikuje a jeho kopie se vloží do cílového místa[16].

Transpozony jsou významným faktorem šíření rezistence na antibiotika, protože jsou na nich lokalizovány geny kódující enzymy, které inaktivují antibiotika. Přenos transpozonu z plazmidu na jiné plazmidy nebo z chromozomu DNA do plazmidu a naopak způsobuje přenos genů rezistence na antibiotika u bakterií. Některé transpozony jsou však vždy zachovány v místě inserce v genomu. Většina transpozonů je deaktivována a v důsledku toho se nemůže pohybovat.[14] Transpozony jsou také důležité při mobilizaci integronů[13].

Na základě způsobu transpozice se transpozony dělí do dvou hlavních skupin, retrotranspozony a DNA transpozony. Bakteriální transpozony (např. kompozitní transpozony nebo transponovatelné bateriofágy) patří do DNA transpozonů, které jsou obvykle nositeli genů pro antibiotickou rezistenci.[14]

Retrotranspozony

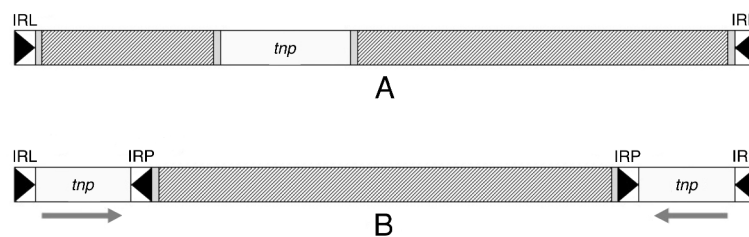
Retrotranspozony se přenáší procesem nazývaným retropozice. Retropozice je transpozice zpětnou transkripcí katalyzovaná enzymem reverzní transkriptáza. Retrotranspon se v donorovém místě přepíše do RNA, která se reverzní transkriptázou přepíše do DNA vkládající se do cílového místa.[14]

Retrotranspozony a další transpozony založené na transpozičních proteinech se dělí do dvou skupin, autonomní a neautonomní. Autonomní retrotranspozony mají otevřený čtecí rámec a geny, které kódují transpoziční proteiny. Neautonomní retrotranspozony nejsou schopny tyto proteiny kódovat, a proto využívají dostupné proteiny autonomních retrotranspozonů.[14]

Retrotranspozony byly nalezeny v bakteriích, archaea i eukaryotech. Často se nacházejí v genomu eukaryot. Bakteriální retrotranskriptáza byla nalezena v retronech a některých intronech. Retrony jsou sekvence DNA, které kódují reverzní transkriptázu. Byly nalezeny v genomu mnoha bakterií. Zatím není jasné, jestli jsou retrony mobilními elementy.[14]

Kompozitní transpozony

Kompozitní transpozony jsou transpozony ohraničené dvěma inzerčními sekvencemi, obvykle v opačné orientaci. Inzerční sekvence poskytují kompozitnímu transpozonu enzym transpozázu a invertované repetice. Obecná struktura transpozonu obsahuje gen transpozázy *tnp* ohraničený genetickou informací nepotřebnou k transpozici transpozonu a levou a pravou invertovanou repeticí IRL a IRP. Kompozitní transpon obsahuje genetickou informaci nepotřebnou k transpozici transpozonu, jsou to např. geny rezistence, s opačně orientovanými inzerčními sekvencemi na koncích (obrázek 1.4.2).[13]



Obr. 1.4: Struktura transpozonů. A - obecná struktura transpozonu, B - struktura kompozitního transpozonu, *tnp* - gen transpozázy, IRL - invertovaná repetice levá, IRP - invertovaná repetice pravá, GR - generované repetice, šedivě jsou zobrazeny místa obsahující genetickou informaci nepotřebnou k transpozici transpozonu (např. geny rezistence).[13]

Konjugativní transpozony

Konjugativní transpozony jsou mobilní genetické elementy schopné intercelulárního přenosu konjugací. Při excizi transpozonu z dárcovské molekuly DNA dojde k vytvoření kruhové dvoušroubovice DNA, která je podobná plazmidu, ale není schopná autonomní replikace. Poté je jednoho vlákno DNA konjugativně přeneseno do recipientní buňky. Tato konjugace není shodná s běžným konjugativním přenosem plazmidů, protože na konjugativních transpozonech nejsou kódovány geny pro sex pilus a ostatní proteiny potřebné k běžnému konjugativnímu přenosu. Za konjugaci transpozonů jsou pravděpodobně zodpovědné peptidové feromony. Po přenosu DNA dochází v dárcovské i recipientní buňce k replikaci druhého vlákna DNA a vytvoření původní kruhové dvoušroubovice DNA transpozonu, který je integrován do genomu v obou buňkách.[13] Konjugativní transpozony mohou mobilizovat plazmidy, které nemají obvyklé vlastnosti mobilizovatelných plazmidů.[17]

Transponovatelné bakteriofágy

Transponovatelné bakteriofágy jsou viry, které mohou transponovat svou DNA do bakteriálního chromozomu, plazmidu nebo profága, během tohoto procesu často duplikují sekvenci obklopující jejich místo vložení. Tyto temperované fágy mohou zůstat ve svých hostitelských genomech jako latentní profágy (lysogenní cyklus) nebo se aktivně replikovat (lytický cyklus).[17]

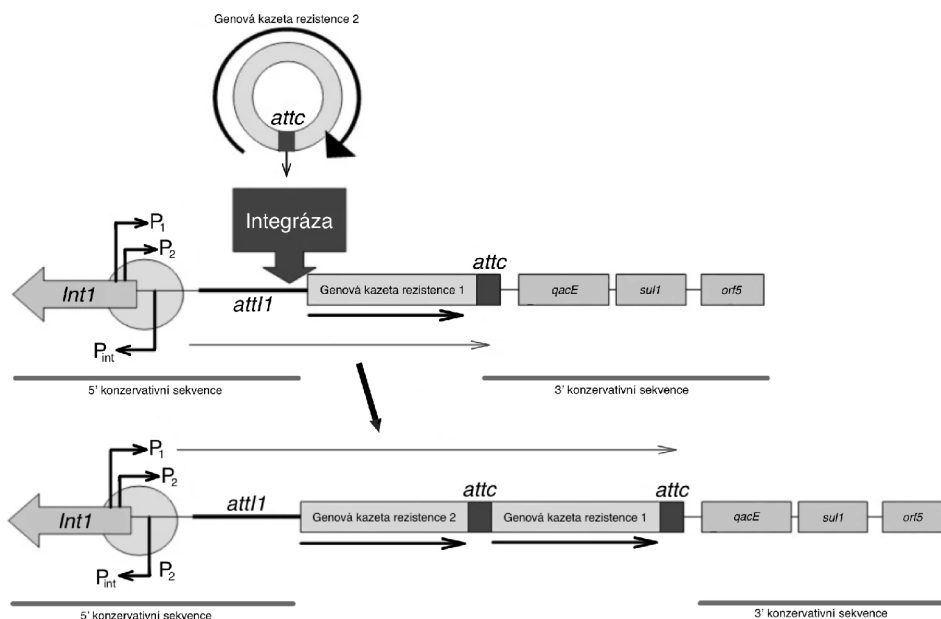
Inzerce tohoto elementu do genu, nebo jeho regulační sekvence, může vést k inaktivaci genu. Mutace vytvořené těmito elementy mají polární účinek. To znamená, že mutace v určité oblasti genu ovlivňuje čtení nebo regulaci genů v jeho blízkém okolí. Geny ve stejném operonu budou tedy také inaktivovány. Operon je úsek DNA pakaryot, který obsahuje více genů, ale funguje jako jedna transkripční jednotka.[17]

Transponovatelné bakteriofágy mohou vyvolat tvorbu různých genomových přeuspořádání, různé velikosti delecí nebo inverzí nebo tvorbu ko-integrátů, které vznikají kombinací dvou nebo více genetických elementů. Mohou také stimulovat mobilitu jiných bakteriofágů nebo vyvolat rekombinaci mezi transponovatelnými prvky.[17]

K přenosu genu rezistence dochází během lytického cyklu transponovatelného bakteriofágu, kdy jsou některé geny bakteriální rezistence zabaleny ve fágu. Pokud jsou jiné bakterie infikovány fágem, jsou tyto geny rezistence přenášeny.[14]

1.4.3 Integrony

Integrony jsou genetické elementy, které sdružují otevřené čtecí rámce a konvertují je do funkčních genů, poskytují jim silné promotory. Struktura integronu je zobrazena na obrázku 1.5.



Obr. 1.5: Struktura integronu a integrace genových kazet. Genová kazeta rezistence se integruje mezi 5' a 3' konzervativní sekvence integronu, *Int1* - gen pro integrázu, P_{int} - promotor genu pro integrázu, *attI1* - vazebné místo pro integrázu na integronu, P_1 a P_2 - promotory genových kazet integrovaných v integronu, *attC* - vazebné místo pro integrázu na genové kazetě, *qacE*, *sul1* a *orf5* - geny rezistence a otevřené čtecí rámce.[18]

Integrony obsahují gen pro integrázu (*Int1*) a promotor genu pro integrázu P_{int} . Integráza je enzym, který umožňuje integraci genových kazet do integronu. Vazebné místo integrázy na integronu je *attI1*. [18]

Genové kazety jsou významnou složkou, která se podílí na struktuře integronu. Jsou to mobilními elementy s kruhovou DNA, které se mohou vyskytovat volně v cytoplazmě. Narozdíl od plazmidů ale nejsou schopny autonomní replikace. Jsou tvořeny otevřeným čtecím rámcem a vazebným místem pro integrázu, *attC*. [13]

Genové kazety antibiotické rezistence jsou integrovány mezi 5' a 3' konzervativní sekvence integronu. Genové kazety nemají promotor před kódující sekvencí. Promotor umístěný v 5' konzervativní sekvenci integronu je tedy nezbytný pro expresi genových kazet. V integronech obsahuje 5' konzervativní sekvence dva potenciální promotory, P_1 a P_2 . [18]

Integrony se dělí do dvou hlavních skupin, superintegrony (SI), které čítají mnoho desítek genových kazet a jsou součástí chromozomu některých bakterií, a integrony, které sdružují geny rezistence na antibiotika (RI). Integrony se od sebe liší strukturou integrázy. Integrony nejsou mobilními elementy, ale mohou být mobilizovány, obvykle jako součást transpozonů. Všechny RI byly popsány jako součást mobilních elementů, proto byly samy integrony dříve považovány za mobilní elementy. [13]

Funkce integronů jako expresních jednotek je významná v šíření antibiotické rezistence, protože umožňuje současnou expresi všech integrovaných genových kazet najednou. Důsledkem toho je sdružená selekce rezistence ke všem antibiotikům, jejichž genové kazety jsou integrovány v integronu.[13]

1.4.4 Plazmidy

Plazmidy jsou molekuly dvouvláknové kruhové DNA schopné samostatné replikace v cytoplazmě bakterií. Obsahují geny nutné pro zahájení a kontrolu vlastní replikace a geny zajišťující trvalou přítomnost plazmidů u daného kmene, například gen, který synchronizuje replikaci plazmidu s buněčným dělením nebo geny, které umožňují konjugaci. Tyto geny nejsou pro bakterie esenciální, proto jsou plazmidy pro základní funkce bakterie postradatelné.[13] Některé plazmidy jsou kryptické a nemají žádné rozpoznatelné účinky na bakteriální buňky, které je obsahují.[8]

Proces horizontálního přenosu plazmidů se nazývá konjugace, aktivní (závislý na energii) kontaktní přenos genetické informace. Konjugace plazmidů je nejdůležitější způsob výměny genetické informace u bakterií. Konjugativní plazmid je takový plazmid, na kterém je kódován celý bílkovinný aparát, který je při konjugaci nutný pro přenos z dárcovské buňky do recipientní. Plazmidy, které jsou schopné konjugace pouze za účasti pomocného konjugativního plazmidu, se nazývají mobilizovatelné.[13]

Mnoho plazmidů kontroluje medicínsky důležité vlastnosti patogenních bakterií, včetně rezistence na jedno nebo několik antibiotik, produkce toxinů a syntézy struktur buněčného povrchu potřebných pro adheenci (interakce mezi strukturami na povrchu bakterií) nebo kolonizaci. Plazmidy, které určují rezistenci na antibiotika, se často nazývají R-plazmidy (R-faktory).[8]

2 Rezistence na antibiotika a těžké kovy

Šíření antimikrobiální rezistence prostřednictvím horizontálního přenosu genetické informace patří mezi jednu z nejzávažnějších výzev současné medicíny. Kvůli rostoucí rezistenci na antibiotika u patogenů se onemocnění způsobená těmito patogeny stávají vážnou hrozbou, jsou-li způsobena klony rezistentními na antibiotika.[19]

Zatímco antimikrobiální léky jsou nejvíce studovanými a běžně diskutovanými antimikrobiálními sloučeninami, bakterie mohou také obsahovat geny které kódují rezistenci na biocidy, které jsou důležité jako dezinfekční prostředky, a na kovy, které mají antimikrobiální účinky (např. měď, zinek).[20]

2.1 Zdroje genů rezistence na antibiotika

Antibiotika jsou kritickou složkou moderní medicíny, vyvinutá tak, aby vyvolávala prospěšné účinky na infekce vyvolané patogeny. Antibiotika běžně podléhají metabolismu nebo biotransformačním procesům v organismu, aby byly účinněji eliminovány. K vylučování těchto sloučenin a jejich metabolitů dochází hlavně močí, stolicí nebo kombinací obojího.[21]

Primárními zdroji kontaminace antibiotiky z lidské perspektivy proto jsou domácí, městské, nemocniční a průmyslové odpadní vody, akvakultura a intenzivní chov dobytka. Vzhledem k tomu, že stávající čistírny odpadních vod je nedokážou účinně odstraňovat, mohou se v detekovatelném množství dostávat do přírodních povrchových a podzemních vod. Slabou účinnost městských čistíren odpadních vod prokazuje detekce látek výhradně pro lidské použití. Kontaminace podzemních vod byla prokázána přítomností látek ve studních. Antibiotika, kromě jiných kontaminantů, mohou narušit mikroekosystém rozšířením rezistence vůči nim.[21]

Antibiotika nejsou striktně selektivní pouze vůči patogenům, takže při každém použití antibiotik je ovlivněn a podroben selekci na rezistentní klony nejen cílový patogen, ale také komenzální mikroflóra. Jakékoli použití antibiotik k potlačení infekce způsobené jedním bakteriálním patogenem tedy nevyhnutelně vede k selekci stovek komenzálních druhů, které jsou také vůči jejich působení odolné. Rezistentní komenzálové mohou později působit jako rezervoáry genů antibiotické rezistence.[19]

Střevní mikrobiota představuje komplexní ekosystém s ohromnou diverzitou. Skládá se z více než tisíce druhů bakterií, počet všech jejich genů (mikrobiom) převyšuje více než stokrát počet genů v lidském genomu.[22] Jedná se tedy o velmi rozsáhlé rezervoáry genů antibiotické rezistence. Testy provedené u kuřat prokázaly, že mikroflóra kuřecího střeva představuje důležitý zdroj genů rezistence na antibiotika[19]

2.2 Dopady antimikrobiální rezistence

Antimikrobiální látky umožnily pokrok v mnoha oblastech lékařské praxe. Úspěšné výsledky mnoha chirurgických zákroků a imunosupresivní léčby závisí na antibiotické profylaxi a schopnosti léčit infekční komplikace. Antimikrobiální rezistence proto představuje vážnou hrozbu pro velkou část zdravotnictví, jak ji známe. Má významné důsledky pro zdraví populace i ekonomické dopady.[23]

Počet případů infekcí, které jsou způsobeny bakteriemi zcela nebo téměř zcela rezistentními k antibiotikům, stále stoupá. Vzhledem k rostoucímu podílu populace ve věku nad 65 let, jejich vyšší míře komorbidit, a rostoucímu používání stále komplikovanějších lékařských postupů s přidruženými riziky infekce, může potřeba antimikrobiálních látek jen narůstat. Méně účinná antibiotika způsobí prodloužení doby hospitalizace, negativní dopad na výsledky mnoha chirurgických zákroků a imunosupresivní léčby a nedostatek perorálních antibiotik k léčbě infekcí v komunitě. Rostoucí počet neléčitelných infekcí a snížená účinnost empirických antibiotických režimů povede ke zvýšené morbiditě a mortalitě. Odhaduje se, že tisíce lidí ročně zemře na infekce získané v nemocnici způsobené rezistentními bakteriemi.[23]

Zatímco se antimikrobiální rezistence stále zvyšuje, zdroje pro vývoj nových antibiotik docházejí. Po osmi desetiletích používání antibiotik se bakteriální infekce, které byly kdysi snadno léčitelné, stávají neléčitelnými. Antibiotická rezistence koreluje s užíváním antibiotik, takže lepší antimikrobiální dohled s lepší prevencí a diagnostikou infekcí může pomoci zachovat v současnosti dostupná antimikrobiální činidla. Má-li vývoj nových antiinfekčních přípravků držet krok se zvyšující se rezistencí, je zapotřebí významných globálních akcí a investic, z veřejných i soukromých zdrojů.[23]

2.3 Detekce genů rezistence a horizontálního přenosu

Antimikrobiální rezistence (AMR) je hrozbou pro globální veřejné zdraví a identifikace genetických determinant antimikrobiální rezistence je kritickou součástí epidemiologických šetření.[20]

Kontrola a zkoumání antimikrobiální rezistence je velmi náročná i proto, že mnoho genetických determinant může být přeneseno a rozšířeno mezi různé typy bakterií. Pochopení, případně předvídání, vzniku a přenosu antimikrobiální rezistence vyžaduje charakterizaci determinant rezistence mezi patogeny a nepatogeny. Charakterizace celého mikrobiomu je nyní možná díky sekvenování nové generace (next-generation sequencing, high-throughput sequencing), které poskytuje přístup ke genetickým determinantům pro antimikrobiální rezistenci napříč všemi mikrobiálními genomy ve vzorku (tj. metagenomu). Rychlý vývoj v technologiích sekvenování nové

generace zpřístupnil tuto technologii výzkumníkům a mikrobiologům v rutinních laboratořích po celém světě.[24] Tento přístup zdůraznil význam mikrobiální ekologie při vzniku a perzistenci antimikrobiální rezistence, včetně interakce mezi procesy, jako je horizontální přenos genetické informace a křížová selekce.[20]

2.3.1 Databáze pro detekci genů rezistence a horizontálního přenosu

Existuje řada různých databází genů AMR. Různé databáze a jejich různá řešení splňují různé potřeby, proto se do značné míry spíše doplňují, než že si konkurují.[24]

Detekce a charakterizace genetických prvků, které zprostředkovávají AMR, je problematická z mnoha důvodů, z nichž mnohé souvisí s problémy s databázemi AMR. Jednak může mít evoluční dynamika AMR za následek rychle se hromadící sekvenční variace, což znamená, že databáze AMR musí být neustále aktualizovány, aby zahrnovaly nové varianty.

Kromě toho jsou genetické mechanismy antimikrobiální rezistence různé, což znamená, že identifikace genu AMR v rámci metagenomických datových souborů musí brát v úvahu geny AMR na základě mechanismu. Například databáze musí rozlišovat bodové mutace, které modifikují cílová místa antibiotik, oproti genovým sekvencím plné délky, které poskytují rezistenci, když jsou vyjádřeny jako funkční proteiny. To musí být zohledněno v rámci bioinformatického přístupu k identifikaci a počítání determinant rezistence v rámci metagenomických dat. Databáze by podle toho měla být strukturovaná.[20]

ACLAME

ACLAME (A CLAssification of Mobile genetic Elements, Klasifikace mobilních genetických elementů) (<http://aclame.ulb.ac.be>) je databáze prokaryotických mobilních genetických elementů z různých zdrojů, zahrnující velké množství známých genomů bakteriofágů, plazmidů a transpozonů. Kromě informací o kompletních genomech a genetických entitách si klade za cíl vytvořit komplexní klasifikaci funkčních modulů MGE na úrovni proteinů a genů.[25]

První verze této databáze obsahuje rozsáhlou klasifikaci 5069 proteinů ze 119 DNA bakteriofágů do více než 400 funkčních rodin. Tato klasifikace byla vytvořena automaticky a následně byla ručně upravena. Hlavními cíli jsou shromáždit a racionálně organizovat komplexitu spojenou s MGE, rozšířit a zlepšit nedostatečné anotace související s MGE a prověřit známé genomy pro validaci a objev nových MGE.[25]

AcrDB

Bakterie vyvíjejí obranu, aby se chránily před fágů, a fágů vyvíjejí protistrategie, jak tyto obrany obejít. Adaptivní imunitní systém CRISPR-Cas je mechanismus, kterým se bakterie chrání před fágovou infekcí. V reakci na CRISPR-Cas si fágů vyvinuly proteinové inhibitory známé jako anti-CRISPR neboli Acr proteiny. Jsou vytvářeny fágů a dalšími mobilními genetickými elementy, aby inhibovaly systémy CRISPR-Cas jejich hostitelů.[26]

AcrDB (<https://bcb.unl.edu/AcrDB/>) je databáze výpočetně predikovaných anti-CRISPR (Acr) a Acr-asociovaných (Aca) operonů. Ve srovnání s podobnými předchozími databázemi, mezi které patří Anti-CRISPRdb (<https://crisprdb.org>), CRISPRminer (<http://www.microbiome-bigdata.com/CRISPRminer/>) a AcrCatalog (<http://acrcatalog.pythonanywhere.com/catalog/>), se jedná o databázi na genomové úrovni s nejrozsáhlejší sbírkou dat, celkem 39 799 Acr-Aca operonů. Nabízí webové rozhraní s různými funkcemi pro procházení, grafické zobrazení, vyhledávání a stahování Acr-Aca operonů. Zaměřuje se na genomický kontext kandidátů na Acr a Aca místo jednotlivých rodin Acr proteinů a sbírá data pomocí 3 nezávislých unikátních programů, AcrFinder, AcRanker a PaCRISPR.[26]

AcrHub

AcrHub (<http://pacrispr.erc.monash.edu/AcrHub/>) je další databáze, která poskytuje komplexní řešení pro studium, predikci a mapování Acr proteinů. AcrHub obsahuje 339 neredundantních experimentálně ověřených Acr proteinů a více než 70 000 předpovězených Acr proteinů, které byly extrahovány z genomových sekvencí různých prokaryotických organismů a jejich virů.[27]

BacAnt

BacAnt (<http://bacant.net/BacAnt/>) je kombinovaný anotační server pro bakteriální DNA sekvence k identifikaci genů antibiotické rezistence, integronů a transponovatelných elementů. Jako jeden z mála poskytuje databázi pro vyhledávání a anotaci přenosových elementů a genů rezistence současně. Výsledky se dají různě vizualizovat.[28]

Pro BacAnt jsou použity 3 databáze, ResDB (databáze sekvencí genů rezistence), IntegronDB (databáze sekvencí integronů) a TransposonDB (databáze sekvencí transposonů). Do ResDB je shromážděno 5029 sekvencí z databáze referenčních genů bakteriální antimikrobiální rezistence NCBI, do IntegronDB je zahrnuto 1094 sekvencí z INTEGRALL (<http://integrall.bio.ua.pt>) a do TransposonDB je zahrnuto 234 sekvencí z The Transposon Registry (<https://transposon.lstmed.ac.uk>).[28]

BacFITBase

BacFitBase (<http://bacfitbase.tartaglialab.com>) je manuálně spravovaná databáze bakteriálních genů, která poskytuje informace o jejich významu při infekci hostitelských organismů, měřeném pomocí transpozonové mutagenese. Shromažďuje více než 90 000 záznamů o vlivu genů na bakteriální zdatnost během infekce hostitelských organismů. Data pocházejí z 15 studií zaměřených na transpozonovou mutagenesi a zahrnují 15 patogenních bakterií a 5 hostitelských obratlovců v různých tkáních.[29]

Důležitým cílem je pochopení mechanismů bakteriální infekce pro vývoj nových antimikrobiálních terapií. Transpozonová mutagenese umožňuje měřit zdatnost jednotlivých genů a identifikovat ty, které jsou klíčové pro infekci. Při transpozonové mutagenesi se vytvářejí mutace, které jsou následně testovány při infekci hostitelských organismů. Získávají se data o míře úspěšnosti mutací při infekci, což pomáhá určit, které geny jsou klíčové pro infekci. Zdatnost genů se měří jako poměr frekvence mutací před a po infekci.[29]

BacFITBase poskytuje cenný zdroj informací pro výzkum infekčních procesů. Shromažďuje data z různých studií, která pomáhají lépe porozumět bakteriální invazi do hostitelských tkání a identifikovat geny s významnou rolí v infekčních procesech. Identifikace klíčových genů zapojených do infekčních procesů může vést k novým cílům pro terapeutické zásahy.[29]

CARD

CARD (The Comprehensive Antibiotic Resistance Database, Komplexní databáze antibiotické rezistence) (<http://arpcard.mcmaster.ca>) je bioinformatická databáze genů rezistence, jejich produktů a souvisejících fenotypů. Obsahuje geny rezistence na antibiotika, dezinfekční prostředky a biocidy. Je široce využívána ve výzkumu, diagnostice a monitorování rezistence bakterií.

CARD obsahuje nástroje pro analýzu molekulárních sekvencí, včetně BLAST a softwaru RGI. RGI (Resistance Gene Identifier, Identifikátor genu rezistence) (<https://card.mcmaster.ca/analyze/rgi>) lze použít k predikci rezistomů z proteinových nebo nukleotidových dat na základě homologie a SNP modelů. Rezistom je termín, který se používá pro soubor všech genů a mechanismů, které umožňují mikroorganismům odolat účinkům antimikrobiálních látek. Tvoří širokou škálu genů, enzymů a transportních systémů, které jsou kódovány buď na chromozomu mikroorganismu nebo na mobilních genetických elementech.[30]

CicerSpTEdb

CicerSpTEdb (<http://cicersptedb.easyomics.org>) je webová databáze pro genomovou identifikaci transponovatelných elementů u druhů Cicer, což zahrnuje například druhy cizrny. Tato databáze poskytuje informace o specifických mobilních genetických elementech a genech rezistence, které jsou přítomné v genomických datech těchto rostlin. CicerSpTEdb je užitečným nástrojem pro studium interakcí mezi geny rezistence a mobilními elementy v rámci tohoto konkrétního rodu.[31]

Důležitou součástí databáze je analýza genomického kontextu mobilních genetických elementů. To znamená zkoumání, jak se tyto prvky nacházejí v blízkosti jiných genů, a jaké jsou jejich potenciální interakce. Díky datům o mobilních elementech a genech rezistence z rodu Cicer umožňuje CicerSpTEdb studium evoluce těchto struktur a jejich biologických funkcí. To může přinést cenné poznatky do oblasti interakcí mezi rostlinami a bakteriemi. Bakterie mohou interagovat s rostlinami a být součástí kořenového mikrobiomu. Geny rezistence, které jsou přítomné v těchto bakteriích, mohou být vystaveny tlaku selekce způsobenému použitím pesticidů nebo antibiotik v zemědělské praxi. To může vést k horizontálnímu přenosu těchto genů mezi bakteriemi v rostlinném prostředí.[31]

ConTEdb

ConTEdb (A Comprehensive database of Transposable Elements in conifers, Komplexní databáze transponovatelných elementů u jehličnanů) (<http://genedenovoweb.ticp.net:81/conTEdb/index.php>) je komplexní databáze transponovatelných elementů u jehličnatých stromů. Jehličnaté stromy představují největší a nejrozšířenější skupinu nahosemenných rostlin, které mají značný ekologický a ekonomický význam. Nicméně jejich obrovské a složité genomy komplikují proces sekvenování a zkoumání genů jehličnatých stromů. Webová databáze ConTEdb, která slouží výzkumníkům v tomto oboru, obsahuje celkem 413 423 transponovatelných elementů z různých jehličnanů. Tato identifikace byla provedena kombinací různých metod a vedla k jejich zařazení do 11 133 různých rodin. ConTEdb umožňuje snadno procházet, vyhledávat a stahovat sekvence transponovatelných prvků z databáze. Obsahuje řadu nástrojů pro analýzu, které usnadňují sběr dat o transponovatelných elementech.[32]

GyDB

Databáze GyDB (<https://gydb.org>) má za cíl analyzovat a klasifikovat rozmanitost mobilních genetických elementů, zejména Gypsy retrotranspozonů, v kontextu

evolučních profilů a stromu života. Projekt se zaměřuje na vztahy mezi viry, mobilními genetickými elementy a opakujícími se sekvencemi v genomu. Důraz je kladen na odhalování molekulárních změn, které ovlivňují horizontální přenos, infekce a onemocnění.[33]

Zvláštní pozornost je věnována neredundantním elementům vykazujícím určitou míru rozdílnosti a tomu, jak je lze kolektivně zarovnat nebo propojit. To umožňuje odvodit evoluci elementů, které jsou klasifikovány na základě společné vnitřní oblasti. GyDB tedy umožňuje pochopení taxonomie a evoluční historie prvků, které klasifikuje.[33]

ICEberg

ICEberg (<http://db-mml.sjtu.edu.cn/ICEberg/>) je databáze, která poskytuje komplexní informace o bakteriálních integrativních a konjugativních elementech (ICE). Mnoho ICE nese pravděpodobné determinanty virulence, geny antibiotické rezistence a geny kódující další vlastnosti prospěšné pro bakterie. ICEberg nabízí organizovaný, snadno prozkoumatelný archiv jak předpokládaných, tak experimentálně podporovaných ICE dat. Současně obsahuje podrobnosti o 1032 ICE. ICE často mobilizují integrativní a mobilizovatelné elementy (IME) a cis-mobilizovatelné elementy (CIME), které jsou v databázi také zahrnuty.[37]

ICEberg také poskytuje online nástroj ICEfinder pro predikci ICE nebo IME v sekvencích bakteriálního genomu. ICEfinder byl navržen tak, aby usnadnil rychlou detekci ICE a IME typu T4SS v sekvencích bakteriálního genomu. ICEfinder nejprve detekuje sekvence rekombinačních a konjugačních modulů. Hledá také oblast *oriT*. Poté lokalizuje, filtruje a seskupuje odpovídající gen. Prvky nesoucí gen integrázy, gen relaxázy a genové klastry T4SS jsou považovány za ICE, zatímco prvky bez T4SS, ale s integrázou a relaxázou, jsou označeny jako domnělé IME. ICEfinder se také pokouší detekovat některé konkrétní IME s integrázou a *oriT*, ale bez relaxázy.[37]

ICEberg poskytuje dobrou možnost pro pochopení biologických vlastností ICE, zejména proto, že jejich interakce s mobilizovatelnými elementy může dále podporovat horizontální přenos genů.[37]

INTEGRALL

INTEGRALL (<http://integrall.bio.ua.pt>) je volně dostupný textový vyhledávací systém vyvinutý s cílem shromažďovat a organizovat informace o integronech v jediné databázi. Aktuální verze obsahuje přes 4800 integronových sekvencí a poskytuje veřejné genetické repozitáře pro sekvenční data a nomenklaturu, které vědcům

nabízí snadný a interaktivní přístup k sekvencím DNA integronu, jejich molekulárnímu uspořádání a jejich genetickým souvislostem.[38]

ISfinder

ISfinder (<https://isfinder.biotoul.fr>) je databáze, která poskytuje seznam inzerčních sekvencí izolovaných z bakterií a archea. Obsahuje jejich obecné vlastnosti, jejich DNA a potenciální proteinové sekvence. Zahrnuje také určité transpozony. Vyhledávač umožňuje vyhledání a zobrazení jednotlivých IS nebo skupin IS na základě kombinace jejich obecných vlastností. ISfinder také zahrnuje prohlížeč genomů ISbrowser, který umožňuje vizualizaci IS v určitých genomech, a ISSaga navržený tak, aby pomáhal při anotaci genomu IS.

mMGE

mMGE (<https://mgedb.comp-sysbio.org>) je komplexní databáze extrachromozomálních mobilních genetických elementů (eMGE), zahrnující plazmidy a fágy z lidských metagenomických vzorků. Mohou se pohybovat mezi různými mikroby, hrají důležitou roli v evoluci genomu a formování struktury mikrobiálních komunit. mMGE zahrnuje 517 251 jedinečných eMGE, z toho 424 759 fágů a 92 492 plazmidů. Databáze poskytuje anotace, taxonomickou příslušnost, hostitele a prevalenci eMGE. Projekt má za cíl zvýšit porozumění eMGE a jejich vlivu na mikrobiom.[39]

MobileElementFinder

MobileElementFinder (<https://cge.food.dtu.dk/services/MobileElementFinder/>) je nástroj k identifikaci mobilních genetických elementů a jejich vztahů ke genům antimikrobiální rezistence a faktorům virulence. Byl vyvinut tak, aby umožnil rychlou detekci MGE a jejich genetického kontextu v zarovnaných sekvenčních datech.[34]

MobileElementFinder je provozován jako online služba na serveru CGE. Online verze se liší od instalovatelné tím, že jsou v ní integrované ResFinder, VirulenceFinder nebo PlasmidFinder. Databáze obsažená v MobileElementFinder byla vytvořena z nukleotidových sekvencí shromážděných z dat v NCBI (National Center for Biotechnology Information, Národní centrum pro biotechnologické informace) RefSeq a GenBank s výjimkou syntetických a částečně mobilních prvků. Anotace a metadata byly shromážděny z databází The Transposon Registry (<https://transposon.lstmed.ac.uk>), ISfinder (<https://isfinder.biotoul.fr>) a ICEberg (<http://db-mm1.sjtu.edu.cn/ICEberg/>). MGE jsou detekovány na základě sekvenční podobnosti s databází 4452 známých elementů.[34]

MEGARes

MEGARes (<https://megares.meglab.org>) je komplexní databáze genů antimikrobiální rezistence. Zahrnuje sekvence rezistence pro antimikrobiální léčiva a sekvence pro determinanty rezistence na kovy a biocidy.[20]

Důležitým aspektem databáze MEGARes je specifická anotace genů, které vyžadují přítomnost jednonukleotidových polymorfismů (SNP, single nucleotide polymorphisms) ve specifických lokusech ke vzniku rezistence. SNP je variace v jediném nukleotidu, která je v populaci přítomna v jisté patrné míře. Informace o SNP lze použít k integraci RGI (resistance gene identifier, identifikátor genu rezistence) pro potvrzení přítomnosti aminokyselinových zbytků potřebných k udělení rezistence.[20] RGI lze použít k predikci rezistomů z proteinových nebo nukleotidových dat na základě homologie a SNP modelů[35]. To usnadňuje analýzu SNP pro tyto geny a zároveň zlepšuje přesnost identifikace genu AMR z metagenomických dat[20].

MEGARes je strukturován jako relační databáze. Obsahuje sekvence antimikrobiální rezistence získané stažením a začleněním nových sekvencí z nejnovějších dostupných verzí CARD, Bacterial Antimicrobial Resistance Reference Gene Database (Databáze referenčních genů bakteriální antimikrobiální rezistence) a ResFinder. Výsledná databáze MEGARes 2.0 obsahuje 7 868 unikátních sekvencí antimikrobiální rezistence. Uzly hierarchické ontologie zahrnují čtyři typy antimikrobiálních sloučenin, 57 tříd, 220 mechanismů rezistence a 1 345 skupin genů, které klasifikují 7 868 přístupů.[20]

PHASTER

PHASTER (<https://phaster.ca>) je webový server pro rychlou identifikaci a anotaci profágových sekvencí v bakteriálních genomech a plazmidech. Fágy jsou viry infikující bakterie a jejich studium je důležité nejen z evolučního hlediska, ale také pro porozumění mikrobiálním komunitám a možným léčebným aplikacím. PHASTER je významným vylepšením webového serveru PHAST, kroky v procesu identifikace fágových prvků v PHASTER zůstávají z velké části stejné jako v původním PHAST, ale řada softwarových a hardwarových vylepšení nyní činí PHASTER rychlejším, efektivnějším, vizuálně a uživatelsky přívětivějším.[40]

PlasmidFinder

PlasmidFinder (<https://cge.food.dtu.dk/services/PlasmidFinder/>) je webový nástroj pro detekci a charakterizaci celogenomových sekvencí a celých plazmidových sekvenčních dat z Enterobacteriaceae. Tyto nástroje usnadní bakteriální typizaci na základě návrhů genomů Enterobacteriaceae rezistentních na více léčiv rychlou detekcí známých typů plazmidů. Sekvence replikonu byly shromážděny z 559

plně sekvenovaných plazmidů přítomných v nukleotidové databázi NCBI spojené s Enterobacteriaceae. Databáze PlasmidFinder v současnosti obsahuje 116 replikonových sekvencí, které se shodují alespoň na 80 % nukleotidové identity se všemi replikonovými sekvencemi identifikovanými v 559 plně sekvenovaných plasmidech. PlasmidFinder je provozován jako online služba na serveru CGE.[41]

ResFinder

ResFinder (<https://cge.food.dtu.dk/services/ResFinder/>) je online služba provozována na serveru CGE (Center for Genomic Epidemiology, Centrum pro genomickou epidemiologii). Jejím původním cílem bylo umožnit uživatelům s omezenými bioinformatickými zkušenostmi identifikovat geny antimikrobiální rezistence v datech získaných sekvenováním nové generace.[24]

Aktuální verze ResFinder je postavena na třech Git repozitářích, oddělujících databáze získaných genů (databáze ResFinder), bodové mutace (databáze PointFinder) a aplikaci ResFinder (software ResFinder). To umožňuje uživatelům s bioinformatickými zkušenostmi používat ResFinder offline a lokálně na svých počítačích a také ji podle potřeby začlenit do svých vlastních výzkumů. Pro splnění původního cíle ResFinder je poskytován také uživatelům s omezenými bioinformatickými schopnostmi prostřednictvím jednoho z webových nástrojů CGE.[24]

ResFinder se skládá z různých databází se známými geny, jako základ byla použita databáze CARD. Na základě literárních rešerší byla upravena se snahou zahrnout pouze geny AMR, které byly získány horizontálním přenosem genů. Neustále se vyvíjí a nejnovější funkce zahrnují předpovědi fenotypů na základě identifikovaných genů AMR. Rovněž byly implementovány druhově specifické predikce pro vybrané antimikrobiální látky, aby byly prezentovány pouze klinicky relevantní fenotypy pro tyto druhy a také aby byly zahrnuty fenotypy zprostředkované mutacemi. Srovnání, které bylo provedeno pro ResFinder 4.0, ukázalo shodu mezi genotypově předpovězenými a fenotypově detekovanými fenotypy v celkovém průměru nad 95 %.[24]

TADB

TADB (<http://bioinfo-mml.sjtu.edu.cn/TADB2/>) je databáze poskytující komplexní informace o bakteriálních lokusech toxin-antitoxin (TA) typu II. Obsahuje 6193 lokusů TA typu II v 870 replikonech bakterií a archeí, zahrnujících 105 experimentálně ověřených lokusů TA. Pomáhá také zkoumat genomický kontext předpovězených lokusů TA pro možné faktory virulence, determinanty antimikrobiální rezistence a mobilní genetické elementy pomocí porovnávání s konkrétními veřejnými databázemi. Nástroj TAFinder-Compare umožňuje porovnávat přítomnost daných TA lokusů napříč genomovými sekvencemi.[42]

Toxin-antitoxin systém je mikrobiální mechanismus, ve kterém se vyskytují dva spojené geny, gen pro toxin a gen pro antitoxin. Tyto dva geny jsou často kódovány blízko sebe na stejné či blízké lokaci v genomu bakterií a jiných mikroorganismů. Toxin je obvykle škodlivý protein, který může buňku poškodit nebo zastavit její normální funkce, což vede k různým biologickým účinkům, jako je například zastavení buněčného růstu nebo způsobení buněčné smrti. Na druhé straně antitoxin je protein nebo RNA molekula, která má schopnost neutralizovat působení toxinu. Toxiny a antitoxiny jsou vzájemně stabilizovány a vytvářejí takový systém, který může regulovat různé biologické procesy v mikrobiálních buňkách.[42]

Toxin-antitoxin systémy jsou důležité pro adaptaci a přežití mikroorganismů v různých podmínkách a mohou hrát klíčovou roli v regulaci různých biologických procesů v těchto buňkách. V bakteriích mohou ovlivňovat rezistenci na antibiotika tím, že produkují toxiny, které narušují buněčné procesy citlivé na antibiotika. Tyto systémy také mohou hrát roli v horizontálním přenosu genů, když jsou součástí mobilních genetických elementů, stabilizujících přenesené geny v hostitelských buňkách.[42]

TnCentral

TnCentral (<https://tncentral.ncc.unesp.br>) je databáze anotovaných prokaryotických transponovatelných elementů získaných z různých zdrojů. Obsahuje informace o různých typech transpozonů, včetně jejich strukturálních charakteristik, sekvencí a funkcí, a umožňuje přístup k těmto informacím pro studium evoluce genomů, horizontálního přenosu genů a dalších genetických procesů. Aktuálně obsahuje přes 400 pečlivě anotovaných transponovatelných elementů, zahrnující transpozony rodin Tn3, Tn7, Tn402 a Tn554, složené transpozony, integrony a související inzerční sekvence. Tyto elementy přenáší geny, včetně genů rezistence na antibiotika a těžké kovy, nebo geny zodpovědné za patogenezi u rostlin, páry genů toxin/antitoxin, transkripční faktory a geny účastnící se metabolismu. Přínos TnCentralu spočívá v jeho schopnosti přispět k porozumění mechanismů transpozice transponovatelných elementů, což je klíčové pro vývoj strategií boje proti bakteriální patogenezi.[43]

WASPS

WASPS (<https://archaea.i2bc.paris-saclay.fr/wasps/>) je webový nástroj umožňující vyhledávání podobností proteinových a DNA sekvencí v databázi obsahující všechny kompletně sekvenované přirozené plazmidy bakteriálního, archeálního a eukaryotního původu. Specializované databáze plazmidů jsou vzácné, často zahrnují pouze bakteriální plazmidy a umožňují přístup pouze k vyhledávání podobných sekvencí.[44]

3 Detekce a charakterizace genů rezistence

3.1 Analyzovaný dataset

Do analýzy bylo zahrnuto celkem 452 izolátů bakteriálních anaerobů získaných ze zemědělských zvířat, kuřat a prasat, a jsou to komenzální bakterie, které nejsou cílem při užívání antibiotik. Slouží ale jako rozsáhlé rezervoáry genů rezistence na antibiotika a těžké kovy. Jedná se o draft genomy, genom organismu je tedy sestaven ze scaffoldů, jejichž přesnou polohu v genomu neznáme. To znamená, že v něm některé segmenty chybí, nebo v něm jsou naopak vícekrát, že jsou ve špatném pořadí, nebo v opačné orientaci, a že genom není kompletní.

Většina těchto genomových sekvencí je publikována v NCBI. Seznam všech bakteriálních izolátů je uveden v tabulce `bakterialni_izolaty.csv` v elektronické příloze, ke každému izolátu je uvedeno oficiální jméno v NCBI, BioSample, ID projektu, ID izolátu, říše, kmen, třída, řád a čeleď.

Bakteriální izoláty byly zařazeny do 9 různých kmenů, *Firmicutes* (245 izolátů), *Bacteroidota* (113 izolátů), *Actinobacteriota* (65 izolátů), *Proteobacteria* (14 izolátů), *Fusobacteriota* (7 izolátů), *Desulfobacterota* (5 izolátů), *Verrucomicrobiota* (1 izolát), *Synergistota* (1 izolát) a *Elusimicrobiota* (1 izolát).

3.2 Testování databází

K detekci bylo vyzkoušeno 20 různých databází. Byly vybrány tak, aby pokryly širokou škálu genetických a funkčních aspektů spojených s rezistencí bakterií a s horizontálním přenosem genů. Seznam těchto databází, co je s jejich pomocí možné detekovat a odkaz, je uveden v tabulce 3.1.

Databáze byly otestovány na 5 bakteriálních izolátech, které patří do 4 různých kmenů, *Alistipes onderdonkii* An90 (SAMN06473769) z kmenu *Bacteroidota*, *Bifidobacterium boum* An918 (SAMN15872570) z kmenu *Actinobacteriota*, *Clostridium spiroforme* An158 (SAMN06473626) z kmenu *Firmicutes*, *Escherichia coli* An190 (SAMN06473654) z kmenu *Proteobacteria* a *Streptococcus alactolyticus* ET54 (SAMN15872590) z kmenu *Firmicutes*.

Tab. 3.1: Databáze pro detekci genů rezistence a mobilních genetických elementů

Databáze	Detekce	Odkaz
ACLAME	mobilní genetické elementy, fágy	http://aclame.ulb.ac.be
AcrDB	Acr-Aca proteiny	https://bcb.unl.edu/AcrDB/
AcrHub	Acr-Aca proteiny	http://pacrispr.erc.monash.edu/AcrHub/
BacAnt	geny rezistence, transpozony, integrony	http://bacant.net/BacAnt/
BacFITBase	bakteriální geny během infekce hostitele	http://bacfitbase.tartaglialab.com
CARD	geny rezistence	https://card.mcmaster.ca
CicerSpTEdb	mobilní genetické elementy z rodu Cicer	http://cicersptedb.easyomics.org
ConTEdb	mobilní genetické elementy z jehličnanů	http://genedenovoweb.ticp.net:81/conTEdb/index.php
GyDB	mobilní genetické elementy	https://gydb.org
ICEberg	ICE, IME	http://db-mml.sjtu.edu.cn/ICEberg/
INTEGRALL	integrony	http://integrall.bio.ua.pt
ISfinder	inzerční sekvence	https://isfinder.biotoul.fr
mMGE	mobilní genetické elementy	https://mgedb.comp-sysbio.org
MEFinder	mobilní genetické elementy	https://cge.food.dtu.dk/services/MobileElementFinder/
MEGARes	geny rezistence	https://megares.meglab.org
PHASTER	fágy	https://phaster.ca
PlasmidFinder	plazmidy	https://cge.food.dtu.dk/services/PlasmidFinder/
TADB	toxin-antitoxin systémy	http://bioinfo-mml.sjtu.edu.cn/TADB2/
TnCentral	transpozony	https://tncentral.ncc.unesp.br
WASPS	plazmidy	https://archaea.i2bc.paris-saclay.fr/wasps/

3.2.1 Stažené databáze

MyDbFinder

MyDbFinder (<https://cge.food.dtu.dk/services/MyDbFinder/>) je nástroj na serveru CGE, který identifikuje geny a MGE z vaší vlastní databáze v úplně nebo částečně sekvenovaných izolátech bakterií. Nahraná databáze musí být FASTA soubor se sekvencemi DNA. Nastaví se práh pro shodu mezi geny nebo MGE v databázi a ve vstupní genomové sekvenci a také délka genu/MGE v genomu, který pokrývá délku genu/MGE v databázi.

Na výstupu je poté tabulka, ve které jsou názvy genů/MGE, které byly získané z FASTA hlaviček sekvencí, proti kterým byl zarovnán vstupní genom, procento identity ve srovnání mezi nejlépe odpovídajícím genem/MGE v nahrané databázi a odpovídající sekvencí ve vstupním genomu, dále je zde uvedena délka genu/MGE v databázi a délku zarovnání mezi genem/MGE v databázi a odpovídající sekvencí v genomu, informace o názvu scaffoldu, ve kterém byl gen/MGE nalezen, a jeho pozice na daném scaffoldu. Různé barvy řádků slouží k vizualizaci shody mezi sekvencemi. Tmavě zelená značí 100 % shodu a pokrytí celé délky referenční sekvence, světle zelená znamená, že délka zarovnání se shoduje s délkou sekvence v databázi, ale není mezi nimi 100 % shoda, a šedá barva znamená, že zarovnání je kratší než délka sekvence v databázi, ale je mezi nimi 100 % shoda. Červená barva značí, že nebyl nalezen žádný gen/MGE se shodou přes nastavený práh. Rozšířený výstup ukazuje detailní zarovnání všech shodných sekvencí, kde zelená značí shodující se nukleotidy, červená nesrovnalosti a šedá absence nukleotidů v zarovnání. Tabulka může být stažena ve formátu TSV, dále je k dispozici FASTA soubor obsahující nejlépe odpovídající sekvence z testovaného genomu a TXT soubor obsahující výslednou tabulku a zarovnání sekvencí.

BacAnt, MEGARes

BacAnt je webový server, který, jako jeden z mála, umožňuje anotaci genů rezistence, integronů a transponovatelných elementů současně. Na vstup ale nelze nahrát multi-FASTA soubor s více sekvencemi, pouze FASTA soubor s 1 sekvencí. Rozdělit multi-FASTA soubor by bylo vzhledem k tomu, že je potřeba otestovat draft genomy s desítkami sekvencí, velmi nevýhodné. Proto byly z BacAnt serveru staženy databáze Integron.fasta (<http://bacant.net/ftp/v3.1/IntegronDB>) a Transposon.fasta (<http://bacant.net/ftp/v3.1/TransposonDB>), které poskytují informace o integronech a transpozonech. MEGARes je databáze genů rezistence na těžké kovy, biocidy a antibiotika, která neposkytuje žádný webový nástroj pro detekci těchto genů, databáze proto byla také stažena. Poté byl na detekci genů rezistence a MGE

z těchto databází použít MyDbFinder. Práh pro shodu mezi genem/MGE v databázi a ve vstupní genomové sekvenci byl nastaven na identitu 90 % a délka genu/MGE v genomu, který pokrývá délku genu/MGE v databázi, na 60 %.

Z databáze BacAnt bylo u bakterie *Escherichia coli* An190 detekováno 6 transpozonů, u bakterie *Streptococcus alactolyticus* ET54 byl detekován 1 transpozon, u ostatních testovaných bakterií nebyly detekované žádné transpozony, ani nebyly detekovány žádné integrony. Z databáze MEGARes byl u bakterie *Alistipes onderdonkii* An90 detekován gen *tet(Q)* a u bakterie *Bifidobacterium boum* An918 gen *tet(W)*, které kódují rezistenci na tetracykliny, u bakterie *Clostridium spiroforme* An158 byly detekovány 2 geny *dfrF* a *ANT6* kódující rezistenci na trimethoprimy a aminoglykosidy, u bakterie *Escherichia coli* An190 bylo detekováno 177 různých genů rezistence na různé druhy antibiotik, těžkých kovů a biocidů, a u bakterie *Streptococcus alactolyticus* ET54 bylo detekováno 7 různých genů rezistence na různé druhy antibiotik.

3.2.2 Online databáze

CARD

RGI databáze CARD je bioinformatická databáze genů rezistence, jejich produktů a souvisejících fenotypů. Na vstup může být nahrán soubor jak ve FASTA formátu, který může obsahovat DNA i proteinové sekvence, tak GenBank formátu. Velikou výhodou této databáze je, že je možné nahrát na vstup několik FASTA souborů najednou, což při analýze velkého souboru genomů významně usnadňuje práci a šetří čas. Poté se zvolí, jak přesné výsledky jsou požadované. Byly zvoleny pouze přesné shody, vysoká kvalita a coverage.

Na výstupu je poté tabulka, která může být stažena v JSON a TXT souboru. Tabulka obsahuje sloupce s informacemi o přesnosti detekce, názvu genu rezistence, podle jakého kritéria byl gen detekován, do jaké skupiny AMR genů gen patří, dále antibiotikum, na které gen kóduje rezistenci, mechanismus rezistence, a nakonec procento shody zarovnaných sekvencí a procento délky sekvence oproti referenčnímu genu.

Byly vybrány pouze výsledky, které vykazovaly shodu mezi sekvencemi nad 90 %. U bakterie *Alistipes onderdonkii* An90 byl detekován gen *tet(Q)* kódující rezistenci na tetracykliny pomocí modifikace cílové struktury antibiotika, u bakterie *Bifidobacterium boum* An918 byly detekovány 2 geny *tet(W)* a *rpoB* kódují rezistenci na tetracykliny a rifamyciny, u bakterie *Clostridium spiroforme* An158 byl detekován gen *dfrF* kódující rezistenci diaminopyrimidiny, u bakterie *Escherichia coli* An190 bylo detekováno 53 různých genů kódujících rezistenci na 19 různých skupin antibiotik a u bakterie *Streptococcus alactolyticus* ET54 bylo detekováno 5 různých genů

na 4 různé skupiny antibiotik.

ICEberg

Databáze ICEberg poskytuje komplexní informace o bakteriálních integrativních a konjugativních elementech (ICE). Na vstupu je soubor s nukleotidovými sekvencemi ve FASTA formátu, nahrán může být také GenBank soubor, poté se jen zvolí, o genom jakého organismu se jedná.

Na výstupu je souhrnná tabulka se seznamem všech predikovaných ICE a IME v bakteriálním genomu, s jejich lokací, délkou a typem ICE nebo IME. Po rozkliknutí konkrétních ICE nebo IME v tabulce se zobrazí detailnější informace pro konkrétní element, kde je zobrazena také mapa pro vizualizaci umístění detekovaných ICE a IME v rámci genomu, seznam genů obsažených v predikovaných ICE a IME a proteinové a DNA sekvence těchto genů. Výsledky se ale bohužel nedají stáhnout v žádném takovém formátu, aby se s nimi mohlo dále efektivně pracovat, a aby bylo možné je používat pro další analýzu.

U bakterie *Alistipes onderdonkii* An90 byly detekovány 2 IME a 1 ICE s T4SS, u bakterií *Bifidobacterium boum* An918 a *Clostridium spiroforme* An158 byly detekovány 2 ICE s T4SS, u bakterie *Escherichia coli* An190 byly detekovány 2 ICE s T4SS a 1 IME a u bakterie *Streptococcus alactolyticus* ET54 byl detekován 1 mobilní ICE s T4SS.

ISfinder

Databáze ISfinder poskytuje seznam inzerčních sekvencí izolovaných z bakterií a archea. Na vstupu je soubor s nukleotidovými sekvencemi ve FASTA formátu. Dále je zde možné nastavit, v jakém formátu budou výsledky zobrazeny, jestli v podobně různě formátované tabulky, nebo zarovnání sekvencí ukazující shodu mezi nimi.

Na výstupu je zobrazena tabulka s detekovanými inzerčními sekvencemi, jejich původ a E-hodnota, která udává pravděpodobnost, že podobnost mezi predikovaným genomem a databází je náhodná, a není důkazem opravdové biologické podobnosti nebo homologie, čím nižší má hodnotu, tím relevantnější výsledky jsou. Dále zde podle vstupního nastavení uživatele může být zobrazeno zarovnání sekvencí ukazující shodu mezi nimi. Výsledky z této databáze se ale bohužel nedají stáhnout v žádném takovém formátu, aby se s nimi mohlo dále efektivně pracovat, a aby bylo možné je používat pro další analýzu.

V průběhu analýzy testovaných bakterií bylo identifikováno velké množství inzerčních sekvencí. Nicméně pouze u menšího počtu z nich byly dosaženy E-hodnoty, které naznačovaly statisticky významné a relevantní výsledky.

MobileElementFinder

MobileElementFinder je nástroj k identifikaci mobilních genetických elementů. Na vstupu je soubor s nukleotidovými sekvencemi ve FASTA formátu. Poté se zde můžou zvolit databáze pro anotaci doplňkových genů, na výběr je ResFinder pro geny antimikrobiální rezistence a VirulenceFinder pro geny virulence. Pro každý nástroj jsou použity defaultní parametry. Ve výchozím nastavením jsou vybrány všechny databáze.

Na výstupu je umožněno filtrovat výsledky několika způsoby, které zahrnují omezení výsledků na určité typy MGE, minimální povolenou shodu sekvencí, coverage zarovnání a maximální povolené zkrácení. Výsledky zobrazují, které MGE byly predikovány pro daný scaffold. Predikované plazmidy, geny rezistence a virulence jsou zobrazeny pod hlavičkou MGE spolu s jejich kvalitou predikce a metadaty. Informace o detekovaných MGE zahrnují synonymní jména, informace o skupině, do které patří, a také odkazy na sekvence a externí databáze, ze kterých byly identifikovány. Dále je zde coverage zarovnání, identita s referenční sekvencí, počet substitucí a také umístění na scaffoldu, pokud byl element zkrácen, zobrazí se úroveň zkrácení. Predikce nízké kvality jsou označeny a ve výsledcích je zvýrazněna kvalita předpovědi. Detekované MGE je možné stáhnout jako tabulku ve formátu TSV, která obsahuje název MGE, metodu predikce, typ MGE, identitu sekvencí a coverage, scaffold a konkrétní pozici na scaffoldu. Jejich DNA sekvence také mohou být staženy ve FASTA formátu.

U bakterií *Bifidobacterium boum* An918 a *Streptococcus alactolyticus* ET54 byla detekována 1 inzerční sekvence, u bakterie *Escherichia coli* An190 bylo detekováno 19 různých MGE, které zahrnovaly různé inzerční sekvence, transpozony, kompozitní transpozony a miniaturní invertované repetice, a u bakterií *Alistipes onderdonkii* An90 a *Clostridium spiroforme* An158 nebyl detekován žádný MGE.

PHASTER

PHASTER umožňuje identifikaci a anotaci profágových sekvencí. Na vstupu je soubor s nukleotidovými sekvencemi ve FASTA formátu, pokud se ale v souboru nachází více sekvencí, je nutné zaškrtnout políčko, které informuje o tom, že se jedná o multi-FASTA soubor. Je možné nahrát také GenBank soubor.

Na výstupu se zobrazí tabulka, ve které je uvedena délka sekvence, predikce, zda sekvence obsahuje celý nebo neúplný profág na základě různě vypočítaných kritérií a skóre, počet proteinů přítomných v sekvenci, které obsahuje predikovaný fág, počáteční a koncová poloha na scaffoldu, na kterém se predikovaný fág nachází, poté jsou zde uvedeny fágy s nejvyšším počtem proteinů nejpodobnějších těm v sekvenci, a nakonec procento GC nukleotidů. Tato tabulka může být stažena v TXT

formátu. Dále je možné stáhnout PNG obrázek s cirkulárním zobrazením genomu, na kterém jsou vyznačeny predikované fágy, a jiná genomová zobrazení.

U bakterií *Alistipes onderdonkii* An90 a *Clostridium spiroforme* An158 byl predikován 1 neúplný fág, u bakterie *Bifidobacterium boum* An918 nebyl predikován žádný fág, u bakterie *Escherichia coli* An190 bylo predikováno 10 fágů, z toho jeden byl označen jako celý, a u bakterie *Streptococcus alactolyticus* ET54 byly detekovány 3 nekompletní fágy. Většina predikovaných fágů byla neúplných, to je u draft genomů běžné.

PlasmidFinder

PlasmidFinder je webový nástroj, který umožňuje detekci plazmidů. Na vstupu je soubor s nukleotidovými sekvencemi ve FASTA formátu. Práh pro shodu mezi sekvencemi v databázi a ve vstupní genomové sekvenci byl nastaven na 90 % a délka elementu v genomu, která pokrývá délku elementu v databázi, byla nastavena na 60 %.

Na výstupu je poté tabulka, ve které je seznam detekovaných plazmidů, je zde uvedena databáze, ze které byl plazmid detekován, název plazmidu, procento identity ve srovnání mezi nejlépe odpovídajícím plazmidem v databázi a odpovídající sekvencí ve vstupním genomu, délka plazmidu v databázi a délka zarovnání mezi plazmidem v databázi a odpovídající sekvencí v genomu, dále je zde uvedena informace o názvu scaffoldu, ve kterém byl plazmid nalezen, a pozice na daném scaffoldu. Rozšířený výstup ukazuje detailní zarovnání všech shodných sekvencí. Barevné značení výsledků nese stejnou informaci jako u MyDbFinder a jiných nástrojů provozovaných na serveru CGE.

U bakterie *Escherichia coli* An190 byly detekovány 4 plazmidy, u bakterie *Streptococcus alactolyticus* ET54 byl detekován 1 plazmid, u ostatních testovaných bakterií žádné plazmidy nalezené nebyly.

TnCentral

TnCentral je databáze anotovaných prokaryotických transponovatelných elementů. Na vstupu je soubor s nukleotidovými sekvencemi ve FASTA formátu. Dále je zde možné nastavit, v jakém formátu budou výsledky zobrazeny, jestli v podobně různě formátované tabulky, nebo zarovnání sekvencí ukazující shodu mezi nimi.

Na výstupu je zobrazena tabulka s detekovanými transponovatelnými elementy, jejich původ a E-hodnota, která udává pravděpodobnost, že podobnost mezi predikovaným genomem a databází je náhodná, a není důkazem opravdové biologické podobnosti nebo homologie, čím nižší má hodnotu, tím relevantnější výsledky jsou.

Dále zde podle vstupního nastavení uživatele může být zobrazeno zarovnání sekvencí ukazující shodu mezi nimi. Výsledky z této databáze se ale nedají stáhnout v žádném takovém formátu, aby se s nimi mohlo dále efektivně pracovat, a aby bylo možné je používat pro další analýzu.

V průběhu analýzy testovaných bakterií byly identifikovány desítky různých transponovatelných elementů. Pouze u menšího počtu z nich byly dosaženy E-hodnoty, které naznačovaly statisticky významné a relevantní výsledky.

BacFITBase, GyDB, INTEGRALL, mMGE, WASPS

Databáze BacFITBase poskytuje informace o významu bakteriálních genů při infekci hostitelských organismů, GyDB a mMGE o různých mobilních genetických elementech, INTEGRALL o integronech a WASPS o různých plazmidech. Tyto informace jsou významné v kontextu horizontálního přenosu genetické informace mezi bakteriemi.

Nicméně nahrávání vstupních sekvencí do těchto databází vykazuje různá omezení a požadavky. Databáze BacFITBase a INTEGRALL vyžadují vložení jediné sekvence ve formátu FASTA. U databáze GyDB je sice deklarován formát multi-FASTA, při nahrávání souboru se ale vždy objeví chybová zpráva, že je soubor příliš velký. U databáze mMGE lze nahrát FASTA soubor s 1 DNA sekvencí a u databáze WASPS je možné nahrát FASTA soubor s 1 DNA nebo proteinovou sekvencí.

Tato omezení lze obejít například rozdělením multi-FASTA souboru na soubory s jednotlivými scaffolds, to je ale při analýze rozsáhlého souboru genomů velice neefektivní. V těchto případech je vhodné zvážit alternativní přístupy nebo využití jiné nástroje a databáze, které jsou schopny lépe zpracovat a analyzovat rozsáhlé soubory genomických dat.

ACLAME, AcrHub, CicerSpTEdb, ConTEdb

Výsledky se nepovedlo získat ze všech uvedených databází. Během opakovaných pokusů o přístup k databázím ACLAME, AcrHub, CicerSpTEdb a ConTEdb se objevily technické potíže, které zabránily úspěšnému načtení stránek nebo připojení k serverům. Důvody nejsou zcela jisté. Webové stránky a databáze mohou být ovlivněny různými faktory, včetně změn v provozu, technických problémů nebo přesunů serverů. Stane se, že tvůrci databáze změní své umístění nebo infrastrukturu, což může mít vliv na dostupnost a fungování stránky.

3.3 Metody detekce a charakterizace genů rezistence

3.3.1 Vybrané databáze

Detekce genů rezistence a horizontálního přenosu u souboru 452 bakteriálních genomů byla provedena pomocí 6 databází, CARD, MEGARes, MobileElementFinder, PlasmidFinder, BacAnt a PHASTER. Tyto databáze byly vybrány s cílem pokrýt co největší spektrum různých genů rezistence a různých mobilních genetických elementů. Dalším kritériem byla možnost provádět analýzu na multi-FASTA souborech a možnost výsledky z databází stáhnout a využít tak pro další analýzu a charakterizaci.

Kombinací těchto databází můžeme získat komplexní pohled na horizontální přenos genů rezistence v bakteriální populaci. Tato strategie nám umožní provést analýzu a charakterizaci genů rezistence na antibiotika a těžké kovy a jejich interakcí s různými typy mobilních genetických elementů v anaerobních bakteriích.

K detekci genů rezistence na antibiotika a těžké kovy byly použity databáze CARD a MEGARes. Databáze CARD byla použita k detekci genů rezistence na antibiotika, databáze MEGARes byla použita k detekci genů rezistence na těžké kovy a biocidy, ale také na antibiotika.

K detekci různých genetických elementů a fágů byly použity databáze MobileElementFinder, PlasmidFinder, BacAnt a PHASTER. Databáze MobileElementFinder byla použita k detekci různých MGE, jako inzerční sekvence, transpozony, integrativní konjugativní a mobilizovatelné elementy a miniaturní invertované repetice. Databáze PlasmidFinder byla použita k detekci plazmidů. Databáze BacAnt byla použita k detekci transpozonů a integronů. A databáze PHASTER byla použita k detekci fágů.

3.3.2 Unifikace výsledků

Po získání výsledků z vybraných databází byla navržena metoda, pomocí které se výsledky s různými formáty unifikovaly do jednotného formátu. Navržená metoda byla implementována v programovacím jazyce Python.

Výsledky z databází CARD a MEGARes byly unifikovány do tabulky AMR_CARD_MEGARes_done.csv pomocí kódu unifikace_vysledku_AMR_geny.py. V tabulce je uveden název bakterie, která gen obsahuje, scaffold a konkrétní start a stop pozice na scaffoldu, na kterém se gen nachází, název genu rezistence, druh látky a konkrétní látka, na kterou gen poskytuje rezistenci, a nakonec názvy databází, ze kterých byl gen detekován. Téměř všechny geny detekované pomocí databáze CARD se shodovaly s geny detekovanými pomocí databáze MEGARes, která ale v porovnání s databází CARD poskytla mnohonásobně větší množství výsledků.

Výsledky z databází MobileElementFinder, PlasmidFinder, BacAnt a PHASTER byly unifikovány do tabulky MGE_MEF_PF_PHASTER_BacAnt_done.csv pomocí kódu unifikace_vysledku_MGE_fag.py. V tabulce je uveden název bakterie, která element obsahuje, scaffold a konkrétní start a stop pozice na scaffoldu, na kterém se element nachází, typ a název elementu, a nakonec název databáze, ze které byl element detekován. Několik transpozonů detekovaných pomocí databáze MobileElementFinder a BacAnt se shodovalo, mimo to ale každá databáze poskytovala unikátní výsledky s informacemi o různých typech MGE, které se doplňovaly.

Tabulky a kódy jsou součástí elektronické přílohy.

3.3.3 Přenos genů rezistence

Po unifikaci výsledků byla navržena metoda pro analýzu a charakterizaci přenosu genů rezistence mezi bakteriemi. Navržená metoda byla implementována v programovacím jazyce Python ve skriptu prenos_genu.py, který je součástí elektronické přílohy.

Přenos genů rezistence byl hodnocen dvěma přístupy. První metoda zkoumá, zda jsou detekované geny rezistence obsaženy v identifikovaných mobilních genetických elementech, které jsou schopny tyto geny přenášet. Druhý přístup analyzuje umístění genů rezistence na scaffoldu. Pokud se tyto geny nacházejí na stejném scaffoldu za sebou a tato struktura se opakuje u více bakterií, předpokládáme, že se geny přenášejí společně.

Kód vytvoří TXT soubor se seznamem všech genů, které jsou součástí mobilních genetických elementů nebo fágů, a seznamem všech genů, které se nachází na scaffoldu těsně za sebou, a které se pravděpodobně přenášejí společně. Pro bezchybné fungování kódu je nutné mít ve stejné složce uložen skript prenos_genu.py a soubory AMR_CARD_MEGARes_done.csv a MGE_MEF_PF_PHASTER_BacAnt_done.csv, které obsahují unifikované výsledky z vybraných databází. Kompletní sada funkcí je popsána přímo ve skriptu. Aby byl vygenerován výsledný soubor ve formátu TXT, stačí skript spustit. Vygenerovaný soubor prenos_genu.txt je také součástí elektronické přílohy.

3.3.4 Přehled získaných výsledků pro vybranou bakterii

V programovacím jazyce Python byl navržen kód, který vytvoří TXT soubor s veškerými informacemi získanými pomocí databází CARD, MEGARes, MobileElementFinder, PlasmidFinder, BacAnt a PHASTER. Pro bezchybné fungování kódu je nutné mít ve stejné složce uložen skript informace_o_bakterii.py, a soubory AMR_CARD_MEGARes_done.csv a MGE_MEF_PF_PHASTER_BacAnt_done.csv, které obsahují unifikované výsledky z vybraných databází. Kompletní sada funkcí

je popsána přímo ve skriptu. Aby byl vygenerován výsledný soubor ve formátu TXT, stačí zadat na vstup funkce main jméno bakterie, pro kterou chceme tyto výsledky získat. Součástí elektronické přílohy je pro ukázkou přiložen TXT soubor `info_o_bakterii_An190.txt` vygenerovaný pro bakterii *Escherichia coli An190*

Vytvořený TXT soubor, který je vždy pojmenovaný podle ID bakteriálního izolátu, obsahuje následující informace: po uvedení jména bakterie následuje seznam všech detekovaných genů rezistence na antibiotika a těžké kovy, pro každý gen je zde uvedena jeho poloha na scaffoldu, druh látky a konkrétní látka, na níž gen poskytuje rezistenci. Dále následuje seznam identifikovaných MGE, kde je uvedena poloha MGE na scaffoldu, jeho jméno a typ. Na závěr jsou prezentovány informace o přenosu genů rezistence. Ty zahrnují seznam MGE, které mají ve své struktuře geny rezistence, které přenášejí, a také seznam genů rezistence, které se nacházejí za sebou na stejném scaffoldu, a pravděpodobně tedy putují mezi bakteriemi spolu.

3.4 Výsledky detekce a charakterizace genů rezistence

3.4.1 Geny rezistence

Porovnání s databázemi CARD a MEGARes ukázalo, že téměř polovina bakteriálních izolátů (221 izolátů, 48,9 % z celkových 452 izolátů) obsahovala alespoň 1 gen rezistence na antibiotika nebo těžké kovy.

V 221 bakteriálních izolátech bylo celkově detekováno 867 genů rezistence na antibiotika a těžké kovy. Jednalo se o 252 různých genů. Gen *tet(W)* byl nejčastěji detekovaný gen ze všech. Byl detekován v 73 izolátech. Poskytuje rezistenci na tetracykliny. Další hojně detekované geny byly *tet(Q)* (ve 39 izolátech), *ANT6* (ve 32 izolátech), *Mef(En2)* (ve 26 izolátech), *lnuA* (ve 26 izolátech) a *tet(O)* (ve 24 izolátech), *lnuC* (ve 23 izolátech) a *ANT9* (ve 20 izolátech). Ostatní geny se vyskytovaly v méně než 13 izolátech.

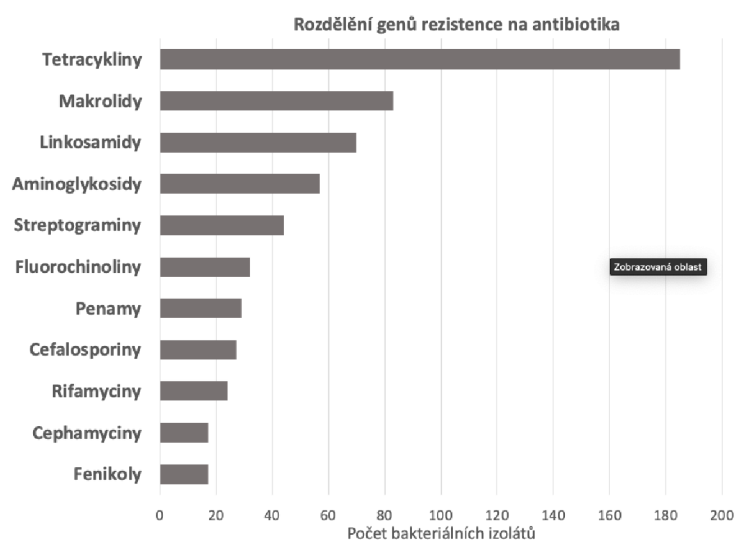
Nejvíce genů rezistence bylo detekováno v bakteriálních izolátech *Escherichia*, které patří do kmene *Proteobacteria*, třídy *Gammaproteobacteria*, řádu *Enterobacterales* a čeledi *Enterobacteriaceae*. Izolát *Escherichia coli An190* obsahoval 180 různých genů rezistence, izolát *Escherichia coli An786* obsahoval 158 různých genů rezistence a izolát *Escherichia fergusonii ET78* obsahoval 106 různých genů rezistence na různé skupiny antibiotik, těžkých kovů a biocidů. Ostatní izoláty obsahovaly méně než 17 různých genů rezistence.

Je nutné podotknout, že to, že je v genomu bakterie detekovaný gen rezistence, nemusí znamenat, že je gen funkční. Aby byl gen aktivní, musí být exprimovaný. Exprese genu by se měla experimentálně ověřit. To se ale ve většině genomických studiích nedělá, *in silico* analýzy to pouze nějakým způsobem predikují. Naopak to,

že v genomu bakterie nebyl detekovaný žádný gen rezistence, nemusí znamenat, že bakterie nemá žádný takový gen. Znamená to jen, že nic v použité databázi genů rezistence rezistenci nenaznačuje, a že gen není součástí databáze.

3.4.2 Antibiotika a těžké kovy

V bakteriálních izolátech byly detekovány geny, které poskytovaly rezistenci na 53 různých skupin antibiotik, těžkých kovů a biocidů. 221 izolátů obsahovalo gen rezistence na antibiotika, 6 izolátů obsahovalo gen rezistence na více druhů látek, 3 izoláty obsahovaly gen rezistence na těžké kovy a 3 izoláty obsahovaly gen rezistence na biocidy. Zastoupení antibiotik, na které jsou bakterie nejčastěji rezistentní, je zobrazeno v grafu na obrázku 3.1,



Obr. 3.1: Rozdělení genů rezistence na antibiotika. V grafu jsou zobrazeny jednotlivé skupiny antibiotik, v porovnání s počtem bakteriálních izolátů, u kterých byly detekovány geny, které kódují rezistenci na tato antibiotika.

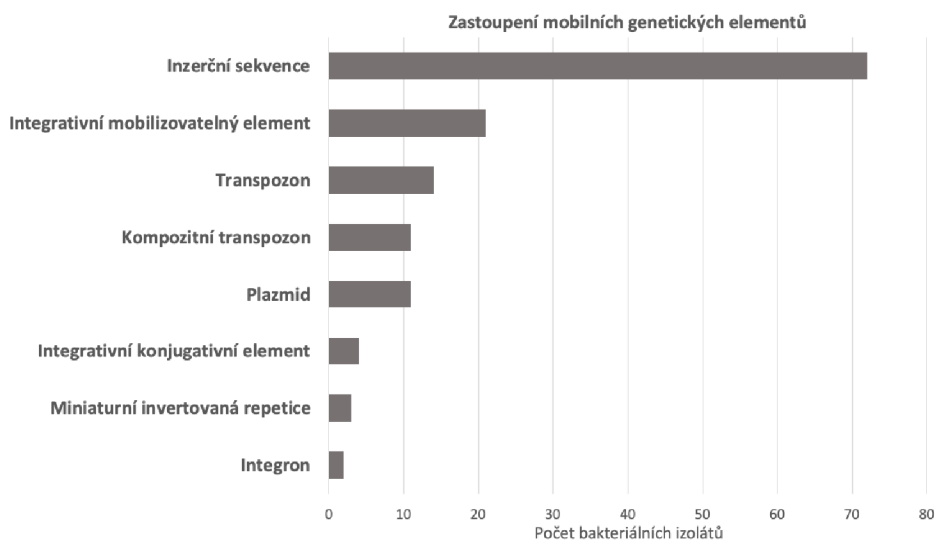
Nejvíce detekovaných genů poskytovalo rezistenci na tetracykliny. Tetracykliny jsou skupinou širokospektrálních antibiotik, která se používají k léčbě infekcí způsobených různými bakteriemi. Patří mezi semisyntetická antibiotika odvozená z přírodního zdroje. Inhibují bakteriální růst tím, že se vážou na ribozomy. Interferují tak s procesem syntézy proteinů. Blokováním ribozomální funkce zabraňují bakteriím v tvorbě nových proteinů a tím brání jejich růstu a množení.[45]

Celkem 185 bakteriálních izolátů (40,9 % z celkových 452) obsahovalo alespoň 1 gen rezistence na tetracykliny. Rezistenci na tetracykliny poskytovalo 32 různých genů. Gen *tet(W)* byl detekován v 73 izolátech, *tet(Q)* byl detekován ve 39 izolátech, *tet(O)* byl detekován ve 24 izolátech, *tet(O/W)*, *tet(40)* a *tet(A)* byly detekovány v 6

izolátech a *tet(M)* a *tet(B(P))* byl detekován v 5 izolátech, zbytek genů rezistence na tetracykliny byl detekován v méně než 4 izolátech.

3.4.3 Mobilní genetické elementy

Porovnání s databázemi MobileElementFinder, PlasmidFinder, BacAnt a PHAS-TER ukázalo, že téměř všechny bakteriální izoláty (424 izolátů, 93,8 % z celkových 452 izolátů) obsahovaly alespoň 1 mobilní genetický element nebo fág. V bakteriálních izolátech bylo detekováno 7 typů MGE. Jejich zastoupení je zobrazeno v grafu na obrázku 3.2.



Obr. 3.2: Zastoupení mobilních genetických elementů. V grafu jsou zobrazeny jednotlivé typy mobilních genetických elementů, v porovnání s počtem bakteriálních izolátů, u kterých byly detekovány.

Nejčastěji detekovaným typem MGE byly inzerční sekvence, které byly detekovány u 72 bakteriálních izolátů. Inzerční sekvence neobsahují geny rezistence. Často se ale nacházely před nebo za detekovaným genem rezistence. Jejich invertované repetice slouží jako silné promotory. Zvyšují tak expresi genů nacházejících se za inzerčními sekvencemi. Tím hrají významnou roli v šíření antimikrobiální rezistence. Mohou také podporovat mobilitu vytvářením kompozitních transpozonů.

Integrativní mobilizovatelné elementy byly detekovány ve 21 izolátech, transpozony byly detekovány ve 14 izolátech, kompozitní transpozony a plazmidy byly detekovány v 11 izolátech, integrativní konjugativní elementy ve 4 izolátech, miniaturní invertované repetice ve 3 izolátech a integrony ve 2 izolátech. Celkem 419 bakteriálních izolátů obsahovalo predikovaný fág, téměř všechny z nich ale byly necelé, což je u draft genomů běžné.

3.4.4 Přenos genů rezistence

Jedním z přístupů, pomocí kterého byl analyzován přenos genů rezistence, bylo zkoumání, zda jsou detekované geny rezistence součástí detekovaných MGE.

Kompozitní transpozon *cn_4228_ISBbi1* obsahoval 3 geny rezistence, *sul2*, *APH3* a *APH6*, transpozon *Tn10* obsahoval také 3 geny rezistence, *tet(B)*, *tet(D)* a *tet(R)*, u bakterie *Escherichia coli An190* byl detekován intaktní fág, který také obsahoval 3 geny rezistence, *cpxA*, *CpxAR* a *sodA*. Dále byly detekovány různé transpozony, jeden integron a několik neúplných fágů, které ve své struktuře také obsahovaly geny rezistence.

Nejčasteji detekovaným MGE s genem rezistence byl integrativní mobilizovatelný element *MTnSag1*, který přenáší gen *lnuC*. Byl detekován u 21 různých bakterií.

Dalším přístupem, pomocí kterého byl analyzován přenos genů rezistence, bylo zkoumání polohy genů rezistence na scaffoldu, které geny rezistence se nachází na scaffoldech poblíž sebe a jestli se tyto nalezené paterny opakují u více bakterií.

Geny *Mef(En2)* a *lnuA* se nacházely za sebou v těsné blízkosti na scaffoldu, tento patern byl zaznamenán u 26 různých bakterií. Geny tedy s určitou pravděpodobností putují mezi bakteriemi společně. Geny *tet(A)* a *tet(B(P))* byly detekovány na scaffoldu u 4 různých bakterií. Dále bylo nalezeno 6 takových paternů, které se opakovaly u 3 různých bakterií, 9 různých paternů nalezených u 2 různých bakterií, 46 různých paternů nalezených pouze u 1 bakterie

Všechny tyto výsledky jsou zaznamenány v souboru *prenos_genu.txt*, který byl vygenerovaný pomocí skriptu *prenos_genu.py*, v externí příloze.

4 Diskuze

Šíření antimikrobiální rezistence horizontálním přenosem genetické informace je jednou z nejzávažnějších výzev současné medicíny. Tato práce se zabývala šířením genů antimikrobiální rezistence mezi komenzálními bakteriemi izolovanými ze zemědělských zvířat. Do analýzy bylo zahrnuto celkem 452 izolátů bakteriálních anaerobů, zařazených do 9 různých kmenů, *Firmicutes* (245 izolátů), *Bacteroidota* (113 izolátů), *Actinobacteriota* (65 izolátů), *Proteobacteria* (14 izolátů), *Fusobacteriota* (7 izolátů), *Desulfobacterota* (5 izolátů), *Verrucomicrobiota* (1 izolát), *Synergistota* (1 izolát) a *Elusimicrobiota* (1 izolát).

I přes to, že se jedná o bakterie, které nikdy nebyly cílem antimikrobiální terapie, výsledky ukazují, že téměř polovina bakteriálních izolátů (221 izolátů, 48,9 % z celkových 452 izolátů) obsahovala alespoň 1 gen rezistence na antibiotika nebo těžké kovy. Téměř všechny z nich (424 izolátů, 93,8 % z celkových 452 izolátů) obsahovaly alespoň 1 mobilní genetický element, se kterým se může gen rezistence přenášet. Komenzální bakterie jsou významným rezervoárem genů rezistence.

K detekci bylo vyzkoušeno 20 různých databází. Byly vybrány tak, aby pokryly širokou škálu genetických a funkčních aspektů spojených s rezistencí bakterií a s horizontálním přenosem genů. Jen některé z databází se ale ukázaly jako efektivní nástroje pro bioinformatickou širokospektrální analýzu.

U databází ACLAME, AcrHub, CicerSpTEdb a ConTEdb byl problém s připojením k serveru. Důvod není zcela jistý, webové stránky a databáze mohou být ovlivněny různými faktory, včetně změn v provozu, technických problémů nebo přesunů serverů. Stane se, že tvůrci databáze změní své umístění nebo infrastrukturu, což může mít vliv na dostupnost a fungování stránky.

Databáze BacFITBase, GyDB, INTEGRALL, mMGE a WASPS byly zhodnoceny jako nevhodné, protože se s jejich pomocí nedá analyzovat multi-FASTA soubory s draft genomy bakterie. Toto omezení by se dalo vyřešit rozdělením multi-FASTA souboru na jednotlivé fasta sekvenvece a provádět analýzu postupně u jednotlivých scaffoldů. Tento přístup by ale byl při analýze tak rozsáhlého datasetu draft genomů velmi zdlouhavý a neefektivní.

Databáze ICEberg, ISfinder a TnCentral sice umožňují analýzu multi-FASTA souborů, byly ale také zhodnoceny jako nevhodné, protože se výsledek analýzy nedá exportovat a stáhnout v žádném formátu, který by byl možný využít pro další analýzu a charakterizaci. Výsledky by se musely ručně zkopírovat nebo přepsat, což by opět bylo při analýze tak rozsáhlého datasetu zdlouhavé a neefektivní.

Databáze BacAnt a MEGARes sice neposkytují online službu, která by umožňovala analýzu genomů, databáze se ale dají stáhnout a může být použit MyDbFinder. MyDbFinder je nástroj, který umožňuje analýzu genomů pomocí jakékoli

datbáze v multi-FASTA formátu. Přesnost výsledků může být nastavena podle potřeby. Výsledky se dají stáhnout jako TSV tabulka, která se dobře zpracovává pro další analýzu a charakterizaci. Databáze CARD, MobileElementFinder, PlasmidFinder a PHASTER umožňují analýzu draft genomů, nabízejí export výsledků ve formátech umožňujících další analýzu a charakterizaci, a rovněž zobrazují úroveň přesnosti výsledků. Tyto databáze umožňují rychlou a efektivní bioinformatickou širokospektrální analýzu.

Zkoumáním polohy genů na scaffoldech bylo zjištěno, že se určité geny rezistence šíří napříč bakteriemi společně. Geny *Mef(En2)* a *lnuA* v těsné blízkosti na scaffoldu byly detekovány u 26 různých bakterií. Geny *tet(A)* a *tet(B(P))* byly detekovány v těsné blízkosti na scaffoldu u 4 různých bakterií. O těchto dvojicích genů se již dá s jistou pravděpodobností tvrdit, že putují mezi bakteriemi společně.

Zkoumáním, jestli jsou geny v těsné blízkosti na scaffoldu, je jeden z možných přístupů, mezi geny rezistence se ale často nacházejí také hypotetické geny, jejichž úlohu neznáme, a které potom nejsme schopni detekovat. Těchto hypotetických genů může být obrovské množství.

Zkoumáním mobilních genetických elementů, které mají ve své struktuře geny rezistence, bylo zjištěno, že mezi bakteriemi často putuje integrativní mobilizovatelný element *MThSag1*, který přenáší gen *lnuC* kódující rezistenci na linkosamidy. Byl identifikován u 21 bakteriálních izolátů.

Nejvíce detekovaných genů kódovalo rezistenci na tetracykliny, a to pravděpodobně kvůli širokému použití tetracyklinů jako přísad do krmiv a stimulátorů růstu[19]. Antibiotika, na které detekované geny kódují rezistenci, se do značné míry shodují s antibiotiky ve studii [19], nejčtenější skupiny antibiotik, jako jsou tetracykliny, makrolidy, linkosamidy, streptograminy a aminoglykosidy, se shodují, a to i přes to, že k detekci genů byla použita jiná databáze, konkrétně databáze ResFinder, než v této práci.

Nejčastěji detekovaným genem rezistence byl gen *tet(W)*. Gen *tet(W)* se uvádí jako jeden z nejrozšířenějších genů rezistence na tetracykliny, přítomný v anaerobních bakteriích[49].

Tato čísla mohou být ve skutečnosti mnohonásobně vyšší, protože byly detekovány pouze geny a MGE, které jsou obsaženy v databázích CARD, MEGARes, MobileElementFinder, PlasmidFinder, BacAnt a PHASTER. Evoluční dynamika genů antimikrobiální rezistence má za následek hromadění sekvenčních variací, které mohou být dost odlišné od variant zaznamenaných v databázi. Navíc se jedná o draft genomy. To znamená, že genom není kompletní, že tam některé segmenty chybí, nebo tam jsou naopak vícekrát, že jsou ve špatném pořadí, nebo v opačné orientaci. Při detekci se tedy gen nebo MGE nemusí ukázat, i přes to, že je tam skutečně přítomný.

Závěr

Tato bakalářská práce se zabývala detailní analýzou genů rezistence na antibiotika a těžké kovy a znaků horizontálního přenosu genetické informace u zástupců anaerobních bakterií. Práce se zaměřuje na využití moderních bioinformatických metod a veřejných databází pro detekci a charakterizaci těchto genů z genomických dat.

V teoretické části práce byla provedena literární rešerše zaměřená na stávající poznatky o genech rezistence na antibiotika a těžké kovy u bakterií. Byly prozkoumány různé mechanismy horizontálního přenosu genů mezi bakteriemi. Poté byly popsány veřejné databáze a bioinformatické nástroje určené pro detekci genů rezistence a horizontálního přenosu.

V praktické části práce byla provedena detekce genů rezistence na antibiotika a těžké kovy a horizontálního přenosu genů u zástupců anaerobních bakterií. Bylo vyzkoušeno 20 různých databází, které byly vybrány tak, aby pokryly širokou škálu genetických a funkčních aspektů spojených s rezistencí bakterií a s horizontálním přenosem genů. Byla navržena metoda pro detekci a charakterizaci genů rezistence z genomických dat bakteriálních anaerobů, dílčí části metod byly realizovány a otestovány na 5 bakteriálních izolátech.

Pro analýzu a charakterizaci genů rezistence a horizontálního přenosu bylo použito 6 databází, CARD, MEGARes, MobileElementFinder, PlasmidFinder, BacAnt a PHASTER. Navržená metoda byla implementována v programovacím jazyce Python. Byla otestována na souboru 453 genomů. Výsledky byly vyhodnoceny a diskutovány.

Literatura

- [1] BENCKO, Vladimír a Petr ŠÍMA. Horizontální přenos genetické informace a jeho význam pro vznik antibiotické rezistence. *Časopis lékařů českých*. 2018, 157(3), 141-145.
- [2] BENCKO, Vladimír a Petr ŠÍMA. Antibiotická rezistence a význam horizontálního přenosu genetické informace. *Praktický lékař*. 2018, 98(5), 195-199.
- [3] SOUCY, S. M., J. HUANG a J. P. GOGARTEN. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics*, 2015, 16(8): 472-482.
- [4] FROST, L. S., R. LEPLAE, A. SUMMERS, et al. Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*. 2005, 3(9), 722-732.
- [5] BELLO-LÓPEZ, J. M., O. A. CABRERO-MARTÍNEZ, G. IBÁÑEZ-CERVANTES, et al. Horizontal gene transfer and its association with antibiotic resistance in the genus *Aeromonas* spp. *Microorganisms*. 2019, 7(9), 363.
- [6] LLOSA, M., F. X. GOMIS-RÜTH, M. COLL, et al. Bacterial conjugation: a two-step mechanism for DNA transport. *Molecular microbiology*. 2002, 45(1), 1-8.
- [7] WATERS, Virginia L. Conjugation between bacterial and mammalian cells. *Nature genetics*. 2001, 29(4), 375-376.
- [8] HOLMES, Randall K a Michael G. JOBLING. *Medical Microbiology: Genetics* [online]. 4th edition. Galveston: University of Texas Medical Branch at Galveston, 1996 [cit. 2022-12-06]. ISBN 10: 0-9631172-1-1. Dostupné z: <https://www.ncbi.nlm.nih.gov/books/NBK7908/>
- [9] WONG, J. J. W., J. LU a J. N. M. GLOVER. Relaxosome function and conjugation regulation in F-like plasmids – a structural biology perspective. *Molecular microbiology*. 2012, 85(4), 602-617.
- [10] PAUL, John H. a Sunny C. JIANG. Lysogeny and transduction. *Methods in microbiology*. 2001, Academic Press, Vol. 30, 105-125.
- [11] MAURICE, C. F., C. BOUVIER, R. DE WIT, et al. Linking the lytic and lysogenic bacteriophage cycles to environmental conditions, host physiology and their variability in coastal lagoons. *Environmental microbiology*. 2013, 15(9), 2463-2475.

- [12] MAHILLON, Jacques a Michael CHANDLER. Insertion sequences. *Microbiology and Molecular Biology Reviews*. 1998, 62(3), 725-774.
- [13] HRABÁK, J., A. ZEMANOVÁ a E. CHUDÁČKOVÁ. Mobilní genetické elementy v epidemiologii rezistence bakterií k antibiotikům. *Epidemiologie, Mikrobiologie, Imunologie*. 2010, 59(2), 55-66.
- [14] BABAKHANI, Sajad a Mana OLOOMI. Transposons: the agents of antibiotic resistance in bacteria. *Journal of basic microbiology*. 2018, 58(11), 905-917.
- [15] GORYSHIN, Igor Yu a William S. REZNIKOFF. Tn5 in vitro transposition. *Journal of Biological Chemistry*. 1998, 273(13), 7367-7374.
- [16] HALLET, Bernard a David J. SHERRATT. Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements *FEMS Microbiology Reviews*. 1997, 21(2), 157-178.
- [17] DARMON, Elise a David R. F. LEACH. Bacterial genome instability. *Microbiology and Molecular Biology Reviews*. 2014, 78(1), 1-39.
- [18] CARATTOLI, Alessandra. Importance of integrons in the diffusion of resistance. *Veterinary Research*. 2001, 32(3-4), 243-259.
- [19] JURICOVA H., J. MATIASOVICOVA, T. KUBASOVA, et al. The distribution of antibiotic resistance genes in chicken gut microbiota commensals. *Scientific Reports*. 2021, 11(1), 3290.
- [20] DOSTER E., S. M. LAKIN, C. J. DEAN, et al. MEGARes 2.0: a database for classification of antimicrobial drug, biocide and metal resistance determinants in metagenomic sequence data. *Nucleic acids research*. 2020, 48(D1), D561-D569.
- [21] VIANA, P., L. MEISEL, A. LOPES, et al. Identification of antibiotics in surface-groundwater. A tool towards the ecopharmacovigilance approach: a Portuguese case-study. *Antibiotics*. 2021, 10(8), 888.
- [22] TLASKALOVÁ HOGENOVÁ, H., Z. JIRÁSKOVÁ ZÁKOSTELSKÁ, J. PETANOVÁ, et al. Mikrobiota, imunita a imunologicky mediované choroby. *Vnitřní lékařství*. 2019, 65(2), 98-107.
- [23] MACGOWAN, Alasdair a Emily MACNAUGHTON. Antibiotic resistance. *Medicine*. 2017, 45(10), 622-628.
- [24] FLORENSA, A. F., R. S. KAAS, P. T. L. C. CLAUSEN, et al. ResFinder - an open online resource for identification of antimicrobial resistance genes in

- next-generation sequencing data and prediction of phenotypes from genotypes. *Microbial Genomics*. 2022, 8(1), 000748.
- [25] LAPLAE, R., A. HEBRANT, S. J. WODAK, et al. ACLAME: A CLAssification of Mobile genetic Elements. *Nucleic Acids Research*. 2004, Volume 32, Database issue, D45–D49.
- [26] HUANG L., B. YANG, H. YI, et al. AcrDB: a database of anti-CRISPR operons in prokaryotes and viruses. *Nucleic Acids Research*. 2021, Volume 49, Database issue, D622–D629.
- [27] WANG, J., W. DAI, J. LI, et al. AcrHub: an integrative hub for investigating, predicting and mapping anti-CRISPR proteins. *Nucleic Acids Research*. 2021, 49(D1), D630–D638.
- [28] HUA, X., Q. LIANG, M. DENG, et al. BacAnt: a combination annotation server for bacterial DNA sequences to identify antibiotic resistance genes, integrons, and transposable elements. *Frontiers in microbiology*. 2021, 12: 649969.
- [29] RENDÓN, J. M., B. LANG, G. G. TARTAGLIA, et al. BacFITBase: a database to assess the relevance of bacterial genes during host infection. *Nucleic Acids Research*. 2020, Volume 48, Issue D1, D511–D516.
- [30] ALCOCK, B. P., W. HUYNH, R. CHALIL, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*. 2022, 51(D1), D690–D699.
- [31] MOKHTAR, M. M., A. M. ALSAMMAN, H. M. ABD-ELHALIM, et al. Cicer-SpTEdb: A web-based database for high-resolution genome-wide identification of transposable elements in Cicer species. *PLoS One*. 2021, 16(11): e0259540.
- [32] YI, F., J. LING, Y. XIAO, et al. ConTEdb: a comprehensive database of transposable elements in conifers. *Database (Oxford)*. 2018, 2018: bay131.
- [33] LLORENS, C., R. FUTAMI, L. COVELLI, et al. The Gypsy Database (GyDB) of Mobile Genetic Elements: Release 2.0 *Nucleic Acids Research*. 2011, 39(suppl 1), D70–D74.
- [34] JOHANSSON, M. H. K., V. BORTOLAIA, S. TANSIRICHAIIYA, et al. Detection of mobile genetic elements associated with antibiotic resistance in *Salmonella enterica* using a newly developed web tool: MobileElementFinder. *Journal of Antimicrobial Chemotherapy*. 2021, 76(1), 101–109.

- [35] CARD: Comprehensive Antibiotic Resistance Database [online]. [cit. 2022-12-27]. Dostupné z: <https://card.mcmaster.ca/analyze/rgi>
- [36] XIE, Zhiqun a Haixu TANG. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics*. 2017, 33(21), 3340-3347.
- [37] LIU, M., X. LI, Y. XIE, et al. ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic acids research*. 2019, 47(D1), D660-D665.
- [38] MOURA, A., M. SOARES, C. PEREIRA, et al. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics*. 2009, 25(8), 1096-1098.
- [39] LAI, S., L. JIA, B. SUBRAMANIAN, et al. mMGE: a database for human metagenomic extrachromosomal mobile genetic elements. *Nucleic Acids Research*. 2021, Volume 49, Issue D1, D783–D791.
- [40] ARNDT, D., J. R. GRANT, A. MARCU, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*. 2016, Volume 44, Issue W1, W16–W21.
- [41] CARATTOLI, A., E. ZANKARI, A. GARCÍA-FERNÁNDEZ, et al. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*. 2014, 58(7), 3895-903.
- [42] XIE, Y., Y. WEI, Y. SHEN, et al. TADB 2.0: an updated database of bacterial type II toxin-antitoxin loci. *Nucleic acids research*. 2018, 46(D1), D749-D753.
- [43] ROSS, K., A. M. VARANI, E. SNESRUD, et al. TnCentral: a Prokaryotic Transposable Element Database and Web Portal for Transposon Analysis. *mBio*. 2021, 12(5): e0206021.
- [44] BADEL, C., V. D. CUNHA, R. CATCHPOLE, et al. WASPS: web-assisted symbolic plasmid synteny server. [?]. 2020, Volume 36, Issue 5, 1629–1631.
- [45] CHOPRA, Ian a Marilyn ROBERTS. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and molecular biology reviews*. 2001, 65(2), 232-260.
- [46] WEBBER, M. A. a L. J. V. PIDDOCK. The importance of efflux pumps in bacterial antibiotic resistance. *Journal of antimicrobial chemotherapy*. 2003, 51(1), 9-11.

- [47] BLAIR, J. M. A., M. A. WEBBER, A. J. BAYLAY, et al. Molecular mechanisms of antibiotic resistance. *Nature reviews microbiology*. 2015, 13(1), 42-51.
- [48] SPRATT, Brian G. Resistance to antibiotics mediated by target alterations. *Science*. 1994, 264(5157), 388-393.
- [49] ROBERTS, Marilyn C. Update on acquired tetracycline resistance genes. *FEMS microbiology letters*. 2005, 245(2), 195-203.