



# 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

## ANALÝZA HORIZONTÁLNÍHO PŘENOSU GENETICKÝCH KOMPONENT POMOCÍ STATICKÉ SÍŤOVÉ ANALÝZY

ANALYSIS OF HORIZONTAL TRANSFER OF GENETIC COMPONENTS USING STATIC  
NETWORK ANALYSIS

### BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

### AUTOR PRÁCE

AUTHOR

Anastasiya Labanova

### VEDOUCÍ PRÁCE

SUPERVISOR

Ing. et Ing. Jana Schwarzerová, MSc

BRNO 2023

## Bakalářská práce

bakalářský studijní program **Biomedicínská technika a bioinformatika**

Ústav biomedicínského inženýrství

*Studentka:* Anastasiya Labanova

*ID:* 221520

*Ročník:* 3

*Akademický rok:* 2022/23

### NÁZEV TÉMATU:

Analýza horizontálního přenosu genetických komponent pomocí statické síťové analýzy

### POKYNY PRO VYPRACOVÁNÍ:

- 1) Seznamte se s principy horizontálního přenosu genů a genetických komponent.
- 2) Prostudujte laboratorní techniky používané pro sekvenaci dat.
- 3) Vytvořte si vhodný testovací formát dat, ve kterém můžeme předpokládat výskyt horizontálně přenosných genetických komponent. A identifikujte horizontálně přenosné geny/genetické komponenty ve Vašem testovacím datasetu.
- 4) Vizualizujte horizontální přenos pomocí sítí v programu Cytoscape.
- 5) Proveďte statickou analýzu Vámi vytvořených sítí.
- 6) Proveďte diskusi k výsledkům. Pro splnění semestrálního projektu je potřeba splnit body 1) až 3).

### DOPORUČENÁ LITERATURA:

- [1] SYVANEN, Michael; KADO, Clarence I. Horizontal gene transfer. Academic Press, 2001.
- [2] KEELING, Patrick J.; PALMER, Jeffrey D. Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics, 2008, 9.8: 605-618.
- [3] JURICOVA, Helena, et al. The distribution of antibiotic resistance genes in chicken gut microbiotacommensals. Scientific reports, 2021, 11.1: 1-10.
- [4] MEDVECKY, Matej, et al. Whole genome sequencing and function prediction of 133 gut anaerobes isolated from chicken caecum in pure cultures. BMC genomics, 2018, 19.1: 1-15

*Termín zadání:* 6.2.2023

*Termín odevzdání:* 29.5.2023

**Vedoucí práce:** Ing. et Ing. Jana Schwarzerová, MSc

**Konzultant:** 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 ponížit 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 práce**

Bakalářská práce se věnuje problematice horizontálního přenosu genetických elementů mezi bakteriemi různých kmenů a implementaci softwarové analýzy, která umožní identifikaci genů, přenesených horizontálně. Použité balíčky a nástroje byly testovány na datasetu genomů bakterií několika kmenů. Teoretická část práce je věnována detailnímu popisu přenosu genetických elementů mezi bakteriemi a také jsou v ní popsány některé moderní laboratorní techniky umožňující sekvenaci genomů různými způsoby. V praktické části se práce zabývá předzpracováním genomických souborů za účelem získání vhodného formátu dat pro následnou anotaci. Pro detekci horizontálního přenosu genetických komponent mezi jednotlivými bakteriemi je představen vlastní skript, který anotované bakterie seřazuje do tabulek a hledá v jejich genomech stejné geny, které dle teoretických předpokladů byli přeneseny horizontálním způsobem. Dále je detekovaný přenos genů vizualizován pomocí nástrojů, které graficky znázorňují fylogenetické vztahy mezi bakteriemi. V posledním kroku jsou mobilní genetické elementy vizualizovány pomocí sítí a na základě jejich statické analýzy je provedena diskuse k výsledkům přesnosti a úspěšnosti navržené analýzy.

## **Klíčová slova**

Anotace, Fylogenetická analýza, Horizontální přenos genů, Sekvence, Síťová analýza

## **Abstract**

The bachelor's thesis focuses on the issue of horizontal genetic elements transfer between bacteria of different strains and the software analysis implementation that enables horizontally transferred genes identification. The packages and tools used were tested on a dataset of bacterial genomes from several strains. The thesis' theoretical part provides a detailed description of the genetic components transfer between bacteria and describes modern laboratory techniques that enable genome sequencing in various ways. In the practical part, the thesis deals with the preprocessing of genomic files to obtain suitable data for annotation. To detect the horizontal transfer of genetic elements between bacteria, a script is introduced, which organizes annotated bacteria to tables and searches for the same genes in their genomes that, under theoretical assumptions, were horizontally transferred. Furthermore, the gene transfer is visualized using tools that graphically represent phylogenetic relations between bacteria. In the final step, bacterial genomes are connected into networks, and based on their static analysis, a discussion is conducted on the results accuracy and the success of the proposed analysis.

## **Keywords**

Annotation, Phylogenetic analysis, Horizontal gene transfer, Sequencing, Network analysis

## **Bibliografická citace**

LABANAVA, A. Analýza horizontálního přenosu genetických komponent pomocí statické síťové analýzy. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství, 2023. 36 s., 2 s. příloh. Bakalářská práce. Vedoucí práce: Ing. et Ing. Jana Schwarzerová, MSc.

## Prohlášení autora o původnosti díla

<b>Jméno a příjmení studenta:</b>	<b>Anastasiya Labanova</b>
<b>VUT ID studenta:</b>	<b>221520</b>
<b>Typ práce:</b>	<b>Bakalářská práce</b>
<b>Akademický rok:</b>	<b>2022/23</b>
<b>Téma závěrečné práce:</b>	<b>Analýza horizontálního přenosu genetických komponent pomocí statické síťové analýzy</b>

Prohlašuji, že svou závěrečnou práci jsem vypracoval samostatně pod vedením vedoucí 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 autor uvedené závěrečné práce dále prohlašuji, že v souvislosti s vytvořením této závěrečné práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si 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.

V Brně dne: 29. května 2023

-----  
podpis autorky

## Poděkování

Děkuji vedoucí mé bakalářské práce Ing. et Ing. Janě Schwarzerové, MSc., a také mé konzultantce Mgr. Bc. Darině Čejkové, Ph. D., za účinnou a stálou metodickou, pedagogickou a odbornou pomoc a četné cenné rady při zpracování mé bakalářské práce.

*Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA CZ LM2018140) supported by the Ministry of Education, Youth and Sports of the Czech Republic.*

V Brně dne: 29.5.2023

-----  
podpis autorky

# Obsah

<b>SEZNAM OBRÁZKŮ .....</b>	<b>14</b>
<b>SEZNAM TABULEK.....</b>	<b>15</b>
<b>ÚVOD .....</b>	<b>16</b>
<b>1. PŘENOS GENETICKÉ INFORMACE.....</b>	<b>17</b>
1.1 VERTIKÁLNÍ PŘENOS GENŮ .....	17
1.2 HORIZONTÁLNÍ PŘENOS GENŮ .....	18
1.2.1 <i>Konjugace</i> .....	18
1.2.2 <i>Transdukce</i> .....	19
1.2.3 <i>Přirozená transformace</i> .....	20
1.3 SOUČASNÝ VÝZKUM V OBLASTI HGT .....	21
<b>2. LABORATORNÍ DATA .....</b>	<b>22</b>
2.1 SEKVENAČNÍ METODY A JEJICH ZÁKLADNÍ PRINCIPY.....	22
2.1.1 <i>Illumina sekvenování</i> .....	22
2.1.2 <i>Ion Torrent</i> .....	23
2.1.3 <i>Sekvenace Oxford Nanopore</i> .....	24
2.2 STŘEVNÍ MIKROBIOM HOSPODAŘSKÝCH ZVÍŘAT .....	25
<b>3. METODY ANALÝZY HORIZONTÁLNÍHO GENOVÉHO PŘENOSU .....</b>	<b>26</b>
3.1 ANOTAČNÍ METODY .....	27
3.2 METODY FYLOGENETICKÉ ANALÝZY.....	29
3.3 POPIS IMPLEMENTOVANÉ FUNKCE IDENTIFIKACE MGE .....	30
3.4 METODY SÍTOVÉ ANALÝZY .....	31
<b>4. VIZUALIZACE A INTERPRETACE VÝSLEDKŮ PROVEDENÝCH ANALÝZ .....</b>	<b>32</b>
4.1 VIZUALIZACE ANOTOVANÝCH GENETICKÝCH ELEMENTŮ .....	32
4.2 FYLOGENETICKÁ ANALÝZA.....	33
4.3 STATICKÁ SÍTOVÁ ANALÝZA .....	34
<b>5. ZÁVĚR.....</b>	<b>38</b>
<b>LITERATURA.....</b>	<b>39</b>
<b>SEZNAM PŘÍLOH.....</b>	<b>43</b>

## SEZNAM OBRÁZKŮ

Obrázek 1: Vertikální přenos genů nebo dědičnost (1A), horizontální přenos genů a jeho typy (1B). Upraveno a převzato z [1] .....	17
Obrázek 2: Schématické zobrazení procesu konjugace: v části (A) jsou zobrazeny bakterie spojené pílím, v části (B) jsou zobrazeny výsledné bakterie obsahující konjugativní plazmid. Převzato a upraveno z [1] .....	19
Obrázek 3: Schéma obecné transdukce u bakterií [3] .....	19
Obrázek 4: Sekvence Illumina. Převzato a modifikováno z [21] .....	23
Obrázek 5: Proces čtení nukleotidů pomocí sekvenační metody Ion Torrent. Převzato a modifikováno z [24] .....	24
Obrázek 6: Proces sekvenace nanopórem. Převzato a modifikováno z [27] .....	25
Obrázek 7: Schématické zobrazení navrženého funkčního postupu analýzy horizontálního genového přenosu. Anotace (A), fylogenetická analýza (B), síťová analýza (C) .....	26
Obrázek 8: Schéma nástrojů pipeline UBCG. Převzato a modifikováno z [33] .....	29
Obrázek 9: Popis implementovaného skriptu .....	30
Obrázek 10: Grafické zobrazení lokalice genu <i>mecA</i> v genomech <i>Enterococcus gallinarum</i> (A), <i>Hydrogenoanaerobacterium saccharovorans</i> (B) a <i>Lacrimispora saccharolytica</i> (C) .....	32
Obrázek 11: Zobrazení genomů izolovaných anaerobů z datasetu pomocí fylogenetického stromu online- nástrojem iTOL .....	33
Obrázek 12: Fylogenetický vztah bakteriálních genomů kmenu Firmicutes, nesoucích gen <i>mecA</i> .....	34
Obrázek 13: Zobrazení genomických vztahů v kmenu Actinobacteria pomocí sítí v Cytoscape .....	35
Obrázek 14: Síť kmenu Firmicutes po filtraci .....	37



## SEZNAM TABULEK

Tabulka 1: Seznam nástrojů, které využívá program Prokka.....	27
Tabulka 2: Výstupní soubory z programu Prokka .....	28
Tabulka 3: Výsledky statické síťové analýzy .....	36

# ÚVOD

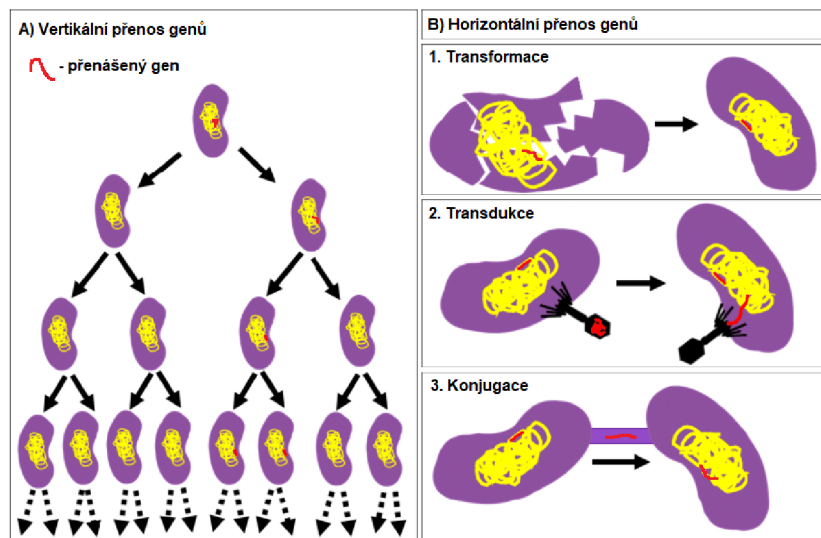
Horizontální přenos genetických elementů mezi bakteriemi hraje významnou roli v jejich evoluci a podílí se na šíření patogenů a rezistencí na antimikrobiální látky. Horizontální přenos genů je tedy důležitý pro udržení genetické rozmanitosti v bakteriálních populacích a může přispět k adaptaci bakterií na nové prostředí. Bakterie tak mohou získat geny pro produkci enzymů, které jim umožňují zpracovávat nové zdroje energie nebo substráty. Dále mohou bakterie získat geny pro produkci toxinů a také geny zvyšující virulenci. Horizontální přenos genů proto představuje důležitý mechanismus v evoluci bakterií, které tak mohou získávat nové genetické informace a vylepšovat své schopnosti. Tento proces, avšak také přispívá k šíření patogenů a může vést až k selekci nových kmenů bakterií, které do svého genomu začlenily různé geny rezistence. To pak často vede až k absolutní odolnosti bakterie vůči všem známým a používaným medikamentům.

Bakalářská práce se zabývá správnou implementací již dostupných softwarů a nástrojů, kterými lze analyzovat genomy bakterií různých kmenů a porovnávat je mezi sebou, za účelem vyhledání společných genů. Zejména těch, které byly přeneseny horizontálně. V první části bakalářské práce jsou popsány způsoby přenosu genetických komponent mezi bakteriemi a jejich význam. Druhá část práce se zabývá popisem několika nyní dostupných sekvenačních technik, umožňujících sekvenaci genomů různými způsoby a technologiemi.

Ve třetí kapitole následuje návrh postupu zpracování genomů, jejich fylogenetická analýza, anotace, a grafické znázornění obsahu genomů pro jejich následné porovnání. Dále se práce zabývá aplikací dostupných balíčků a online-nástrojů, jmenovitě: balíčkem UBCG pro fylogenetickou analýzu, nástrojem Prokka pro anotaci genomů, nástrojem DNAPlotter pro grafické znázornění anotovaných částí genomu a online-nástrojem iTOL, který umožňuje sestavit fylogenetický strom na základě podobnosti zkoumaných genomů. Také nezbytnou částí této kapitoly je návrh vlastní funkce pro identifikaci mobilních genetických elementů a nalezení podobností mezi genomy. Pro grafické znázornění této podobnosti se v práci používá nástroj Cytoscape, umožňující genomy jednoho kmenu zobrazit v podobě síti a propojit je na základě sdílených genů.

# 1. PŘENOS GENETICKÉ INFORMACE

Genetické komponenty se mohou mezi buňkami přenášet různými způsoby, které lze rozdělit do dvou základních kategorií: na horizontální a vertikální přenos. Vertikální přenos je proces přenosu genetické informace, ve kterém se daná informace přenesou z rodiče na potomka, viz Obrázek 1A. Naopak, horizontální přenos je proces přenosu genetické informace mezi jedinci nepříbuzných generací, viz Obrázek 1B. [1]



Obrázek 1: Vertikální přenos genů nebo dědičnost (1A), horizontální přenos genů a jeho typy (1B). Upraveno a převzato z [1]

## 1.1 Vertikální přenos genů

Vertikálním přenosem genetické informace nazýváme přenos, při kterém rodičovská buňka předává své geny dceřině buňce. Tento proces se řídí Mendelovými zákony a principy o dědičnosti [1]:

1. Princip uniformity – tento princip říká, že při vzájemném křížení dvou homozygotů jsou všichni potomci takového křížení genotypově a fenotypově identičtí.
2. Princip segregace – alely určující příznaky se rozdělují v gametech během meiózy, a tak vzniká stejný počet pohlavních buněk.
3. Princip nezávislé segregace – alely jednoho lokusu segregují do gamet nezávislých na alelách jiných lokusů.

Tyto tři základní principy značně ovlivnily celkový výzkum spojený s genetikou. V současnosti tyto principy stále slouží pro popsání základních jevů dědičnosti a detekci genů, které byly přeneseny od rodičů k potomkům neboli vertikálním přenosem genetické informace. [1]

## 1.2 Horizontální přenos genů

Horizontální přenos genů (angl. *horizontal gene transfer* (HGT)) je proces, kdy organismus přenáší genetický materiál do jiného organismu bez geneticky příbuzenského vztahu např. v rámci stejné generace. Tento proces je znám a převážně zkoumán u bakterií. [1]

Právě schopnost bakterií přizpůsobit se novému prostředí v rámci evoluce je nejčastěji výsledkem získávání nových genů prostřednictvím HGT, nikoli mutacemi. Bakterie poskytovatele genu se nazývá dárcovská neboli *donor*, bakterie přijímající gen nazýváme příjemcem neboli *recipientem*. Bakterie mohou obsahovat kromě bakteriálního chromozomu i malé nechromozomální molekuly DNA, které se nazývají plazmidy. Plazmidy jsou tvořeny obvykle 5 až 100 geny a nejsou pro život bakterií nezbytné. Avšak, právě plazmidy uchovávají evolučně výhodné bakteriální geny, jako například geny rezistence k antibiotikům. [1]

Horizontální přenos genů lze obecně rozdělit do tří základních typů: konjugace, transdukce a naturální neboli přirozená transformace. Konjugace je proces genetické rekombinace, při které dochází k přenosu DNA z donoru do recipientu kontaktem mezi buňkami. Transdukce je přenos fragmentu bakteriální DNA z jedné bakterie do druhé pomocí bakteriofága. Přirozená transformace je příjem neboli integrace extracelulární DNA, ke které může dojít za přirozených podmínek života bakterií.

### 1.2.1 Konjugace

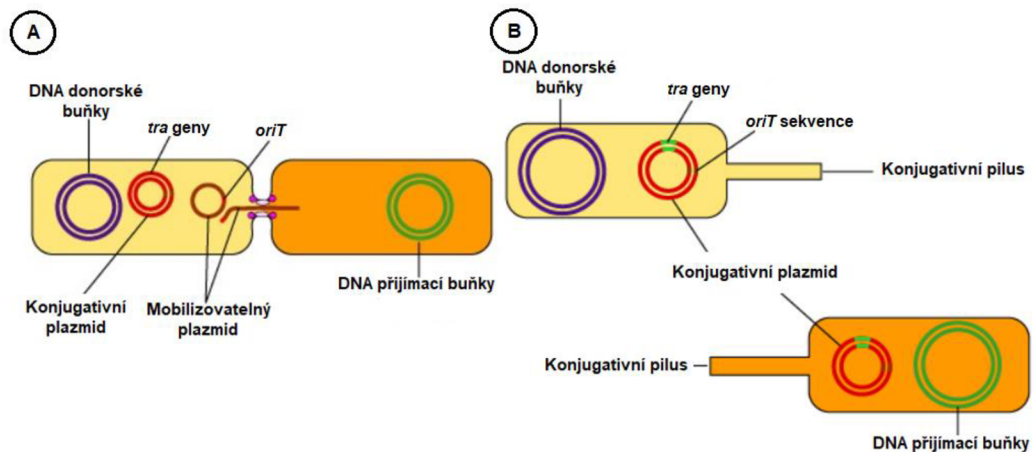
Konjugace je mechanismus horizontálního přenosu genů mezi bakteriemi, který zahrnuje přímý kontakt mezi dvěma buňkami, během něhož se přenáší genetický materiál. Tento proces umožňuje přenos plazmidů, kruhových molekul DNA obsahujících genetické informace, mezi bakteriemi.

Pro konjugaci jsou důležité dva hlavní prvky: *oriT* (origin of transfer) a *tra* geny. *oriT* je specifická sekvence DNA na plazmidu, která funguje jako místo, ze kterého začíná přenos genetického materiálu. *oriT* obsahuje informace, které jsou rozpoznány konjugativním aparátem bakterie, který je zapojen do přenosu plazmidu. *Tra* geny jsou geny kódované na plazmidu, které jsou zapojeny do procesu konjugace. Tyto geny obsahují informace pro tvorbu struktur a enzymů, které jsou nezbytné pro konjugaci. [1]

Prvním krokem konjugace je vazba *konjugativním pílím*<sup>1</sup>, který je vytvořen konjugativními proteiny kódovanými *tra* geny. Pilus umožňuje kontakt mezi buňkami a vytváří kanál pro přenos plazmidu, viz Obrázek 2A. Jakmile je vytvořen kontakt, přenos plazmidu probíhá rozvolněním plazmidové DNA, kdy je jeden řetězec plazmidu přenesen na recipienta. Po přenosu je druhý řetězec syntetizován v recipientovi, pak je plazmid kompletní v obou buňkách, viz Obrázek 2B. [1]

---

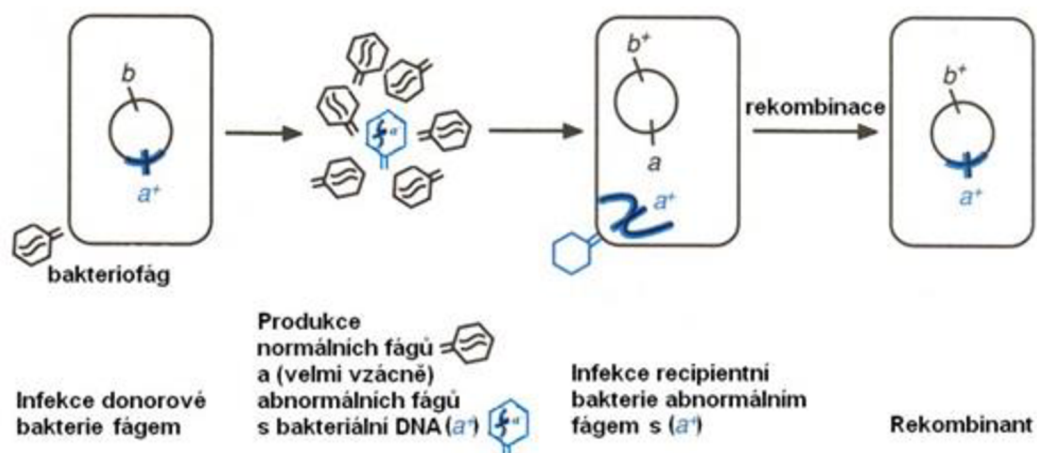
<sup>1</sup> Konjugativní pilus – vlasovitý útvar na povrchu bakterií, sloužící ke spojení dvou bakterií při konjugaci [3].



Obrázek 2: Schématické zobrazení procesu konjugace: v části (A) jsou zobrazeny bakterie spojené pílím, v části (B) jsou zobrazeny výsledné bakterie obsahující konjugativní plazmid. Převzato a upraveno z [1]

### 1.2.2 Transdukce

Transdukce je jeden z mechanismů horizontálního přenosu genů mezi bakteriemi, při kterém jsou geny přenášeny z jedné bakterie do druhé prostřednictvím bakteriálních virů nazývaných bakteriofágy. Bakteriofágy jsou viry, které infikují bakterie a reprodukují se v jejich hostitelských buňkách. Během tohoto procesu mohou bakteriofágy náhodně zachytit geny z genetického materiálu hostitelské bakterie. Základní princip transdukce spočívá v několika krocích. Zaprvé se bakteriofág připojuje k povrchu hostitelské bakterie pomocí specifické signalizace mezi fágem a receptory na povrchu bakterie. Některé fágy injektují svůj genetický materiál přes buněčnou stěnu do hostitelské buňky, zatímco jiné se celé vstřebají do buňky. Po proniknutí dovnitř hostitelské buňky se genetický materiál fága replikuje. To zahrnuje kopírování DNA fága a syntézu nových fágových proteinů, viz Obrázek 3 [1]. Existují dvě formy transdukce: obecná a specifická.



Obrázek 3: Schéma obecné transdukce u bakterií [3]

Obecná transdukce je náhodný proces, při kterém jsou geny náhodně zachyceny z genomu hostitelské bakterie a přeneseny do genomu bakteriofágu. Tyto zachycené geny jsou poté přeneseny do nové bakterie, pokud je fág infikuje. Konkrétním příkladem obecné transdukce je možné uvést tvorbu virových částic, kdy fág zachytí část dárcovské DNA a může předat cizí geny jiným buňkám, čímž je infikuje. Jakmile je recipientní buňka infikována částicí fága, je do ní zaveden genetický materiál dárcovské buňky, který se rekombinuje s homologní částí chromozomu od příjemce a vytvoří stabilního rekombinanta. [1]

Specifická transdukce je děj, při kterém jsou specifické geny přenášeny z jedné bakterie na druhou pomocí specializovaných bakteriofágů. Tyto fágy jsou schopny přenášet pouze určité geny nebo sekvence z hostitelského genomu. Vzhledem k tomu, že ve většině středně aktivních bakteriofágů se integruje genetické informace do bakteriálního chromozomu na specifických úsecích, přenášejí pouze specifickou část bakteriální DNA. DNA defektního fága se rekombinuje s DNA přijímající buňky pomocí rekombinace specifické pro dané místo a rekombinant se díky vnesenému genu stane *merodiploidem*<sup>2</sup>. Tento proces můžeme pozorovat například při přenosu genu *gal* P1 fágem do *Escherichia coli* za účelem předání schopnosti modifikace galaktózy, jak se popisuje ve studii Waldora a spol. [5]

### 1.2.3 Přirozená transformace

Jedná se o jedinou formu genetické rekombinace, při které se fragment z mrtvé rozložené bakterie dostane do jiné kompetentní recipientní bakterie a je vyměněn za část DNA recipienta. Transformace zahrnuje pouze rekombinaci homologních úseků DNA, které mají téměř stejné sekvence nukleotidů. Obvykle se jedná o podobné bakteriální kmeny. [1]

Za účelem přirozené transformace, se bakteriální buňce musí nejprve vyvinout regulovaný fyziologický stav, který zahrnuje přibližně 20 až 50 proteinů. Kompetence tohoto stavu pak závisí na přítomnosti specifických proteinů v buněčné membráně, které jsou afinitní k DNA. Transformační aktivitu má pouze dvouřetězcová molekula DNA, jelikož do příjemce proniká pouze jedno vlákno DNA. Druhé vlákno podléhá degradaci na buněčné membráně. Tato degradace uvolní dostatek energie k tomu, aby se první vlákno dostalo do buňky. Vysoká molekulová hmotnost transformační DNA zvyšuje pravděpodobnost rekombinace. Integrace s chromozomem následně vyžaduje homologní místa v transformující se DNA. K výsledné rekombinaci dochází na jednom vlákně, čímž vzniká heteroduplexní molekula, jejíž jedno vlákno má genotyp příjemce a druhé vlákno genotyp rekombinanta. [5]

---

<sup>2</sup> Merodiploid – je částečně diploidní bakterie, která má vlastní chromozomový komplement a fragment vnesený konjugací, transformací nebo transdukci. [6]

### 1.3 Současný výzkum v oblasti HGT

Rozpoznání mobilizovatelných elementů se stalo výchozím bodem pro současný výzkum přenosu genetických komponent. Jak již bylo výše zmíněno, HGT je hlavním zdrojem fenotypových změn u bakterií. Do genomů se vmísí radikálně odlišné genotypy ze vzdálených populací obsazujících stejnou niku<sup>3</sup>, nebo dokonce nové geny s novými funkcemi. Navíc horizontální přenos genů u bakterií, které nesou rezistentní informaci vůči antibiotikům, je jedním z příkladů mající velký dopad na oblast výzkumu pro lidské zdraví[8] [9].

Studie od Gary Kaisera [1] uvádí, že pokud je v prostředí přítomno specifické antibiotikum, bakterie nesoucí gen její rezistence mají evoluční výhodu vůči ostatním bakteriím. Navíc, multirezistentní bakterie mohou být odolné vůči řadě známých antibiotik, například *Staphylococcus aureus* rezistentní vůči meticilinu a *Enterococcus faecalis* rezistentní vůči vankomycinu představují jedno z hlavních rizik pro veřejné zdraví.

Schwarzerová J. a spol. [10] uvádí novou metodologii pro identifikaci genů, přispívající k identifikaci HGT mezi organismy osidlující střevní trakt kuřete. Zmíněná analýza pomáhá objevit a lépe pochopit vlastnosti genů předávaných horizontálně.

Ve studii J. Huanga [11] o horizontálním přenosu genů je popsán model slabého spojení (angl. *weak-link model*), který naznačuje, že HGT se pravděpodobněji vyskytuje mezi blízkými příbuznými organismy. Tento model se zakládá na předpokladu, že v populaci mikroorganismů existují jedinci, kteří mají "slabé vazby" (*weak links*), tedy mají větší pravděpodobnost přijetí cizího genetického materiálu a jeho následného přenesení na další jedince v populaci. Tato schopnost může být dána různými faktory, jako jsou přítomnost specifických receptorů na povrchu buňky, přítomnost specifických enzymů nebo defektů v mechanismech, které by jinak bránily přenosu cizího genetického materiálu.

Model slabých vazeb předpokládá, že HGT se s větší pravděpodobností vyskytuje mezi organismy, které mají nedávného společného předka nebo se vyvinuly v úzkém symbiotickém vztahu. HGT se tím pádem může s větší pravděpodobností vyskytovat v prostředích, kde je vysoká rozmanitost mikroorganismů, například v půdě nebo ve střevním mikrobiomu eukaryot. Model slabých vazeb je sice jednou z teorií, které byly navrženy k vysvětlení vzorců HGT, není však jedinou. V současnosti v této oblasti probíhá výzkum a diskuse o faktorech, které ovlivňují HGT, a o mechanismech, kterými k němu dochází. Tento model může být užitečný při studiu evoluce bakterií a šíření rezistence na antibiotika, protože umožňuje lépe porozumět procesům horizontálního přenosu genů a identifikovat faktory, které ovlivňují jejich rozšíření v populaci. [11]

---

<sup>3</sup> Nika – souhrn podmínek prostředí nutných k přežití a rozmnožování jedinců.[8]

## 2. LABORATORNÍ DATA

Daná kapitola popisuje sekvenační techniky, které se v dnešní době využívají pro získání dat v oblasti výzkumu horizontálního genového přenosu. V první části této kapitoly jsou popsány metody sekvenování. Jedná se o metody nové generace, mezi které patří Illumina, Ion Torrent a metoda třetí generace Oxford Nanopore technologie (ONT). Pomocí techniky Illumina-NextSeq 500 byla sekvenována data, použitá v praktické části této bakalářské práce [12]. V druhé části této kapitoly, je dále uveden podrobnější popis těchto dat.

### 2.1 Sekvenační metody a jejich základní principy

Sekvenace DNA je metoda určování pořadí nukleotidů v molekule DNA a je nepostradatelným krokem v bioinformatických analýzách. Sekvenační technologie se během posledního desetiletí staly dostupnějšími především díky snížení nákladů a zvýšení rychlosti sekvenace kompletních genomů.

Základem pro rozvoj dnešních technik sloužilo Sangerovo sekvenování vyvinuté Frederikem Sangerem a jeho kolegy. Daná metoda sekvenování je založena na reakci polymerizace a kapilární elektroforéze a používala se pro sekvenaci kratších úseků DNA člověka [13]. V současnosti se do popředí sekvenačních technik dostaly metody nové generace (angl. *Next Generation Sequencing* (NGS) [14]). Tyto sekvenační techniky umožňují sekvenaci několika různých řetězců DNA současně [15].

#### 2.1.1 Illumina sekvenování

Jedná se o technologii vyvinutou v roce 2006 a její základem je princip sekvenace syntézou ssDNA [16]. Na začátku jsou fragmenty výchozí DNA hybridizovány v reakční komoře na povrchu opticky průhledné pevné látky. Běžné sekvenování syntézou využívá DNA polymerázu k paralelnímu prodlužování několika řetězců ssDNA, tím se vytváří shluk (angl. *cluster*), viz Obrázek 4. Nukleotidy nebo krátké oligonukleotidy jsou dodávány po jednom, nebo modifikovány identifikačními markery tak, aby typ báze bylo možné určit v průběhu prodlužování sekvence. [17]

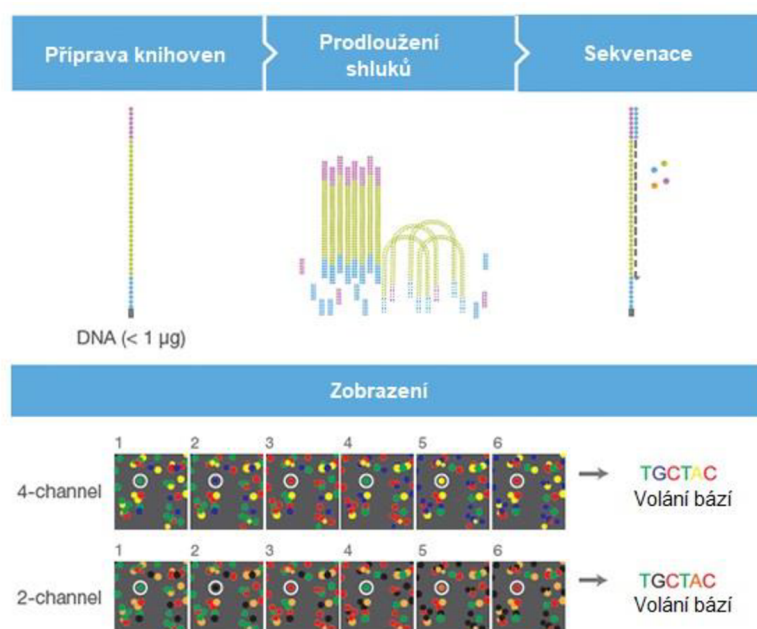
#### **Illumina NextSeq500**

NextSeq 500 je vysoce výkonným počítačovým sekvenátorem, který nabízí sekvenování celého genomu a transkriptomu. NextSeq 500 má pokročilé funkce jako možnost vzdáleného nastavování experimentu, vložení a odebrání reagentů, jednoduché uživatelské rozhraní a přednastavený algoritmus sekvenace.



Řetězce ssDNA jsou amplifikovány na povrchu průtokové destičky NextSeq procesem zvaným můstková PCR<sup>4</sup>. Každé vlákno je amplifikováno vícekrát, a tímto vznikají miliony klonálních shluků. Během jednoho sekvenačního cyklu jsou vzorky DNA čteny začleněním jednoho značeného nukleotidu. Po každém cyklu zde zdroj světla excituje značený nukleotid a tím způsobí vyzáření fluorescenčního signálu. Uživatel může určit počet cyklů, aby bylo dosaženo vhodné délky čtení pro jeho experiment. V tomto paralelním procesu se provádějí stovky milionů čtení najednou. [18]

Integrované počítače provádějí primární analýzu dat včetně volání bází a hodnocení kvality. Sekvenační data jsou pak okamžitě přenesena, analyzována a uložena v BaseSpace – genomickém výpočetním prostředí Illumina [20]. Analýza dat zahrnuje aplikaci vyvinuté společností pro zarovnávání a detekci variability, anotaci, vizualizaci a interpretaci. [18]



Obrázek 4: Sekvence Illumina. Převzato a modifikováno z [21]

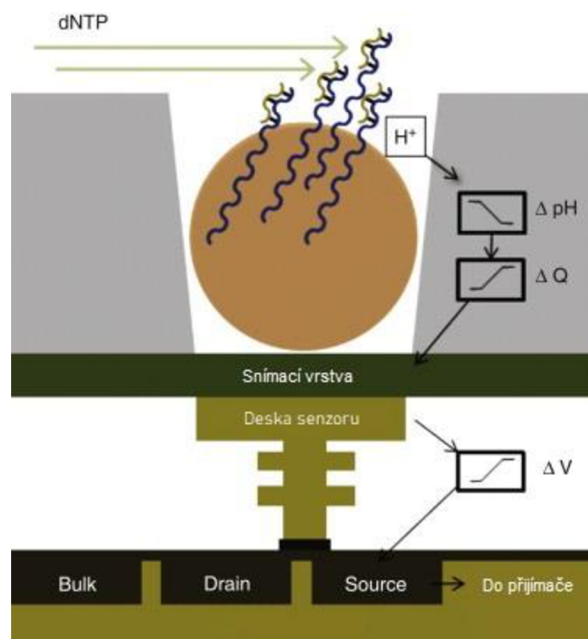
### 2.1.2 Ion Torrent

Ion Torrent je metoda vyvinuta společností Life Technologies. Chemie sekvenování Ion Torrent je založena na konceptu, že DNA polymeráza katalyzuje syntézu kovalentních vazeb prodlužujícího se vlákna DNA, čímž se uvolňuje pyrofosfát a proton vodíku. S uvolňováním protonu vodíku klesá pH prostředí a tato změna je následně detekována za účelem stanovení nukleotidů v sekvenci. [22]

Sekvenátor má na polovodičovém čipu mikrojamky, z kterých každá obsahuje molekulu ssDNA a je připravena k sekvenování. Poté se do jamek opakovaně přidávají nemodifikované deoxynukleosidtrifosfáty (dNTP) a DNA polymeráza. Při přidání

<sup>4</sup> PCR (angl. *Polymerase Chain Reaction*) – enzymatická metoda používaná k amplifikaci (namnožení) sekvence DNA za krátkou dobu. [19]

jednoho druhu dNTP dojde k biologické odezvě, pokud se nukleotid začlení do prodlužujícího se řetězce s uvolněním vodíkových iontů, viz Obrázek 5. V každém cyklu se do vznikajícího řetězce může začlenit více než jeden nukleotid. Pod každou mikrojamkou je umístěn iontově selektivní sensor, který detekuje změny pH vyvolané uvolňováním vodíkových iontů prostřednictvím rozdílů potenciálů a také zaznamenává každé začlenění nukleotidu. Nepřipojené molekuly dNTP se před zahájením dalšího cyklu odstraňují. Cyklus se pak opakuje s následujícími řetězci dNTP. [23]



Obrázek 5: Proces čtení nukleotidů pomocí sekvenační metody Ion Torrent.  
Převzato a modifikováno z [24]

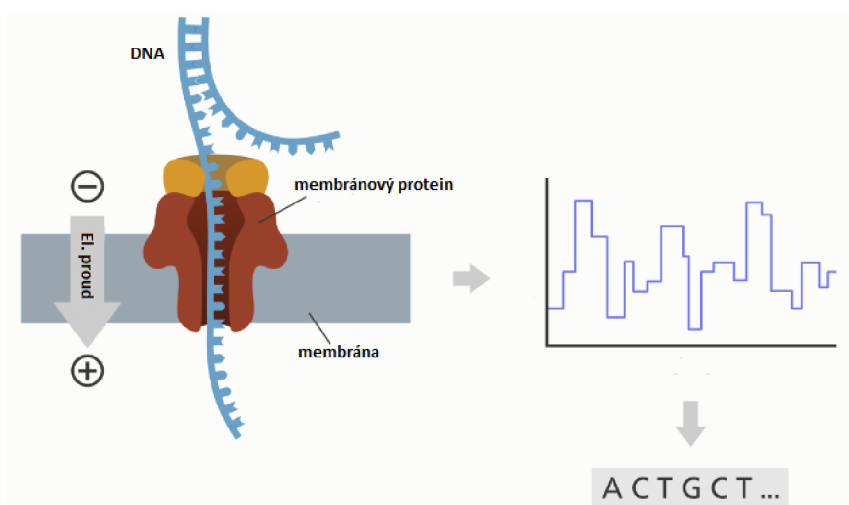
### 2.1.3 Sekvence Oxford Nanopore

Oxford Nanopore sekvenace je unikátní, škálovatelná technologie, která umožňuje přímou analýzu dlouhých fragmentů DNA nebo RNA v reálném čase [25]. Nanopóry jsou kanály o šířce nanometrů, které mohou být tří typů [26]:

- Biologické – póry tvořeny proteinem formujícím póry v membráně, jako například alfa-hemolysin,
- Pevné látky – póry tvořené syntetickým materiálem nebo odvozené chemickou cestou,
- Hybridní – póry tvořené biologickým materiálem, který je zapouzdřený v syntetickém materiálu.

Označování nebo identifikace nukleotidů není u nanoporového sekvenátoru DNA nutná. Podstatou této metody je kontrola iontového proudu, který vzniká při pohybu molekuly DNA nanoporem, viz Obrázek 6. Tento postup pomáhá mimo jiné určit délku, průměr a tvar molekuly. Zpočátku se nanopórem propouští známá hodnota iontového proudu. Vzhledem k tomu, že různé nukleotidy blokují proud po různě dlouhou dobu díky svému proměnlivému odporu, je možné měřením této doby určit sekvenci dané molekuly.

Metoda může být podrobena dalšímu zdokonalení, které povede k vytvoření rychlé metody sekvenování DNA založené na nanopórech. [15]



Obrázek 6: Proces sekvenace nanopórem. Převzato a modifikováno z [27]

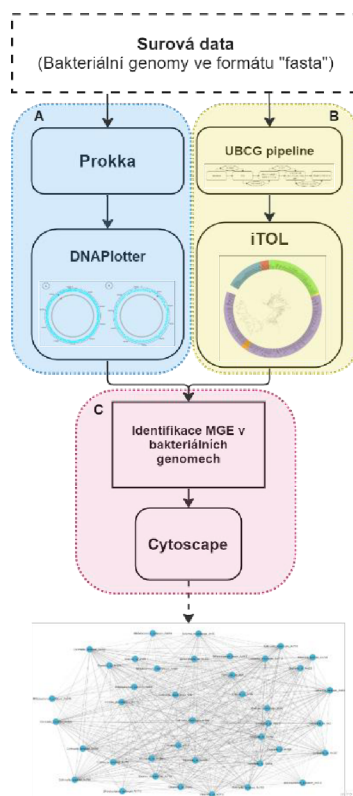
## 2.2 Střevní mikrobiom hospodařských zvířat

Dataset, který je použit v této bakalářské práci, se sestavuje ze 275 genomů bakterií. 204 genomy byly publikovány ve studii Medveckého a spol. [28], a od té doby je tato sbírka neustále rozšiřována. Ve studii se kultivovaly stovky izolátů anaerobů z kuřecího slepého střeva. Kuřata v komerční produkci se líhnou v čistém prostředí bez kontaktu s dospělými slepicemi a jejich kolonizace je závislá na zdrojích prostředí.

Komerčně vylíhnutá kuřata jsou velmi citlivá na kolonizaci různými patogenními organismy, např. *Salmonella Enteritidis*, nicméně jejich odolnost lze zvýšit tím, že se jim poskytne komplexní mikrobiota dospělých kuřat [29]. Všechny analyzované genomy izolátů v této bakalářské práci byly sekvenovány pomocí platformy Illumina NextSeq 500. Jelikož některé izoláty vykazovaly skoro absolutní identitu jejich genomů, proto konečný počet izolátů klesl na identifikovaných 275 (publikované v Bioprojektu PRJNA377666 – *Whole genome sequencing of chicken gut anaerobes*, registrovaném dne 2.3.2017 a stále doplňovaném) kmenů patřících do osmi různých kmenů – Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, Fucobacteria, Verrucomicrobia, Elusimicrobia a Synergistetes.

### 3. METODY ANALÝZY HORIZONTÁLNÍHO GENOVÉHO PŘENOSU

Pro celkovou analýzu v této práci byla použita řada nástrojů seřazených do funkčního postupu analýzy dat, viz Obrázek 7. Předzpracování genomových sekvencí bylo provedeno pomocí shell-skriptů (viz Příloha A -), spouštěných na výpočetních zdrojích dostupných na portálu Metacentrum VO – virtuální organizace [30]. Pro nalezení konzervovaných genů a selekci genů přenesených vertikálním genetickým přenosem, v první části práce (sekce A) byla provedena anotace genomů pomocí nástroje Prokka v1.14.5 [32] a následně vizualizována v programu DNA plotter v18.2.0 [47]. Ve druhé části (sekce B) byla provedena fylogenetická analýza genomů pomocí UBCG v3.0 (angl. *Up-to-date bacterial core gene pipeline*) [33] sady nástrojů, které umožňují zarovnat a vizualizovat genomické sekvence do fylogenetického stromu. Fylogenetický strom byl zrekonstruován pomocí online nástroje iTOL v6.7.5 [31]. V poslední části (sekce C) byly identifikovány mobilní genetické komponenty pomocí vlastního skriptu a jejich následné zobrazení pomocí nástroje Cytoscape v3.9.1 [51].



Obrázek 7: Schematické zobrazení navrhnutého funkčního postupu analýzy horizontálního genového přenosu. Anotace (A), fylogenetická analýza (B), síťová analýza (C)

### 3.1 Anotační metody

Pro vlastní anotaci genomů byl v této práci použit nástroj Prokka v 1.14.5 [32]. Prokka zahrnuje řadu softwarových metod pro získání detailní a důvěryhodné anotace genomových sekvencí bakterií. Nástroj Prokka je vhodný pro integraci do pipeline genomického softwaru a modelů sekvenční analýzy. Vstupem tohoto nástroje jsou assemblované sekvence DNA či scaffoldy<sup>5</sup> ve formátu *fasta*. Prokka se při anotování genomických vlastností v scaffoldech spoléhá na externí nástroje pro predikci, které jsou vypsány v Tabulka 1.

První nástroj Prodigal v 2.6.3 [34] je určen k identifikaci mikrobiálních genů. Cílem tohoto nástroje je zvýšit citlivost při detekci existujících genů, tj. předpovídat místa iniciace translace a snížit četnost falešně pozitivních předpovědí. Princip dalšího nástroje RNAmmer [42] spočívá v analýze sekvencí DNA s cílem identifikovat a lokalizovat geny kódující ribosomální RNA (rRNA). Výstupem nástroje RNAmmer je seznam predikovaných rRNA genů a informace o jejich umístění v sekvenci DNA.

Aragorn [43] předpovídá sekundární strukturu tRNA a tmRNA na základě homologie s rozpoznávanými konsenzuálními sekvencemi pomocí heuristických algoritmů. Další nástroj SignalP v 4.0 [44] se používá na predikci průniku *signálních peptidů*<sup>6</sup> v proteinech. Během analýzy program vyhodnocuje různé vlastnosti a charakteristiky proteinů, jako je přítomnost hydrofobních oblastí a specifických aminokyselin. Na základě těchto informací SignalP přiřazuje pravděpodobnostní skóre pro přítomnost signálních peptidů. Jeho použití přispívá k porozumění struktuře a funkci proteinů a poskytuje důležité informace pro studium buněčného metabolismu a signalizace. Nástroj Infernal v 1.0 [46] má za účel vyhledávání a analýzu struktur RNA pomocí sekvenčních dat. Jeho princip je založen na porovnání zadaných sekvencí RNA s databázemi existujících struktur RNA, známé jako kovariační modely (CM). Tento princip umožňuje zjistit, jaké funkce může RNA plnit v buňce a jak se liší mezi různými organismy.

Tabulka 1: Seznam nástrojů, které využívá program Prokka

Nástroj	Princip
Prodigal	Identifikace genů kódujících proteiny
RNAmmer	Predikce rRNA genů
Aragorn	Identifikace genů tRNA a tmRNA
SignalP	Predikce signálních peptidů
Infernal	Vyhledávání a analýza struktur RNA

<sup>5</sup> Scaffoldy – uspořádaná řada čtení oddělených mezerami, jejich přibližná délka je známa. [41]

<sup>6</sup> Signální peptidy – krátké sekvence aminokyselin v proteinech, sloužící jako značky pro transport proteinů přes membrány buněčných organel nebo pro jejich transport ven z buňky. [45]

Výstupem anotačního nástroje Prokka jsou anotované genomy v různém datovém formátu. Přehled dokumentů a jejich obsah jsou popsány v Tabulka 2.

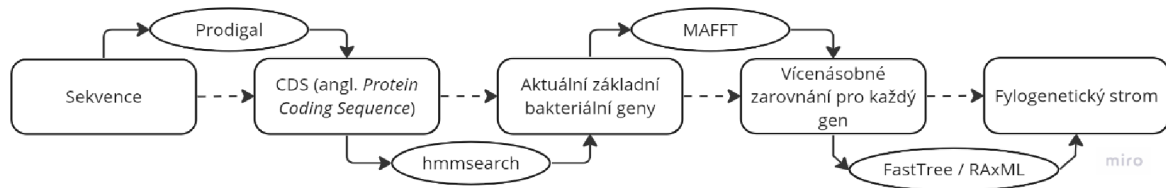
Tabulka 2: Výstupní soubory z programu Prokka

<b>Zkratka formátu</b>	<b>Popis obsahu</b>
<b>.gff</b>	Hlavní výstup anotace obsahující sekvence a jejich popis. Lze ho zobrazit v programu Artemis nebo DNA Plotter
<b>.gbk</b>	Standardní soubor Genbank odvozený z hlavního souboru
<b>.fna</b>	Soubor FASTA obsahující nukleotidy vstupních kontigů
<b>.faa</b>	Soubor FASTA obsahující proteiny přeložených CDS
<b>.ffn</b>	Soubor se všemi predikovanými transkripty (CDS, rRNA, tRNA, tmRNA)
<b>.sqn</b>	Soubor „Sequin“ pro uložení do Genbank
<b>.fsa</b>	Většinou má stejný obsah jako <i>.fna</i> , ale používá se k vytvoření <i>.sqn</i>
<b>.tbl</b>	Podílí se na vytvoření souboru <i>.sqn</i>
<b>.err</b>	Seznam neprovedených anotací – zpráva NCBI o nesrovnalostech
<b>.log</b>	Záznam celého procesu anotace včetně případných chyb
<b>.txt</b>	Statistické údaje o nalezených anotovaných prvcích
<b>.tsv</b>	Soubor se všemi prvky anotace, rozdělenými tabulátory

Vizualizace výsledné anotace se dále provádí pomocí nástroje DNAPlotter v18.2.0 [47]. DNAPlotter umožňuje zobrazit genomické údaje ve formě cirkulárního grafu, který umožňuje snadné porovnávání genomů různých organismů a vyhledávání specifických oblastí v sekvencích DNA. Vhodnými vstupními daty pro DNAPlotter jsou hlavní výstupní soubory nástroje Prokka formátu *.gff*, které jsou hlavními výstupními soubory obsahující sekvence a jejich kompletní anotace. Vizualizace dané anotace je dále provedena v podkapitole 4.1

## 3.2 Metody fylogenetické analýzy

Pro fylogenetickou analýzu dat byl použit balíček nástrojů UBCG pipeline. Jedná se o sadu nástrojů pro základní analýzu, která umožňuje vygenerovat a vyhodnotit fylogenetické stromy bakteriálních sekvencí. Metoda byla dříve použita k odvození fylogenomických vztahů *Escherichii spp.* a jejich příbuzných taxonů [33]. Obecně ji lze použít pro analýzu souboru bakteriálních genomů na jakémkoli taxonomickém ranku. Nástroje obsažené v této pipeline jsou uvedeny ve schématu, viz Obrázek 8.



Obrázek 8: Schéma nástrojů pipeline UBCG. Převzato a modifikováno z [33]

První použitý nástroj Prodigal v2.6.3 [34] byl podrobněji popsán v podkapitole 3.1, jelikož je obsažen i v rámci sady nástrojů Prokka. Jeho účelem je identifikace mikrobiálních genů. Cílem tohoto nástroje je zvýšit citlivost při detekci existujících genů, tj. předpovídat místa iniciace translace a snížit četnost falešně pozitivních předpovědí.

Další nástroj je HMMsearch v3.1b2 [35]. Jedná se o softwarový toolbox, který poskytuje nástroje pro vytváření pravděpodobnostních modelů doménových rodin proteinů a DNA sekvencí. Tyto pravděpodobnostní modely jsou dále použity například k anotaci nových sekvencí a k provádění hloubkových vícenásobných zarovnání sekvencí.

Pro vícenásobné zarovnání byl dále použit nástroj MAFFT v7.310 [36]. MAFFT vychází z předpokladu, že zarovnávané sekvence sdílejí jednu společnou sekvenci předka. MAFFT vždy sestaví vícenásobné zarovnání, které obsahuje všechny znaky vstupních sekvencí. Pořadí znaků v každé sekvenci je shodné s pořadím vstupní sekvence, ačkoli sekvence mohou být seřazeny podle podobnosti.

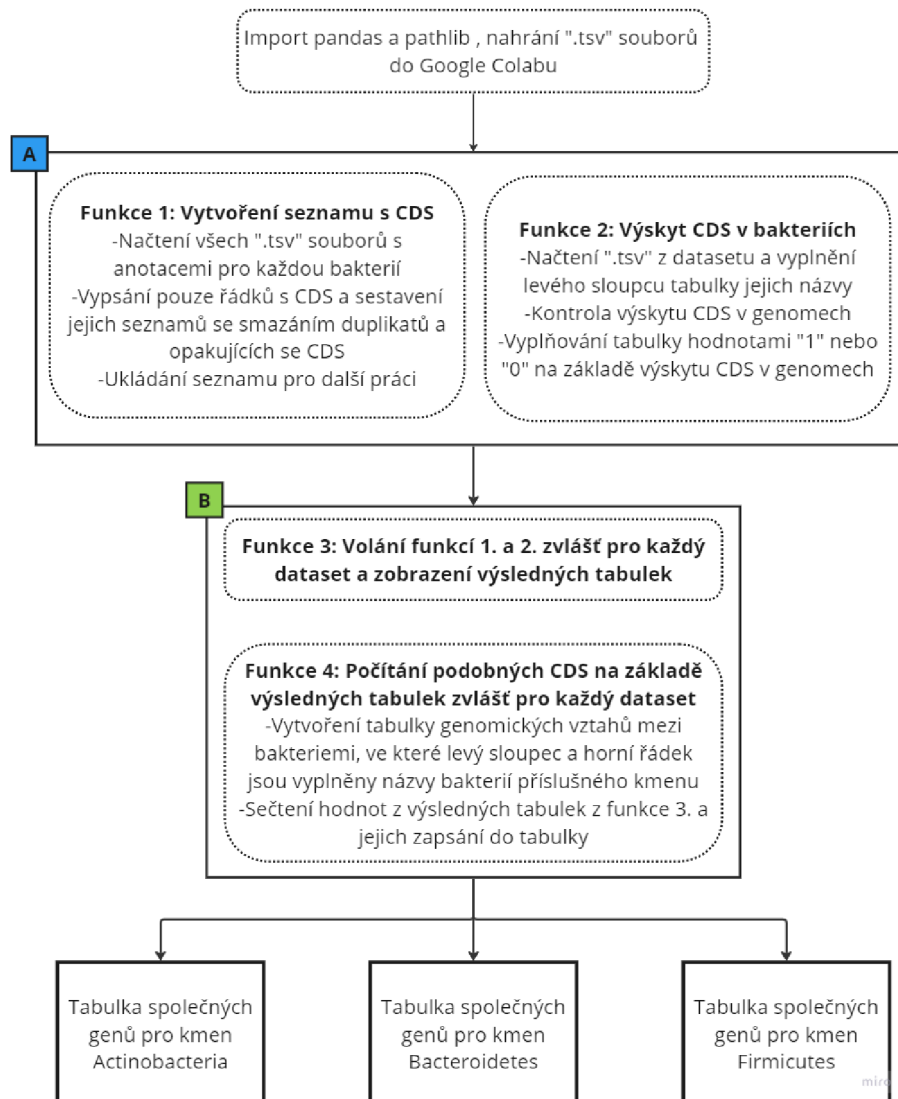
Nástroj RAxML v8 [37] je populární program používaný pro fylogenetickou analýzu velkých souborů dat podle maximální podobnosti. Největší výhodou tohoto nástroje je rychlý algoritmus, který objeví podobnosti, jež představují fylogenetické stromy. Implementuje standardní neparametrický bootstrap [38] a také rychlý bootstrap [39]. Kromě dat o DNA a proteinech podporuje RAxML také binární, vícevrstevná morfologická data a data o sekundární struktuře RNA. U všech výše uvedených typů dat dokáže korigovat zkreslení při určování.

Nástroj FastTree v2.1.10 [40] je založen na principu „minimální evoluce“, tzn. provádí optimalizaci, která minimalizuje součet délek větví. FastTree používá heuristickou variantu spojování sousedů k rychlému nalezení počátečního stromu a ke zpřesnění topologie používá výměny nejbližších sousedů. S přístupem maximální pravděpodobnosti je evoluce explicitně modelována pomocí matice rychlosti přechodu

a pomocí stromu, který nejlépe vizualizuje shluky v stromové struktuře použitých dat, tj strom s nejvyšší pravděpodobností – je vybrán jako nejlepší strom.

### 3.3 Popis implementované funkce identifikace MGE

K identifikaci horizontálně přenesených genů byl vytvořen skript v programovacím jazyce Python v online-prostředí Google Colab [48], přiložený k práci v rámci elektronické přílohy (B). Daný skript lze rozdělit na několik sekcí, viz Obrázek 9. Na začátku byly importovány balíčky a moduly *pandas* v2.0.1 [49] a *pathlib* v1.0.1 [50] nezbytné pro práci s tabulkami a pro rychlou manipulaci s adresáři. Jako výchozí data byla použita hlavní výstupní anotace ve formátu „.tsv“, které jsou popsány v podkapitole 3.1 a se kterými dále pracuje celý implementovaný skript. Soubory byly rozděleny na vhodné vstupní datasety pro analýzu HGT na základě rozdělení kmenů bakterií – Actinobacteria, Bacteroidetes a Firmicutes, viz Obrázek 11.



Obrázek 9: Popis implementovaného skriptu



Sekci A v Obrázku 9 lze rozdělit do dvou obecných funkcí. V první funkci se procházejí všechny soubory „.tsv“ a vypisují se řádky s CDS, které se následně zapisují do seznamu a ukládají se pro další práci. Ve druhé funkci se načítají všechny soubory „.tsv“ a na základě seznamu CDS, z první funkce, se identifikuje výskyt daných CDS v každém anotovaném souboru. Výsledkem daného kroku analýzy je binární klasifikace přítomnosti (1) nebo nepřítomnosti (0) CDS v souborech. Výstup analýzy se zapisuje do nové prázdné tabulky, ve které první sloupec je vyplněn názvy bakterií zkoumaného kmenu a prvním řádkem tabulky je seznam CDS z výstupu první funkce.

V sekci B se volají obě funkce ze sekce A, ale přitom se nastavuje výchozí adresář s anotovanými soubory bakterií příslušného kmenu, aby byl obdržen výsledek pro každý kmen zvlášť. Výsledkem třech běhů skriptu jsou tři tabulky, které obsahují informace o přítomnosti daných CDS v bakteriálních genomech, každá v rámci jednoho kmenu. Dále následuje poslední krok analýzy, ve kterém se na základě výstupních tabulek sestavují nové konečné tabulky relací mezi bakteriemi v rámci kmenu. Podkladem pro analýzu slouží hodnoty, které značí přítomnost CDS v genomech bakterií (viz sekce A). Na základě principu porovnání „každý s každým“ funkce prochází tabulku po párech bakterií, sečítá všechny hodnoty reprezentující přítomnost CDS a zapisuje výsledné číslo do nové tabulky na pozici pro porovnané genomy. Postup se opakuje pro každý kmen zvlášť.

Celkovým výstupem této části jsou tři tabulky relací mezi bakteriemi každého kmenu. Tyto tabulky jsou vhodně předzpracovány (např. vynulováním diagonály reprezentující geny pro tentýž genom) pro následnou síťovou analýzu pomocí programu Cytoscape v3.9.1 [51].

### 3.4 Metody síťové analýzy

Horizontální přenos je v rámci práce vizualizován pomocí sítí sestrojených v programu Cytoscape v3.9.1 [51]. Cytoscape je nástroj s otevřeným přístupem pro vizualizaci a analýzu sítí. Je používán především pro zobrazování a analýzu biologických sítí, jako jsou proteomové, genomové nebo sítě proteinových interakcí. Cytoscape umožňuje importovat síťová data z různých zdrojů a zobrazit je ve formě grafů s uzly a hrany. Uzly představují jednotlivé entity (v této práci se jedná o genomy bakterií konkrétního kmene) a hrany představující vztahy mezi nimi (tj. společně sdílené geny, respektive počet přenesených mobilních genetických elementů). Cytoscape je užitečným nástrojem pro biologický výzkum, kde bývá použit k porozumění složitým biologickým systémům a jejich interakcím.

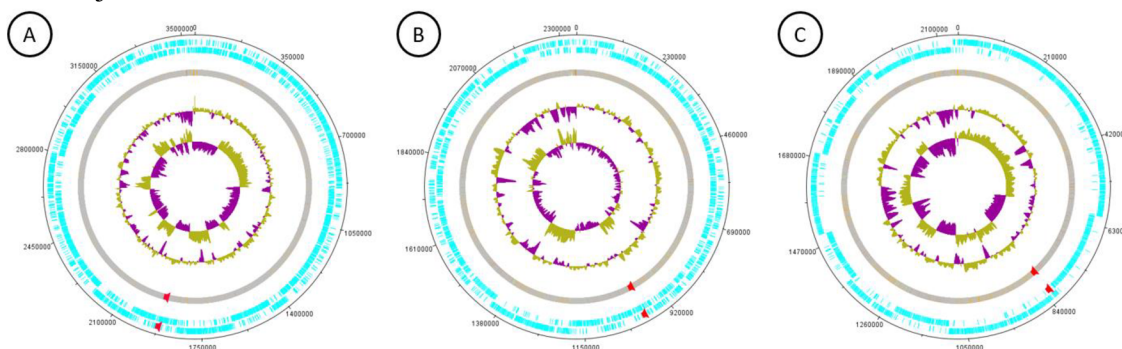
Pro vizualizaci sítí do programu Cytoscape byly nahrány tabulky z výstupu skriptu pomocí pluginu MatReader [52] sloužícím k načítání biologických dat ve formátu matice. Jako řada dalších doplňkových rozšíření Cytoscape, tento plugin se musí stáhnout přes záložku „Apps“ v horním panelu.

## 4. VIZUALIZACE A INTERPRETACE VÝSLEDKŮ PROVEDENÝCH ANALÝZ

Kapitola uvádí vizualizace všech provedených analýz dle navrženého schématu v kapitole 3. Jelikož výskyt HGT můžeme předpokládat s největší pravděpodobností u jedinců s největším počtem identifikovaných MGE, pro všechny provedené analýzy byly vybrány nejrozsáhlejší kmeny Actinobacteria, Bacteroidetes a Firmicutes. Následující podkapitoly obsahují výsledky a diskuse k detekci MGE ve vybraných kmenech a ověření nalezení daných MGE na základě publikovaných biologických vědeckých studií, u kterých již byl prokázán horizontální přenos na základě laboratorních experimentů.

### 4.1 Vizualizace anotovaných genetických elementů

Hlavní výstupní „gff“ soubory reprezentující anotaci genomů popsané v podkapitole 3.1 byly nahrány a vizualizovány pomocí nástroje DNAPlotter [47]. Během analýzy byla nalezena řada společně sdílených genetických komponent. Ve studii Lamberta P. [53] o bakteriální rezistenci k antibiotikům se uvádí gen *mecA*, který se podílí na rezistenci bakterií k methicilinu. Tento gen kóduje penicilin-vazebný protein, který patří mezi enzymy zapojené do syntézy peptidoglykanové složky bakteriální buněčné stěny, a předpokládá se, že se přenáší horizontálně. V průběhu anotace daný gen byl nalezen u třech jedinců kmene Firmicutes – *Enterococcus gallinarum*, *Lacrimispora saccharolytica* a *Hydrogenoanaerobacterium saccharovorans*. Lokalizace daného genu ve všech třech genomech je označena červeně na Obrázek 10.



Obrázek 10: Grafické zobrazení lokalizace genu *mecA* v genomech *Enterococcus gallinarum* (A), *Hydrogenoanaerobacterium saccharovorans* (B) a *Lacrimispora saccharolytica* (C)

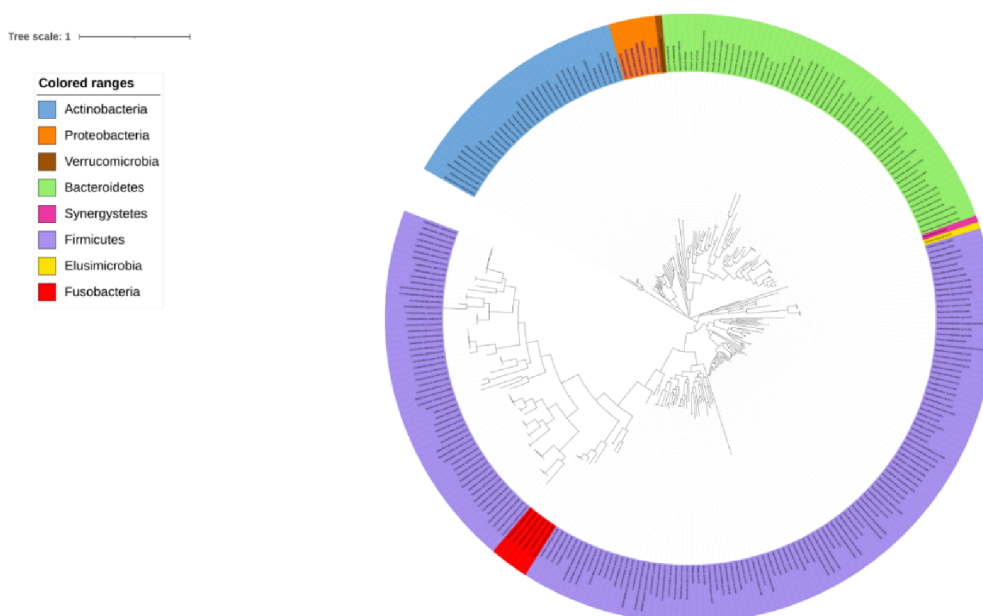
V genomech náležejících do kmenů Bacteroidetes byly také nalezeny genetické elementy, u kterých se podle studie Shoemakera N. B. a spol. [54] taktéž předpokládá horizontální přenos. Jsou to geny *tetQ* (gen rezistence k tetracyklinu), *ermF* a *ermG* (geny rezistence k erytromycinu), které se podle výzkumu přenáší na konjugativních

transpozonech. U řady jedinců z kmenu Actinobacteria byl nalezen gen *mepA*, podílející se na biosyntéze melaninu podle analýzy ve výzkumu Chizhevské a spol [55].

Zmíněné bakterie, u kterých dané geny byly nalezeny, nejsou geneticky příbuzné. Ověřit danou skutečnost je možné pomocí fylogenetické analýzy, popsané v následující podkapitole.

## 4.2 Fylogenetická analýza

Zkoumání taxonomií souboru dat odhalilo společné genomické vlastnosti mezi větším počtem sekvencí. Po aplikaci UBCG balíčku na počáteční dataset genomů byly výsledné soubory následně použity pro sestavení fylogenetického stromu pomocí online nástroje iTOL – angl. *Interactive Tree of Life*, viz Obrázek 11.

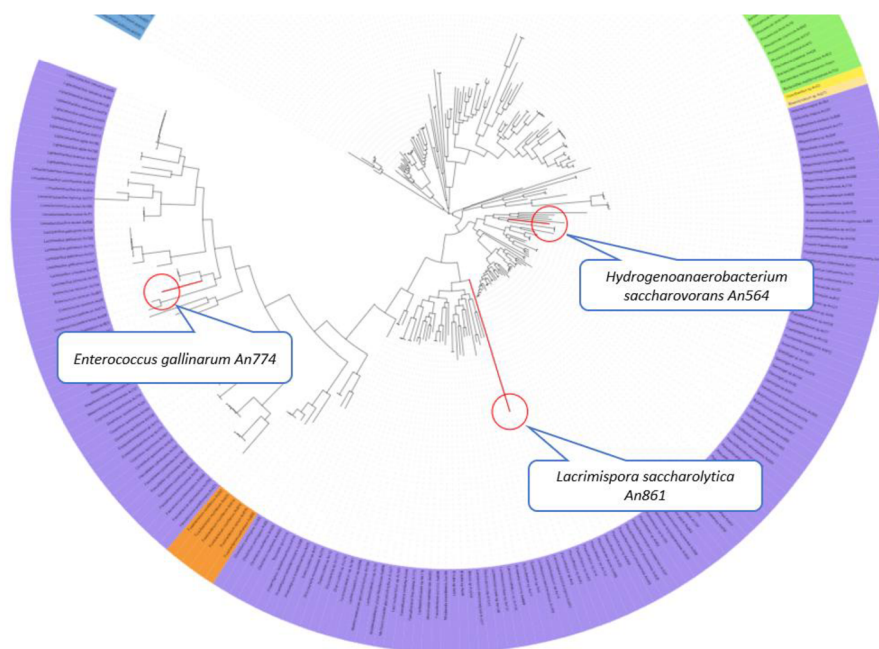


Obrázek 11: Zobrazení genomů izolovaných anaerobů z datasetu pomocí fylogenetického stromu online-nástrojem iTOL

Genomy jsou rozděleny do 8 různých kmenů – Actinobacteria, Bacteroidetes, Firmicutes, Fusobacteria, Proteobacteria, Synergystetes, Elusimicrobia a Verrucomicrobia byli zvláště zvýrazněny na stromu – *Akkermansia muciniphila* (hnědá), *Cloacibacillus sp* (růžová), *Elusimicrobium sp* (žlutá). Ke kmenu Actinobacteria bylo přiřazeno 40 genomů, 56 genomů k Bacteroidetes, 163 genomů k Firmicutes, 7 k Proteobacteria a 6 k Fusobacteria. Nejvýraznější větvení lze pozorovat u fylů z kmenu Firmicutes, mezi kterými se začlenili genomy z kmenu Fusobacteria, jelikož jsou genomicky příbuzní.

Daná analýza přispívá k lepšímu pochopení fylogenetických vztahů mezi jedinci, pomáhá ověřit genomickou podobnost organismů a hodnotit pravděpodobnost výskytu

horizontálního přenosu genetických komponent. Jak již bylo uvedeno v podkapitole 4.1, v rámci každého z nejrozsáhlejších zkoumaných kmenů byla nalezena řada společně sdílených genetických komponent. Hypotéza, že uvedená komponenta byla přenesena horizontálně, je možné ověřit vyhledáním zkoumaných organismů na fylogenetickém stromu. Genomy bakterií z kmenu Firmicutes uvedené v předchozí podkapitole na Obrázek 10 byly také vyhledány na fylogenetickém stromu, viz Obrázek 12.



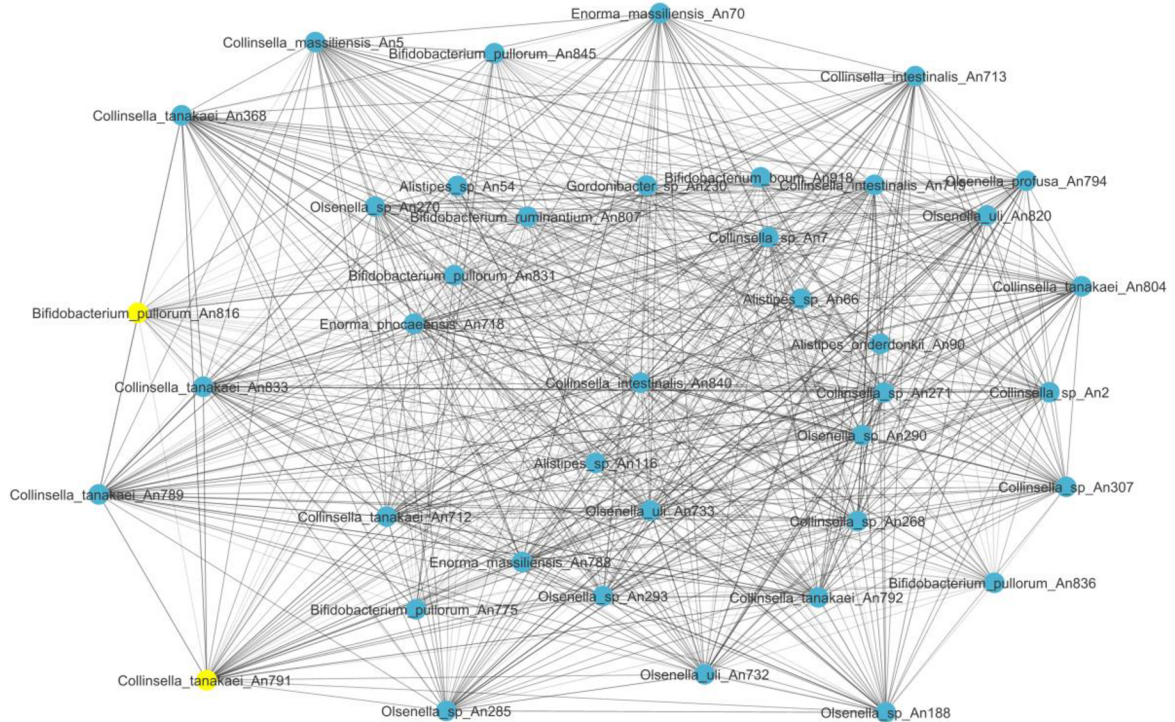
Obrázek 12: Fylogenetický vztah bakteriálních genomů kmenu Firmicutes, nesoucích gen *mecA*

Podle informací, získaných pomocí fylogenetické analýzy, u genu *mecA* je možné předpokládat horizontální přenos, jelikož zkoumané organismy jsou genomicky vzdálené a náleží do různých podvětví sestaveného stromu. Naopak, některé genomy bakterií z kmenu Actinobacteria, u kterých byl dříve nalezen gen *mepA*, se nacházejí těsně vedle sebe a náleží do stejné větvi. Tato skutečnost znamená, že gen *mepA* mezi nimi mohl být přenášen od společného předka, avšak hypotézu horizontálního přenosu nelze vyloučit, jelikož daný gen byl také nalezen v genomech jedinců z kmenů Firmicutes a Bacteroidetes společně s geny *tetQ*, *ermG* a *ermF*, které se podílejí na rezistenci k antibiotikům a vyskytují se nejvíc u jedinců z kmenu Bacteroidetes.

### 4.3 Statická síťová analýza

Síťová analýza v rámci této bakalářské práce se zabývá propojením genomů organismů do sítí na základě společně sdílených genů. Jelikož v rámci jednoho kmenu je společně sdílených genů více, z důvodu vertikálního přenosu pro každý kmen byla vytvořena vlastní síť, viz Obrázek 13. Vizualizace byla provedena pomocí nastavení

parametru průhlednost hran na základě počtu společně sdílených příznaků. Například u *Bifidobacterium pullorum* An816 a *Collinsella tanakaei* An791 (na Obrázek 13 jsou zvýrazněné žlutě) je číslo společně sdílených MGE malé, proto hrana mezi nimi je téměř neviditelná.



Obrázek 13: Zobrazení genomických vztahů v kmenu Actinobacteria pomocí sítí v Cytoscape

Pro každou vytvořenou síť byla dále provedena statická analýza určena několika základními parametry, viz Tabulka 3. Počet uzlů (angl. *number of nodes*) určuje počet bakteriálních genomů v síti. Počet hran (angl. *number of edges*) udává celkový počet interakcí mezi propojenými genomy. Průměrný počet sousedů (angl. *average number of neighbours*) uvádí průměrný počet sousedních uzlů spojených s každým uzlem v síti. Jelikož zkoumané genomy sdílí velké množství společných genů, všechny genomy v síti jsou spojené mezi sebou a každý má tak počet sousedů o jeden méně, než je celkový počet genomů v síti. Koeficient shlukování (angl. *clustering coefficient*) udává míru shlukování nebo vzájemné propojenosti sousedních uzlů v síti.

Charakteristická délka cesty (angl. *characteristic path length*) vyjadřuje průměrný počet hran, které je třeba překonat, aby se dosáhlo z jednoho uzlu na druhý. Menší hodnoty charakteristické délky cesty naznačují, že síť je dobře propojená a informace se mohou rychle šířit mezi uzly. Síťová hustota (angl. *network density*) vyjadřuje, jaké procento možných hran je ve skutečnosti přítomno v síti. Hodnota síťové hustoty se pohybuje v rozmezí 0 až 1.

Tabulka 3: Výsledky statické síťové analýzy

Parametry sítě	Statická síťová analýza kmenů		
	Actinobacteria	Bacteroidetes	Firmicutes
Počet uzlů	40	56	163
Počet hran	1560	3080	26406
Průměrný počet sousedů	39,000	55,000	162,000
Koeficient shlukování	1,000	1,000	1,000
Charakteristická délka cesty	1,000	1,000	1,000
Síťová hustota	1,000	1,000	1,000

Podle výsledků statické síťové analýzy počet uzlů v každé síti odpovídá počtu zkoumaných genomů každého kmenu ve výchozím datasetu. Kvůli přítomnosti velkého množství společných genetických elementů u bakterií včetně přenesených vertikálním způsobem, byly všechny genomy v síti navzájem propojeny. Tuto skutečnost udávají hodnoty „1“ parametrů koeficientu shlukování (každá síť je jedním shlukem), síťové hustoty a charakteristické délky cesty. Proto i počet přímých sousedů pro každý genom odpovídá počtu zbylých genomů, avšak takové rovnoměrné rozložení a parametry sítě nepřispívají k nalezení specifických vlastností genomů anebo k spekulaci o výskytu horizontálního přenosu MGE. Pro odhalení daných vlastností se vyžaduje manuální aplikace filtrů a další úpravy sítě pro lepší přehlednost (např. nastavení průhlednosti hran na základě počtu sdílených MGE). Dalším možným řešením daného problému je úprava vstupních dat (tabulek relací – viz podkapitola 3.3) takovým způsobem, aby vyskytující se v každém genomu MGE (nebo geny zděděné po společném předku) se neuvažovaly při sestavení sítě. Po takové úpravě dat by se síť následně sestavovala pouze na základě genových rozdílů a počet propojených genomů by výrazně klesl. V takovém případě by hrany nesly více konkrétních informací o pravděpodobném výskytu HGT nebo jiných genetických změnách, které nebyly ovlivněny běžnou dědičností.

Pomocí provedené statické síťové analýzy je možné odhalit zkoumané vlastnosti za použitím filtrace zobrazovaných dat. Kvůli velkému množství hran (tj. 26406) v rámci sítě reprezentující nejrozsáhlejší kmen Firmicutes byl pro vizualizaci použit filtr umožňující zobrazit pouze hrany s největším počtem společně sdílených MGE. Po aplikaci filtru většina hran zmizela a zbyly pouze ty, u kterých počet sdílených MGE odpovídal mezím filtru – 900 až 1560, viz. Obrázek 14.



Obrázek 14: Síť kmenu Firmicutes po filtraci

Zároveň po filtraci byl ověřen předpokládaný výskyt horizontálního přenosu genu *mecA* (viz Obrázek 12) u třech jedinců Firmicutes, kteří ve výsledné síti zůstali propojeni i po odstranění většího počtu hran (jsou zvýrazněny žlutě – červenou na Obrázek 14). Obdobným způsobem byl nalezen výskyt HGT genů *tet(Q)* a *ermF* u bakterií kmenu Bacteroidetes. Tyto skutečnosti potvrzují správnou identifikaci MGE mezi danými bakteriemi.

Na základě provedených analýz a jejich výsledků je možné konstatovat, že celkový postup navržené a implementované analýzy v této bakalářské práci je úspěšný a je možné tuto analýzu reprodukovat na dalších datech pro správnou identifikaci MGE související s HGT.

## 5. ZÁVĚR

Bakalářská práce se zabývá problematikou horizontálního přenosu genetických komponent, který má velký význam jak v evoluci bakterií, tak i v rozvoji bakteriálních onemocnění u člověka. Teoretická část práce se zabývá popisem způsobů přenosu genetických komponent mezi bakteriemi. Následně se práce věnuje popisu několika nyní dostupných sekvenačních technik. Konkrétně jsou zde popsány techniky Illumina, Ion Torrent a sekvenace Nanopore.

V praktické části práce je představen návrh algoritmu zpracování genomu bakterií různých kmenů. Podle daného algoritmu se postupně prováděly jednotlivé analýzy.

Zprvce byla provedena anotace bakteriálních genomů pomocí softwarového balíčku nástrojů Prokka a následně fylogenetická analýza pomocí balíčku nástrojů UBCG. Dále byly výsledky provedených analýz graficky zobrazeny pomocí příslušných nástrojů. Pro grafické zobrazení genomických struktur a porovnání genomů bakterií byl použit nástroj DNAPlotter, pomocí kterého se vyhledaly specifické geny, které za teoretických předpokladů byly přeneseny horizontálním způsobem. Následné použití online-nástroje iTOL pro sestavení fylogenetického stromu odhalilo skutečnost, že zmíněné genomy náleží do různých větví stromu, tím se potvrdila pravděpodobnost HGT.

Nezbytnou součástí dané analýzy byl návrh vlastního skriptu v programovacím jazyce Python, který seřadil bakteriální genomy do tabulek a vyhledával společně sdílené geny. Na základě výstupních dat genomy byly rozříděny do příslušných kmenů a byly rekonstruovány do sítí pomocí programu Cytoscape. Následně byla provedena statická analýza daných sítí a manuální úpravy pro lepší zobrazení. Vzhledem k tomu, že se sítě sestavovaly na základě všech společných MGE v rámci kmenu včetně přenesených vertikálně, všechny bakteriální genomy byly propojené mezi sebou, a proto byly na rekonstruované sítě použité filtry umožňující zobrazování podle vybraných vlastností. Daná analýza potvrdila teoretické předpoklady horizontálního přenosu genů tetracyklinové a erytromycinové rezistence *tet(Q)* a *ermF*, u Firmicutes se potvrdil přenos genu *mecA*, podílející se na rezistenci bakterií k methicilinu.

Celkové hodnocení správnosti navrženého a implementovaného algoritmu spočívalo v ověření teoretických předpokladů výskytu HGT u bakteriálních genomů z použitého datasetu. Navržená analýza odhalila výskyt HGT v rámci jednotlivých kmenů a umožnila daný přenos znázornit graficky.



## LITERATURA

- [1] Kaiser, Gary. Microbiology [online]. Dostupné z: [https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A\\_Microbiology\\_\(Kaiser\)](https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A_Microbiology_(Kaiser))
- [2] Hedman, Hayden & Vasco, Karla & Zhang, Lixin. (2020). A Review of Antimicrobial Resistance in Poultry Farming within Low-Resource Settings. *Animals*. 10. 1264. 10.3390/ani10081264.
- [3] Bursová, Šárka, Lenka Necidová, and Marta Dušková. *Mikrobiologie potravin a mikrobiologické laboratorní metody: Obecná mikrobiologie*. Veterinární a farmaceutická univerzita Brno, 2014.
- [4] Holada, Karel. Genetika bakterií, antibiotika [online]. Dostupné z: <http://fb.cuni.cz/Data/files/UIM/materials/doc/Osetrovatelstvi%20rocnik/4-genetika-bakterii.pdf>
- [5] Ho TD, Waldor MK. Enterohemorrhagic *Escherichia coli* O157:H7 gal mutants are sensitive to bacteriophage P1 and defective in intestinal colonization. *Infect Immun*. 2007 Apr;75(4):1661-6. doi: 10.1128/IAI.01342-06. Epub 2006 Dec 11. PMID: 17158899; PMCID: PMC1865682.
- [6] merodiploid. *Oxford Reference*. Retrieved 28 Dec. 2022, from <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100151596>.
- [7] Thomas, C., Nielsen, K. Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria. *Nat Rev Microbiol* 3, 711–721 (2005). <https://doi.org/10.1038/nrmicro1234>
- [8] Plesník, Jan. "Příroda jako proudící mozaika." *Co přinesly novější poznatky ekosystémové ekologie*. *Ochrana přírody* 65.3 (2010): 27-30.
- [9] Hiramatsu K, Cui L, Kuroda M, Ito T (2001) The emergence and evolution of methicillin-resistant *Staphylococcus aureus*. *Trends Microbiol*. 9:486–93 PMID: 11597450
- [10] SCHWARZEROVÁ, J.; ZEMAN, M.; RYCHLIK, I.; WECKWERTH, W.; PROVAZNIK, I.; DOLEJSKA M.; CEJKOVA C.; Systems biology approach for analysis of mobile genetic elements in chicken gut microbiome. In 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2022 – in print.
- [11] Huang J. 2013 Horizontal gene transfer in eukaryotes: the weak-link model. *Bioessays* 35, 868–875.
- [12] PAIJMANS, Johanna LA, et al. Sequencing single-stranded libraries on the Illumina NextSeq 500 platform. arXiv preprint arXiv:1711.11004, 2017.
- [13] Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463–5467
- [14] Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed*. 2013 Dec;98(6):236-8. doi: 10.1136/archdischild-2013-304340. Epub 2013 Aug 28. PMID: 23986538; PMCID: PMC3841808.

- [15] Verma M, Kulshrestha S, Puri A. Genome Sequencing. *Methods Mol Biol.* 2017; 1525:3-33. doi: 10.1007/978-1-4939-6622-6\_1. PMID: 27896715.
- [16] Hu Z, Cheng L, Wang H. The Illumina-solexa sequencing protocol for bacterial genomes. *Methods Mol Biol.* 2015; 1231:91-7. doi: 10.1007/978-1-4939-1720-4\_6. PMID: 25343860.
- [17] Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett.* 2013 Nov 1;340(2):284-95. doi: 10.1016/j.canlet.2012.11.025. Epub 2012 Nov 19. PMID: 23174106; PMCID: PMC5739311.
- [18] NextSeq 500 technology [online]. Dostupné z: <https://www.illumina.com/systems/sequencing-platforms/nextseq.html>
- [19] Rahman, M. T., Uddin, M. S., Sultana, R., Moue, A., & Setu, M. (2013). Polymerase Chain Reaction (PCR): A Short Review. *Anwer Khan Modern Medical College Journal*, 4(1), 30–36.
- [20] BaseSpace Sequence Hub [online]. Dostupné z: [https://support.illumina.com/sequencing/sequencing\\_software/basespace/training.html](https://support.illumina.com/sequencing/sequencing_software/basespace/training.html)
- [21] Illumina. Faster sequencing and data processing [online]. Dostupné z: <https://www.illumina.com/science/technology/next-generation-sequencing/sequencing-technology/2-channel-sbs.html>
- [22] Hui P. (2014) Next – generation sequencing: chemistry, technology and application. *Top CurrChem* 336:1-18
- [23] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, Bettman Betal (2010). Real-time DNA sequencing from single polymerasemolecules. *Science*323:133-138
- [24] F. Chang, G.L. Liu, C.J. Liu, M.M. Li. Somatic diseases (cancer) amplification-based next-generation sequencing. S. Kulkarni, J. Pfeifer (Eds.), *Clinical Genomics: A Guide to Clinical Next Generation Sequencing*, Elsevier/Academic Press, Cambridge, MA (2015), pp. 298-317
- [25] Venkatesan BM, Bashir R (2011) Nanopore Sensors for nucleic acid analysis. *Nat Nanotechnol* 6:615-624
- [26] Clarke J, Wu HC, Jayasinghe L, Patel A, Reid S, Bayley H (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 4:265–270
- [27] What is Oxford Nanopore (ONT) sequencing [online]. Dostupné z: <https://www.yourgenome.org/facts/what-is-oxford-nanopore-technology-ont-sequencing/>
- [28] MEDVECKY, Matej, et al. Whole genome sequencing and function prediction of 133 gut anaerobes isolated from chicken caecum in pure cultures. *BMC genomics*, 2018, 19.1: 1-15.
- [29] Varmuzova K, Kubasova T, Davidova-Gerzova L, Sisak F, Havlickova H, Sebkova A, Faldynova M, Rychlik I. Composition of gut microbiota. Influences

- resistance of newly hatched chickens to Salmonella Enteritidis infection. *Front Microbiol.* 2016; 7:957.
- [30] MetaCentrum (MetaVO) – virtuální organizace pro celou akademickou obec [online]. Dostupné z: <https://metavo.metacentrum.cz/>
- [31] Ivica Letunic, Peer Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Research*, Volume 49, Issue W1, 2 July 2021, p.W293–W296, <https://doi.org/10.1093/nar/gkab301>
- [32] Torsten Seemann, Prokka: rapid prokaryotic genome annotation, *Bioinformatics*, Volume 30, Issue 14, 15 July 2014, Pages 2068–2069, <https://doi.org/10.1093/bioinformatics/btu153>.
- [33] NA, Seong-In, et al. UBCG: up-to-date bacterial core gene set and pipeline for phylogenomic tree reconstruction. *Journal of Microbiology*, 2018, 56.4: 280-285.
- [34] HYATT, D., et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *bmc bioinformatics* 11, 119–119, doi: 10.1186. 2010
- [35] Johnson, L.S., Eddy, S.R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* 11, 431 (2010). <https://doi.org/10.1186/1471-2105-11-431>
- [36] Katoh, K., Standley, D.M. (2014). MAFFT: Iterative Refinement and Additional Methods. In: Russell, D. (eds) *Multiple Sequence Alignment Methods. Methods in Molecular Biology*, vol 1079. Humana Press, Totowa, NJ. [https://doi.org/10.1007/978-1-62703-646-7\\_8](https://doi.org/10.1007/978-1-62703-646-7_8)
- [37] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014 May 1;30(9):1312-3. doi: 10.1093/bioinformatics/btu033. Epub 2014 Jan 21. PMID: 24451623; PMCID: PMC3998144.
- [38] Kozlov, A.M.; Stamatakis, A. Using RAxML-NG in Practice. Preprints 2019, 2019050056 (doi: 10.20944/preprints201905.0056.v1).
- [39] Alexandros Stamatakis, Paul Hoover, Jacques Rougemont, A Rapid Bootstrap Algorithm for the RAxML Web Servers, *Systematic Biology*, Volume 57, Issue 5, October 2008, Pages 758–771, <https://doi.org/10.1080/10635150802429642>
- [40] PRICE, Morgan N.; DEHAL, Paramvir S.; ARKIN, Adam P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS one*, 2010, 5.3: e9490.
- [41] scaffold. Oxford Reference. Retrieved 28 Dec. 2022, from <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100444102>.
- [42] K. Lagesen, P. Hallin, E. A. Rødland, Hans-Henrik Stærfeldt, T. Rognes, David W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes, *Nucleic Acids Research*, Vol. 35, Issue 9, 1 May 2007, p. 3100–3108, <https://doi.org/10.1093/nar/gkm160>.
- [43] D. Laslett, B. Canback, ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences, *Nucleic Acids Research*, Volume 32, Issue 1, 1 January 2004, Pages 11–16, <https://doi.org/10.1093/nar/gkh152>.

- [44] Almagro Armenteros, J.J., Tsirigos, K.D., Sønderby, C.K. *et al.* SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* **37**, 420–423 (2019). <https://doi.org/10.1038/s41587-019-0036-z>
- [45] Hajar Owji, Navid Nezafat, Manica Negahdaripour, Ali Hajiebrahimi, Younes Ghasemi. A comprehensive review of signal peptides: Structure, roles, and applications. *European Journal of Cell Biology*, vol. 97, issue 6, 2018, p. 422-441, ISSN 0171-9335, <https://doi.org/10.1016/j.ejcb.2018.06.003>.
- [46] Diana L. Kolbe, Sean R. Eddy, Fast filtering for RNA homology search, *Bioinformatics*, Volume 27, Issue 22, 15 November 2011, Pages 3102–3109, <https://doi.org/10.1093/bioinformatics/btr545>.
- [47] Tim Carver, Nick Thomson, Alan Bleasby, Matthew Berriman, Julian Parkhill, DNAPlotter: circular and linear interactive genome visualization, *Bioinformatics*, Volume 25, Issue 1, 1 January 2009, Pages 119–120, <https://doi.org/10.1093/bioinformatics/btn578>
- [48] Bisong, E. (2019). Google Colaboratory. In: Building Machine Learning and Deep Learning Models on Google Cloud Platform. Apress, Berkeley, CA. [https://doi.org/10.1007/978-1-4842-4470-8\\_7](https://doi.org/10.1007/978-1-4842-4470-8_7)
- [49] Pandas documentation [online]. Dostupné z: <https://pandas.pydata.org/docs/index.html>
- [50] Pathlib documentation [online]. Dostupné z: <https://pathlib.readthedocs.io/en/pep428/#>
- [51] Lukasz Salwinski, David Eisenberg, The MiSink Plugin: Cytoscape as a graphical interface to the Database of Interacting Proteins, *Bioinformatics*, Volume 23, Issue 16, 15 August 2007, Pages 2193–2195, <https://doi.org/10.1093/bioinformatics/btm304>
- [52] aMatReader: A Cytoscape Reader for Adjacency Matrices [online]. Dostupné z: <https://www.cgl.ucsf.edu/cytoscape/utilities3/amatreader.shtml>
- [53] LAMBERT, P. Bacterial resistance to antibiotics: Modified target sites. *Advanced Drug Delivery Reviews*, 57(10), 1471–1485. doi:10.1016/j.addr.2005.04.003
- [54] Shoemaker, N. B., Vlamakis, H., Hayes, K., & Salyers, A. A. (2001). Evidence for Extensive Resistance Gene Transfer among *Bacteroides* spp. and among *Bacteroides* and Other Genera in the Human Colon. *Applied and Environmental Microbiology*, 67(2), 561–568. doi:10.1128/aem.67.2.561-568.2001
- [55] Chizhevskaya, E. P., Naidenova, E. A., Onishchuk, O. P., Andronov, E. E., & Simarov, B. V. (2018). The Melanin Biosynthesis Gene from the CA15-1 Strain of Alfalfa Nodule Bacteria: Molecular Analysis and Phylogeny. *Russian Journal of Genetics*, 54(8), 925–932. doi:10.1134/s1022795418080045
- [56] Su G, Morris JH, Demchak B, Bader GD. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics*. 2014 Sep 8;47:8.13.1-24. doi: 10.1002/0471250953.bi0813s47. PMID: 25199793; PMCID: PMC4174321.

## **SEZNAM PŘÍLOH**

<b>PŘÍLOHA A - SHELL SKRIPTY PRO METACENTRUM.....</b>	<b>44</b>
<b>PŘÍLOHA B - ELEKTRONICKÁ PŘÍLOHA .....</b>	<b>45</b>

# Příloha A - Shell skripty pro Metacentrum

## A.1 Shell skript pro nástroj Prokka

```
##*****  
*****  
#přidat modul na condu (metacentrum)  
module add conda-modules-py37  
#aktivace prokky v conde  
conda activate prokka  
#for cyklus pro celý dataset genomů  
cd DatasetFaa  
for i in *.faa  
do  
prokka --kingdom Bacteria --outdir Anotace${i%.faa} --force --prefix  
${i%.faa} --addgenes --cpus 8 $i  
done
```

## A.2 Shell skript pro pipeline UBCG

```
##*****  
*****  
#Instalace modulů popořadě  
module avail prodigal  
module add prodigal-2.6.3  
module avail hmmer  
module add hmmer-3.1b1-intel  
module avail mafft  
module add mafft-7.487  
module avail fasttree  
module add fasttree-2.1.8  
module avail raxml  
module add raxml-7.3.0  
module add jdk-8  
#for cyklus  
cd Dataset  
array=(ls -d *.fasta)  
for ((i=0;i<("${#array[@]}");i+=1));  
do java -jar UBCG.jar extract -i Dataset/${array[i]} -bcg_dir bcg_files  
-label ${array[i]%%.*}  
done  
java -jar UBCG.jar align -bcg_dir bcg_files -t 8
```

## **Příloha B - Elektronická příloha**

Seznam souborů v elektronické příloze

- Vlastní skript formátu „*ipynb*“: Skript\_BP.ipynb
- Složka Cytoscape s výslednými sítěmi formátu „*cyt*“
- Tabulka formátu „*csv*“ se všemi CDS nalezenými v genomech: Lists CDS.csv