

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

CHARAKTERISTIKA ČREVNEJ
MIKROFLÓRY U LABORATÓRNYCH
ZVIERAT POMOCOU SEKVENOVANIA
TRETEJ GENERÁCIE

BAKALÁRSKA PRÁCA

2023

MONIKA JANKECHOVÁ

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

CHARAKTERISTIKA ČREVNEJ
MIKROFLÓRY U LABORATÓRNYCH
ZVIERAT POMOCOU SEKVENOVANIA
TRETEJ GENERÁCIE

BAKALÁRSKA PRÁCA

Študijný program: Bioinformatika
Študijný odbor: Informatika a biológia
Školiace pracovisko: Katedra informatiky
Školiteľ: Mgr. Barbora Konečná, PhD.

Bratislava, 2023
Monika Jankechová



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Monika Jankechová
Študijný program: bioinformatika (Medziodborové štúdium, bakalársky I. st., denná forma)
Študijné odbory: informatika
biológia
Typ záverečnej práce: bakalárska
Jazyk záverečnej práce: slovenský
Sekundárny jazyk: anglický

Názov: Charakteristika črevnej mikroflóry u laboratórných zvierat pomocou sekvenovania tretej generácie
Characterization of intestinal microflora in laboratory animals using third-generation sequencing

Anotácia: Cieľom práce je charakterizovať odlišnosti v zastúpení mikrobov v črevách zvierat pred a po užívaní antibiotík a taktiež zistiť vplyv pitia odpadovej vody na črevnú mikroflóru.

Vedúci: Mgr. Barbora Konečná, PhD.
Katedra: FMFI.KI - Katedra informatiky
Vedúci katedry: prof. RNDr. Martin Škoviera, PhD.
Dátum zadania: 27.10.2022

Dátum schválenia: 27.10.2022

doc. Mgr. Bronislava Brejová, PhD.
garant študijného programu

.....
študent

.....
vedúci práce

Pod'akovanie:

Touto cestou by som sa chcela poďakovať vedúcej bakalárskej práce Mgr. Barbore Konečnej, PhD. za rady a správne nasmerovanie pri vypracovaní bakalárskej práce. Poďakovanie by som chcela vyjadriť aj všetkým, ktorí mi umožnili využiť ich hardvér na spracovanie dát. Na záver by som chcela poďakovať svojej rodine a Filipovi za oporu pri práci.

Abstrakt

Črevný mikrobióm organizmov obsahuje vysoké množstvo mikroorganizmov, ktoré ovplyvňujú fungovanie jedinca. Na identifikáciu mikrobiálnych spoločenstiev je možné použiť mnoho rôznych nástrojov a softvérov. V tejto práci sme využili sekvenovanie tretej generácie spolu s basecallerom Dorado a workflowmi Oxford Nanopore Technologies. Dospeli sme k záveru, že tieto nástroje sú vhodné na identifikáciu mikrobiálnych spoločenstiev. Basecaller Dorado produkuje kvalitné čítania porovnateľné s už zaužívaným basecallerom pre Oxford Nanopore čítania - Guppy. Workflow wf-metagenomics využívajúci nástroj Kraken 2 a Bracken poskytuje kvalitnú identifikáciu taxónov prítomných vo vzorkách.

Cieľom tejto práce bolo zistiť, ako užívanie antibiotík alebo odpadovej vody dokáže ovplyvniť zastúpenie mikroorganizmov v črevách a tým narušiť správne fungovanie jedinca. Pomocou experimentálnych modelov s laboratórnymi zvieratami sme zistili, že antibiotiká znižujú diverzitu črevnej mikrobioty u myší a odpadová voda výrazne mení zastúpenie jednotlivých druhov mikroorganizmov v črevách u potkanov. Po užívaní odpadovej vody sa výrazne zvyšuje zastúpenie baktérií kmeňa *Firmicutes*. Mierny nárast hojnosti bol pozorovaný aj u kmeňov *Bacteroidota*, *Proteobacteria* a *Actinobacteria*.

Kľúčové slová: črevný mikrobióm, odpadová voda, antibiotiká, wf-metagenomics, Dorado, Kraken 2, sekvenovanie tretej generácie

Abstract

The gut microbiome of organisms contains a high number of microorganisms that affect the functioning of the individual. Many different tools and software can be used to identify microbial communities. In this work, we used third-generation sequencing together with Dorado basecaller and Oxford Nanopore Technologies workflows. We have come to a conclusion that these tools are suitable for the identification of microbial communities. Basecaller Dorado produces high quality reads comparable to the already widely used basecaller for Oxford Nanopore reads - Guppy. The wf-metagenomics workflow using Kraken 2 and Bracken provides high-quality identification of taxons present in the samples.

The aim of this work was to find out how antibiotics or wastewater can affect the abundance of microorganisms in the intestines and thereby disrupt the proper functioning of an individual. Using experimental models with laboratory animals, we found that antibiotics reduce the diversity of intestinal microflora of mice, and wastewater significantly changes the abundance of individual types of microorganisms in the intestines of rats. After using wastewater, the abundance of *Firmicutes* bacteria increases significantly. A slight increase in abundance was also observed in the phylum *Bacteroidota*, *Proteobacteria* and *Actinobacteria*.

Keywords: gut microbiota, wastewater, antibiotics, wf-metagenomics, Dorado, Kraken 2, third-generation sequencing

Obsah

Úvod	1
1 Základné pojmy	3
1.1 Mikrobióm	3
1.2 Sekvenovanie	4
1.2.1 Nanopórové sekvenovanie	4
1.3 Taxonómia	5
1.3.1 Mikroorganizmy črevného mikrobiómu	6
1.3.2 Taxonomická klasifikácia metagenomických dát	7
1.4 Diverzita mikrobiálnych spoločenstiev	8
1.4.1 Index počtu taxónov	9
1.4.2 Shannon-Weaverov index diverzity	9
1.4.3 Simpsonov index divezity	9
2 Metodika práce	11
2.1 Zdroj vzoriek	11
2.2 Spracovanie vzoriek	11
2.3 Spracovanie a analýza dát	13
2.3.1 Basecalling a demultiplexing	13
2.3.2 Porovnanie basecallerov	15
2.3.3 Analýza a vizualizácia metagenomických dát	17
3 Výsledky	19
3.1 Porovnanie behu a výsledkov basecallerov	19
3.1.1 Zarovnanie čítaní k referenčnému genómu	21
3.2 Myši a antibiotiká	22
3.3 Potkany a odpadová voda	26
4 Diskusia	31
Illumina MiSeq verusus MinION sekvenátor	31
Basecalling - Dorado verusus Guppy	32

MinION a rôzne bioinformatické nástroje	32
Pozitívne kontroly	33
Vplyv odpadovej vody na potkany	33
Záver	35
Príloha A	41

Úvod

Odpadovú vodu môžeme definovať ako použitú vodu z domácností, priemyslu či poľnohospodárstva, ktorá je týmto použitím znečistená. Najčastejšie putuje kanalizačným potrubím do čistiarní odpadových vôd (ČOV). V dnešnej dobe rozvinutého priemyslu a poľnohospodárstva sme svedkami čoraz väčšieho znečisťovania vody. Do odpadových vôd sa napríklad priemyselnými praktikami či samotným pôsobením jednotlivca uvoľňujú mnohé zlúčeniny, ktoré ju znečisťujú a všeobecne narušajú vodný cyklus. Viacero vedeckých článkov sa v posledných rokoch zaoberá týmito zlúčeninami nazývanými aj *emerging pollutants*, čo môžeme v preklade pochopiť ako novoobjavujúce sa znečisťujúce látky [1]. Problémom vznikajúcich znečisťujúcich látok je nedostatok vedomostí o ich vplyve na ľudské zdravie, životné prostredie a vodné prostredie. Tieto látky sú aktívne skúmané, rovnako aj ich možný dopad na voľne žijúce živočíchy a ľudské zdravie.

Bežným znečisťovateľom vody sú napríklad farmaceutické zlúčeniny [1]. Patria sem hormóny, antidepresíva, antibiotiká, protizápalové antiepileptiká atď. Využívajú sa v domácnostiach i nemocniciach a po podaní sa tieto liečivá vylúčia ako pôvodné alebo metabolizované. Mnoho štúdií poukázalo na prítomnosť liekov v komunálnych odpadových vodách, povrchových vodách i v niekoľkých podzemných a pitných vodách. V rokoch 2015 až 2016 bola realizovaná štúdia o monitorovaní širokej škály antibiotík v odpadovej vode z čistiarní [2]. Jednalo sa o prvú štúdiu na medzinárodnej úrovni a bolo do nej zapojených 7 európskych krajín. Z 53 sledovaných antibiotík bolo vo výtoku z ČOV aspoň raz detegovaných 17. Hoci čistiarne odpadových vôd využité v tejto štúdii boli v súlade s legislatívou EÚ, bolo zistené, že pozorované antibiotiká v koncentráciách v rozmedzí od 3 do 598 ng/l sa nepretržite vypúšťajú do sladkovodných a morských ekosystémov. Toto môže ovplyvniť životné prostredie a látky sa môžu dostať až do pitnej vody. Pravidelné pitie odpadovej vody by mohlo prispieť k vývoju rezistencie na antibiotiká. Aj keď koncentrácia antibiotík vo vode nemusí dnes znamenať nebezpečie, v budúcnosti sa situácia môže rapídne zhoršiť. Zapríčiniť to môže napríklad rast populácie a tým spôsobená vyššia spotreba antibiotík.

Ako konkrétny príklad znečistenia vody uvádzame laboratórnu analýzu odpadovej vody z čistiarne v Trnave (ČOV Zeleneč), ktorá poukázala na množstvo látok nachádzajúcich sa vo vode aj po jej dôkladnom vyčistení. Presný zoznam látok je prezentovaný v tabuľke Príloha A. Mnohé z vypísaných látok sú vo vyšších koncentráciách vysoko

návykové, hlavne drogy a lieky proti úzkosti a bolesti s tlmivým účinkom. Medzi takéto látky patria hlavne:

- lieky - alprazolam (anxiolytikum), oxazepam, haloperidol (antipsychotikum), tramadol (analgetikum), sertralín (antidepresívum)
- drogy a opiáty - MDMA (extáza), morfín, katinón, kokaín, kofeín (alkaloidy),
- metabolity - norketamín, kanabinol, norsertralín

Cieľom tejto práce je identifikovať zmenu črevnej mikroflóry (spoločenstva organizmov v črevách) u potkanov po užívaní odpadovej vody a u myší po užívaní antibiotík. Zmena bude pozorovaná predovšetkým na úrovni kmeňov a rodov baktérií v črevách myší a potkanov. Zameriame sa aj na konkrétne počty nájdených druhov. Ak nastane pokles, či vzrast hojnosti určitých druhov baktérií u laboratórnych zvierat, je možné, že podobný vplyv môže mať odpadová voda aj na tráviaci systém človeka.

Ďalším cieľom je priblíženie celého procesu charakteristiky črevnej mikroflóry. Získanie vzoriek mikroflóry, premietnutie ich do dát a následné spracovanie a analýza sú typické súčasti takéhoto procesu. Dôraz bude kladený na overenie použitých nástrojov, či sú vhodné na charakteristiku mikrobiálnych spoločenstiev.

V prvej kapitole si najskôr popíšeme základné pojmy potrebné na pochopenie našej práce. V ďalšej kapitole si ujasníme aké vzorky budeme používať, ich zdroj, spôsob ich spracovania, použité metódy a softvér. Ďalej v kapitole Výsledky budú odprezentované získané poznatky a pozorovania našej analýzy. Nakoniec sa v záverečnej kapitole budeme venovať diskusii k použitým nástrojom a získaným výsledkom.

Kapitola 1

Základné pojmy

1.1 Mikrobióm

Pojem mikrobióm môžeme definovať ako spoločenstvo všetkých mikróbov - baktérií, húb, vírusov, spolu s ich prirodzeným prostredím v ktorom žijú [4]. Mikróby sú zväčša viditeľné len pod mikroskopom, no aj napriek svojej malej veľkosti dokážu silne ovplyvniť zdravie jedinca. U mnohých organizmov, napríklad aj človeka, poznáme rôzne druhy mikrobiómu - orálny, kožný a črevný. Každý mikrobióm má svoje typické zastúpenie mikróbov.

Organizmy črevného mikrobiómu, nazývané aj črevná mikroflóra sa nachádzajú v tráviacom trakte živočíchov, vrátane človeka. V ľudských črevách žije približne 10^{13} baktérií, ktoré tvoria až 60% stolice [5]. Jedna z najznámejších črevných baktérií je *Escherichia coli*. Celkovo môžu črevá obsahovať vyše 500 druhov baktérií [6]. Spolu so svojim hostiteľom vytvárajú tieto mikroorganizmy obojstranne výhodný vzťah - *mutualizmus*. Medzi hlavné prínosy mikrobiómu v tele patrí:

- zamedzenie rastu patogénnych druhov mikroorganizmov pomocou produkcie látok, ktoré škodia týmto patogénom [7],
- dodávanie zdroja energie pre vývoj čriev,
- formovanie imunitného systému tým, že tvorí štruktúry a metabolické produkty, ktoré regulujú imunitné bunky [8],
- fermentovanie celulózy a nestrávených organických látok v črevách [9]

Na to, aby mikrobióm správne fungoval a plnil spomenuté úlohy musí byť v rovnováhe. Črevnú mikroflóru môže ovplyvniť a teda narušiť jeho rovnováhu mnoho faktorov ako strava a diéty, antibiotiká, probiotiká, vek, stres, atď. Tieto narušenia dokážu napríklad prínosné baktérie zmeniť na patogénne a tým ovplyvniť zdravie jedinca. Rôznymi

štúdiami boli dokázané ochorenia asociované s narušením črevného mikrobiómu. Medzi takéto ochorenia patrí obezita, diabetes mellitus typu II, cirhóza pečene, rakovina čreva [10]. Črevný mikrobióm môžeme považovať za významný článok, ktorý ovplyvňuje správny vývoj a zdravie živočíchov.

Kolekciu genómov a génov zložiek mikrobiómu nazývame metagenóm [4]. Metagenóm teda obsahuje celkovú genetickú informáciu mikrobiómu. Metagenomika je veda zaoberajúca sa štúdiom genómov mikrobiálnych komunit. Výsledkom tohto druhu štúdia sa najčastejšie zisťuje zloženie a potenciálna funkcia mikrobiómu.

1.2 Sekvenovanie

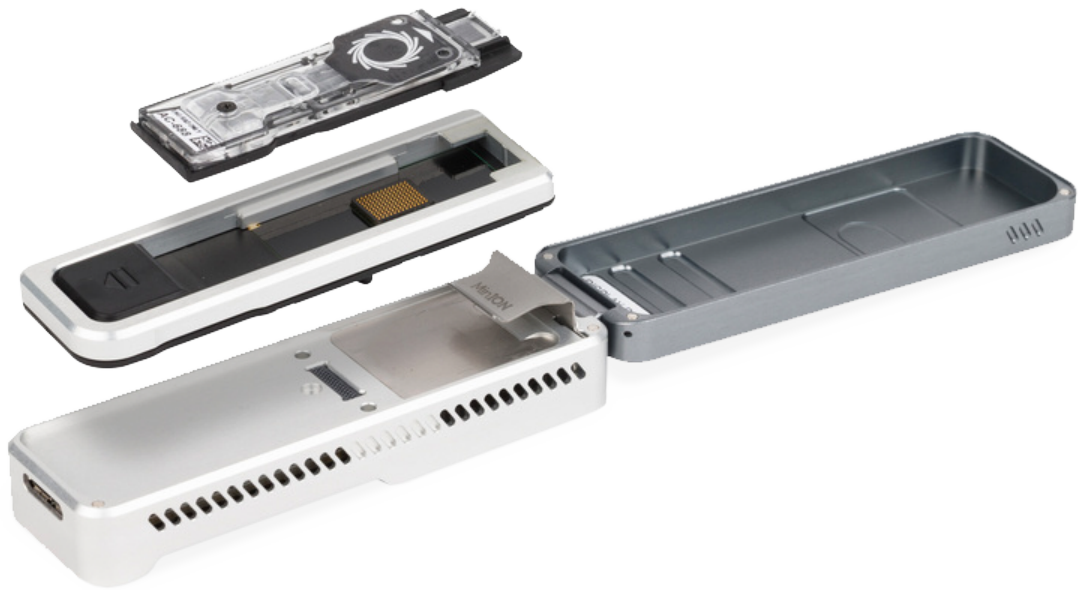
Proces, pri ktorom sa určuje primárna štruktúra DNA, teda poradie nukleotidov v konkrétnych reťazcoch DNA sa nazýva sekvenovanie. Sekvenovaním celého genómu získame poradie nukleotidov celého organizmu. Pre získanie metagenómu je potrebné osekvenovať celý mikrobióm v jednom kroku.

Proces sekvenovania môže prebiehať využitím rôznych techník. Medzi prvé techniky sekvenovania patrila Sangerova metóda a Maxam-Gilbertova metóda. Tieto metódy boli založené na enzymatických a chemických vlastnostiach DNA. Výsledky bolo nutné ručne analyzovať. Neskôr pribudli ďalšie modernejšie metódy, spadajúce pod druhú a neskôr tretiu generáciu sekvenovania. Automatické analyzátory DNA a počítačom riadený zber dát umožnili simultánne sekvenovať niekoľko miliónov fragmentov v jednom behu.

1.2.1 Nanopórové sekvenovanie

Nanopórové sekvenovanie je technológia spadajúca pod tretiu generáciu sekvenovania. Umožňuje priamu analýzu dlhých fragmentov DNA/RNA v reálnom čase, čo poskytuje okamžitý prístup k výsledkom. Taktiež je možné zastavenie sekvenovania po dosiahnutí výsledku. Práve produkcia dlhých fragmentov umožňuje identifikáciu dlhých repetitívnych sekvencií, dokáže zlepšiť metagenomickú analýzu blízko príbuzných druhov. Pomocou nanopórového sekvenovania dokážeme sekvenovať celé mikróby v jednotlivých čítaniach, čo značne uľahčuje ich následnú identifikáciu[11]. Prvým nanopórovým sekvenátorom je MinION od firmy Oxford Nanopore Technologies (obrázok 1.1). Medzi novšie ONT sekvenátory patrí GridION či PromethION.

Nanopórové sekvenovanie funguje tak, že monitoruje zmeny elektrického prúdu, keď nukleové kyseliny v podobe jednovláknových reťazcov prechádzajú cez proteínový nanopór [12]. Nanopóry sú vložené do elektricky odolnej membrány (obrázok 1.2). Stovky až tisíce takýchto nanopórov sa nachádzajú na tzv. flowcelle (prietoková bunka). Záporne nabitú nukleotidovú reťazec sú preháňané cez nanopóry zo záporne nabitej strany

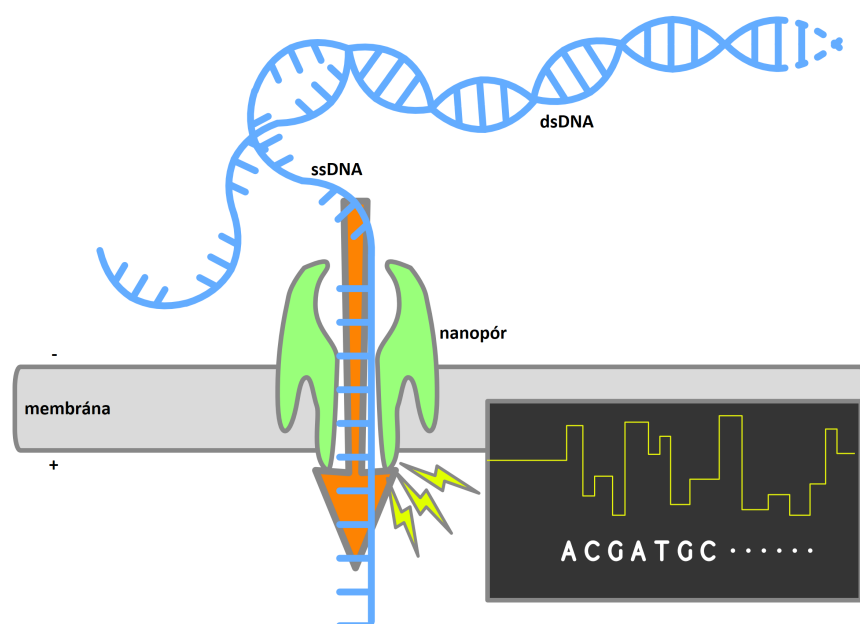


Obr. 1.1: MinION sekvenátor of firmy ONT. Pomocou pipety sa vnesie roztok obsahujúci DNA/RNA. Výsledné dáta sú z prístroja priamo posielané do zariadenia. Zdroj: <https://nanoporetech.com/products>

membrány do kladne nabitej strany. Reľazec je poháňaný cez póry pomocou motor proteínu (molekulárneho motora), ktorý slúži aj na rozpletenie dvojvláknovej DNA na jednovláknovú pomocou svojej helikázovej aktivity. Každá z nukleotidových báz (A, C, T/U, G) po prechode nanopórom je identifikovaná svojim charakteristickým narušením elektrického prúdu. Výsledný signál sa dekóduje pomocou algoritmov, aby sa získala špecifická sekvencia DNA alebo RNA. Tento proces nazývame aj basecalling (volanie báz). Basecalling je možné spustiť priamo popri sekvenovaní, ale aj dodatočne cez rôzne softvéry. Oddelenie sekvenovania od basecallingu je často výhodou. Basecalling vyžaduje veľa výpočtových prostriedkov, robí veľké množstvo zápisov a čítaní, čo môže viesť k strate sekvenovaných dát. Medzi známe basecallery od firmy ONT patrí napríklad Guppy, Bonito či Dorado.

1.3 Taxonómia

Biologická klasifikácia alebo aj taxonómia popisuje abstraktné úrovne skupín organizmov - taxónov. Základný hierarchický systém taxonómie pozostáva z úrovní: doména, ríša, kmeň, trieda, rád, čeľad, rod a druh. Tri hlavné domény života sú: *Bacteria* (bakté-



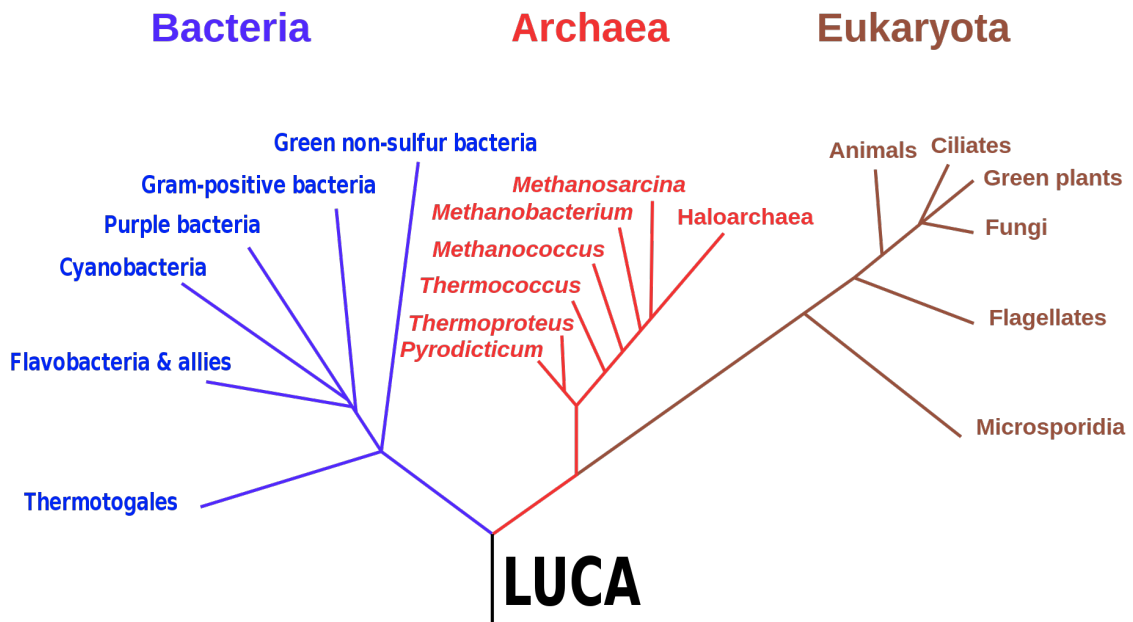
Obr. 1.2: Princíp nanopórového sekvenovania. Zdroj: <https://doi.org/10.7875/togopic.2020.01>

rie), *Archaea* (archeóny) a *Eukarya* (eukaryoty). Ich vývoj a členenie popisuje obrázok 1.3.

1.3.1 Mikroorganizmy črevného mikrobiómu

Z 55 kmeňov, ktoré tvoria doménu *Bacteria*, sa sedem až deväť nachádza v ľudskom čreve [13]. Najväčšie zastúpenie majú kmene *Bacteroidetes* a *Firmicutes*. Iné kmene identifikované v ľudskom čreve zahŕňajú *Proteobacteria*, *Actinobacteria*, *Fusobacteria* a *Verrucomicrobia*. Bakteriálne druhy nachádzajúce sa v lumene sa líšia vo svojom zložení od slepého čreva po konečník. V slepom čreve spôsobuje pomalá doba prechodu a nedostatok jednoduchých cukrov prevažovanie fermentačných anaeróbov degradujúcich polysacharidy. Nachádzajú sa tu najmä rody *Prevotella*, *Roseburia*, *Faecalibacterium*, *Lachnospira* a *Eubacterium*. V úsekoch hrubého čreva sú bežné mukolytické (degradujúce mucín) a proteolytické druhy (produkujúce proteázy), napríklad rodu *Bacteroides*, *Ruminococcus*, *Akkermansia*, *Bifidobacterium*, *Methanobrevibacter*, *Desulfovibrio*, *Proteus* a *Escherichia*.

Zatiaľ bolo zistených veľmi málo druhov *Archaea*. Doména *Archaea* pozostáva z jednobunkových mikroorganizmov, ktorých štruktúra je podobná baktériám [14]. Archeóny boli nájdené v rôznych extrémnych prostrediach - veľmi kyslé, slané, horúce či studené prostredia. Metanogénne archeóny dokážu produkovať metán a boli nájdené aj v gastrointestinálnom trakte cicavcov (väčšinou *Methanobrevibacter smithii*).



Obr. 1.3: Domény života. Všetky tri domény majú nejakého univerzálneho spoločného predka - LUCA (last universal common ancestor). Ďalej sa delia na rôzne ríše, kmene, triedy atď. Zdroj: https://en.wikipedia.org/wiki/Three-domain_system

1.3.2 Taxonomická klasifikácia metagenomických dát

Jednou z najbežnejších metód používaných na taxonomickú identifikáciu druhovej diverzity prokaryotov (*Bacteria* a *Archaea*) je sekvenovanie génu, ktorý kóduje malú podjednotku ribozomálnej RNA (16S rRNA) [13]. Ďalšou metódou výzkumu nezávislom na kultúre je sekvenovanie celého genómu, ktoré poskytuje kolekciu všetkých génov prítomných vo vzorke. Sekvenovanie celého genómu tiež umožňuje analýzu funkčných a metabolických sietí, a taktiež detekciu génov, ktoré patria nebakteriálnym členom mikrobiálnej komunity, ako sú vírusy, kvasinky a protista.

Po sekvenovaní je potrebné jednotlivým čítaniam určiť ich taxonomické zaradenie. To znamená, že je potrebné porovnať sekvencie čítaní s referenčnou databázou. V posledných rokoch bolo možné vidieť niekoľko prístupov, ako túto úlohu splniť časovo efektívnym spôsobom. Jedným z takýchto nástrojov je Kraken 2 [15]. Jeho algoritmus funguje na základe vyberania k-merov a hešovania. Kraken 2 je nástupca nástroja Kraken, pričom algoritmus Kraken 2 zredukoval využitie pamäte o 85% voči jeho predchodcovi.

Na začiatku sú z každého čítania referenčnej knižnice vybrané k-mery. Predovolená dĺžka $k = 36$. Každý k-mer je klasifikovaný pomocou nájdenia jeho minimalizátora (minimizer), ktorý je prednastavený na dĺžku $l = 31$ bp. Hodnoty k a l je možné meniť. Minimalizátor l bp k-meru ($l \leq k$) definujeme ako lexikograficky najmenší kanonický l -mer nachádzajúci sa v k -meri. Minimalizátor - l -mer sa nazýva kanonický, ak je lexi-

kograficky menší alebo rovný jeho reverz komplementu.

Následne sa využije maska (s), ktorá aplikuje medzery na minimalizátor. Hodnota masky s je číslo, ktoré určuje koľko z pozícií báz minimalizátora bude maskovaných. Maskovanie bázy znamená, že daná báza nebude použitá v ďalších krokoch. Pomocou hešovacej funkcie sa minimalizátory zahešujú do kompaktnej hešovacej tabuľky (CHT, compact hash table) v podobe jednotiek a núl. Takto upravenému k-meru je priradený najnižší taxón posledného spoločného predka (LCA, last common ancestor) v podobe jeho kompaktného heš kódu. V CHT sú teda uložené dvojice kľúč-hodnota pričom kľúč predstavuje zahešovaný minimalizátor (15 bitov) a hodnota predstavuje konkrétny LCA taxón (17 bitov). Jeden záznam v CHT predstavuje pole fixnej dĺžky 32 bitov. Zjednodušený popis postupu je možné vidieť na obrázku 1.4.

Po vytvorení CHT je v nej možné začať vyhľadávať. Pre vstupné sekvencie báz sa vytvoria ich k-mery, následne minimalizátory, ktoré sa zahešujú a ich kompaktný hešovací kód je použitý ako vyhľadávací dotaz v CHT. CHT sa prehľadáva lineárne. Pri klasifikácii sa zaznamenáva, koľko k-mer hitov bolo namapovaných na každý taxón, čím sa vytvorí orezaný klasifikačný strom. Algoritmus priradí každému čítaniu najšpecifickejšie označenie, aké môže [16]. Napríklad, ak je čítanie určitej dĺžky 100% identické s dvoma rôznymi druhmi, Kraken 2 ho priradí k ich najnižšiemu spoločnému predkovi (LCA), ktorý môže byť na úrovni rodu alebo vyššej. Pre vzorku obsahujúcu viac veľmi podobných druhov to znamená, že počet druhovo špecifických čítaní môže byť oveľa nižší, ako sa očakávalo. Na vyriešenie tohto problému sa využíva nástroj Bracken (Bayesian Re-estimation of Abundance after Classification with Kraken) [17]. Bracken používa Bayesovský algoritmus a premieňa výstup Krakena na odhady početnosti na úrovni druhov.

1.4 Diverzita mikrobiálnych spoločenstiev

Pod pojmom diverzita budeme chápať bohatosť a početnosť organizmov na úrovni druhu (biodiverzita). Početnosť je definovaná ako miera počtu alebo frekvencie jedincov toho istého druhu (kvalita)[18]. Bohatosť alebo aj rôznorodosť určuje počet rôznych organizmov (kvantita). V rámci tejto práce budeme používať aj mieru nazývanú hojnosť (angl. abundance). Hojnosť predstavuje počet - zastúpenie jedného taxónu vo vzorke.

Biodiverzitu je možné skúmať na viacerých úrovniach. Definujeme ich ako alfa, beta a gama diverzitu [19].

Alfa diverzita sa vzťahuje na diverzitu v lokálnom meradle. Opisuje bohatosť druhov v rámci funkčného spoločenstva, napríklad mikrobiómu, či iných ekologických jednotiek - rybník, úsek pozemku, a iné. Beta diverzita popisuje rozdielosť zastúpenia druhov medzi komunitami, napríklad v rôznych menších lokalitách. Gamma diverzita opisuje

celkovú druhovú diverzitu v rámci spoločenstiev vo väčšej geografickej oblasti. Často sa vyjadruje bohatosť druhov medzi krajinami a bioregiónmi.

V prípade alfa diverzity vieme vystihnúť mnohé aspekty spoločenstiev jedinou hodnotou - indexami diverzity [20]. Tieto indexy sú založené buď na empirickom prístupe, alebo matematických teóriach, napríklad Shannonovej informačnej teórii.

1.4.1 Index počtu taxónov

Najjednoduchším indexom diverzity je určenie počtu taxónov, nachádzajúcich sa v spoločenstve. Ide o hodnotu spadajúcu do kategórie indexov založených na početnosti druhov. Tento index v sebe nesie informáciu o celkovom počte nájdených druhov:

$$S = \text{počet taxónov}$$

1.4.2 Shannon-Weaverov index diverzity

Shannon-Weaverov index diverzity je založený na pomere početnosti druhov a vychádza z informačnej teórie. Jeho hodnota vyjadruje, koľko rovnako početných druhov by vytvorilo index s rovnakou hodnotou. Zvyčajne nadobúda hodnoty 1,5 až 4,5. Teoreticky môžeme dostať aj hodnotu indexu = 0 pri absolútnej dominancii jedného druhu [21]. Základný vzťah na výpočet Shannon-Weaverovho indexu je:

$$H = - \sum_{i=1}^S p_i \ln p_i$$

$$p_i = \frac{n_i}{N}$$

kde S je celkový počet taxónov, p_i je počet jedincov i -teho druhu a N celkový počet jedincov.

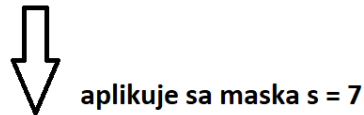
1.4.3 Simpsonov index diverzity

Simpsonov index diverzity vyjadruje ako bohatosť populácie, tak aj rovnomernosť. Jeho hodnota sa nachádza v rozpätí 0 až 1. Čím vyššiu má index hodnotu, tým vyššia je diverzita spoločenstva. Je silne závislý na najpočetnejších druhoch a menej citlivý ku vzácnym druhom. S jeho zvyšujúcou sa hodnotou stúpa dominancia a klesá vyrovnanosť spoločenstva. Hodnotu Simpsonovho indexu vypočítame pomocou vzťahu:

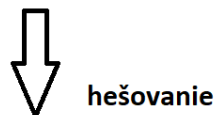
$$D = - \sum_{i=1}^S \frac{n_i(n_i - 1)}{N(N - 1)}$$

kde S je celkový počet taxónov, p_i je počet jedincov i -teho druhu a N celkový počet jedincov.

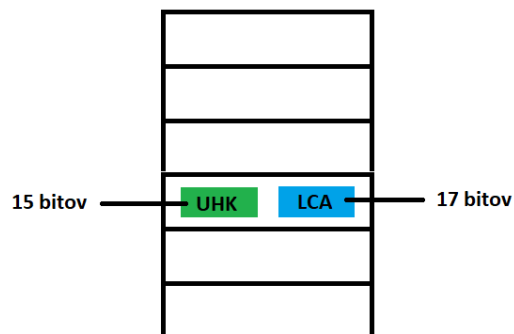
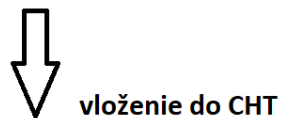
GCATTGTAGCATCGACGTGCCACTGCCGAGTCAGA - *k-mer* (35)
 - *l-mer minimizer* (31)



TTGTAGCATCGACGTGC-A-T-C-G-G-C-G



univerzálny heš kód



Obr. 1.4: Ukážka postupu algoritmu Kraken 2. *k-mer* má dĺžku 35 báz, *l-mer* 31 bp. *l-mer* sa zahešuje do univerzálneho heš kódu pomocou hešovacej funkcie. Do CHT sa ukladá univerzálny heš kód minimalizátora spolu LCA ako dvojica kľúč-hodnota.

Kapitola 2

Metodika práce

2.1 Zdroj vzoriek

Najskôr sa ako vhodné modelové organizmy zvolili myši a potkany. Tieto laboratórne zvieratá boli chované vo zverinci Lekárskej fakulty UK, vedenom Ústavom molekulárnej biomedicíny. Experimenty boli schválené etickou a veterinárnou komisiou. Zvieratá mali ad libitum prístup k štandardnej strave a vode. S výnimkou potkanov v skupine, kde bola namiesto obvyčajnej kohútikovej vody podávaná odpadová voda z čističky odpadových vôd v Trnave (ČOV Zeleneč) po dobu 8 mesiacov. Analýza vody z ČOV Zeleneč je uvedená v úvode práce a v Prílohe A. Myšiam bol podávaný koktejl antibiotík po dobu 2 týždňov. Myšiam bola stolica odobratá pred a po podaní antibiotík. Potkanom z kontrolnej aj z experimentálnej skupiny bola stolica odobratá na konci experimentu. Manipulácia so zvieratami prebiehala výlučne vo zverinci pod dohľadom pracovníkov zverinca. Vzorky na analýzu boli po danom čase zmrazené a prebrané na ďalšie spracovanie a analýzu v laboratóriách Slovenskej akadémie vied (SAV).

2.2 Spracovanie vzoriek

Jedným z kľúčových bodov spracovania vzoriek bola izolácia čistej DNA z výkalov. Na tento proces sa využili tzv. kity. V našom prípade konkrétne QIAamp PowerFecal Pro DNA Kit (obrázok 2.1) od firmy QIAGEN (Holandsko). Tento kit je presne určený na izolovanie DNA zo stolice a vo výsledku poskytuje vysokú bakteriálnu diverzitu[22]. Vzorky spracovávané pomocou súpravy kitu sa pridali do skúmavky so špeciálnymi guľôčkami (bead beating tube) a následne prebehla homogenizácia pomocou chemických a mechanických metód. Akonáhle boli bunky lyzované, odstránili sa inhibítory (polysacharidy, zlúčeniny hému) a výsledná genómová DNA bola zachytená pomocou silikónovej membrány v spin-column skúmavke. V poslednom kroku sa premyla a elovala. Takto pripraveným vzorkám bola zistená koncentrácia DNA laboratórnymi pra-

covníkmi. Na základe hodnoty koncentrácie sa vybrali najvhodnejšie vzorky (tabuľka 2.1).



Obr. 2.1: Ukážka PowerFecal kitu. Zdroj: <https://www.qiagen.com/us>

V ďalšom kroku nasledovala príprava vzoriek na sekvenovanie pomocou Rapid Sequencing Kit of firmy Oxford Nanopore Technologies (ONT). Tento kit slúži na prípravu knižnice a DNA barcoding. Pomocou krátkych úsekov DNA zo špecifických génov sme si takto identifikovali druhy vzoriek, s ktorými narábame. Celkovo sa takto označilo 6 vzoriek potkanej DNA - 3 kontrolné a 3 ovplyvnené odpadovou vodou, 5 vzoriek myšacej DNA - 3 kontrolné pred antibiotikami a 2 po užívaní antibiotík a jedna vzorka *Escherichia coli* ako pozitívna kontrola. Presné pridelenie barkódov ku vzorkám je uvedené v tabuľke 2.1.

V poslednom kroku sa výsledný obsah skúmavky so všetkými druhmi osekvenoval. Ako sekvenátor sa použil MinION Mk1B od firmy ONT s flowcellou FLO-MIN106. Report zo sekvenovania uvádza, že bolo vygenerovaných 1,5 milióna čítaní, ktoré spolu obsahujú približne 4 Gb.

Tabuľka 2.1: Tabuľka vzoriek s pridelenými barkódmi. Vysvetlivky: WW - waste water vzorka; CTRL - kontrolná vzorka; ATB - vzorka po podaní antibiotík; za číslom vzorky je pri potkaních vzorkách uvedené, či ide o samca (M), alebo samicu (F)

Barcode	Druh vzorky	Číslo vzorky	Organizmus
1	CTRL	7M	potkan
2	CTRL	3M	potkan
3	CTRL	3F	potkan
4	WW	3M	potkan
5	WW	1F	potkan
6	WW	4M	potkan
7	CTRL	1	myš
8	CTRL	2	myš
9	CTRL	4	myš
10	ATB	6	myš
11	ATB	5	myš
12	-	-	E.coli

2.3 Spracovanie a analýza dát

2.3.1 Basecalling a demultiplexing

Sekvenovaním bol získaný výstup - tzv. raw dáta. Tieto dáta boli uložené vo FAST5 súboroch v podobe signálov, ktoré zachytili nanopóry pri čítaní báz. V ďalšom kroku sa spracovali dané signály a identifikovali sekvencie prečítaných báz pomocou dvoch basecallerov - Dorado a Guppy.

Guppy je ONT basecaller založený na báze neurónových sietí. Existuje v dvoch verziách - GPU a CPU, ktoré sa líšia v tom, či program využíva centrálnu procesorovú jednotu alebo grafickú. GPU verzia je značne rýchlejšia, no vyžaduje aj výkonný hardvér. Pre spustenie je potrebné zadať cestu k priečinku obsahujúci vstupné súbory (input path), cestu k výstupnému priečinku (save path) a buď názov konfiguračného súboru (config file), alebo názov flowcelly (flowcell name) použitej pri sekvenovaní spolu s názvom kitu (kit name) použitom pri príprave vzoriek na sekvenovanie. Typické príkazy na spustenie môžu vyzeráť nasledovne:

```
#pomocou konfiguračného súboru
$ guppy_basecaller -i <input path> -s <save path> -c <config file>

#pomocou použitej flowcelly a kitu
```



```
$ guppy_basecaller -i <input path> -s <save path>
--flowcell <flowcell name>
--kit <kit name>
```

V našom prípade bol použitý príkaz s názvom flowcellly FLO-MIN106 a kitu SQK-RBK004. Pri takomto spustení si program dokáže vybrať správny konfiguračný súbor, inak nazývaný aj model pre basecalling. Ak by sme chceli využiť prvú možnosť spustenia softvéru, zoznam kitov a flowcell s ich kompatibilnými modelmi je možný nájsť v príručke softvéru.

Okrem basecallingu dokáže Guppy aj filtrovať čítania podľa skóre priradenému bázam. Výstupom sú potom v tomto prípade 2 priečinky - `fail` a `pass`, kde v priečinku `pass` dostaneme čítania vyhovujúce zadanej minimálnej hranici skóre a vo `fail` priečinku ostatné. Automaticky je skóre nastavené na 9 a je možné ho zrušiť pomocou parametra `--disable_qscore_filtering`, alebo zmeniť na vlastnú hodnotu pomocou parametra `--min_qscore` za ktorým by nasledovalo číslo nového minimálneho skóre. Pri GPU verzii je možné určiť napríklad parametrom `-x cuda:0`, kde si program vyberie iba GPU zariadenie číslo 0. V našom prípade bola spustená GPU verzia a bolo zrušené filtrovanie na základe skóre báz. Výsledné čítania sú uložené vo FASTQ súboroch. Ukážka FASTQ súboru:

```
@005991ec-48db-4c1c-94ba-31e7c7186af4
TGCCGTTCCATTTTCAGATACAAGGTGTTTA
+
%\$\&\$\$\$\$\%'\&\$\$\$\%'\(\%\$\#\#\$\$\$\%\%\%\%*\')+ .27
@0011ed2f-c1f4-4c63-8814-696b73ce7540
AAGTATGCTTCGTTTCATTCAATTTGGGTGTTTAACCAAGC
+
,) (+ . . ***+95499, \&'%\&\%\&\&\&()* ,23611. (\&\&())2
```

Jedno čítanie sa skladá zo štyroch riadkov. Prvý riadok obsahuje za znakom '@' označenie čítania. V druhom riadku sa nachádza samotná sekvencia nukleotidov čítania za ktorou nasleduje riadok so znakom '+'. V poslednom riadku je pre každú bázu z čítania zapísaná jej odhadovaná kvalita pomocou ASCII znakov. ASCII kód každého znaku predstavuje kvalitu jednej bázy v DNA. Vyššia kvalita znamená nižšiu pravdepodobnosť chyby pri sekvenovaní. Hodnota kvality Q (ASCII kód) sa počíta pomocou vzorca

$$Q = 33 + (-10 \log_{10} p)$$

kde p je pravdepodobnosť, že báza je určená nesprávne.

Na rovnakých vstupných dátach bol spustený basecall pomocou softvéru Dorado. Dorado je vysokovýkonný, open source basecaller určený hlavne pre Oxford Nanopore čítania[24]. Najvyšší výkon má predovšetkým pri POD5 vstupných súboroch, no funguje aj na FAST5 súbory. Tento softvér je stále v alfa verzii. Na jeho spustenie je potrebný špecifický hardvér - vyžaduje grafické karty typu Nvidia s aspoň 8GB pamäťou. Typický príklad spustenia basecallera Dorada má syntax:

```
$ dorado basecaller <model name> fast5s/ > output.sam
```

Pre ďalšiu prácu bol potrebný FASTQ výstup, ktorý bol dosiahnutý pridaním parametra `--emit-fastq`. Ako model sa použil `dna_r9.4.1_e8_hac@v3.3`. Tento HAC model spadá medzi high accuracy models (modely s vysokou presnosťou) a rovnaký model bol použitý pri basecallingu pomocou Guppy. Okrem HAC sú dostupné aj SUP (super accuracy) a FAST modely. S presnosťou modelu rastie aj čas behu basecallera.

Aby bolo možné rozlíšiť jednotlivé druhy vzoriek z čítaní, bolo potrebné spustiť tzv. barcoding alebo aj demultiplexing. Je to proces, pri ktorom sa nám jednotlivé čítania rozdelia do priečinkov na základe barkódov, ktorými sme si ich v predošlých krokoch označili. Výsledné dáta teda budú v dvanástich priečinkoch, kde každý priečinok reprezentuje jeden organizmus. Na barkódovanie bol použitý softvér Guppy na oba typy výstupných dát z basecallerov. Dorado vo svojej terajšej verzii ešte funkciou barcoding nedisponuje. Príkaz na barcoding má syntax:

```
$ guppy_barcode -i <input path> -s <save path>
  --barcode_kits <kit name>
```

2.3.2 Porovnanie basecallerov

Oba basecallery bežali na rovnakom hardvéri a s rovnakými parametrami. Taktiež oba využívali GPU ako procesorovú jednotku. V tejto podkapitole je cieľom porovnať rozšírený a bežne používaný basecaller Guppy s novým basecallerom Dorado, ktorý je zatiaľ v alfa verzii. Na porovnanie boli využité skripty v jazyku Python a softvéry Nanoplot [25] a Nanostat[26]. Skripty obsahujú mini kód na porovnanie odhadu kvality báz vyprodukovaných basecallerom. V prípade Guppy sa dá využiť súbor `sequencing_summary.txt`. Je to automaticky vyprodukovaný súbor softvérom pri volaní báz a obsahuje okrem iného aj priemer kvality báz pre každé čítanie. Dorado zatiaľ nedisponuje takýmto výstupným súborom. Preto bolo potrebné pre každé čítanie z výstupného FASTQ súboru vypočítať priemer kvality báz. Na prácu s FASTQ súbormi bola použitá knižnica `pysam.FastxFile` [27]. Použitím tejto knižnice je možné čítať FASTQ súbor po jednotlivých čítaniach. Obsahuje funkcie:

- **name** - vracia názov čítania,

- **quality** - vracia reťazec kvality čítania,
- **sequence** - vracia sekvenciu nukleotidov čítania

Pomocou vzorca na výpočet kvality báz, ktorý sme si predstavili v podkapitole Basecalling a demultiplexing, je možné spočítať priemernú kvalitu pre každé čítanie.

Barkód 12 obsahuje vzorku *Escherichia coli*. Práve túto vzorku využijeme ako pozitívnu kontrolu na zarovnanie s referenčným genómom. Cieľom je zistiť, ktorý z basecallerov vyprodukoval lepšie sekvencie - sekvencie s lepším zarovnaním. Na zarovnanie čítaní s referenčným genómom *E.coli* bol použitý program Minimap2 [28]. Referenčný genóm bol stiahnutý z NCBI databázy [29]. Výsledné zarovnanie boli uložené v SAM súboroch, upravené pomocou balíčka Samtools [30] na BAM súbory. Neskôr boli tieto súbory analyzované softvérmi NanoPlot a NanoStat. Príkazy na spustenie programov majú syntax:

minimap2:

```
#zarovnanie čítaní Dorado/Guppy s referenčným genómom
#uložené v SAM formáte
$ minimap2 -a -x map-ont --secondary=no -t 1
    <referencny_genom.fasta> <citania.fasta> > <output.sam>
```

samtools:

```
#uloženie zarovnaní do BAM súboru
#sort a indexovanie BAM súboru pre lepšiu manipuláciu
$ samtools view -S -b <output.sam> -o - |
    samtools sort - -o <output.bam>
$ samtools index <output.bam>
```

NanoPlot:

```
$NanoPlot --bam <output.bam>
```

NanoStat:

```
$NanoStat --bam <output.bam>
```

2.3.3 Analýza a vizualizácia metagenomických dát

Sekvencie báz boli analyzované využitím workflow-ov ONT, ich vyvinutými nástrojmi a už dostupnými algoritmi. Ďalšie výstupné súbory boli vizualizované skripami v jazyku Python.

V prvom kroku analýzy bol použitý workflow wf-metagenomics [31] od firmy ONT. Tento workflow slúži na identifikáciu pôvodu jednotlivých čítaní metagenomických dát. Pre bežanie workflowu je potrebné mať nainštalovaný Nextflow [32] na správu výpočtových a softvérových zdrojov a platformu Docker [33] na izoláciu softvéru od infraštruktúry počítača. Wf-metagenomics môže bežať v dvoch módoch - kraken2 a minimap2. Tieto dva módy sa líšia v postupe spracovania dát, no obe majú základ vo využití k-merov a minimalizátorov. Algoritmus módu kraken2 je popísaný v predchádzajúcej kapitole. Práve tento mód bol využitý pri spustení workflowu. Je prednastavený ako predvolený režim a je rýchlejší v spracovaní dát ako minimap2. Minimap2 je považovaný za presnejší mód, no má vyššiu pravdepodobnosť produkcie falošne pozitívnych výsledkov. Typický príkaz na spustenie workflowu má syntax:

```
$ nextflow run epi2me-labs/wf-metagenomics
  --fastq <input path>
  --out_dir <save path>
```

Pri zmene módu na minimap2 je nutné pridať parameter `--classifier minimap2`. Pomocou reportu a súborov z behu wf-metagenomics si dokáže používateľ vizualizovať klasifikáciu počtov a početnosti druhov v špecifickej mikrobiálnej komunite. Pre ešte rýchlejšie zobrazenie diverzity, je možné pridať parameter `--bracken_level` a požadovanú taxonomickú úroveň: 'D' = Domain (doména), 'P' = Phylum (kmeň), 'C' = Class (trieda), 'O' = Order (rad), 'F' = Family (čeľaď), 'G' = Genus (rod), 'S' = Species (druh). V takomto prípade bude výsledok diverzity zameraný len na zvolenú taxonomickú úroveň. Štandardne je prednastavená hodnota 'S' (úroveň druhu), ktorá bola použitá v tejto práci.

Workflow bol spustený pre všetky skupiny organizmov naraz. Okrem reportu, ktorý popisuje základné charakteristiky obsahu mikrobióty, je výstupom workflowu aj textový súbor s názvami druhov aj ich frekvencie v jednotlivých barkódoch. Pomocou týchto súborov a Python knižnice `scikit-bio` boli vizualizované alfa diverzity jednotlivých druhov organizmov. Funkcie balíčka využité na určenie alfa diverzity:

- **shannon** - vracia Shannonov index diverzity populácie,
- **simpson** - vracia Simpsonov index diverzity populácie

Na ďalšiu analýzu boli použité výstupné súbory algoritmu kraken2, ktoré obsahovali všetky nájdené taxonomické úrovne spolu s počtami ich výskytu. Pre každý barkód

bol vyprodukovaný jeden takýto výstup. Následne sa výstupy spracovali a vizualizovali skriptami v jazyku Python. Zo súborov boli samostatne vzaté taxóny na úrovni kmeňa a rodu. Na vizualizáciu sa využili knižnice `seaborn`, `matplotlib`, `pandas` a `numpy`.

Kapitola 3

Výsledky

3.1 Porovnanie behu a výsledkov basecallerov

Po spustení oboch basecallerov na rovnakých dátach a s rovnakými parametrami sme dostali výsledný rovnaký počet čítaní. Počas behu teda nenastali žiadne náhodné straty dát. Celkový čas behu Guppy bol približne 22 hodín a Dorado 30 hodín. Výsledné čítania sa mierne líšili v svojej dĺžke a kvalite (viď ukážka súborov v tabuľkách 3.1 a 3.2).

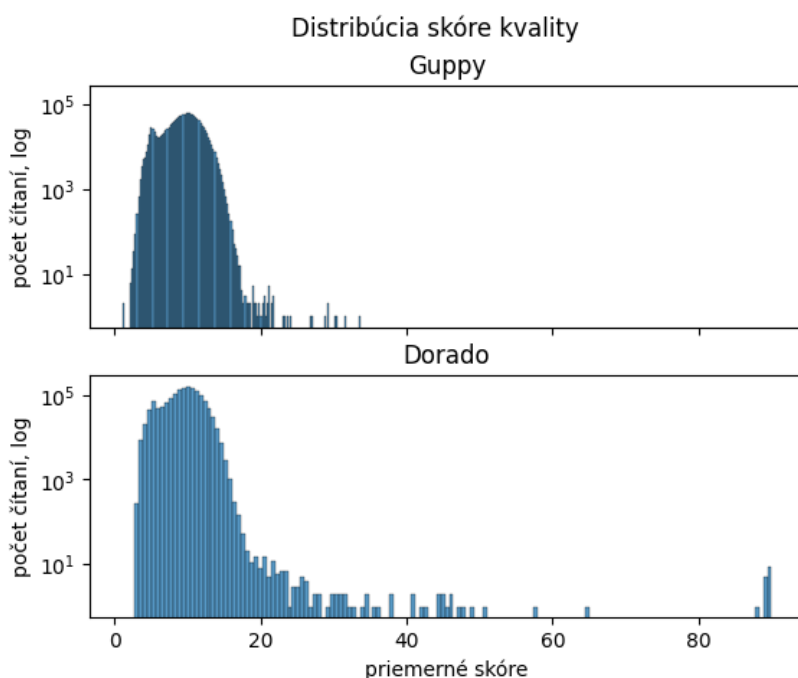
Tabuľka 3.1: Ukážka dĺžky a priemernej kvality čítaní - Guppy

read_id	sequence_length	mean_qscore
0000028d-6a93-4a4c-b8b6-3ae5d3e30844	3144	9.263274
000019c1-d2a5-4139-8687-aa3aeb57d23c	4597	7.247869
00001a88-b6a2-4b28-96f2-38a02c5362d8	4251	7.470748
00002722-9e73-48da-9edd-b8bb6744fd68	2355	11.214777
00004cba-4ed2-430d-ad88-5ca6f3da8809	1126	9.515033

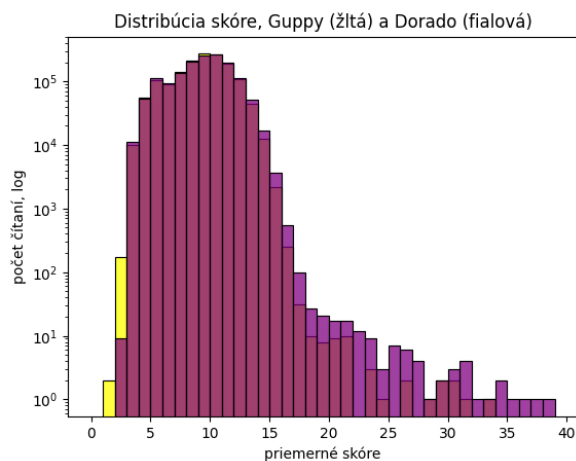
Tabuľka 3.2: Ukážka dĺžky a priemernej kvality čítaní - Dorado

read_id	sequence_length	mean_qscore
0000028d-6a93-4a4c-b8b6-3ae5d3e30844	3168	9.531521
000019c1-d2a5-4139-8687-aa3aeb57d23c	4559	7.084873
00001a88-b6a2-4b28-96f2-38a02c5362d8	4358	7.346786
00002722-9e73-48da-9edd-b8bb6744fd68	2329	11.326157
00004cba-4ed2-430d-ad88-5ca6f3da8809	1096	9.924255

Celková distribúcia skóre bola medzi Guppy a Dorado veľmi podobná (obrázok 3.1), za hodnotou skóre približne 40 má Dorado navyše čítania. Bližší pohľad na hodnoty skóre do 40 poskytuje obrázok 3.2.



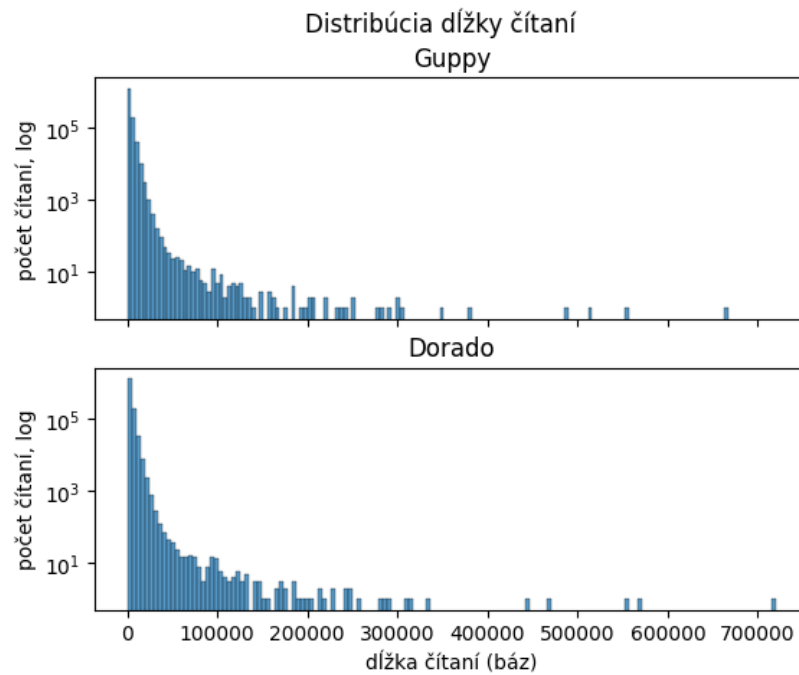
Obr. 3.1: Celková distribúcia priemernej kvality čítaní oboch basecallerov



Obr. 3.2: Distribúcia kvality skóre basecallerov do hodnoty 40. Guppy čítania sú označené žltou farbou a Dorado fialovou. Ružová prevládajúca farba reprezentuje prekrývajúce sa hodnoty oboch basecallerov.

Vo všeobecnosti sa priemerné skóre a medián skóre oboch basecallerov líši o malú hodnotu. Priemerné skóre Guppy čítaní je približne 9.3, pre Dorado 9.4. Medián skóre má Guppy 9.5 a Dorado 9.6.

Dĺžky čítaní boli v oboch prípadoch veľmi podobné (obrázok 3.3)



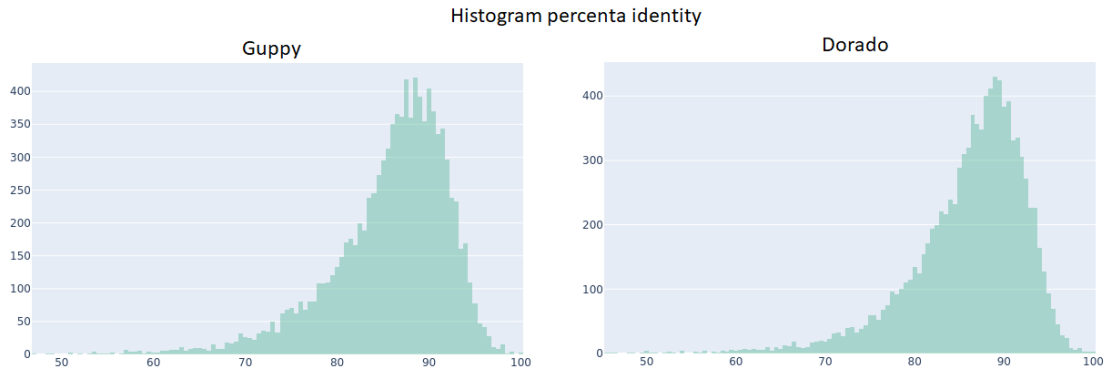
Obr. 3.3: Distribúcia dĺžky čítaní basecallerov.

3.1.1 Zarovnanie čítaní k referenčnému genómu

Po zarovnaní čítaní barkódu 12 jednotlivých basecallerov k referenčnému genómu *Escherichia coli* bola vykonaná analýza zarovnaní. Oba typy čítaní pokryli celý genóm *E.coli*. Výsledky štatistiky sú zobrazené v tabuľke 3.3. Bližší pohľad na početnosť zastúpení percent identity poskytuje obrázok 3.4.

Tabuľka 3.3: Štatistika zarovnaní Guppy a Dorado čítaní. Hodnoty boli vypočítané softvérom NanoStat.

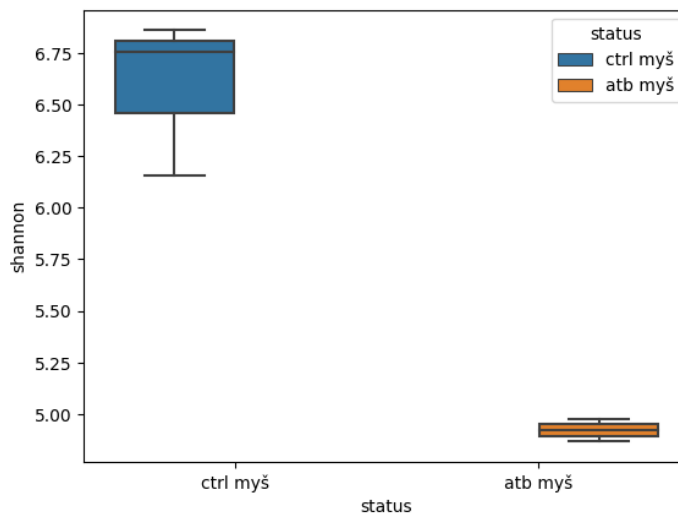
	Dorado	Guppy
Počet čítaní	134,017.0	140,507.0
Priemerná dĺžka čítaní	3,530.5	3,504.4
Medián dĺžky čítaní	2,162.0	2,145.0
Počet báz	473,151,818.0	492,396,475.0
Počet zarovnaných báz	380,062,545.0	394,745,981.0
Podiel zarovnaných báz	0.8	0.8
Priemerné percento identity	85.9	86.0
Medián percenta identity	87.1	87.3



Obr. 3.4: Histogram rozdelenia percenta identity zarovania - Dorado a Guppy. Grafy boli vytvorené softvérom NanoPlot

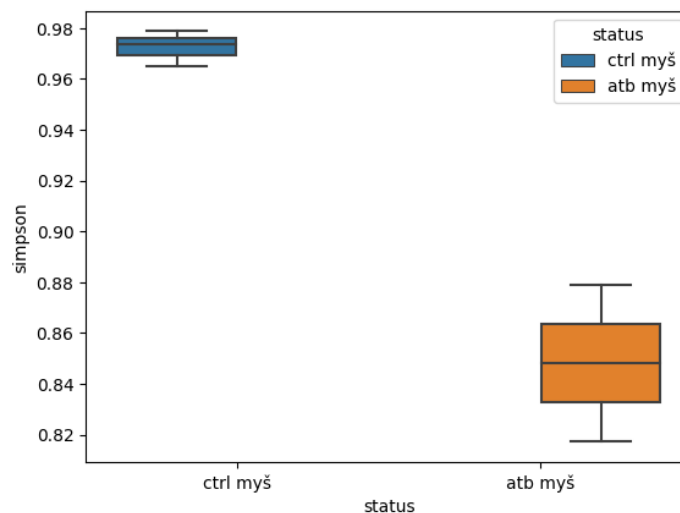
3.2 Myši a antibiotiká

Počet druhov nájdených vo vzorkách myší pred antibiotikami sa pohyboval v priemere okolo hodnoty 250. U myší po užívaní antibiotík počet druhov klesol na priemernú hodnotu 186. Alfa diverzita myší pred a po antibiotikách je zobrazená pomocou Shannonovej diverzity na obrázku a pomocou Simpsonovej diverzity na obrázku 3.6. Záznam myší pred užívaním antibiotík reprezentuje označenie `ctrl` myš. V tomto zázname sú obsiahnuté čítania barkódu 7, 8 a 9. Záznam myší po užívaní antibiotík reprezentuje označenie `atb` myš. V tomto zázname sú obsiahnuté čítania barkódu 10 a 11.



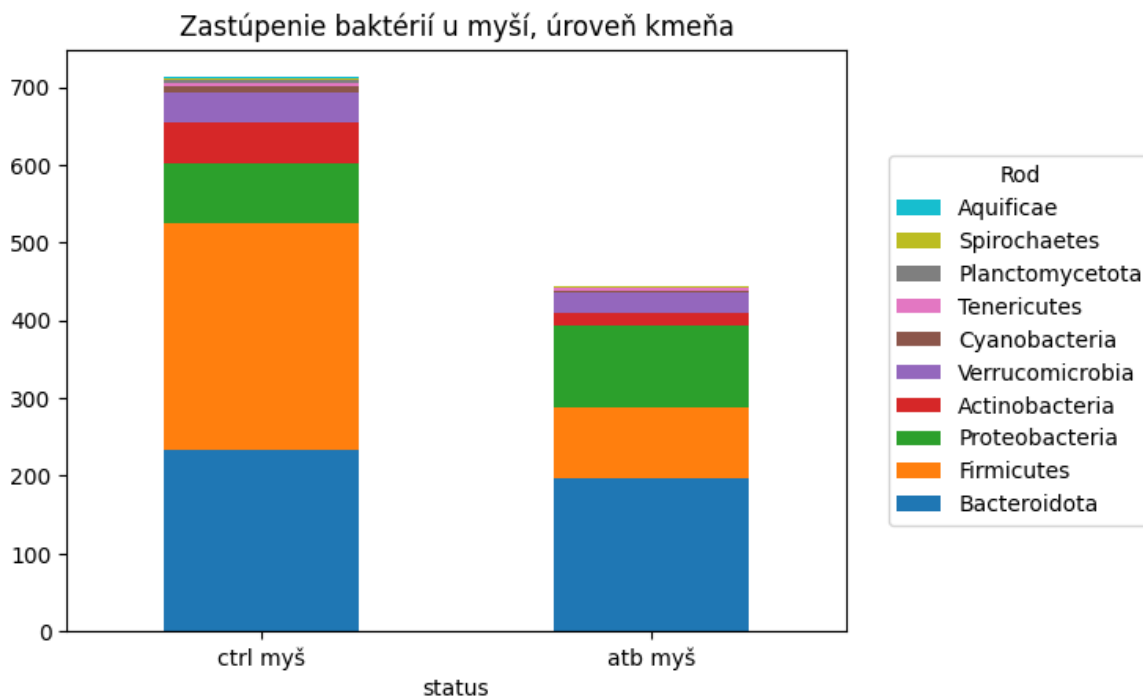
Obr. 3.5: Boxplot Shannonovej diverzity u myší pred a po užívaní antibiotík. Diverzita po užívaní antibiotík je výrazne nižšia.

Na obrázkoch 3.7 a 3.8 môžeme pozorovať rozdielnosť zastúpenia kmeňov baktérií pred a po antibiotikách. Obrázok 3.7 sa zameriava na všeobecné porovnanie hojnosti

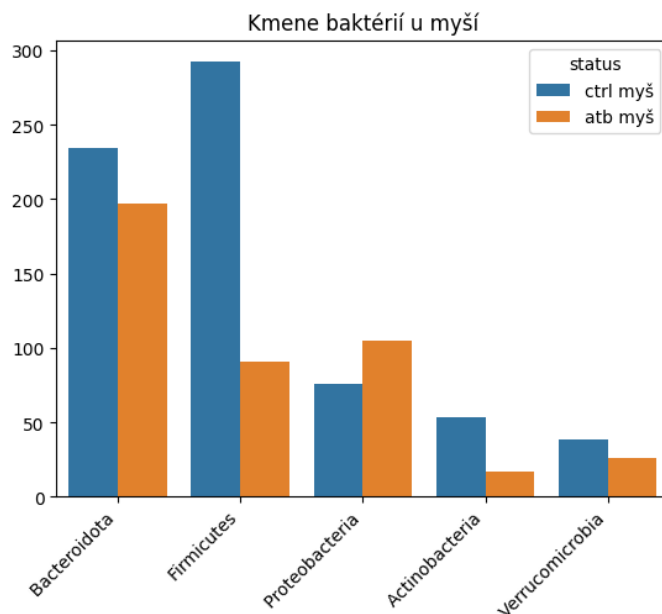


Obr. 3.6: Boxplot Simpsonovej diverzity u myši pred a po užívaní antibiotík. Diverzita po užívaní antibiotík je výrazne nižšia.

kmeňov vizualizované skladaným stĺpcovým grafom. Kmene boli zoradené podľa hodnoty ich priemernej hojnosti a následne boli vybrané tie najsignifikantnejšie. Obrázok 3.8 sa sústreďuje na 5 najhojnejších kmeňov baktérií a porovnáva ich hojnosť medzi atb a ctrl myšami stĺpcovým grafom.

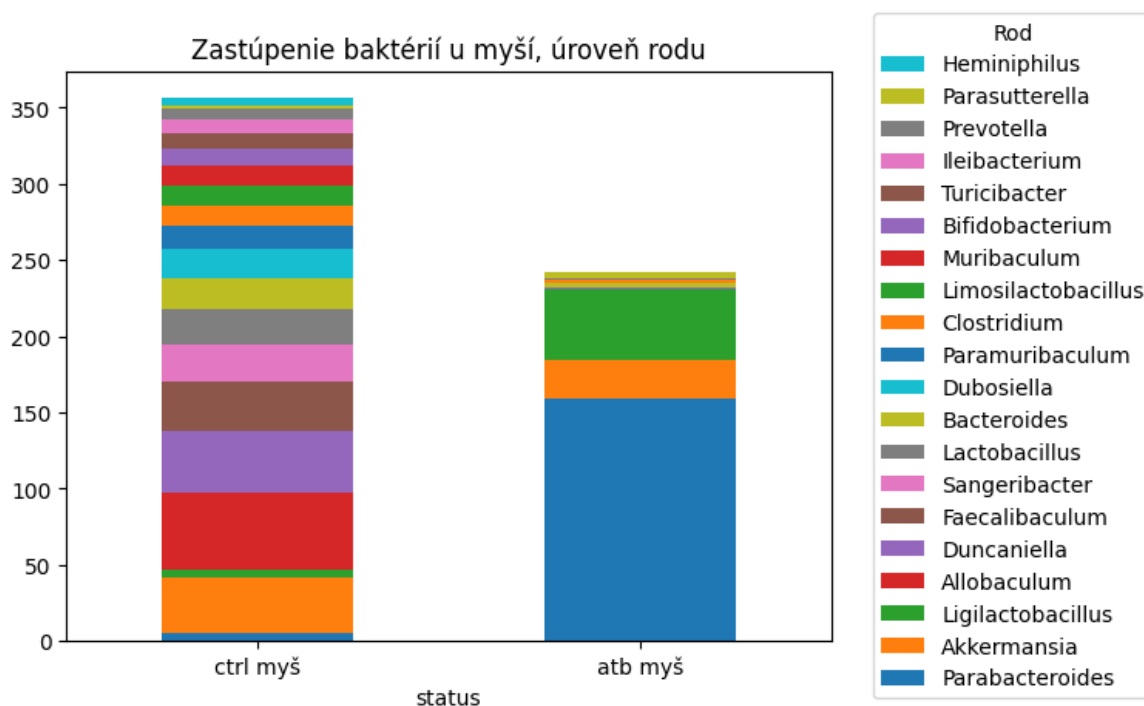


Obr. 3.7: Skladaný graf hojnosti kmeňov baktérií u myši pred a po užívaní antibiotík. Zobrazených je prvých 10 najhojnejších kmeňov. Hodnoty hojnosti sú vypočítané ako priemery z jednotlivých barkódov.

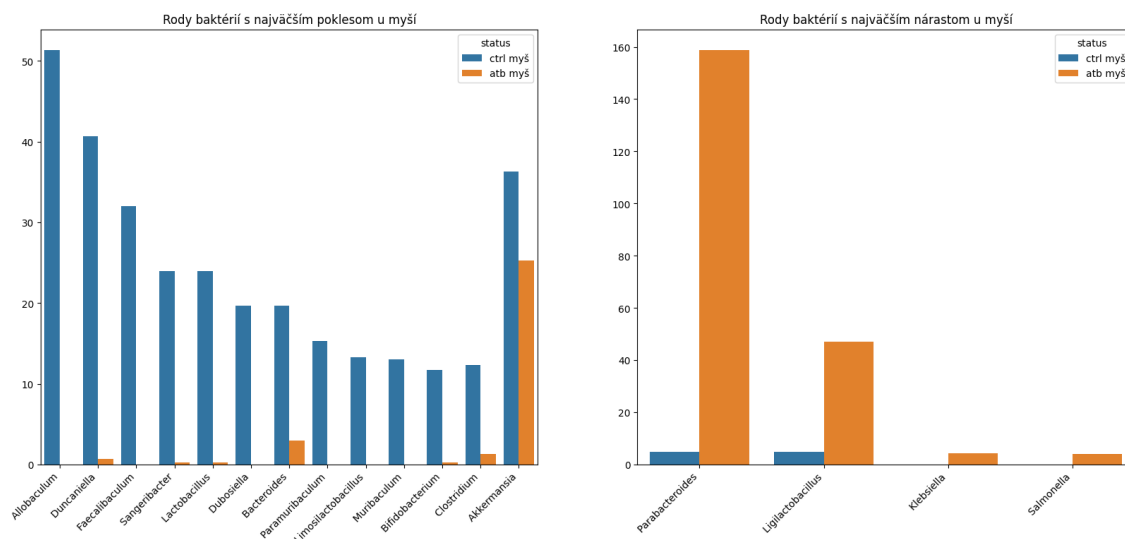


Obr. 3.8: Stĺpcový graf hojnosti kmeňov baktérií u myší pred a po užívaní antibiotík. Zobrazených je prvých 5 najzastúpenejších kmeňov. Hodnoty hojnosti sú vypočítané ako priemery z jednotlivých barkódov. U všetkých kmeňov okrem *Proteobacteria* nastal pokles hojnosti. Najsignifikantnejší pokles nastal u kmeňa *Firmicutes*.

Podrobnejší pohľad na rozdielnosť baktérií u myší prezentujú obrázky 3.9 a 3.10. Obrázok 3.9 zobrazuje hojnosti baktérií na úrovni rodu v podobe skladaného grafu. Obrázok 3.10 sa sústreďuje na podrobnejšie porovnanie rodov, u ktorých došlo ku výraznému poklesu a nárastu hojnosti rodov v podobe stĺpcového grafu.



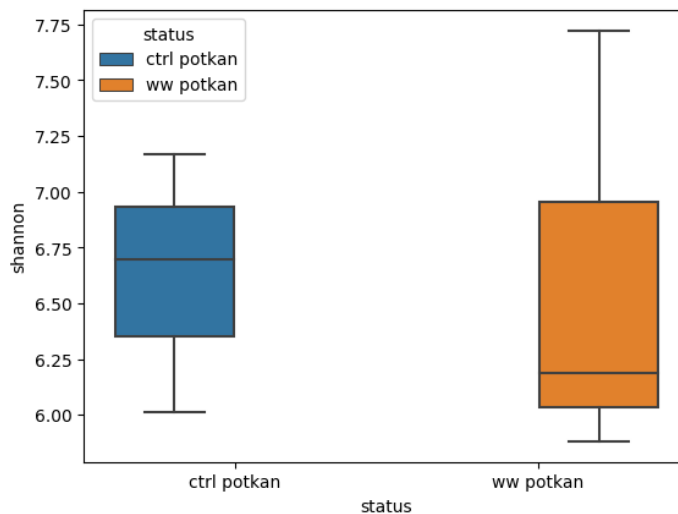
Obr. 3.9: Skladaný graf hojnosti rodov baktérií u myší pred a po užívaní antibiotík. Zobrazených je prvých 20 najhojnejších rodov. Hodnoty hojnosti sú vypočítané ako priemery z jednotlivých barkódov.



Obr. 3.10: Sltĺpcový graf hojnosti rodov baktérií u myší pred a po užívaní antibiotík. Zobrazené sú rody s najvyšším poklesom a vzrastom hojnosti. Na ľavom grafe sa nachádzajú rody baktérií u ktorých nastal pokles hojnosti, na pravom grafe nárast. Hodnoty hojnosti sú vypočítané ako priemery z jednotlivých barkódov.

3.3 Potkany a odpadová voda

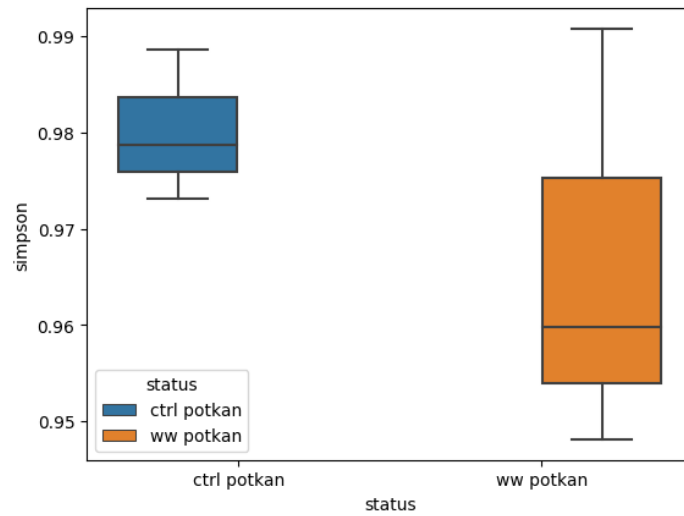
Priemerný počet druhov u kontrolných potkanov je 152 a u ovplyvnených odpadovou vodou 219. Bohatosť druhov oproti myšiam vzrástla. Indexy určujúce rôzne alfa diverzity kontrolných potkanov verzus ovplyvnených odpadovou vodou možno vidieť na obrázku 3.11 a obrázku 3.12. Záznam potkanov, ktoré užívali štandardnú vodu reprezentuje označenie **ctrl potkan**. V tomto zázname sú obsiahnuté čítania barkódu 1, 2 a 3. Záznam potkanov, ktoré užívali odpadovú vodu reprezentuje označenie **ww potkan**. V tomto zázname sú obsiahnuté čítania barkódu 4, 5 a 6.



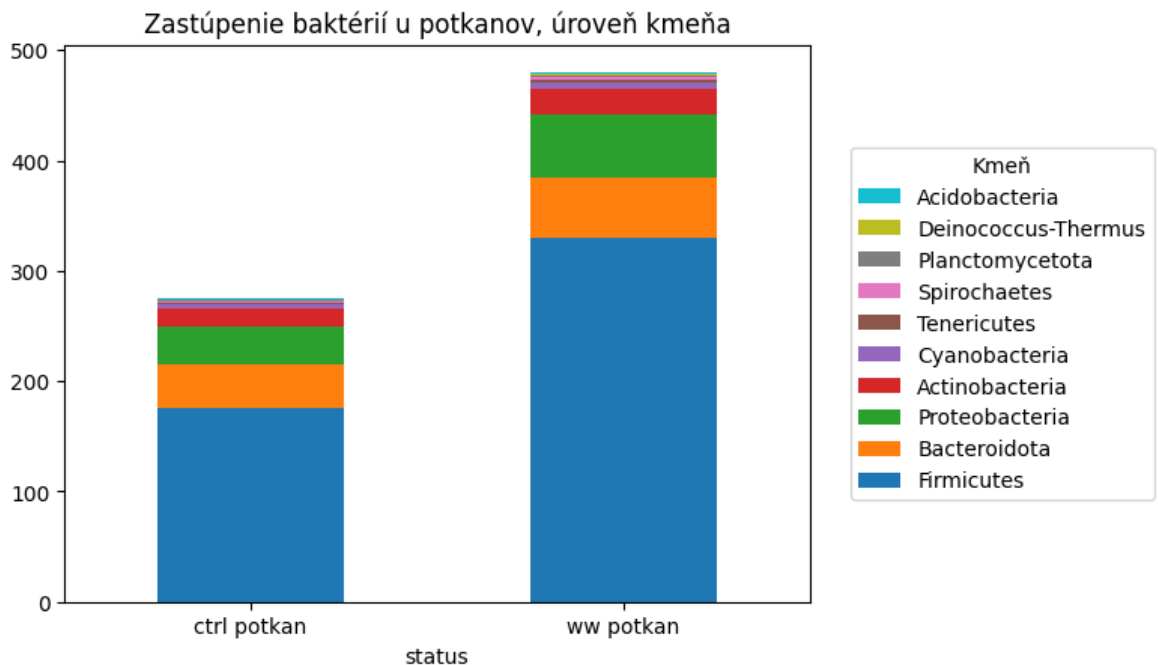
Obr. 3.11: Boxplot Shannonovej diverzity u potkanov, ktoré užívali odpadovú vodu verzus diverzita u potkanov so štandardnou vodou. Nenastal výrazný pokles ani vzrast diverzity.

Na obrázkoch 3.13 a 3.14 môžeme pozorovať rozdielnosť zastúpenia kmeňov baktérií u ctrl vs ww potkanov. Obrázok 3.13 sa zameriava na všeobecné porovnanie hojnosti kmeňov vizualizované skladaným stĺpcovým grafom. Kmene boli zoradené podľa hodnoty ich priemernej hojnosti a následne boli vybrané tie najsignifikantnejšie, rovnakým postupom ako u myší. Obrázok 3.14 sa sústreďuje na 4 najhojnejšie kmene baktérií a porovnáva ich hojnosť medzi potkanmi stĺpcovým grafom.

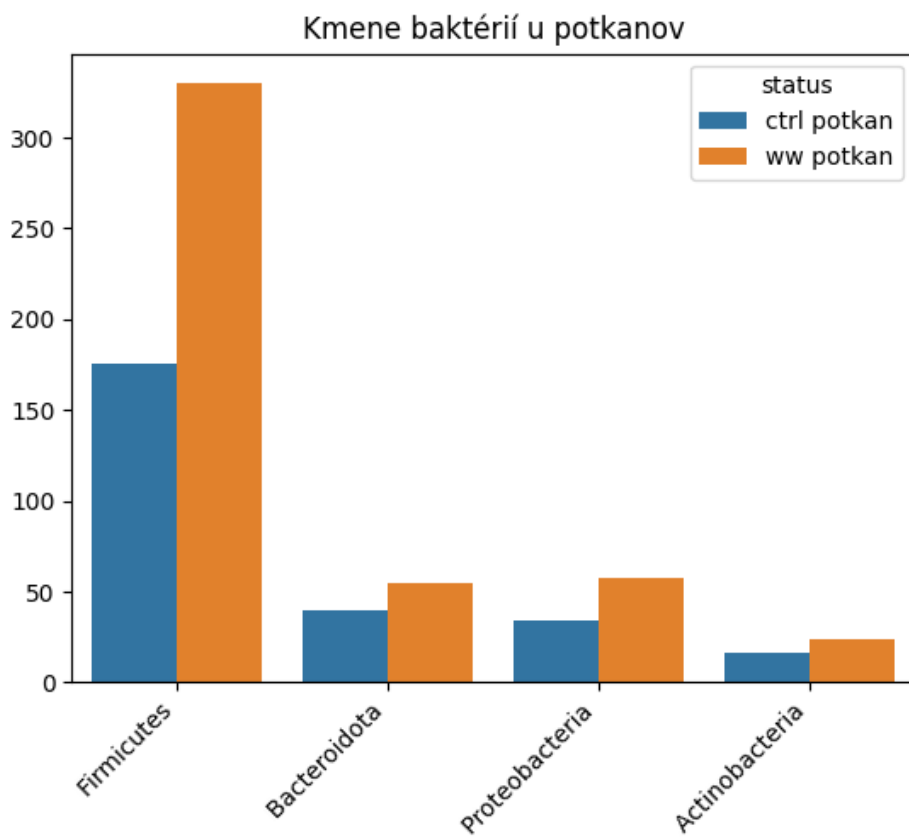
Podrobnejší pohľad na rozdielnosť baktérií u potkanov prezentujú obrázky 3.15 a 3.16. Obrázok 3.15 zobrazuje hojnosti baktérií na úrovni rodu v podobe skladaného grafu. Obrázok 3.16 sa sústreďuje na podrobnejšie porovnanie najhojnejších rodov, u ktorých došlo najmä k výraznému nárastu hojnosti rodov v podobe stĺpcového grafu.



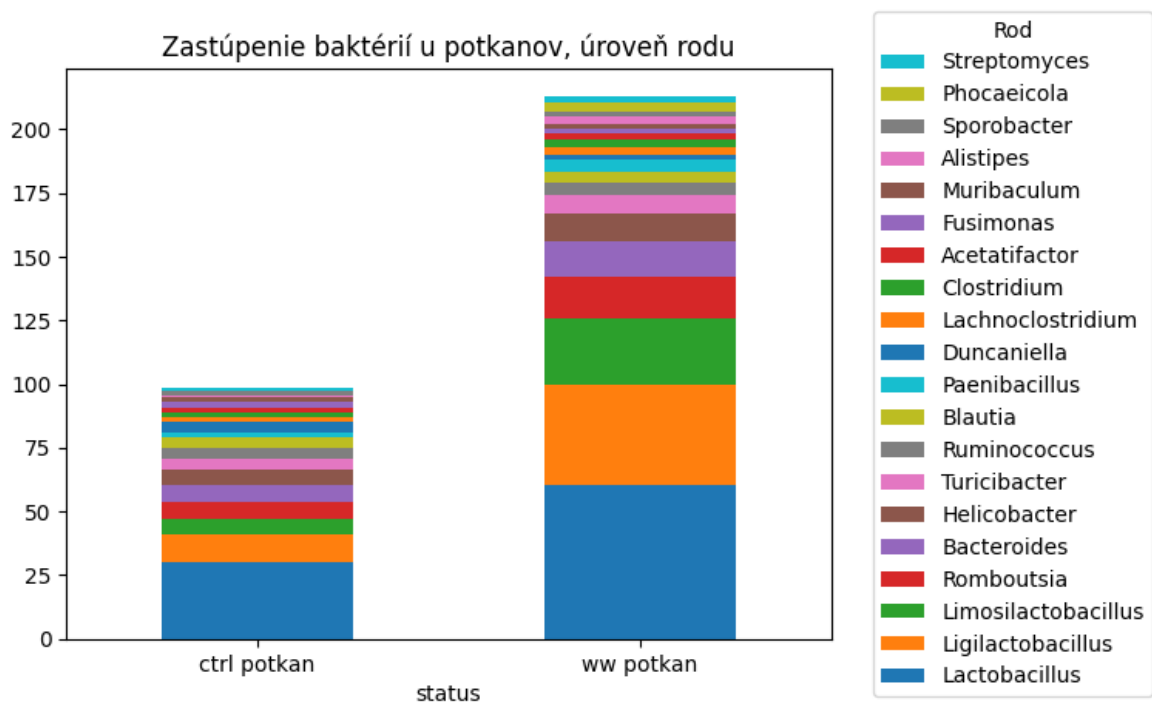
Obr. 3.12: Boxplot Simpsonovej diverzity u potkanov, ktoré užívali odpadovú vodu verzus diverzita u potkanov so štandardnou vodou. Nenastal výrazný pokles ani vzrast diverzity.



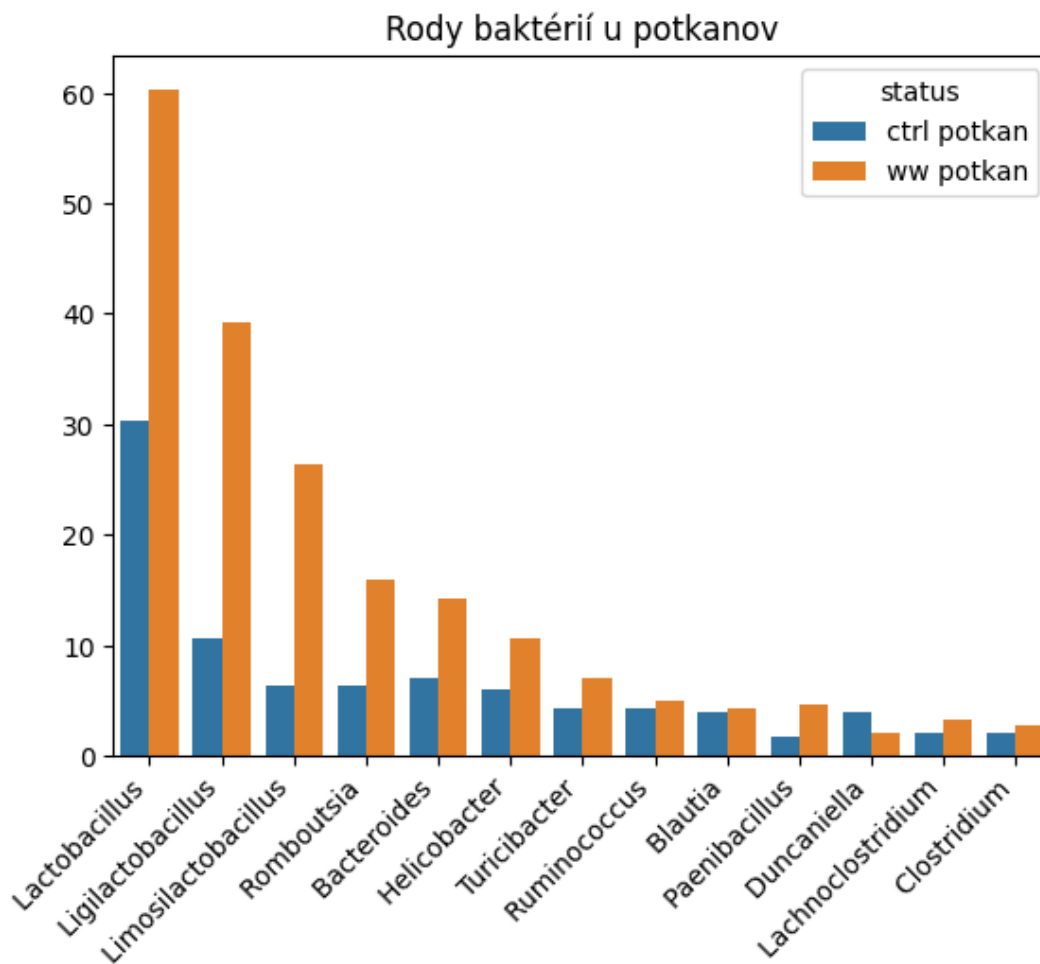
Obr. 3.13: Skladaný graf hojnosti kmeňov baktérií u potkanov. Zobrazených je prvých 10 najhojnejších kmeňov. Hodnoty hojnosti sú vypočítané ako priemery z jednotlivých barkódov.



Obr. 3.14: Stĺpcový graf hojnosti kmeňov baktérií u potkanov. Zobrazené sú prvé 4 najhojnejšie kmene. Hodnoty hojnosti sú vypočítané ako priemery z jednotlivých bar-kódov. Najväčší nárast zastúpenia baktérií je viditeľný u kmeňa *Firmicutes*.



Obr. 3.15: Skladaný graf hojnosti rodov baktérií u potkanov. Zobrazených je prvých 20 najhojnejších rodov. Hodnoty hojnosti sú vypočítané ako priemery z jednotlivých barkódov.



Obr. 3.16: Stĺpcový graf hojnosti rodov baktérií u potkanov. Zobrazených je 13 rodov s najvyššou hodnotou hojnosti. Hodnoty hojnosti sú vypočítané ako priemery z jednotlivých barkódov. Najväčší nárast zastúpenia baktérií je viditeľný u laktobacilov.

Kapitola 4

Diskusia

Illumina MiSeq verzus MinION sekvenátor

V tejto práci sme sa zamerali na nanopórové sekvenovanie tretej generácie, ktoré produkuje hlavne dlhé čítania (10,000-30,000 bp). Výhodou dlhších čítaní je jednoduchšie zostavenie genómu, presnejšia identifikácia zriedkavých variantov a jasnejšie rozlišovanie opakujúcich sa sekvencií. Druhá generácia sekvenovania, do ktorej patrí technológia zvaná MiSeq od Illumina, je oproti MinION-u značne pomalšia a produkuje prevažne krátke čítania (50-300 bp). Má však výhodu, a to presnosť čítania. Pri sekvenovaní pomocou MinION-u dostaneme čítania s presnosťou 87-99% a pri sekvenovaní pomocou platformy Illumina sa presnosť pohybuje okolo hodnoty 99% [35]. Oba prístupy majú teda svoje výhody i nevýhody. Z daných informácií vyplýva, že ak by sme použili sekvenovanie druhej generácie, dostali by sme presnejšie sekvencie báz a teda by sme mohli dostať aj lepšiu charakteristiku črevnej mikrobioty. V nasledujúcich odsekoch si ukážeme, že nižšia presnosť nanopórového sekvenovania nemusí byť v našom prípade problémom.

Výsledkom vedeckej štúdie [34] je porovnávacie hodnotenie sekvenačných platforiem produkujúcich krátke čítania (Illumina) a dlhé čítania (ONT). Sekvenovaná bola ribozomálna RNA, ktorá je súčasťou malej podjednotky prokaryotických ribozómov (16S rRNA), z rovnakej dávky celkovej genómovej DNA extrahovanej zo vzoriek črevnej mikrobioty. Gén 16S rRNA má približne 1500 bázových párov a skladá sa z deviatich variabilných oblastí, ktoré sú prepojené s vysoko konzervovanými oblasťami. V rámci vedeckej štúdie sa rôzne oblasti génu 16S amplifikovali, čiarkovo kódovali a sekvenovali pomocou Illumina MiSeq a MinION sekvenátorov. Mapovanie sekvenovaného amplikónu pomocou MinION na celý gén 16S rRNA sa analyzovalo pomocou cloudového algoritmu EPI2ME. MiSeq čítania boli zarovnané a analyzované použitím CLC softvéru. Viac ako 90% z čítaní generovaných pomocou odlišných sekvenátorov bolo presne klasifikovaných podľa rodu, alebo podľa druhu. Nesprávna klasifikácia čítaní na

úrovni druhov medzi týmito dvoma prístupmi nebola významná. Z vykonaného experimentu porovnávajúceho výsledky rôznych postupov vyplýva, že sekvenačná platforma MinION spojená s príslušným algoritmom by mohla fungovať ako presná a výhodná stratégia pri charakteristike zloženia mikrobiómu.

V tejto práci bol sekvenovaný celý genóm, nie oblasti génu 16S rRNA. Náš postup teda viedol k veľa neklasifikovaným čítaniam po zarovnaní s referenčnou databázou.

Basecalling - Dorado verzus Guppy

Guppy je jeden z najbežnejšie používaných basecallerov na ONT čítania. Do popredia sa dostáva nový ONT basecaller Dorado, ktorý sme využili v tejto práci. V porovnaní s Guppy má veľmi podobné výsledky, čo sa týka kvality a použiteľnosti výsledných sekvencií. Po zarovnaní čítaní barkódu 12 s referenčným genómom *E.coli* sme sa dopracovali k výsledku, že podiel zarovnaných báz ku genómu bol v prípade oboch basecallerov rovnaký a to 0,8. Priemerná dĺžka Dorado čítaní bola 3530,5 báz, pričom u Guppy bol priemer rovný 3504,4 báz. Na druhú stranu, počet báz v čítaniach bol vyšší v prípade Guppy dát. Priemerné percento identity čítaní oboch basecallerov sa pohybovalo okolo veľmi podobnej hodnoty - 85,9% Dorado a 86,0% Guppy. Distribúcia priemerného skóre a dĺžky čítaní basecallerov vykazovala veľmi nízke odlišnosti. Dorado však obsahovalo viac kvalitnejších čítaní za hodnotou skóre 15. Týmto vysoko kvalitným čítaniam neprikladáme veľkú signifikantnosť, pretože značne neovplyvnili výslednú kvalitu zarovnania s referenčným genómom. Prišli sme však k záveru, že Dorado už vo svojej alfa verzii je vhodný na použitie v našej práci.

MinION a rôzne bioinformatické nástroje

MinION ako sekvenátor je vhodná voľba, ak chceme získať čítania z mikrobiálnej DNA, potrebné na ďalšie bioinformatické spracovanie, špeciálne analýzu druhov a rodov. Nástroj EPI2ME ponúka rôzne workflowy na spracovanie dát. Medzi nich patrí aj wf-metagenomics, ktorý sme použili na analýzu dát v našej práci. Dlhé čítania boli vo vedeckej štúdii [34] analyzované pomocou dvoch rôznych nástrojov: EPI2ME a CLC genomic workbench. Podľa štúdie umožnila aplikácia nástroja EPI2ME s referenciami z databázy NCBI 16S klasifikáciu mikrobiálnych komunití až po úroveň rodu a druhu, zatiaľ čo používanie modulu CLC spojeného s referenciami z databázy ribozomálnych sekvencií SILVA[36] umožnilo klasifikáciu črevnej mikroflóry väčšinou len na úrovni rodu. Pomocou modulu CLC spojeného s databázou SILVA aj nástroja EPI2ME s databázou NCBI 16S bola na úrovni rodu väčšina mapovaných čítaní korektne klasifikovaná a bol k nim priradený podobný taxonomický profil. Na úrovni druhov bol viditeľný mierny pokles správne zaradených čítaní so súčasným nárastom počtu neza-

radených čítaní pri použití oboch algoritmov.

Sekvenčné prístupy krátkeho a dlhého čítania umožňujú rýchlu, vysoko výkonnú a presnú klasifikáciu spoločenstiev s identifikáciou génov 16S rRNA. Napriek tomu, metódika sekvenovania, dĺžka sekvenovaného čítania a analytický pracovný postup týkajúci sa sekvenovania krátkeho a dlhého čítania sú odlišné. Na základe najnovších postupov je možné rôzne druhy čítaní spracovať rovnakými nástrojmi a teda porovnať ich účinnosť a kompatibilitu. Z porovnania nástrojov je zrejmé, že výsledky EPI2ME analýzy sú významnejšie a teda celkovo výsledky MinION-u spojené s cloudovou platformou EPI2ME sú použiteľné na analýzu črevných mikrobiálnych spoločenstiev.

Pozitívne kontroly

Jednou pozitívnou kontrolou bola identifikácia druhu *Escherichia coli* v sekvenciách barkódu 12. Sekvencie sa úspešne namapovali na referenčný genóm pri analýze basecallerov a taktiež bol tento organizmus nájdený aj pomocou workflowu wf-metagenomics.

Ďalšou pozitívnou kontrolou v našej práci boli myši. Z našich výsledkov alfa diverzity je zrejmé, že diverzita myši po užívaní antibiotík sa signifikantne znížila. Značný pokles zastúpenia baktérií bol zaznamenaný u kmeňa *Firmicutes*, menej výrazný pokles u kmeňov *Bacteroidota*, *Actinobacteria* a *Verrucomicrobia*. Vzrast hojnosti sme pozorovali u kmeňa *Proteobacteria*. Na úrovni rodu sa najviac znížila hojnosť napríklad u baktérií *Allobaculum*, *Duncaniella*, *Sangeribacter*, *Faecalibaculum*, *Lactobacillus*, *Bacteroides* a *Bifidobacterium*. Naopak, výrazný vzrast hojnosti sme pozorovali pri rode *Parabacteroides*. Práve baktérie rodu *Parabacteroides* patria pod kmeň *Bacteroidota*. V našej analýze *Bacteroidota* ako kmeň klesol, no jeho časť - rod *Parabacteroides* sa zvýšil.

Z mnohých štúdií vieme, že rozmanitosť baktérií črevného mikrobiómu sa po užívaní antibiotík zmení. Výsledky vedeckej štúdie [37] o vplyve antibiotík na črevnú mikrobiótu človeka hovoria, že okrem zníženia diverzity mikrobióty, by mal nastať pokles v zastúpení baktérií kmeňa *Actinobacteria* a nárast u kmeňa *Bacteroidetes* (synonymum *Bacteroidota*). Iná štúdia naznačuje rovnako pokles diverzity ako aj pokles hojnosti kmeňov *Bacteroidetes* a *Firmicutes* [38].

Možeme teda konštatovať, že naše výsledky sa do veľkej miery podobajú na výsledky z vedeckých štúdií. Ďalšia literatúra považuje rod *Parabacteroides* za baktérie s priaznivým, no aj patogénnym účinkom na ľudské zdravie [39]. Vplyvom antibiotík mohol nastať nárast počtu práve patogénnych druhov.

Vplyv odpadovej vody na potkany

Na rozdiel od myší užívajúcich antibiotiká, sa diverzita potkanov po užívaní odpadovej vody výrazne nezmenila. Pri Shannonovom indexe diverzity je stredná hodnota nižšia u potkanov užívajúcich odpadovú vodu, no celkový rozsah boxplotu je takmer identický s boxplotom u kontrolných potkanov. Túto zmenu teda nemôžeme považovať za signifikantnú. Rovnako aj pri Simpsonovom indexe diverzity nenastala výrazná zmena medzi diverzitami potkanov. Zmenila sa však hojnosť niektorých baktérií. Medzi kmene, ktorým narástla hodnota výskytu patria *Firmicutes*, *Bacteroidota*, *Proteobacteria* a *Actinobacteria*. Na úrovni rodu sa najviac zvýšila hojnosť baktérií *Lactobacillus*, *Ligalactobacillus*, *Limosilactobacillus*, *Romboutsia*, *Helicobacter* a *Ruminococcus*. Jediný pokles bol pozorovaný pri baktériách rodu *Duncaniella*, no nebol až tak signifikantný.

Mnohé z baktérií, ktorým sa zvýšila hojnosť sú prospešné pre zdravie jedinca. Tieto baktérie môžu napríklad čerpať priveľa živín a tým oslabovať jedinca. Niekoľko štúdií naznačuje koreláciu medzi možným poklesom i nárastom prospešných baktérií a následnou zmenou ich pomeru s rôznymi ochoreniami [39]. Štúdie potvrdili a odhalili rozdielnosti v kmeňoch a rodoch baktérií u pacientov s IBS (Syndróm dráždivého čreva) v porovnaní so zdravými kontrolami. Medzi znaky IBS patrilo zvýšenie pomeru *Firmicutes* ku *Bacteroidetes*, zníženie niektorých rodov *Firmicutes* (*Lactobacilli*, *Faecalibacterium*) a *Actinobacterium* (bifidobaktérie) a nárast niektorých čeladi *Firmicutes* (*Streptococci*, *Ruminococcus*) a proteobaktérií. Tieto zistenia štúdie preukázali stratu mikrobiálnej bohatosti, ktorá sa môže podieľať na syntéze aminokyselín, na integrite bunkových spojení a na zápalovej odpovedi, čo naznačuje slabosť funkcií epitelovej bariéry, ktorá čiastočne vysvetľuje symptómy IBS.

Dospeli sme k záveru, že ovplyvniť zdravé fungovanie jedinca môže aj nárast hojnosti prospešných baktérií.

Záver

Sekvenovanie tretej generácie spolu s využitím primeraných nástrojov je vhodný postup na klasifikáciu mikrobiálnych spoločenstiev. Pre kvalitnejšie výsledky experimentu by bolo vhodné sekvenovať iba regióny 16S rRNA. Vyhli by sme sa tak veľkému množstvu neklasifikovaných čítaní. Basecaller Dorado vyprodukoval kvalitné čítania z osekvenovaných dát. Následne workflow wf-metagenomics spadajúci pod cloudovú platformu EPI2ME identifikoval očakávané rozdielnosti medzi kontrolnými a atb vzorkami u myší. Diverzita myší po užívaní antibiotík klesla, spolu so zastúpením niektorých kmeňov *Bacteroidota*, *Firmicutes* a *Actinobacteria*.

Diverzita u potkanov užívajúcich odpadovú vodu sa značne nezmenila. Vplyv odpadovej vody na črevný mikrobióm nemusí byť tak závažný ako pri konzumovaní antibiotík, no pozorovaná zmena pomerov jednotlivých baktérií môže znamenať narušenie mikrobiómu. Medzi kmene, ktorým vzrástla hojnosť patrili *Firmicutes*, *Bacteroidota*, *Proteobacteria* a *Actinobacteria*. Na úrovni rodu sa najviac zvýšila hojnosť baktérií *Romboutsia*, *Helicobacter*, *Ruminococcus* a mnohých laktobacilov. Aj nárast hojnosti môže negatívne pôsobiť na jedinca. Výsledkom tejto práce však nemôžeme vylúčiť negatívny vplyv odpadovej vody na životné prostredie.

Literatúra

- [1] Deblonde T. et al. Emerging pollutants in wastewater: A review of the literature. *International Journal of Hygiene and Environmental Health*. 214, 6, 2011, [online] Dostupné na internete: <https://www.sciencedirect.com/science/article/pii/S1438463911001325>
- [2] Rodriguez-Mozaz S. et al. Antibiotic residues in final effluents of European wastewater treatment plants and their impact on the aquatic environment. *Environment International*. 140, 2020, [online] Dostupné na internete: <https://www.sciencedirect.com/science/article/pii/S016041201934886X>
- [3] Senta I. et al. Wastewater analysis to monitor use of caffeine and nicotine and evaluation of their metabolites as biomarkers for population size assessment. *Water Research*. 74, 2015, [online] Dostupné na internete: <https://www.sciencedirect.com/science/article/pii/S0043135415000755>
- [4] Marchesi J. R., Ravel J. The vocabulary of microbiome research: a proposal. *Microbiome*. 3, 31, 2015, [online] Dostupné na internete: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4520061/>
- [5] Sender R., Fuchs S., Milo R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell*. 164,3, 2016, [online] Dostupné na internete: <https://pubmed.ncbi.nlm.nih.gov/26824647/>
- [6] Stephen A.M., Cummings J.H. The microbial contribution to human faecal mass. *J Med Microbiol*. 13, 1, 1980, [online] Dostupné na internete: <https://pubmed.ncbi.nlm.nih.gov/7359576/>
- [7] Jandhyala S.M. et al. Role of the normal gut microbiota. *World J Gastroenterol*. 21, 29, 2015, [online] Dostupné na internete: <https://pubmed.ncbi.nlm.nih.gov/26269668/>
- [8] D'Amelio P., Sassi F. Gut Microbiota, Immune System, and Bone. *Calcif Tissue Int*. 102, 4, 2018, [online] Dostupné na internete: <https://pubmed.ncbi.nlm.nih.gov/28965190/>

- [9] Robert C., Bernalier-Donadille A. The cellulolytic microflora of the human colon: evidence of microcrystalline cellulose-degrading bacteria in methane-excreting subjects. *FEMS microbiology ecology*. 46, 1, 2003, [online] Dostupné na internete: <https://pubmed.ncbi.nlm.nih.gov/19719585/>
- [10] Kumar D., Antonarakis S., *Medical and Health Genomics*. Academic Press. 2016, [online] Dostupné na internete: <https://www.sciencedirect.com/science/article/pii/B9780124201965090018>
- [11] Oxford Nanopore technologies. Nanopore DNA sequencing. [online] Dostupné na internete: <https://nanoporetech.com/applications/dna-nanopore-sequencing>
- [12] Wang Y., Zhao Y., Bolas A. et al. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol*. 39, 2021, [online] Dostupné na internete: <https://www.nature.com/articles/s41587-021-01108-x>
- [13] Ramirez J. et al. Antibiotics as Major Disruptors of Gut Microbiota. *Front. Cell. Infect. Microbiol*. 10, 2020, [online] Dostupné na internete: <https://doi.org/10.3389/fcimb.2020.572912>
- [14] DasSarma S., Coker J.A., DasSarma P., *Archaea (overview)*. Academic Press, 2009, [online] Dostupné na internete: <https://doi.org/10.1016/B978-012373944-5.00108-5>
- [15] Wood D.E., Lu J., Langmead B., Improved metagenomic analysis with Kraken 2. *Genome Biology*. 20, 257, 2019, [online] Dostupné na internete: <https://doi.org/10.1186/s13059-019-1891-0>
- [16] microBEnet: the microbiology of the Built Environment network. Why use Bracken instead of Kraken? [online] Dostupné na internete: <https://microbe.net/2017/04/27/why-use-bracken-instead-of-kraken/>
- [17] Bracken. [online] Dostupné na internete: <https://ccb.jhu.edu/software/bracken/>
- [18] Travlos I.S. et al. Weed-Species Abundance and Diversity Indices in Relation to Tillage Systems and Fertilization. *Front. Environ. Sci*. 6, 2018, [online] Dostupné na internete: <https://doi.org/10.3389/fenvs.2018.00011>
- [19] Andermann T. et al. Estimating Alpha, Beta, and Gamma Diversity Through Deep Learning. *Front. Plant. Sci*. 13, 2022, [online] Dostupné na internete: <https://doi.org/10.3389/fpls.2022.839407>
- [20] Jarkovský J., Littnerová S., Dušek L. *Statistické hodnocení biodiverzity*. AKADEMICKÉ NAKLADATELSTVÍ CERM, s.r.o. Brno, 2012

- [21] Divíšek J., Culek M., Biogeografie. Kapitola 7.1.1 Měření alfa diverzity., [online] Dostupné na internete: <https://is.muni.cz/do/rect/el/estud/prif/ps13/>
- [22] QIAGEN. QIAamp PowerFecal Pro DNA Kits. [online] Dostupné na internete: <https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/qiaamp-powerfecal-pro-dna-kit>
- [23] Oxford Nanopore Technologies. How basecalling works. [online] Dostupné na internete: <https://nanoporetech.com/how-it-works/basecalling>
- [24] Dorado. [online] Dostupné na internete: <https://github.com/nanoporetech/dorado>
- [25] NanoPlot. [online] Dostupné na internete: <https://github.com/wdecoster/NanoPlot>
- [26] NanoStat. [online] Dostupné na internete: <https://github.com/wdecoster/nanostat>
- [27] Pysam dokumentácia. [online] Dostupné na internete: <https://buildmedia.readthedocs.org/media/pdf/pysam/latest/pysam.pdf>
- [28] Minimap2 dokumentácia. [online] Dostupné na internete: <https://lh3.github.io/minimap2/minimap2.html>
- [29] NCBI databáza. Referenčný genóm E.coli. [online] Dostupné na internete: https://www.ncbi.nlm.nih.gov/genome/167?genome_assembly_id=368386
- [30] Samtools dokumentácia. [online] Dostupné na internete: <http://www.htslib.org/doc/samtools.html>
- [31] EPI2ME Labs. wf-metagenomics documentation. [online] Dostupné na internete: <https://labs.epi2me.io/workflows/wf-metagenomics/>
- [32] Nextflow. [online] Dostupné na internete: <https://www.nextflow.io/>
- [33] Docker. [online] Dostupné na internete: <https://www.docker.com/>
- [34] WEI P.-L. et al. Characterization of Fecal Microbiota with Clinical Specimen Using Long-Read and Short-Read Sequencing Platform. International Journal of Molecular Sciences. 21, 19, 2020, [online] Dostupné na internete: <https://www.mdpi.com/1422-0067/21/19/7110>
- [35] Bridge Informatics. Battle of the Sequencers: Illumina vs Nanopore Sequencing. [online] Dostupné na internete: <https://bridgeinformatics.com/battle-of-the-sequencers-illumina-vs-nanopore-sequencing/>
- [36] SILVA databáza. [online] Dostupné na internete: <https://www.arb-silva.de/>

- [37] Pennycook J.-H., Scanlan P.-D. Ecological and Evolutionary responses to Antibiotic Treatment in the Human Gut Microbiota. *FEMS Microbiol Rev.* 45, 5, 2021 [online] Dostupné na internete: <https://academic.oup.com/femsre/article/45/5/fuab018/6212228>
- [38] Rinninella E. et al. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms.* 7, 1, 2019, [online] Dostupné na internete: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6351938/>
- [39] Ezeji J.-C. et al. Parabacteroides distasonis: intriguing aerotolerant gut anaerobe with emerging antimicrobial resistance and pathogenic and probiotic roles in human health. *Gut Microbes.* 13, 1, 2021, [online] Dostupné na internete: <https://pubmed.ncbi.nlm.nih.gov/34196581/>

Príloha A: výsledky analýzy odpadovej vody

V tejto prílohe uvádzame tabuľku s podrobnými údajmi analýzy odpadovej vody ČOV Trnava Zeleneč. Štyri analýzy v rôznych dátumoch prebehli v laboratóriách FCHPT STU (Bratislava).

dátum analýzy	5.5.2021	11.11.2021	2.12.2021	27.1.2022
jednotky	ng/L			
látka				
2-oxo-3-hydroxy-LSD	<2.2	<1.8	<1.9	<1.3
6-acetylmorphine	<1.2	<0.98	<1	<0.72
Alprazolam	4.9	3.2	5.9	0.82
Amitriptyline	2.3	2.9	5.2	1.6
Amphetamine	120	41	90	78
Benzoylcegonine	210	79	110	21
Caffeine	52000	30000	45000	9700
Cannabinol	<3	<2.6	<2.9	<2
Carbamazepine	520	870	460	92
Catinone	<1.6	<1.5	<1.3	<1.3
Citalopram	35	32	50	15
Clomipramine	<0.65	<0.47	<0.54	<0.28
Clonazepam	<1.2	<1.2	<1.1	<0.96
Cocaine	25	52	53	15
Codeine	65	39	64	16

Dihydro CBZ	3.4	4.6	2.7	<0.73
Donepezil	<4	<2.9	<3.2	<1.7
Epoxy CBZ	57	35	67	6.4
Haloperidol	<1.5	<1.1	<1.3	<0.65
Ketamine	<4.4	<3.4	<3.7	<2.3
Lamotrigine	240	100	240	38
Maprotiline	<1.4	<0.87	<1.1	<0.55
MDA	<10	<8.7	<9.3	<6.8
MDEA	<4.6	<3.5	<3.9	<2.2
MDMA	18	6.5	14	2.8
Mephedrone	<1.9	<1.5	<1.6	<0.99
Metamphetamine	1800	870	1400	270
Methadone	<2.3	1.6	2.2	<0.96
Mianserin	<1.3	<0.93	<1.1	<0.55
Mirtazapine	24	18	39	7.8
Morphine	340	160	300	48
N-Desmethylocitalopram	19	21	33	8.3
Norketamine	<2.9	<2.2	<2.4	<1.5
Norsertaline	<0.98	<0.71	<0.82	<0.42
O-Desmethylenlafaxine	820	470	670	150
Oxazepam	19	16	28	7.6
Oxcarbazepine	49	35	58	12
Oxycodone	<5	<4.2	<4.5	<3.2
Sertraline	<4.5	<3.3	6.5	2.1
THC-COOH	79	80	290	46
Tramadol	400	230	300	80
trans-dihydro-dihydroxy CBZ	1000	590	780	190
Trazodone	39	25	76	10

Venlafaxine	220	140	180	50
Vortioxetine	<0.81	<0.59	<0.68	<0.35