

Svoluji k zapůjčení své diplomové práce ke studijním účelům a prosím, aby byla vedena přesná evidence vypůjčovatelů. Převzaté údaje je vypůjčovatel povinen řádně ocitovat.



**Univerzita Karlova**  
**Přírodovědecká fakulta**

Studijní program: Biologie

Studijní obor: Genetika, molekulární biologie, virologie



**Bc. Kateřina Poláčková**

Sekvenování nové generace v klinické virologii: optimalizace metody a použití na vzorcích  
s neznámým původcem infekce

Next Generation Sequencing in Clinical Virology: method optimization and its use for  
samples with unknown infectious agent

Diplomová práce

Vedoucí práce: Mgr. Lenka Kramná, Ph. D.

Praha, 2021



**Prohlášení:**

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

V Praze 11. 8. 2021

.....

Kateřina Poláčková

### **Poděkování:**

V první řadě bych chtěla bych poděkovat vedoucí mé diplomové práce Mgr. Lence Kramné, Ph.D., za skvělé vedení, odborné a cenné rady a její čas, který mi po celou dobu práce věnovala. Poděkování také patří prof. MUDr. Ondřejovi Cinkovi, Ph. D, který mi umožnil vypracovat diplomovou práci v Laboratoři molekulární genetiky, kde se mi i díky skvělému kolektivu dobře pracovalo. A v neposlední řadě bych chtěla poděkovat mým rodičům, příteli a kamarádům, kteří mě v mém studiu po celou dobu podporovali.

## **Abstrakt**

V této diplomové práci bylo testováno použití sekvenátoru MinION (Oxford Nanopore) na vzorcích připravených tak, aby simulovaly infekční vzorky. Testovaný postup má simulovat práci se vzorkem s neznámým patogenem, proto byl také vybrán metagenomický přístup.

Byly testovány tři kity: Rapid Barcoding Sequencing, PCR Barcoding a Premium whole genome amplification, které se lišily v délce a náročnosti přípravy a způsobu amplifikace nukleových kyselin. Pro testování bylo vybráno celkem osm virů s různými délkami a různým typem genomu (5,6 – 152 kb, ss/ds RNA, dsDNA), ze kterých bylo připraveno 10 vzorků simulujících různé infekce (respirační, gastrointestinálního traktu a moči) a jeden vzorek obsahoval pouze vodu, jako negativní kontrolu. Před přípravou vzorků kity od firmy Oxford Nanopore byly vzorky ošetřeny DNázou a RNázou, virová RNA byla náhodně přepsána do DNA a u některých vzorků byla provedena amplifikace k dosažení větší koncentrace nukleových kyselin.

Přípravou Rapid Barcoding Sequencing kit byly detekovány všechny použité viry s největším počtem virových čtení, celkem 4403, o délce 100-250 nt a s dobrým pokrytím virového genomu. PCR Barcoding kitem bylo detekováno pět z osmi virů a počet identifikovaných virových čtení s délkou 100-200 nt výrazně poklesl. Pomocí Premium whole genome amplification kitu bylo detekováno nejméně virů (tři z osmi) s nejdelšími čteními (až 7500 nt), které však obsahovaly jen krátké virové fragmenty opakující se ve dlouhé sekvenci.

V identifikaci patogenů v simulovaných infekčních vzorcích byl neúspěšnější Rapid Barcoding Sequencing kit.

**Klíčová slova:** sekvenování nové generace, Oxford Nanopore, MinION, klinická virologie, neznámý patogen, metagenomické sekvenování

## **Abstract**

The use of the MinION sequencer (Oxford Nanopore) was tested on samples prepared to simulate infectious samples. The tested procedure is to simulate work with a sample with an unknown pathogen. Therefore, a metagenomic approach was chosen.

Three kits were tested: Rapid Barcoding Sequencing, PCR Barcoding and Premium whole genome amplification. Each kit differed in duration, difficulty to prepare and in amplification of nucleic acids. In total it was chosen eight viruses with different genome lengths and with varying types of the genome (5,6 – 152 kb, ss/ds RNA, dsDNA). Ten samples were prepared to simulate different types of infection (respiratory, gastrointestinal tract and urine), and one sample contained pure water as a negative control. Before preparation of the library with Oxford Nanopore's kits, DNase/RNase treatment was used. The viral RNA was transcribed into DNA and in chosen samples were amplified to reach a higher concentration of nucleic acids.

Rapid barcoding sequencing kit detected all selected viruses with the highest number of viral reads (4403) with a length between 100 and 250 nt and quality coverage of viral genomes. PCR Barcoding kit detected five out of eight viruses, and the number of identified reads with a length of 100-200 nt distinctly decreased. Premium whole genome amplification kit detected only three out of eight viruses, but with the longest reads (up to 7500 nt). However, these reads contained only short viral repeating fragments in a long sequence.

The Rapid Barcoding Sequencing kit was the most successful in the identification of pathogens in simulated infection samples.

**Keywords:** next generation sequencing, Oxford Nanopore, MinION, clinical virology, unknown pathogen, metagenomic sequencing



# Obsah

1.	Úvod	7
2.	Cíl práce	8
3.	Přehled literatury	9
3.1	Sekvenování nové generace v mikrobiologii – vysvětlení pojmů	9
3.2	Nejčastěji používané NGS sekvenační platformy – princip	12
3.3	Využití NGS v klinické mikrobiologii	17
3.4	Oxford Nanopore ve virologii	20
4.	Materiál a metody	25
4.1	Simulace infekčních vzorků	25
4.2	Příprava vzorků před použitím kitů od Oxford Nanopore	29
4.2.1	Materiál a přístroje	29
4.2.2	Izolace nukleových kyselin, ověření kvantity virů a smíchání vzorků	30
4.2.3	Reverzní transkripce, syntéza druhého řetězce DNA, amplifikace/fragmentace, přečištění	33
4.3	Metodika příprava knihovny Oxford Nanopore	43
4.3.1	Chemikálie a přístroje	43
4.3.2	Rapid Sequencing kit	43
4.3.3	PCR Barcoding kit	46
4.3.4	Premium whole genome amplification	51
4.4	Bioinformatická analýza	56
4.4.1	„Basecalling“ a „barcoding“	56
4.4.2	Identifikace sekvencí	56
5.	Výsledky	62
5.1	Celkové hodnocení sekvenace	62
5.1.1	Počty a délky čtení pro tři různé přípravy knihovny – analyzováno softwarem Geneious	62
5.1.2	Celkové hodnocení identifikace sekvencí nástrojem MG-RAST	64
5.2	Obsah sekvencí v negativních kontrolách	66
5.3	Identifikace sekvencí v jednotlivých vzorcích – pomocí softwaru Geneious a MG-RAST	67
5.4	Porovnání počtů identifikovaných čtení v neřaděných a řaděných vzorcích	74
5.5	Identifikace dlouhých a krátkých čtení pomocí softwaru Geneious	75
5.5.1	Identifikace řetězcích se sekvencí v jednom čtení ve vzorcích připravených „WGS kitem“	75
5.5.2	Identifikace krátkých artefaktů ve vzorcích připravených „PCR kitem“	76
6.	Diskuze	77

7. Souhrn	81
8. Seznam použité literatury	82

## Seznam zkratek

	<b>anglicky</b>	<b>česky</b>
16S rRNA	16S ribosomal RNA	16S ribosomální RNA
18S rRNA	18S ribosomal RNA	18S ribosomální RNA
bp	base pair	pár bází
cDNA	complementary dexyribonucleic acid	komplementární deoxyribonukleová kyselina
Ct	threshold cycle	prahový cyklus
dAMP	deoxyadenosin monophosphate	deoxyadenosinmonofosfát
DNA		
dNTP	deoxynucleoside triphosphate	deoxynukleosidtrifosfát
dsDNA	double stranded deoxyribonucleic acid	dvouvláknová deoxyribonukleová kyselina
dsRNA	double stranded ribonucleic acid	dvouvláknová ribonukleová kyselina
E30	echovirus 30	echovirus 30
EIA	enzyme immunoassay	enzymová imunoanalýza
EV	enterovirus	
gDNA	genomic deoxyribonucleic acid	genomová deoxyribonukleová kyselina
H3N2	hemagglutinin 3, neuraminidase 2	hemaglutini 3, neuraminidáza 2
HPeV-1	human parechovirus type 1	lidský parechovirus typ 1
HSV-1	herpes simplex virus 1	herpes simplex virus 1
IgM	immunoglobulin M	imunoglobulin M
kb	kilobase	kilobáze
L132	human lung embryonic cell line 132	buněčná linie odvozená z plíc lidského embrya č. 132
MDA	multiple displasment amplification	vícemístná amplifikace
MDCK	Madine Darby Canine kidney cell line	buněčná linie psích ledvinových buněk Madine Darby
MeV	measles virus	virus spalniček
MG-RAST	Metagenomics rapid annotaion using subsystem technology	Rychlá metagenomická anotace využívající technologii podsystémů
mNGS	metagenomic next generation sequencing	metagenomické sekvenování nové generace
NGS	next generation sequencing	sekvenování nové generace
nt	nucleotide	nukleotid
ON	Oxford Nanopore	Oxford Nanopore
PCR	polymerase chain reaction	polymerázová řetězová reakce
qPCR	quantitative polymerase chain reaction	kvantitativní polymerázová řetězová reakce
RNA	ribonucleic acid	ribonukleová kyselina
RSV B	respiratory sycytial virus type B	respirační sycytiální virus typu B
RT	reverse transcription	reverzní transkripce

RT-PCR	reverse transcription polymerase chain reaction	reverzně transkriptázová polymerázová řetězová reakce
SARS-CoV-2	severe acute respiratory syndrome-related coronavirus 2	těžký akutní respirační syndrom vyvolaný koronavirem 2
SMRT	single molecule real-time sequencing	sekvenování jedné molekuly v reálném čase
ssDNA	single stranded deoxyribonucleic acid	jednovláknová deoxyribonukleová kyselina
ssRNA	single stranded ribonucleic acid	jednovláknová ribonukleová kyselina
UTR	untranslated region	nepřekládaná oblast
VIPIE	Virome pipeline extraction	(nástroj určený k analýze dat viromu)

# 1. Úvod

Sekvenování nové generace (next generation sequencing, NGS) se stalo důležitou součástí výzkumu jak v oblasti lidské a bakteriální genomiky, metagenomiky, tak při studiu genomu virů a jejich variant. Jeho význam se ukázal při nedávných pandemiích způsobených viry SARS-CoV-2 (Wu et al. 2020; Zhu et al. 2020), ale také při epidemii viru Eboly (Quick et al. 2016), Ziky (Quick et al. 2017a) apod., kdy pomohl k identifikaci viru a stále napomáhá k detekci jeho nových variant (Wawina-Bokalanga et al. 2019). Kromě toho, že musíme čelit novým výzvám jako jsou nové epidemie a pandemie, zůstává v medicíně množství případů, kdy nebyl původce infekce rozpoznán. Tyto případy jsou shrnuty například v review (Kennedy, Quan, a Lipkin 2017), kdy u 60 % případů, u kterých se předpokládala virová encefalitida, nebyl odhalen původce infekce. Toto může být způsobeno buďto tím, že se jedná o novou variantu viru či bakterie, na kterou nejsou optimalizovány běžné konvenční metody (např. kultivace, polymerázová řetězová reakce (polymerase chain reaction, PCR), enzymová imunoanalýza (enzyme-linked immuno sorbent assay, ELISA)), nebo se jedná o takového původce, který s onemocněním není běžně spojován, či mohou být na vině jiné důvody neinfekční povahy. NGS by mohlo pomoci tyto diagnózy objasnit, jelikož se jedná o metodu, která se schopna získat sekvence všeho, co se ve vzorku nachází. S tím se ovšem pojí i mnoho metodických úskalí, jako je množství kontaminujících nukleových kyselin ve vzorku, problém se zachycením patogenu, pokud je kvantita příliš nízká a naposled i vysoká cena, v závislosti na použité sekvenační technice.

V této práci jsme se rozhodli otestovat pro tyto účely techniku sekvenování Oxford Nanopore. Výhodami této technologie je především nízká pořizovací cena (\$1000), není potřeba čekat na nasbírání dostatečného množství vzorků, technologie je schopná sekvenovat čtení dlouhá až 100 kb a sekvenování i analýzu je možné provádět v reálném čase.

## 2. Cíl práce

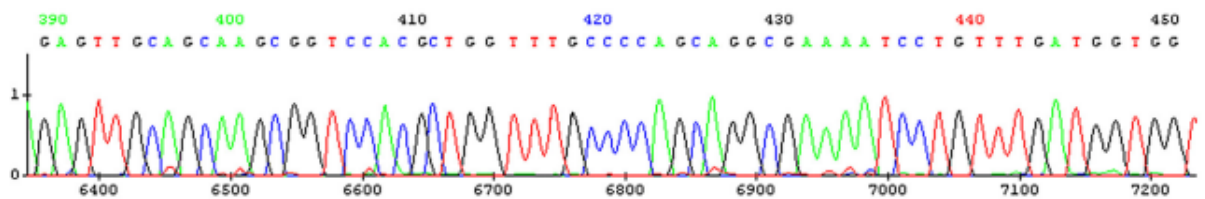
V této diplomové práci byly stanoveny následující cíle:

- na simulovaných infekčních vzorcích obsahující virus, nebo skupiny virů otestovat tři různé přípravy vzorků s použitím tří různých tří kitů od firmy Oxford Nanopore a vybrat nejvhodnější kit pro identifikaci patogenu v klinickém vzorku s neznámou infekcí
- na ředěných infekčních vzorcích otestovat sensitivitu použitých metod
- zhodnotit vliv přítomnosti kontaminujících nukleových kyselin na výsledky sekvenování
- kromě otestování a přípravy protokolu pro laboratorní část otestovat také různé nástroje, jejich výhody a nevýhody pro analýzu dat.

### 3. Přehled literatury

#### 3.1 Sekvenování nové generace v mikrobiologii – vysvětlení pojmů

Sekvenování nové generace (next generation sequencing, NGS) se začalo používat kolem roku 2005 (Margulies et al. 2005), mezi první platformy patřily sekvenátory firmy Roche ([www.roche.com](http://www.roche.com)) a Illumina ([www.illumina.com](http://www.illumina.com)). Jednalo se o možnost poprvé sekvenovat statisíce až miliony čtení, a to v jednom běhu – paralelně. Dříve bylo k dispozici Sangerovo sekvenování, kdy je sekvenátor schopen číst vždy jen jednu sekvenci v reakci (**obrázek 1**).



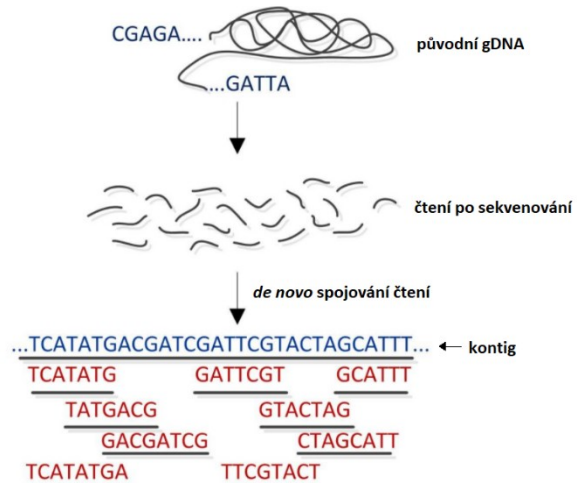
**Obrázek 1:** Elektroforeogram Sangerova sekvenování jedné sekvence DNA (upraveno podle Foret, Klepárník, a Minárik 2017)

První lidský genom byl získán pomocí Sangerova sekvenování (Venter et al. 2001), nicméně až rozvoj NGS umožnil porovnávat lidské genomy za účelem jak výzkumným, tak klinickým (Zhang et al. 2011). Od té doby bylo osekvenováno nejen množství lidských genomů, ale také genomy virů a bakterií. Kromě jednotlivých organismů jsme dnes schopni sekvenovat celé komunity bakterií, virů a eukaryotních organismů ve vzorku. Vzniklo také množství projektů využívajících NGS mapujících sekvence v různých prostředích (např. Human microbiome project (Turnbaugh et al. 2007), Human genome project (Wheeler et al. 2008)).

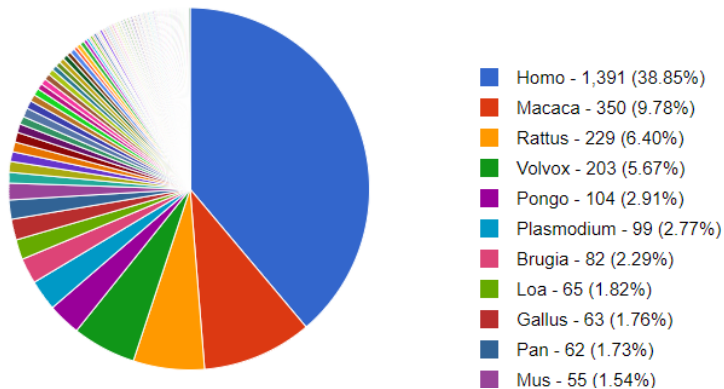
V souvislosti s NGS se používají pojmy, které rozlišují jednak účel sekvenování, s čímž souvisí také způsob přípravy vzorku a následné zpracování sekvencí bioinformaticky. Z oblasti mikrobiologie se jedná zejména o:

- Celogenomové sekvenování – jak už z názvu vyplývá jde o metodu sekvenování celého genomu jednoho organismu. DNA/RNA organismu je získána většinou čistá, bez dalších kontaminujících nukleových kyselin. Na počátku je většinou fragmentována na pšky o délce, kterou umí sekvenátor přečíst. Tyto sekvence jsou nakonec mapovány na referenční sekvenci stejného organismu, pokud se nejedná o sekvenování organismu zcela nového, kde je třeba přistoupit k jinému bioinformatickému zpracování. Získáme tak sekvenci konkrétního organismu, kde lze např. sledovat vzniklé mutace a polymorfismy v jeho genomu.

- Metagenomické sekvenování – tímto přístupem jsou sekvenovány všechny organismy v biologickém vzorku (prostředí). Pokud se jedná o viry, hovoříme o viromu (soubor všech virů ve vzorku), u bakterií o bakteriomu (všechny bakterie) nebo o mikrobiomu (všechny mikroorganismy, často zaměňováno s bakteriomem). Také zde dochází k fragmentování nukleových kyselin na kratší úseky, přičemž genetická informace pochází z více organismů. Výsledná čtení jsou poté spojována bioinformaticky do výchozích sekvencí způsobem *de novo* – na základě podobnosti jedné s druhou. Výsledkem jsou delší fragmenty původních genomů – kontigy (**obrázek 2**). Obsah sekvencí je analyzován různými nástroji, např. softwarem MG-RAST (z angl. Metagenomics Rapid Annotation using Subsystem Technology, příklad výsledku takové analýzy **obrázek 3**).



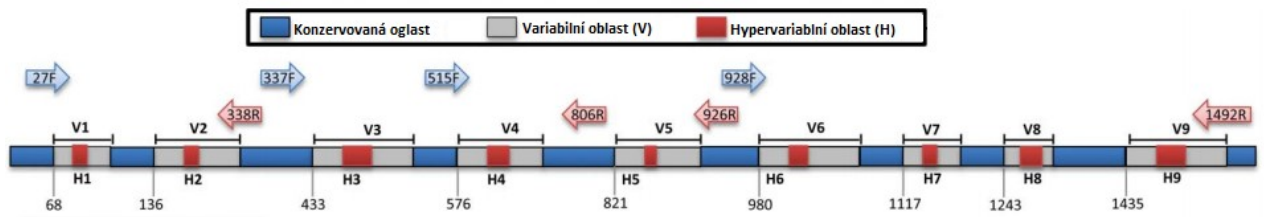
**Obrázek 2:** *De novo* spojování jednotlivých čtení do kontigů (upraveno podle Pavlopoulos et al. 2013)



**Obrázek 3:** Výsledek analýzy softwarem MG-RAST metagenomického sekvenování vzorku nukleových kyselin v lidském aspirátu (příklad z této studie)

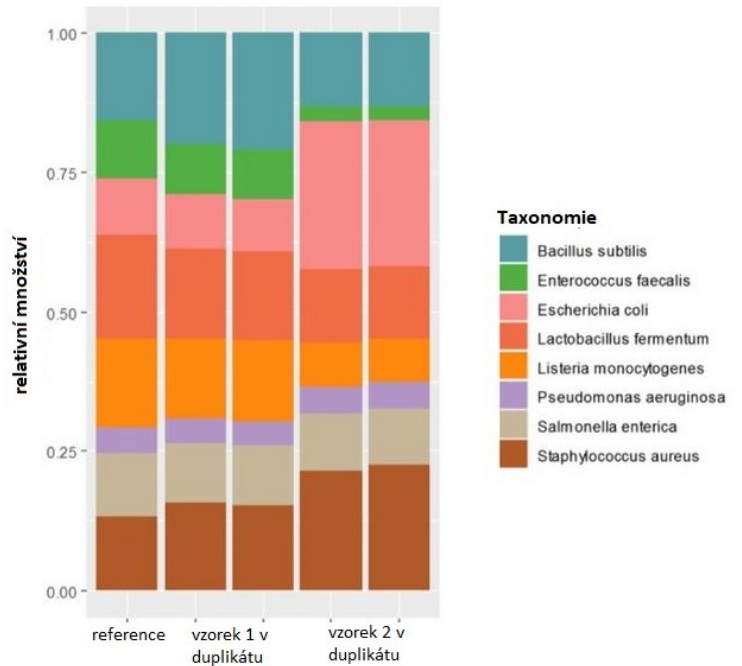
- Metagenomické sekvenování pomocí genu pro 16S rRNA nebo 18S rRNA – gen pro 16S rRNA o délce 1500 bp se vyskytuje u všech bakterií a využívá se pro jejich identifikaci, protože obsahuje několik variabilních oblastí, které jsou obklopeny evolučně konzervovanými oblastmi (**obrázek 4**).





**Obrázek 4:** Gen pro 16S rRNA s barevně vyznačenými oblastmi pro konzervované a (hyper)variabilní oblasti, šipky znázorňují primery (upraveno podle Shahi, Freedman, a Mangalam 2017)

Na tyto konzervované oblasti jsou navrhovány primery tak, aby byly amplifikovány i variabilní oblasti a mohly být následovně sekvenovány pro určení druhů bakterií ve vzorku (příklad takového výsledku **obrázek 5**). Obdobné oblasti jsou i v genu pro 18S rRNA, který nalezneme u hub a eukaryot. Po sekvenování jsou čtení porovnávány v databázích, např. v Ribosomal Database Project ([www.rdp.cme.msu.edu](http://www.rdp.cme.msu.edu)).



**Obrázek 5:** Příklad výsledku metagenomického sekvenování genu 16S rRNA, kde je vidět relativní množství jednotlivých druhů ve vzorcích (upraveno podle Cuscó et al. 2018)

## 3.2 Nejčastěji používané NGS sekvenační platformy – princip

### Illumina

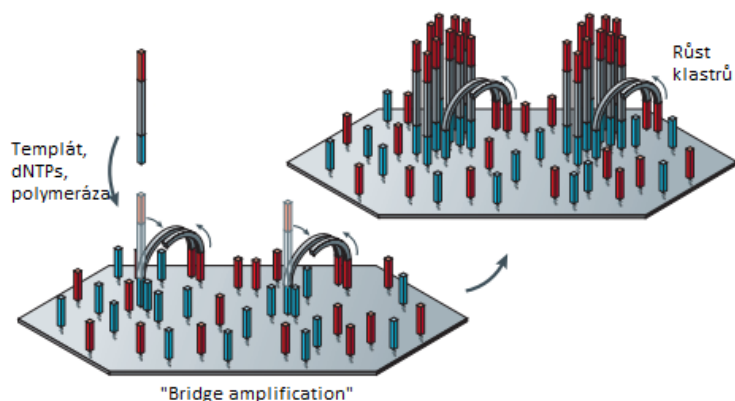
Tato platforma se řadí mezi sekvenátory druhé generace, protože vyžaduje namnožení výchozí knihovny (templátu) díky metodě „bridge amplification“ (přemostění). Namnožení templátu je charakteristické pro sekvenátory druhé generace, které potřebují k přečtení sekvence zesílení signálu. To je zajištěno pomnožením sekvencí, které už dostatečně velký, uniformní signál vydávají (Metzker 2010).

U sekvenátorů od firmy Illumina je během jednoho běhu možné získat jednu miliardu čtení. Délka sekvenovaných fragmentů je 35–300 bp (www.illumina.com).

#### Amplifikace pomocí přemostění („Bridge amplification“)

Po nanesení denaturované knihovny na „flowcellu“ (sekvenační destičku) hybridizují jednovláknové DNA pomocí ligovaných adaptérů k oligonukleotidům navázaných na „flowcelle“. Tyto jsou ve směru „forward“ a „reverse“, stejně jako adaptéry na konci DNA fragmentu. Následně se tato vlákna amplifikují pomocí několikanásobné PCR. 3' konec fragmentu hybridizuje s komplementárním oligonukleotidem na „flowcelle“ a vytvoří tak most a kopíruje vlákno z druhé strany

(obrázek 6). Tvoří se klastry identických molekul (jeden klaster obsahuje přibližně 1000 kopií jedné molekuly (Buermans a den Dunnen 2014)). Jakmile dojde k vytvoření klastrů probíhá vlastní sekvenace.



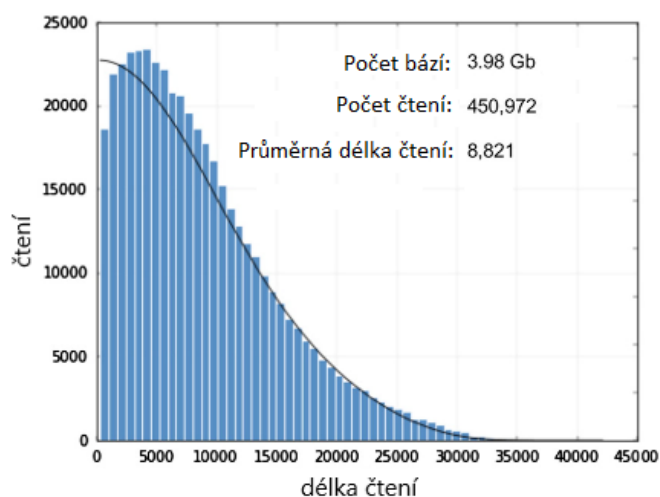
**Obrázek 6:** Amplifikace templátového vlákna do klastrů (upraveno podle Metzker 2010)

### Vlastní sekvenace:

Na destičku jsou přidány sekvenační primery, DNA polymeráza a fluorescenčně značené nukleotidy. Tyto nukleotidy obsahují reversibilní terminační skupinu, která znemožňuje připojení dalšího nukleotidu. Terminační skupina spolu s fluorescenční značkou je odstraněna poté, co je přečten signál a následně může být připojen další nukleotid a proces se opakuje (**obrázek 7**). Přesnost tohoto sekvenování přesahuje 99 % (Stoler a Nekrutenko 2021).

### **Pacific Biosciences**

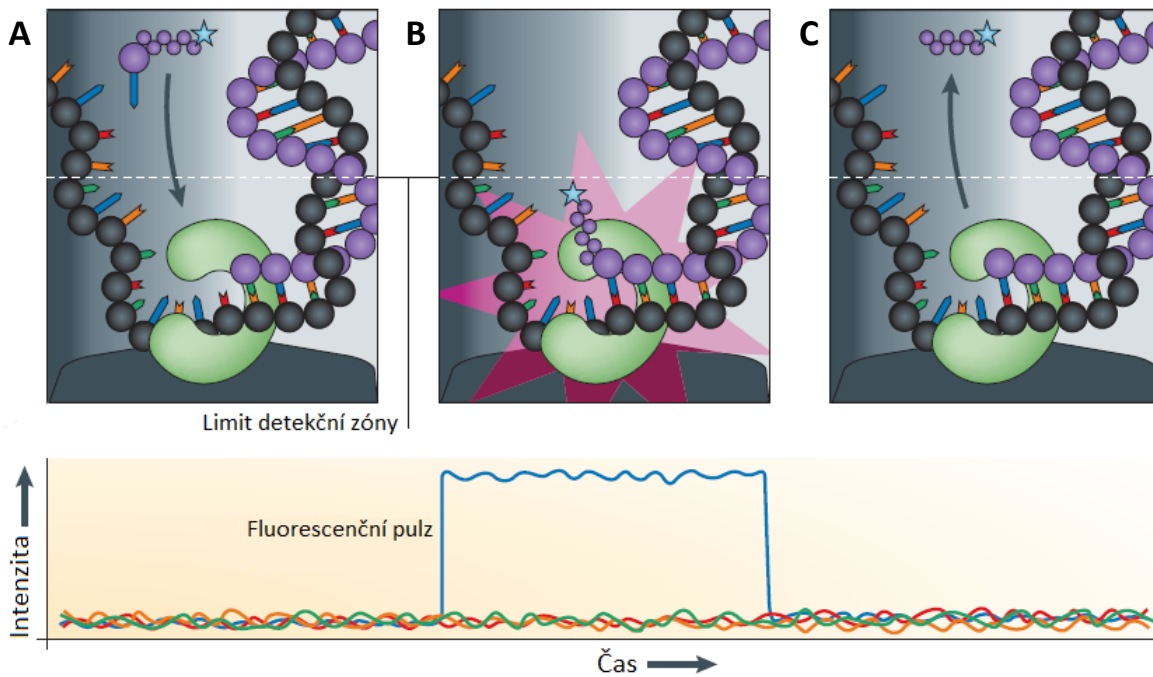
Tato platforma patří do sekvenátorů třetí generace, protože sekvenátor již umí číst signál z jediné molekuly DNA. Použitá metoda sekvenování se nazývá SMRT (z anglického „Single Molecule Real-Time“). Pomocí SMRT lze sekvenovat až 25 kb dlouhá čtení. Rychlost syntézy je 2-4 nukleotidy za sekundu (Eid et al. 2009). Maximální počet čtení, stejně jako jejich maximální délka závisí na zvoleném způsobu přípravy (příklad pro chemii s názvem P5-C3 **obrázek 8**).



**Obrázek 8:** Graf délek a počtu čtení po sekvenování platformou Pacific Biosciences s P5-C3 chemií. Data získaná po sekvenování jedné knihovny v osmi SMRT jamkách (upraveno podle Buermans a den Dunnen 2014)

### Princip SMRT:

Metoda je založena na monitorování polymerázové aktivity během inkorporace nukleotidů k vlákně DNA. Reakce probíhá v malé jamce, kde je k podkladu kovalentně navázána jedna molekula DNA polymerázy. Nukleotidy jsou specificky fluorescenčně značeny na fosfátové skupině. Jakmile jsou v kontaktu s templátovou DNA a polymerázou, tak dojde ke světelnému pulzu, který je detekován senzorem (**obrázek 9B**). Poté je fosfátová skupina i s fluorescenční značkou odstraněna (**obrázek 9C**). Templátem je dsDNA, na jejíž koncích jsou připojeny adaptéry, které vytvářejí smyčky a vytvoří tak cirkulární ssDNA a jakmile polymeráza zreplikuje jedno vlákno může přes adaptér pokračovat replikací druhého vlákna.



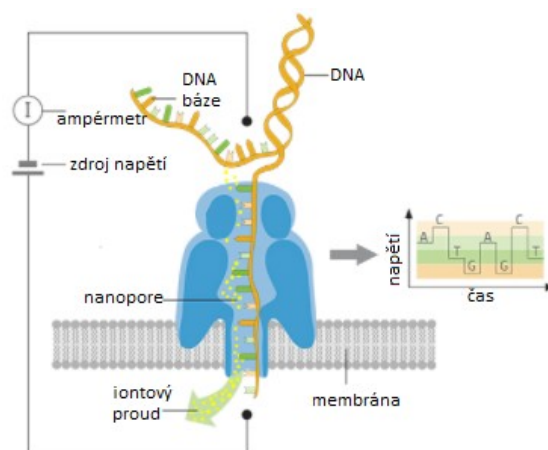
**Obrázek 9:** Princip SMRT, A: přiblížení komplementárního dNTP k vláknu DNA a DNA polymeráze, B: asociace dNTP s templátem a aktivním místem DNA polymerázy a vytvoření světelného pulzu, C: odštěpení fosfátové skupiny s fluorescenční barvou a polymeráza se přemístí na další nukleotid (upraveno podle Metzker 2010)

## Oxford nanopore

V roce 2014 přišla firma Oxford Nanopore (ON) s novou technologií pro sekvenování pomocí nanoporů, kdy lze detekovat signál už jen z jedné molekuly DNA. Proto se technologie řadí do sekvenátorů třetí generace. Nejen sekvenování, ale i analýza dat může probíhat v reálném čase (po zjištění, že máme již dost dat na vlastní analýzu), což snižuje dobu sekvenování z dní na hodiny (hodí se u projektů, kdy je žádoucí znát výsledek co nejdříve). Protože sekvenování není závislé na pomnožení templátové DNA, jako je tomu např. u platformy Illumina, pravděpodobnost vzniku artefaktů se snižuje. Délka sekvenovaného čtení může být až 100 kb, průměrně se pohybuje mezi 250 nukleotidy až 10 kb (Jain et al. 2016). Maximální počet čtení se odvíjí od délky sekvenovaných fragmentů, MinION je schopen produkovat až sto tisíc čtení o průměrné délce 5 kb (Greninger et al. 2015). Mezi hlavní nevýhody této metody patří chybovost, která se průměrně pohybuje mezi 10-15 % (Loit et al. 2019).

### Princip metody:

Během přípravy knihovny jsou ke koncům DNA připojeny sekvenační adaptéry a na 5' konec je pak navázán motorový protein, který se připojí k nanoporu a umožňuje průchod vlákna pórem (**obrázek 10**). Adaptéry také zajišťují koncentraci DNA vláken na povrchu membrány v blízkosti pórů, čímž zvyšují až 1000x vychytávání vlákna (Jain et al. 2016). Poblíž nanoporů jsou na „flowcelle“ také hydrofobní provazce („tether“), které navádí DNA do blízkosti nanoporu.

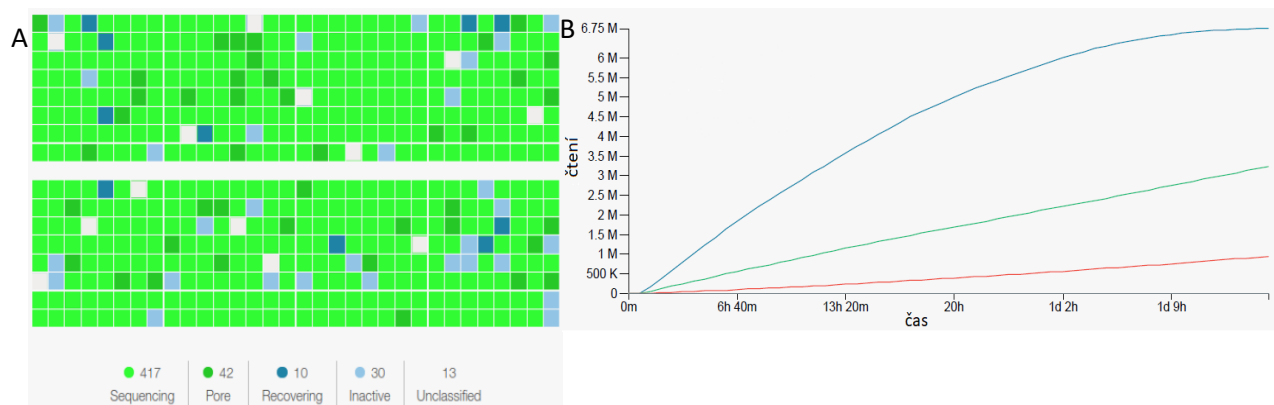


**Obrázek 10:** Vlákno DNA procházející nanoporem (upraveno podle [www.scienceschool.org](http://www.scienceschool.org))

Nanopory tvořené z  $\alpha$ -hemolysinu jsou zanořeny do elektro-rezistentní lipidové membrány. Na každé destičce je 2048 nanoporů. Nanopory prochází ionty s konstantním napětím, které je narušeno průchodem vlákna DNA. Každá jeho báze je jinak veliká a charakteristicky naruší proud iontů a tím se změní i napětí. Každý nanopore je napojen na vlastní elektrodu, která je napojena na kanál a senzorový čip, který měří změnu napětí ([www.nanoporetech.com](http://www.nanoporetech.com)).

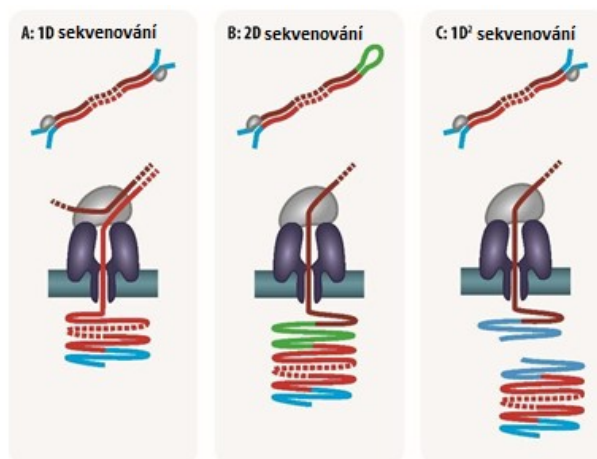
Na každý kanál jsou napojeny čtyři nanopory, napětí se ovšem měří vždy jen z jednoho póru v čase. Všechny kanály jsou aktivní najednou (molekuly DNA jsou sekvenovány zároveň). Jednotlivými póry může procházet i více molekul DNA za sebou, dokud nedojde k jejich zablokování – pór pak již není aktivní. Signál ke kanálu je pak přenášen z dalšího póru. Na začátku každého sekvenování je u všech pórů měřena jejich aktivita (tzv. „Mux scan“). Poté jsou čtveřice pórů rozděleny podle jejich aktivity od neaktivnějšího po nejméně aktivní. Sekvenování začíná nejvíce aktivním pórem po dobu 24 hodin a zbylá doba je rozdělena mezi ostatní tři póry. „Mux scan“ je opakován po celou dobu sekvenování každou 1,5 hodinu (doba lze nastavit) (Lu, Giordano, a Ning 2016). Průběh sekvenace lze sledovat na obrazovce, viz **obrázek 11**. V průběhu sekvenování se může stát, že dojde k nahromadění zotavujících se a nedostupných pórů. Tyto póry mohou být znovu aktivovány promytím „flowcelly“, kdy dojde k odstranění knihovny pomocí DNázy I a poté se může „flowcella“ použít k dalšímu

sekvenování stejné nebo další knihovny. Maximální doba používání „flowcellu pro přístroje MinION a GrindION 48 hodiny, pro přístroj PromethION 64 hodin ([www.nanoporetech.com](http://www.nanoporetech.com)).



**Obrázek 11:** Průběh sekvenace v softwaru MinKNOW, A: Status kanálů v průběhu sekvenování (světle zelená barva – pór právě sekvenuje, tmavě zelená barva – pór je aktivní, ale nesekvenuje, tmavě modrá barva – zotavující se pór, světle modrá barva – inaktivní pór, bílá barva – neklasifikováno, B: graf kumulace počtu čtení v čase, modrá barva – celkový počet čtení, zelená barva – čtení, u kterých byl úspěšný „basecalling“, červená barva – čtení, u kterých nebyl úspěšný „basecalling“

I když je přesnost sekvenování pomocí ON spíše nižší, od představení prvního sekvenátoru MinION v roce 2014, kdy byla jeho přesnost něco málo přes 60 % (Laver et al. 2015), došlo k výraznému zlepšení až na 95 %. Kvalita sekvenování závisí na způsobu přípravy knihovny. Pokud je sekvenováno pouze jedno templátové vlákno (1D sekvenování), je přesnost menší než při sekvenování i komplementárního vlákna (1D<sup>2</sup> sekvenování), které nahradilo dříve používané 2D. Sekvenování obou vláken umožňují speciální adaptéry, které u 2D sekvenování ligovaly na jednom konci obě vlákna k sobě a ty se pak sekvenovaly hned za sebou (**obrázek 11B**). Při 1D<sup>2</sup> sekvenování dochází ke čtení obou vláken zvlášť, což je zabezpečeno adaptéry udržující komplementární vlákno v blízkosti póru (**obrázek 11C**) (Rang, Kloosterman, a de Ridder 2018). Sekvenováním pouze jednoho vlákna dosáhneme sice kvality čtení v rozmezí 80-90 %, ale za to získáme více čtení než při sekvenování obou vláken, kde ale může být přesnost sekvenování až 99 % (Ashikawa et al. 2018; Loit et al. 2019).



**Obrázek 11:** Způsoby sekvenování dvouvláknové DNA, A – 1D sekvenování, B – 2D sekvenování, C – 1D<sup>2</sup> sekvenování (upraveno podle Rang, Kloosterman, a de Ridder 2018)

### 3.3 Využití NGS v klinické mikrobiologii

NGS metody mají velký potenciál pro identifikaci patogenů v klinických vzorcích bez potřeby předchozí znalosti o přítomnosti viru či jiných agens díky získání velkého počtu sekvencí. Zatím se však v běžném klinickém provozu nepoužívají z důvodu jejich vysoké ceny, složitého a dlouhého procesu přípravy vzorku (týká se vzorku, kde chceme sekvenovat virus) a nutnosti nasbírat množství vzorků pro jeden běh, aby se sekvenování vyplatilo. Přesto je význam NGS pro oblast klinické mikrobiologie stále testován, zejména pro případy, kdy měl pacient všechny znaky infekčního onemocnění, ale žádný běžně testovaný patogen se u něj nepodařilo identifikovat. Uvádí se, že u 50 procent případů infekcí nervové soustavy zůstane původce infekce nerozpoznán (Zanella et al. 2019). NGS by v těchto případech mohlo být přínosem vedle konvenčních metod jako je např. PCR, imunologické analýzy nebo sérologie, případně by bylo i možné identifikovat nový, dosud neznámý patogen (Towner et al. 2008; Palacios et al. 2008).

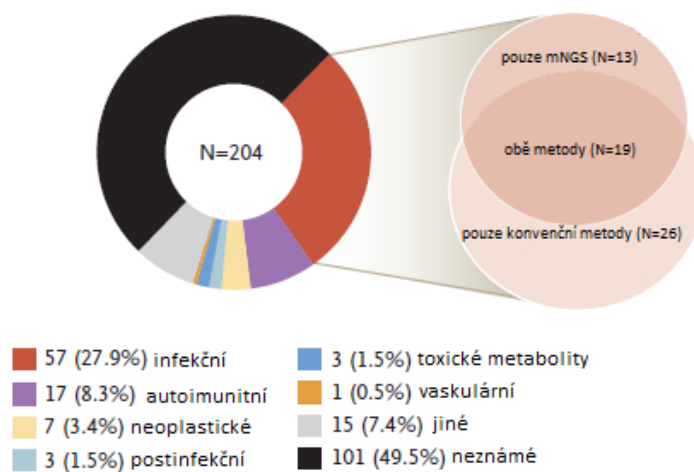
V review Zanella et al. (2019) hodnotili přínos NGS metody při identifikaci původce infekce virových encefalitid, meningoencefalitid a meningitid v klinických vzorcích centrální nervové soustavy (CNS). Review shrnuje výsledky z celkem 29 studií, kde byly použity různé NGS protokoly. Dohromady bylo vyšetřeno 307 vzorků od pediatrických a dospělých, imunokompetentních i imunokompromitovaných pacientů. Vzorky použité k sekvenování byly mozkomíšni mok nebo biopsie mozkové tkáně.

- U 41 pacientů se podařilo identifikovat virus, který potencionálně vyvolal onemocnění.
  - o V deseti případech NGS umožnilo identifikovat viry, u kterých nebylo známo nebo se nepředpokládalo, že by infikovali CNS a proto nebyly konvenčními metodami testovány (např. parvovirus 4, lidský koronavirus OC43, lidský astrovirus), a to hlavně u imunokompromitovaných pacientů.
  - o U pacientů s encefalitou byly objeveny tři nové virové druhy nebo kmeny (Viet-Nam cyclovirus, arenavirus příbuzný s lymfocytárním choriomeningitovým virem a lidský desovirus 1 asociovaný s mozkomíšním mokem).
  - o U 11 pacientů byly nalezeny viry, u kterých je známo, že infikují CNS, ale nebyly běžnými rutinními metodami detekovány nebo vyšetřeny (např. virus Epstein-Barrové, virus klíšťové encefalidity, virus spalniček).

- Ze studie není jasné, u kolika případů byl původce infekce rozpoznán pouze konvenčními metodami.

Autoři konstatují, že chybí standardizace této metody, protože existují různé protokoly pro přípravu knihoven, různé sekvenační technologie a také bioinformatické postupy a databáze, což také znemožňuje objektivní srovnání výsledků. Problematická je také interpretace negativních výsledků, které mohou být ovlivněny nižší sensitivitou metody. I přes to si autoři studie myslí, že by měla být metoda zařazena mezi standardní klinické vyšetření pro případy, protože v tomto review vedlo NGS k odhalení původce infekce u dalších 13 % případů, kdy klasické metody nenašly nic. Hodí se také pro odhalení lokální epidemie. Doba od získání a zpracování vzorků až po bioinformatickou analýzu může trvat 48 hodin až 7 dní, je tedy důležité rozhodnout se pro NGS vyšetření co nejdříve, především v akutních případech.

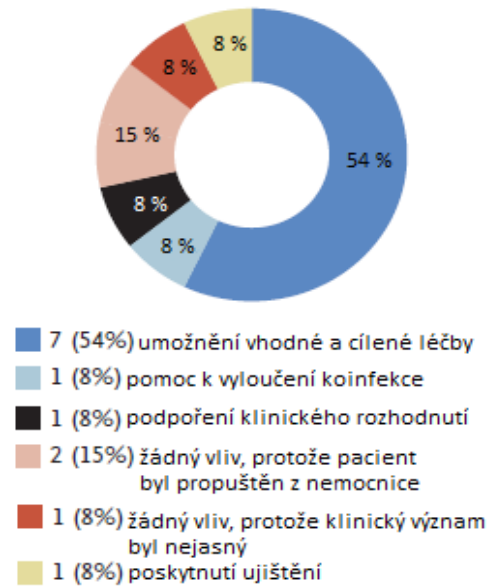
Skupina Wilson et al. (2019) navrhla prospektivní jednorocní studii, ve které sledovali účinnost metagenomického NGS (mNGS) v identifikaci infekčního agens u 204 pacientů hospitalizovaných s infekční meningitidou nebo encefalitidou, výsledky byly ověřeny a porovnány s výsledky konvenčních laboratorních metod. U 50,5 % pacientů byl odhalen původ onemocnění, z toho nejčastější byl infekční původ (27,9 %), dále pak autoimunitní (8,3 %). Celkem bylo identifikováno 58 infekcí u 57 pacientů, 33 % bylo identifikováno pomocí mNGS i konvenčních metod, 26 % pouze konvenčními metodami a 22 % pouze mNGS (**obrázek 12**). Ve třech případech byly patogeny identifikované pouze pomocí mNGS označeny za falešně pozitivní a jejich nález byl přisuzován kontaminaci z prostředí nebo normální lidské flóry. Nicméně kliničtí lékaři využili výsledky z mNGS k vyloučení koinfekce, diagnostice infekčních syndromů a pomohli jim k nasazení správné léčby. U osmi z 13 případů diagnostikovaných pouze pomocí mNGS výsledky příznivě ovlivnily klinické rozhodnutí



**Obrázek 12:** Zastoupení a kategorie stanovených diagnóz (upraveno podle Wilson et al. 2019)



ošetřujících lékařů a v sedmi případech umožnily vhodnou a cílenou terapii (**obrázek 13**), což může představovat další krok v diagnostice meningoencefalitid.



**Obrázek 13:** Vliv na rozhodování o léčbě u případů diagnostikovaných pouze pomocí mNGS (upraveno podle Wilson et al. 2019)

### 3.4 Oxford Nanopore ve virologii

Od vyvinutí technologie používání tohoto sekvenátoru rychle roste. Jeho hlavní výhodou je nízká pořizovací cena (cca \$1000), ale také nízké náklady na jednu sekvenaci (cca \$100). Tomu samozřejmě odpovídá i počet získaných sekvencí, který je výrazně nižší než další dvě již zmiňované platformy (PacBio a Illumina). Toto může být ale výhodou, protože nemusíme čekat na nasbírání velkého počtu vzorků k sekvenaci (což téměř vyřazuje použití obou dvou zmíněných platform pro klinickou virologii), ale lze ekonomicky sekvenovat i jen jeden vzorek (hodí se například u případů akutní infekce, pokud je sekvenování žádoucí).

Při sekvenování technologií ON se používá buď sekvenování amplikonů, nebo sekvenování metagenomické. Sekvenování amplikonů má nespornou výhodu v tom, že pomocí specifických primerů lze ze vzorku tzv. „vytáhnout“ pouze virová čtení, a to pouze od viru, který chceme sekvenovat. To se velmi hodí při sekvenování viru přímo z biologického vzorku, který obsahuje i další kontaminující sekvence. Nevýhodou sekvenování amplikonů však je, že musíme znát virus, který chceme sekvenovat a mít navrženy primery tak, aby nasedaly na jeho konzervativní oblasti. Pokud se vyskytne mutace pod tímto primerem, může dojít k jejímu přehlédnutí, pokud není oblast pokryta dalším PCR produktem. Metagenomickým přístupem naproti tomu sekvenujeme vše ve vzorku, bez předchozí znalosti přítomnosti viru. Bohužel tímto způsobem také získáváme množství nežádoucích sekvencí, které zejména v oblasti virologie často úplně překryjí signál pocházející z viru. Technologie ON byla využita například ve studiích:

- skupina Wang et al. (2015) testovala sekvenování na MinIONu pro rychlou diagnostiku v brzkých fázích pandemie chřipky a rychlému sekvenování virového genomu. RNA viru byla izolována z buněčných kultur a sekvenováno bylo všech osm genů viru influenzy A. Ty byly společně amplifikovány pomocí RT-PCR (z angl. „reverse transcriptase PCR“). K amplifikaci byly použity primery komplementární k vysoce konzervovaným 5' a 3' koncům genů každého segmentu. Knihovna byla připravena gDNA sekvenačním kitem (Oxford Nanopore) a ke koncům amplikonů byly ligovány sekvenační adaptéry. Po čtyřech hodinách sekvenování bylo získáno celkem 118052 čtení, z toho 17108 bylo vysoce kvalitních, protože byly získány z obou vláken DNA (2D sekvenování) a 84,46 % těchto čtení se alignovalo k referenci viru. Bylo dosaženo plného pokrytí všech segmentů influenzy A.
- Quick et al. (2016) vypracovali protokol pro sledování evoluce viru Ebola během propuknutí infekce sekvenováním celého genomu viru Ebola (EBOV) na MinIONu, s

cílenou RT-PCR a navrhli 38 párů primerů, které pokrývaly celý genom viru. Během půl roku osekvenovali celkem 142 vzorků. RNA byla izolována ze vzorku krve, séra nebo moči, která byla poté přepsána do cDNA a amplifikována. Ekvimolární množství ampliconů bylo zpoolováno tak, aby celkové množství DNA bylo 1 µg. Po úpravě konců mohly být připojeny sekvenační adaptéry pro 2D sekvenování a následně byly vzorky sekvenovány na MinIONu. Čtení poté byla alignována k referenci kmene EBOV získaného na začátku propuknutí infekce (EM\_079517), aby byly detekovány varianty genomu. Tímto způsobem byl genom pokryt z 65,7 % s pokrytím alespoň 25 čtení.

- Skupina Quick et al. (2017b) testovali novou metodou přípravy knihovny pomocí multiplexové PCR, kterou by bylo možné cíleně namnožit virový genom obsahující i méně než 50 kopií viru na reakci, a následně sekvenovat MinIONem (Oxford Nanopore). Touto metodou by mělo jít dosáhnout kompletního pokrytí virového genomu z klinického vzorku ve dvou reakcích. Pro pokus vybrali referenční vzorek viru Zika s prahovým cyklem (threshold cycle, Ct) 18-20, ze kterého byla vyizolovaná virová RNA. Následně byla přepsána do cDNA reverzní transkriptázou s náhodnými hexamery. Primery pro amplifikaci byly navrženy tak, aby se vytvořily amplicony o délce 300-500 bp. cDNA byla amplifikována v jedné reakci s poolem primerů (každý o finální koncentraci 0,02 µM). Pro přípravu knihovny byl použit Native Barcoding kit a kit pro přípravu 2D knihovny (Oxford Nanopore). Z celkových 9606 2D čtení se 97,7 % namapovalo na genom viru Zika s pokrytím více jak 25 čtení. Při testování této metody na pěti klinický vzorcích pozitivních na virus Zika s hodnotami Ct v rozmezí 33,6-35,9 osekvenovali více jak 1400000 čtení z každého vzorku, ale pouze jednotky až stovky byly alignovány k referenčnímu genomu. Nejlepšího výsledku bylo dosaženo u vzorku s Ct 33,9, kde našli 223 čtení, které ze 79,2 % pokryly referenční genom viru Zika.
- Brinkmann et al. (2021) zavedl metodu „AmpliCov“ pro rychlou identifikaci varianty viru SARS-CoV-2 pomocí multiplexové PCR a celogenomového sekvenování na přístroji MinION. Metoda je založena na multiplexové amplifikaci ve dvou reakcích s 113 a 111 primery pro amplifikaci celého genomu SARS-CoV-2. Primery byly navrženy podle alignmentu všech dostupných referencí viru v databázi GISAID (k 23. lednu 2020) a jedné reference viru SARS-CoV-1 (NC\_004718) tak, aby pokud možno pokrývaly konzervované sekvence mezi všemi sekvencemi SARS-CoV-2 a SARS-

CoV-1 pro amplifikaci oblastí, kde je zvýšená pravděpodobnost mutací. Délka ampliconů byla v rozmezí 199 a 775 bp. Protokol testovali na třech různých kmenech SARS-CoV-2 izolovaných z buněčných kultur a na 36 klinických vzorcích pacientů s laboratorně potvrzeným onemocněním COVID-19. Extrahovaná RNA byla reverzně transkribovaná s náhodnými hexamery do cDNA. Amplifikace probíhala ve dvou reakcích, protože byly použity dva různé pooly primerů. Následovala příprava knihovny dvěma způsoby (Oxford Nanopore):

- 1) Rapid Barcoding kit: tímto kitem byly připraveny vzorky jak z buněčných kultur, tak od pacientů. Nejprve je DNA fragmentována transpozázou a zároveň jsou k ní připojeny barcodes. Poté jsou vzorky zpoolovány a nakonec připojeny sekvenační adaptéry.
- 2) Ligační sekvenační kit: tento způsob přípravy podstoupily pouze vzorky od pacientů. U celých genomů přepsaných do DNA byly upraveny konce, připojeny barcodes a následně i sekvenační adaptéry.

Připravené knihovny byly sekvenovány na přístroji MinION. U dvou ze tří kmenů izolovaných z buněčných kultur byla identifikována čtení, které se 100% shodovaly s referenčním genomem. Sekvence z každého izolátu (dohromady 75000 čtení) pokrývaly referenční sekvenci z více jak 98 % s hloubkou pokrytí >10. Vzorky získané od pacientů byly sekvenovány do té doby, než bylo získáno 50000 od každého vzorku. Pokrytí genomu bylo vypočítáno jen u oblastí s hloubkou pokrytí >10. Vzorky připravené ligačním sekvenačním kitem s hodnotami Ct 17-23 pokryly genom z 92,6-99,3 %, s Ct 24-29 bylo pokrytí 91,9-99,2 % a s Ct do 35 byl genom pokrytý z 92,2%. Rapid Barcoding kitem připravené vzorky s Ct 17-23 dosahovaly pokrytí genomu 92,0-99,2 % a vzorky s hodnotami Ct >31 nebylo možné osekvenovat.

- Ve studii Greninger et al. (2015) byly metagenomicky sekvenovány Chikunguniya virus (CHIKV), Ebola virus (EBOV) a virus žloutenky C (hepatitis C virus, HCV) a to ve čtyřech různých vzorcích pacientů (CHIKV v plasmě, 2x EBOV v krvi a HCV v séru), u kterých byla laboratorně potvrzena nákaza. Cílem studie bylo poukázat na potenciál technologie v identifikaci patogenů v klinických vzorcích. Koncentrace virů ve vzorku u CHIKV a EBOV byla  $10^7$ - $10^8$  kopií/ml, u HCV  $10^5$  kopií/ml. Tyto vzorky se podařilo osekvenovat bez cílené virové amplifikace, nejprve byla vyizolovaná RNA přepsána do cDNA pomocí náhodných primerů s „tagem“, která byla následně amplifikována

s primery komplementárními k „tagům“. Pro přípravu knihovny byl použit sekvenační kit pro genomickou DNA, pomocí kterého byly upraveny konce cDNA, aby na ně mohly být ligovány sekvenační adaptéry. Sekvence probíhala na přístroji MinION a dohromady byly použity 3 „flowcelly“. Výsledky sekvenování:

- 1) CHIKV: celkem 556 čtení ze sekvenování na první „flowcelle“ se namapovalo na referenci CHIKV. Pokud byly parametry nastaveny na minimální pokrytí 3x byla získána sekvence 90 % genomu. Identita s referencí byla 97-99 %. Pokud bylo sníženo pokrytí na 1x bylo získáno 98 % genomu.
  - 2) EBOV1: po promytí „flowcelly“ byl sekvenován vzorek pozitivní na EBOV, ze kterého bylo získáno pouze 41 čtení z celkových 13090. Tak málo sekvencí bylo získáno nejspíš kvůli přítomnosti velkého množství kontaminujících nukleových kyselin hostitele, protože virus byl izolován z krve.
  - 3) EBOV2: vzorek druhého pacienta pozitivní na EBOV byl sekvenován 4x za sebou na jedné flowcelle. Celkem bylo generováno 335308 čtení a z toho 609 bylo alignováno k EBOV referenci, přičemž 88 % genomu bylo pokryto alespoň jedním čtením.
  - 4) HCV: sekvenací vzorku pozitivního na HCV na nové „flowcelle“ bylo generováno celkem 85647 čtení a pouze 6 z nich bylo alignováno k referenci viru HCV, nejspíš kvůli počátečnímu nízkému titru viru ve vzorku ( $10^5$  kopií/ml).
- Mostafa et al. zjišťovali, jestli je možné detekovat metatranskriptomickým sekvenováním třetí generace koronavirus SARS-CoV-2 ve vzorcích od 40 pacientů, u kterých byl předem pozitivně testován pomocí qPCR (z angl. quantitative PCR) (prahový cyklus v rozmezí 13,9-36,6) a 10 vzorků od pacientů podezřelých, ale negativních na SARS-CoV-2. K tomu ještě charakterizovali možné virové a bakteriální koinfekce. Z nasofaryngeální stěrů byly izolovány veškeré nukleové kyseliny, které byly dále zpracovány dvěma způsoby:
    - 1) sekvenování RNA s poly(A) koncem pro získání sekvencí SARS-CoV-2, která byla přespána do cDNA a k vytvoření knihovny byl použit ligační sekvenační kit, kterým byly upraveny konce cDNA a připojeny sekvenační adaptéry a barcodey.
    - 2) DNA byla přímo zpracována „PCR barcoding kitem“ pro identifikaci koinfekcí. Rozdíl od předchozího kitu byl, že nukleové kyseliny byly nejprve fragmentovány,

poté byly upraveny konce DNA a byly ligovány adaptéry obsahující místa pro navázání primerů pro následnou amplifikaci a barcoding.

Obě knihovny byly poté sekvencovány na přístroji GridION (Oxford Nanopore). U 31 z 40 pozitivních pacientů na SARS-CoV-2 se jim podařilo infekci potvrdit, přitom rozmezí Ct u vzorků, kde se virus nedetekoval, bylo 21-36,6. Autoři se domnívají, že virová čtení mohly být zastíněny hostitelskými. U 10 negativních pacientů se žádná čtení patřící SARS-CoV-2 nenašla. U pěti pozitivních pacientů byly detekovány klinicky závažné organismy s více jak o 50 % větším relativním zastoupením, než jsou normální hodnoty. Např. u dvou pacientů byl nalezen *Haemophilus influenzae* a u jednoho herpes simplex 1.

## 4. Materiál a metody

### 4.1 Simulace infekčních vzorků

K simulaci infekčních vzorků jsme použili viry, bakterie a kvasinku. Jejich taxonomie, charakteristika genomu a informace odkud byly získány jsou shrnuty v **tabulce 1**. Vzorky byly namíchány s ohledem na to, jak se spolu běžně vyskytují i v jaké kvantitě (**tabulka 2**). Snažili jsme se použít viry s různou stavbou i délkou genomu. Vybrané skupiny virů nařazené na požadovanou kvantitu byly smíchány s tzv. „negativním vzorkem“ (shrnuto v **tabulce 2**). Jedná se o vzorek, ve kterém by se běžně testovali a který byl ověřen pomocí kvantitativní PCR na negativitu na všechny testované viry v této studii. Tato strategie byla použita proto, že v negativním vzorku existuje množství tzv. kontaminujících nukleových kyselin pocházející z bakteriálních nebo lidských buněk (zejména ve vzorku stolice). Tento materiál se bude vždy ve vzorku vyskytovat, a proto s ním musíme při testování sekvenční techniky počítat. Použili jsme tři typy negativních vzorků (**tabulka 2**). Negativní vzorky pocházely ze sbírky oddělení virologie, Fakultní nemocnice v Motole. Pro testování bylo použito celkem 10 směsí, 11. vzorek byla voda jako negativní kontrola. U vzorku č. 6 nebyla simulována žádná infekce, byl namíchán tak, abychom otestovali míru a přesnost sekvenace směsi viru, prokaryotního a eukaryotního organismu s různým typem genomu. Vzorky č. 7–10 (**tabulka 2**) vznikly nařazením vzorků 1, 3, 5 a 6 vodou, abychom otestovali míru a přesnost sekvenování technikou Oxford Nanopore u vzorků s nízkou kvantitou.

Virus – taxonomie:			Typ a délka genomu	Odkud jsme je získaly	Materiál použitý k izolaci
Rod	Druh	Genotyp (typ/kmen)			
<i>Enterovirus</i> (EV)	<i>Enterovirus B</i>	Kmen: Echovirus 30 (E30)	+ ssRNA, 7,3 kb	Norwegian Institute of Public Health, Oslo, Norway	supernatant buněčné kultury
<i>Parechovirus</i>	<i>Parechovirus A</i>	Typ: Lidský parechovirus 1 (HPeV-1)	+ ssRNA, 7,3 kb	Norwegian Institute of Public Health, Oslo, Norway	izolováno z původního vzorku (supernatant stolice)
<i>Rotavirus</i>	<i>Rotavirus A</i>	Kmen: G1-4 a P1A[8]	dsRNA (11 segmentů), 18,5 kb	vakcína „RotaTeq“	izolováno z vakcíny
<i>Simplexvirus</i>	<i>Herpes simplex virus 1</i> (HSV-1)		dsDNA, 152 kb	sbírka oddělení virologie, FN Motol, Praha, ČR	izolováno z původního vzorku (aspirát)
<i>Morbillivirus</i>	<i>Virus spalniček</i> ( <i>Measles virus</i> , MeV)		- ssRNA, 15,9 kb	Státní zdravotní ústav v Praze, ČR	izolováno z kultury VERO buněk (celá kultura)

<i>Alphainfluenza</i>	<i>Influenza A virus</i>	Kmen: H3N2	- ssRNA (8 segmentů), 13,5 kb	Státní zdravotní ústav v Praze, ČR	izolováno z kultury MDCK buněk (Madine Darby Canine kidney)
<i>Orthopneumovirus</i>	<i>Respirační syncytiální virus (Respiratory syncytial virus, RSV)</i>	Typ B	- ssRNA, 15,2 kb	Státní zdravotní ústav v Praze, ČR	izolováno z kultury buněk L132 (odvozené z plic lidského embrya, celá kultura)
<i>Betapolyomavirus</i>	<i>BK virus (BKV)</i>		dsDNA, 5,6 kb	sbírka oddělení virologie, FN Motol, Praha, ČR	izolováno z původního vzorku (moč)

**Tabulka 1:** Přehled virů použitých pro přípravu simulovaných infekčních vzorků: jejich taxonomie, typ, délka a organizace genomu, odkud jsme virus získali a z čeho byl izolován



Číslo vzorku	Simulovaná infekce	Vzorek, ze kterého by se běžně infekce testovala (negativní vzorek, do kterého jsme namíchali testované patogeny)	Patogeny ve vzorku	Kvantita viru (Ct*) po namíchání směsi a počet kopií	Kontaminující lidská DNA (kvantita v Ct*) po namíchání směsi**	Kontaminující bakteriální DNA (kvantita v Ct*) po namíchání směsi***
1	Průjmové onemocnění způsobené běžnými viry, gastrointestinální trakt	Vzorek stolice	E30	15 (3*10 <sup>7</sup> kopií/μl)	nebylo detekováno	29 (10 kopií/μl)
			HPeV-1	20 (1,1*10 <sup>6</sup> kopií/μl)		
			<i>Rotavirus A</i> (zahrnující kmeny: G1 – 4, P1A[8])	22		
2	Infekce horních cest dýchacích virem z čeledi <i>Herpesviridae</i>	Vzorek aspirátu	HSV-1	16 (7*10 <sup>5</sup> kopií/μl)	23 (1800 kopií/μl)	32 (1,7 kopií/μl)
3	Infekce dýchacích cest virem chřipky typu A a lidským respiračním syncytiálním virem B (běžnými respiračními viry)	Vzorek aspirátu	H3N2	19,6	30 (40 kopií/μl)	36 (0,15 kopií/μl)
			RSV B	28,6		
4	Infekce virem spalniček	Vzorek moči	MeV	24	37 (1 kopie/μl)	36 (0,15 kopií/μl)
5	Infekce BK virem	Moč	BKV	15 (1,3*10 <sup>6</sup> kopií)	36 (1,7 kopií/μl)	35 (0,3 kopií/μl)
6	Vzorek použit k testování sekvenování různých typů genomů	Směs DNA/ RNA vybraných patogenů ve vodě	E30	31 (200 kopií/μl)	neměřeno	neměřeno
			<i>Pseudomonas aeruginosa</i>	30,5 (3400 kopií/μl)		
			<i>Candida albicans</i>	1700 kopií/μl		
7 (vzorek 1 ředěný 100x)	Průjmové onemocnění způsobené běžnými viry, gastrointestinální (GI) trakt	Vzorek stolice	E30	neměřeno	neměřeno	neměřeno
			HPeV-1			
			<i>Rotavirus A</i> (zahrnující kmeny: G1 – 4, P1A)			

8 (vzorek 3 ředěný 100x)	Infekce dýchacích cest virem chřipky typu A a lidským respiračním syncytiálním virem B (běžnými respiračními viry)	Vzorek aspirátu	H3N2	neměřeno	neměřeno	neměřeno
			RSV B			
9 (vzorek 5 ředěný 100x)	Infekce BK virem	Moč	BKV	neměřeno	neměřeno	neměřeno
10 (vzorek 6 ředěný 100x)	Vzorek použit k testování sekvenování různých typů genomů	Směs DNA/ RNA vybraných patogenů ve vodě	E30	neměřeno	neměřeno	neměřeno
			<i>P. aeruginosa</i>			
			<i>C. albicans</i>			
11	Negativní kontrola	Voda		neměřeno	neměřeno	neměřeno

**Tabulka 2:** Vzorky simulující infekční vzorek s hodnotami Ct pro virus, kontaminující lidskou a bakteriální DNA

\* Ct (threshold cycle), prahový cyklus, používá se ke kvantifikaci pomocí PCR, pokud nemáme k dispozici kalibrační křivku (standarty o známé kvantitě)

\*\* detekováno pomocí genu pro albumin. Hodnoty za prahovým cyklem 35 odpovídají méně než 10 kopiím genu

\*\*\* detekováno pomocí genu pro 16SrRNA. Prahový cyklus 30 odpovídá negativnímu vzorku

## 4.2 Příprava vzorků před použitím kitů od Oxford Nanopore

### 4.2.1 Materiál a přístroje

#### Chemikálie:

Název	Výrobce
DEPC voda	Ambion, USA
AMPure XP	Beckman Coulter, USA
GelRed Nucleic Acid Stain 10 000	Biotinum, USA
Rotiphorese 10× TBE pufr	Carl Roth, Německo
injekční voda	Fresenius kabi, Německo
GeneRuler 100 bp DNA Ladder	Life Technologies, USA
Seakem LE Agarose	Lonza, Švýcarsko
ethanol absolutní p.a.	Penta Chemicals, ČR
bromfenolová modř *	Sigma-Aldrich, USA
sacharóza *	Sigma-Aldrich, USA
xylen-cyanol *	Sigma-Aldrich, USA
ProClin*	Sigma-Aldrich, USA

\*složky nanášecího pufru pro elektroforézu (1,3 g bromfenolové modře, 1,3 g xylen-cyanolu, 150 µ ProClinu, 200 g sacharózy, 250 ml destilované vody)

#### Materiál:

Název	Výrobce
Biosphere Filter Tips	Sarstedt, USA
mikrozskumavky 1,5 ml a 2,0 ml	Sarstedt, USA
Multiply - uStrip Pro 8-strip	Sarstedt, USA

#### Komerční kity:

Název	Výrobce
High Sensitivity DNA Kit	Agilent Technologies, USA
Qubit dsDNA HS Assay Kit	Invitrogen, USA
RNase Coctail Enzym Mix	Invitrogen, USA
DNA Polymerase I Large (Klenow) Fragment	Promega, USA
ImProm-II Reverse Transriptase	Promega, USA
RQ1 RNase-Free DNase	Promega, USA
HotStarTaq polymerase kit	Qiagen, Německo
QIAmp Viral RNA Mini ki	Qiagen, Německo
QuatiTect Probe RT-PCR kit	Qiagen, Německo
AmpliTaqGold DNA Polymerase	Thermo Fischer Scientific, USA
Exonuclease I	Thermo Fischer Scientific, USA
FastAP Thermosensitive Alkaline Phosphatase	Thermo Fischer Scientific, USA

## Přístroje:

Název	Výrobce
Bioanalyzer 2100	Agilent Technologies, USA
elektroforetická vana Midigel XL	Apelex, Francie
ABI 7300 Realtime PCR System	Applied Biosystems, USA
zdroj Power Pac 200	Bio-Rad, USA
S220 ultrasonikátor	Covaris, USA
centrifuga MiniSpin Plus	Eppendorf, Německo
pipeta Research Pro	Eppendorf, Německo
centrifuga Universal 320R	Hettich, Německo
DynaMag - 96 postranní magnet	Invitrogen, USA
Qubit 2.0 Fluorometer	Invitrogen, USA
robot QIAcube	Qiagen, Německo
Vortex Genie 2T	Scientific Industries, USA
LabCycler	Sensquest, Německo
pipeta Finnpiptette, F2	Thermo Fischer Scientific, USA

### 4.2.2 Izolace nukleových kyselin, ověření kvantity virů a smíchání vzorků

Většina vzorků nám byla předána již ve formě vyizolovaných nukleových kyselin. To, na jakém pracovišti byla izolace provedena je shrnuto v **tabulce 3**. Pro účely této studie jsem izolovala pouze tzv. „negativní vzorky“, určené k naředění směsí a vzorek rotavirů z vakcíny RotaTeq.

Virus (či jiné)	Kde	Použitý kit
EV (E30)	Norwegian Institute of Public Health, Oslo, Norway	Vzorek poskytnut pro účely této studie, způsob izolace neznámý
HPeV-1	Norwegian Institute of Public Health, Oslo, Norway	QIAmp Viral RNA Mini kit (Qiagen, Německo), jako bylo popsáno v článku od Cinek et al. (2006)
<i>Rotavirus A</i> - Kmeny G1-4 a P1A[8]	Izolováno nově, pro účely této studie, Laboratoř molekulární genetiky, Praha, ČR	QIAmp Viral RNA Mini kit (Qiagen, Německo)
HSV-1	Oddělení virologie, FN Motol, Praha, ČR	QIAmp Viral RNA Mini kit (Qiagen, Německo)
MeV	Státní zdravotní ústav v Praze, ČR	MagNA Pure Compact Nucleic Acid Isolation Kit I (Roche, Švýcarsko)
Influenza A virus, H3N2	Státní zdravotní ústav v Praze, ČR	MagNA Pure Compact Nucleic Acid Isolation Kit I (Roche, Švýcarsko)

RSV B	Státní zdravotní ústav v Praze, ČR	MagNA Pure Compact Nucleic Acid Isolation Kit I (Roche, Švýcarsko)
BKV	Oddělení virologie, FN Motol, Praha, ČR	QIAmp Viral RNA Mini kit (Qiagen, Německo)
<i>P. aeruginosa</i>	Ústav lékařské mikrobiologie 2. LF UK a FN Motol, Praha, ČR	AMPLICOR Respiratory Specimen Preparation Kit (Roche, Švýcarsko)
<i>C. albicans</i>	Ústav lékařské mikrobiologie 2. LF UK a FN Motol, Praha, ČR	Quick-DNA Fungal/Bacterial Miniprep Kit (Zymo Research, USA)

**Tabulka 3:** Způsob a místo izolace nukleových kyselin z patogenů

### Postup izolace RNA/DNA:

#### 1. Manuální lýze

(vzorek stolice prošel před tímto krokem několikanásobnou centrifugací a filtrací přes 0,22 $\mu$ m filtr a byl odebrán supernatant. Ostatní vzorky nebyly před izolací nijak upravované)

- smíchání 140  $\mu$ l vzorku a 1,4  $\mu$ l nosičové RNA s 560  $\mu$ l AVL (lyzačního) pufru
- 15 sekund vortexování
- inkubace 10 minut při pokojové teplotě

#### 2. izolace v automatickém robotu QIAcube dle protokolu výrobce

Všechny vzorky virů i negativních vzorků (používaných k namíchání směsi virů a simulující vzorek pacienta) byly podrobeny qPCR na všechny testované viry (u negativních vzorků byla ověřena negativita) a nejčastější kontaminující sekvence (lidskou a bakteriální DNA). V případě, že jsme měli více izolátů stejného viru, byl vybrán ten z izolátů, který měl nejvyšší virovou nálož. V případě, že jsme měli na výběr více negativních vzorků byl vybrán ten vzorek, který měl nejnižší podíl kontaminujících sekvencí.

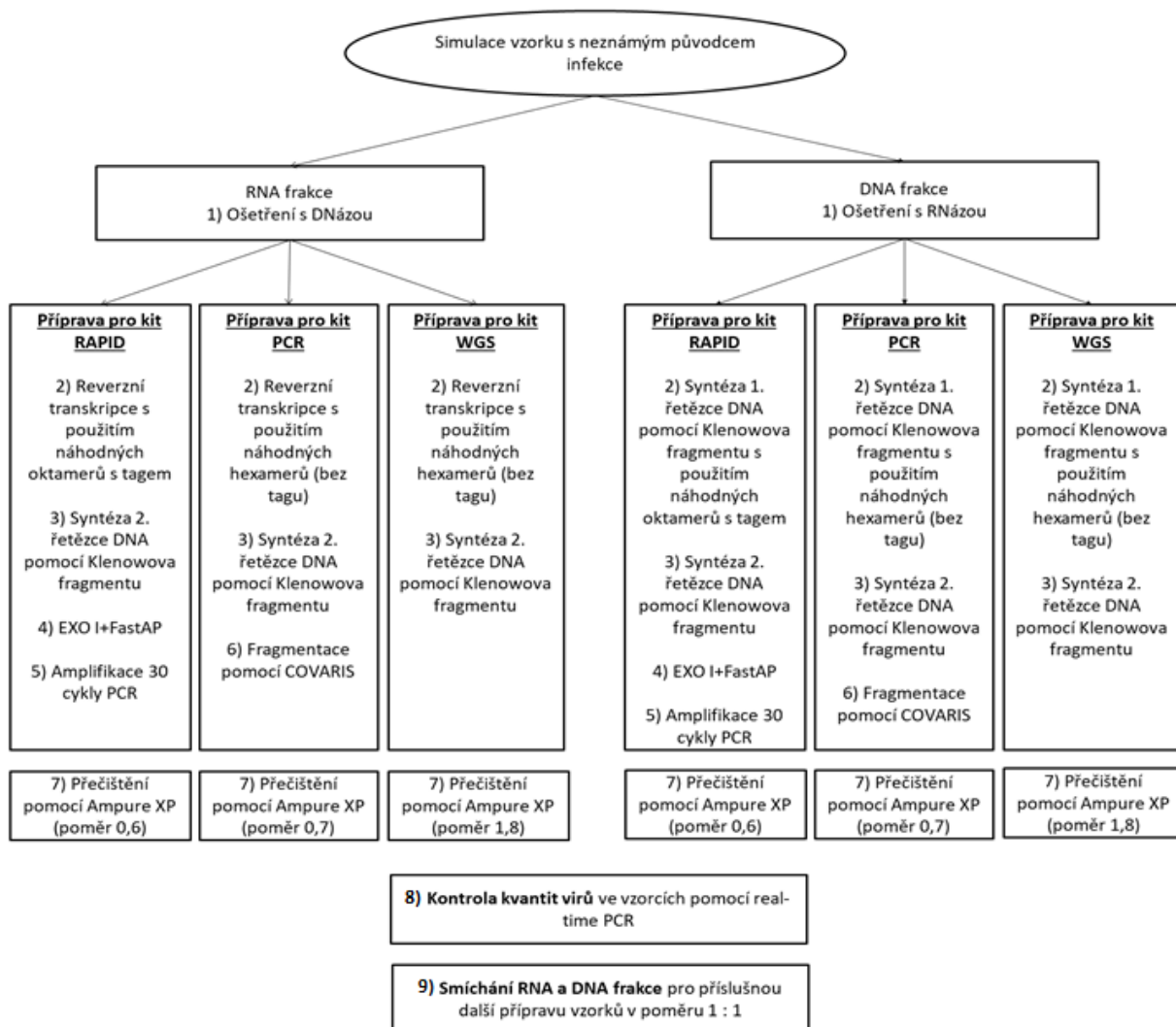
Virus (či jiné)	Použitá chemie	Použitý protokol
EV (E30)	QuatiTect Probe RT-PCR kit	Protokol zavedený v Laboratoři molekulární genetiky, Praha, ČR
HPeV-1	QuatiTect Probe RT-PCR kit	
HSV-1	HotStarTaq DNA Polymerase	
BKV	HotStarTaq DNA Polymerase	
Bakteriální DNA (gen pro 16S rRNA)	HotStarTaq DNA Polymerase	
Lidská DNA (gen pro albumin)	HotStarTaq DNA Polymerase	
<i>Rotavirus A</i> Kmeny G1-4 a P1A[8]	Kvantifikace byla provedena na Oddělení virologie FN Motol, Praha, ČR.	
MeV	V Laboratoři molekulární genetiky nejsou zavedené protokoly, proto byly použity hodnoty naměřené na Státním zdravotnickém ústavu v Praze, ČR.	
<i>Influenza A virus</i> , H3N2		
RSV B		

**Tabulka 4:** Chemie a protokol použitý při kvantifikaci virových a kontaminujících nukleových kyselin

Izolované virové, bakteriální a kvasinkové nukleové kyseliny byly naředěny v příslušném negativním vzorku stolice/ moči/ aspirátu nebo vody tak, abychom získali výsledné kvantitativní množství viru, jaké se při infekcích běžně vyskytuje. Výsledné vzorky jsme podrobili kvantifikaci pomocí PCR a kvantitativní množství jsou uvedeny v **tabulce 2**.

### 4.2.3 Reverzní transkripce, syntéza druhého řetězce DNA, amplifikace/fragmentace, přečištění

Na **obrázku 14** jsou shrnuty všechny kroky přípravy vzorků. Jednotlivé kroky přípravy vzorku, jak jsou popsány na obrázku 14, jsou dále rozepsány.



**Obrázek 14:** Schéma jednotlivých kroků při přípravě vzorků před sekvenováním

Ad1) Nejprve byla směs simulující infekční vzorek rozdělena na dvě poloviny, čímž tak vznikla RNA a DNA frakce. Tento krok byl zařazen proto, že virus v klinickém vzorku, i přes to že může mít vysokou kvantitu, je sekvenačních datech zastoupen většinou málo. Důvodem je to, že kontaminující DNA/RNA pocházející například z bakteriálních či lidských buněk je mnohem delší, což často úplně přehluší virový signál. Tím, že jsme vzorky rozdělili do dvou frakcí lze částečně odstranit tyto kontaminující sekvence pomocí enzymů RNázy a DNázy.

### Ošetření DNázou (RNA frakce)

Reakční směs pro jeden vzorek:

- 1 µl RQ1 RNase-Free DNase
- 1 µl RQ1 RNase-Free DNase 10x reaction buffer
- 8 µl izolované RNA/ DNA

Reakční směs poté byla inkubována 30 minut v termocykleru na 37°C. Po inkubaci byla DNáza inaktivována inkubací s 1 µl RQ1 DNase Stop solution v termocykleru 10 minut na 65°C. Po tomto kroku mohly být vzorky zamraženy a uchovávány při -25°C.

### Ošetření RNázou (DNA frakce)

Reakční směs pro jeden vzorek:

- 1,3 µl RNase Coctail Enzym Mix
- 25 µl izolované RNA/DNA

Ke každé směsi byl přidán koktejl RNáz a spolu se pak inkubovali v termocykleru 1 hodinu při 37 °C a pak hned následovala inaktivace RNáz inkubací v 65 °C po dobu 15 minut.

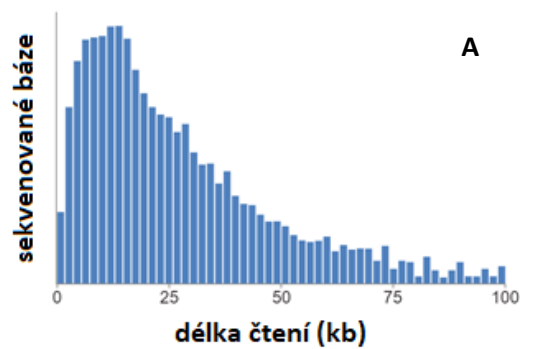
Dále byly obě frakce rozděleny na tři části, u nichž probíhala příprava vzorků různě podle toho, jaký kit pro přípravu k sekvenaci na Oxford Nanopore jsme chtěli použít.



**Základní rozdíl mezi kity byly požadavky na vstupní kvantitu, složitost a časová náročnost v přípravě vzorků a výsledné deklarované délky čtení:**

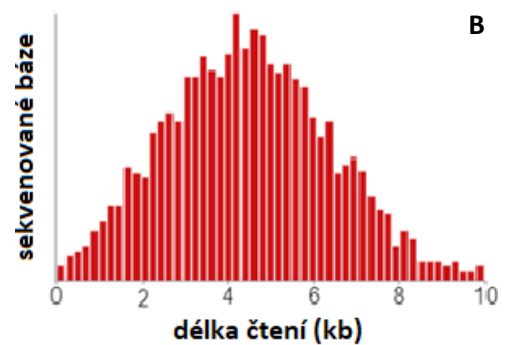
- Rapid Barcoding Sequencing kit (dále jen „Rapid kit“)

- vstupní koncentrace DNA: 54 ng/μl
- doba přípravy: 10 min
- výsledná deklarovaná délka čtení: **obrázek 15A** (v našem případě je ale závislá na délce fragmentů, které připravíme před postupem pomocí kitu „Rapid kit“, což je většinou 200-1000 nukleotidů)
- amplifikaci vzorku je třeba provést před použitím „Rapid kitu“



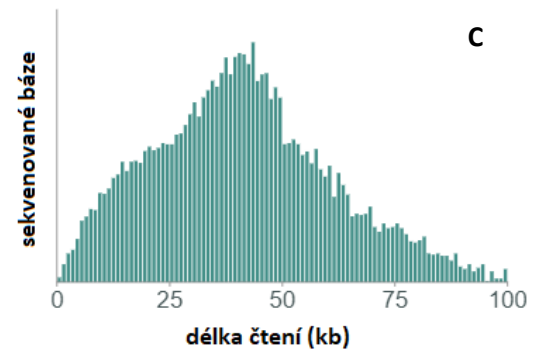
- PCR Barcoding kit (dále jen „PCR kit“)

- vstupní koncentrace DNA: 2 ng/μl
- doba přípravy: 60 min + PCR
- výsledná deklarovaná délka čtení: **obrázek 15B**
- amplifikace vzorku je součástí postupu během přípravy knihovny pomocí „PCR kitu“



- Premium whole genome amplification kit (dále jen „WGS kit“)

- vstupní koncentrace DNA: 2 pg/μl
- doba přípravy: 215 min + 16 hod celogenomová izotermální amplifikace
- výsledná deklarovaná délka čtení: **obrázek 15C**
- amplifikace vzorku je součástí postupu během přípravy knihovny pomocí „WGS kitu“



**Obrázek 15:** Grafy výsledných délek osekvenovaných čtení po přípravě knihovny kity: Rapid Barcoding Sequencing kit (A), PCR Barcoding kit (B), Premium whole genome amplification kit (C) (upraveno podle [www.nanoporetech.com](http://www.nanoporetech.com))

## Ad 2) Reverzní transkripce

Každý vzorek v destičce byl přepsán pomocí reverzní transkripce (RT) do komplementární DNA (cDNA). K tomu byla použita chemie ImProm-II Reverse Transcriptase a dva různé typy primerů:

- náhodné oktamery s „tagem“ (20 nukleotidová sekvence) pro vzorky dále připravované „Rapid kitem“, aby mohly být amplifikovány s použitím primerů komplementárních k sekvenci „tagu“
- náhodné hexamery pro vzorky dále připravované „PCR kitem“ a „WGS kitem“, kdy není nutné vzorky amplifikovat.

Nejprve byl ke každému vzorku přidán 1  $\mu$ l primeru (20 $\mu$ M). Spolu pak byly denaturovány v termocykleru programem:

70°C ... 5 min

4°C ... 2 min

Na ledu pak ke vzorkům bylo přidáno 14,5  $\mu$ l reakční směsi pro RT.

Reakční směs pro jeden vzorek:

- 4,3  $\mu$ l vody
- 4  $\mu$ l 5x reakčního pufru
- 3,2  $\mu$ l MgCl<sub>2</sub> (25 mM)
- 2  $\mu$ l dNTP (5mM)
- 1  $\mu$ l reverzní transkriptázy

Poté byly vzorky inkubovány v termocykleru následujícím programem:

25°C ... 20 minut

42°C ... 1 hod

95°C ... 5 min

4°C ... 5 min

## Ad 2) Syntéza prvního řetězce DNA, kdy je jako templát použita DNA

Pro syntézu prvního řetězce byl použit Klenowův fragment (velký fragment DNA polymerázy I, který má polymerázovou a 3'→5' exonukleázovou aktivitu (Klenow a Henningsen 1970)). Tento krok se zdá být nadbytečný vzhledem k tomu, že pracujeme již s DNA. Nicméně vzorky pro přípravu „Rapid kitem“ je nutné amplifikovat pomocí PCR, abychom dosáhli (alespoň částečně) požadované vstupní kvantity. Proto je potřeba připravit fragmenty, které budou po obou stranách opatřeny sekvencemi „tagů“. Pokud se při přípravě používají bez „tagů“ postupuje se stejným způsobem, aby byly postupy srovnatelné.

Reakční směs pro denaturaci jednoho vzorku:

- 2 µl 10x Klenowův pufr
- 1 µl dNTP, 10mM
- 1 µl primerů
- 10,5 µl vody

Reakční směs byla přidána ke vzorkům a spolu byly denaturovány při 95 °C 5 minut. Poté byly vzorky přeneseny na led, kde k nim byl přidán ředěný Klenowův fragment.

Ředěný Klenowův fragment pro jeden vzorek:

- 0,15 µl 10x Klenowův pufr
- 0,25 µl Klenowův fragment (10U/ml)
- 1,10 µl vody

Reakce poté probíhala v termocykleru následujícím programem:

37°C ... 1 hod

75°C ... 10 min

4°C ... ∞

## Ad 3) Syntéza druhého řetězce DNA (cDNA)

Před syntézou bylo ke všem vzorkům přidáno 0,25 µl 20 µM primerů (náhodné oktamery s tagem nebo náhodné hexamery podle předchozího kroku). Poté byly vzorky připraveny stejným způsobem jako při syntéze prvního řetězce DNA pomocí Klenowova fragmentu (viz. předchozí krok).

#### Ad 4) Exonukleáza I a alkalická fosfatáza

Po reverzní transkripci a syntéze komplementárních řetězců mohou ve vzorcích přebývat nevyužité primery a deoxynukleosidtrifosfáty (dNTP), které inhibují následnou PCR, takže je nutné je před další reakcí odstranit. Exonukleáza I odstraňuje jednovláknovou DNA (single stranded DNA, ssDNA) tak, že napadne 3'-hydroxolovou skupinu ssDNA a odštěpuje z ní 5'-mononukleotidy (Lehman a Nussbaum 1964). Přebytečné dNTP jsou poté degradovány defosforylací alkalickou fosfatázou.

Reakční směs pro jeden vzorek:

- 1 µl Exonukleázy I
- 2 µl FastAP
- 10 µl produktu z předchozí reakce.

Reakční směs byla připravená na ledu a poté byla inkubována v termocykleru následujícím programem:

37°C ... 30 min

85°C ... 15 min

4°C ... 5 min

4°C ... ∞

#### Ad 5) Amplifikace pomocí PCR (pouze pro vzorky dále připravované „Rapid kitem“)

V této fázi byly již připraveny fragmenty dsDNA (z angl. double stranded DNA) opatřených po obou stranách sekvencemi „tagů“. Tyto fragmenty měly různé délky. K amplifikaci byla použita Taq polymeráza a primery komplementární k sekvencím „tagů“.

Reakční směs pro jednu reakci:

- 2,5 µl 10x PCR pufru
- 3 µl MgCl<sub>2</sub>
- 0,25 µl AmpliTaqGold polymerázy
- 0,75 µl dNTP (10mM)
- 1,5 µl primerů (20µM)
- 13,25 µl vody

K reakční směsi pak byly přidány 4 ul vzorku z předchozí reakce a spolu pak byly v termocyklieru inkubovány programem:

iniciální denaturace	95°C ... 5 min
5 cyklů	95°C ... 1 min
	55°C ... 1 min
	72°C ... 1,5 min
25 cyklů	95°C ... 30 sec
	55°C ... 30 sec
	72°C ... 1,5 min (+ 2s/cyklus)
finální syntéza	72°C ... 10 min.

#### Ad 6) Fragmentace pomocí Covaris (pouze pro vzorky dále připravované „PCR kitem“)

Námi vytvořené dvouřetězcové DNA fragmenty byly dále fragmentovány pomocí ultrasonikátoru S220 od firmy Covaris na fragmenty o délce 1500 bází. Je sice deklarováno, že po přípravě vzorků „PCR kitem“ jsme schopni dostat sekvence s mediánem délky 4 kb, takže by se zdál tento krok zbytečný. Naše vzorky (jedná se teď již o směs fragmentů různých délek) a obecně vzorky pro sekvenaci viromu však mají mnohem nižší kvantitu, než je požadovaná. Vzorek je před přípravou samotným „PCR kitem“ ještě přečištěn Ampurem (viz. další krok). Zkušeností Laboratoře molekulární genetiky je, že vzorky mají nižší ztráty, pokud jsou fragmentovány na přibližně stejně dlouhé úseky. Vybraná délka 1500 bází je nejdelší možný fragment pro fragmentaci Covaris.

Nastavení ultrasonikátoru S220 pro fragmentaci na délku 1500 bází:

- typ zkumavky: microTUBE AFA fiber Snap-cap
- objem vzorku: 60 µl
- výkon: 140 W
- délka trvání: 15 s

#### Ad 7) Přečištění pomocí AMPure XP

Před dalšími kroky byly vzorky přečištěny pomocí magnetických kuliček AMPure XP, aby byly zbaveny přebytečných dNTP, primerů a dimerů primerů. Poměr mezi objemem přečišťovaného vzorku a objemem Ampure XP se lišil podle toho, zda chceme pouze odmýt

primery, nebo chceme také odseparovat krátké fragmenty. Bohužel při separaci krátkých fragmentů dochází také k celkové ztrátě DNA vzorku, tento fakt byl také brán na zřetel.

Pro vzorky dále připravované „**Rapid kitem**“ byl poměr mezi množstvím vzorku a kuličkami **1:0,6**.

U vzorků po štěpení na Covarisu pro „**PCR kit**“ byl poměr množství vzorku a kuliček **1:0,7**.

Vzorky dále připravované „**WGS kitem**“ byly přečištěny v poměru vzorku a kuliček **1:1,8**.

Postup:

1. temperace kuliček na pokojovou teplotu, přibližně 30 minut
2. přidání kuliček v příslušném poměru ke vzorku, promíchání pipetováním a inkubace 5 minut při pokojové teplotě
3. přemístění vzorků na magnetickou destičku a vyčkání, dokud není roztok čirý
4. odebrání čirého roztoku, DNA je nyní uchycena na magnetických kuličkách, stále na magnetické destičce
5. přidání 80% ethanolu ke kuličkám a inkubace 30 sekund
6. odebrání ethanolu
7. opakování kroku 5 a 6
8. sušení kuliček při pokojové teplotě 6 minut
9. eluce DNA do vody resuspendováním kuliček v 40  $\mu$ l vody mimo magnet a inkubace 2 minuty
10. přemístění vzorků na magnetickou destičku a vyčkání, dokud není roztok čirý
11. odebrání 35  $\mu$ l eluátu do nových zkumavek

#### Ad 8) Kontrola kvantit

Kvantifikace virů byla provedena pomocí qPCR, hodnoty Ct jsou zaznamenány v **tabulce 5**. K tomu ještě byla změřena koncentrace celkové DNA pomocí přístroje Qubit s kitem High Sensitivity (**tabulka 6**).

Vzorek			Výsledná kvantita virů v jednotlivých frakcích pro přípravu kity:					
			Rapid Sequencing kit (po amplifikaci)		PCR Barcoding kit (bude amplifikováno)		Premium whole genome amplification (bude amplifikováno)	
			RNA frakce (Ct)	DNA frakce (Ct)	RNA frakce (Ct)	DNA frakce (Ct)	RNA frakce (Ct)	DNA frakce (Ct)
1	Infekce viry GI traktu	EV (E30)	12	36	22	nedetek.	21	nedetek.
		HPeV	18	nedetek.	29	nedetek.	28	nedetek.
		Směs* rotavirů	21	neměřeno	30	neměřeno	neměřeno	neměřeno
2	Infekce horních cest dýchacích virem z čeledi <i>Herpesviridae</i>	HSV 1	nedetek.	17	nedetek.	20	32	20
3	Infekce dýchacích cest běžnými respiračními viru	H3N2*	12,5	neměřeno	26	neměřeno	neměřeno	neměřeno
		RSV B*	33,5	neměřeno	33,5	neměřeno	neměřeno	neměřeno
4	Virus spalniček v moči	MeV*	17,7	neměřeno	27,8	neměřeno	neměřeno	neměřeno
5	BK virus v moči	BKV	25	14	38	25	33	25
6	Směs EV, <i>P. aeruginosa</i> a <i>C. albicans</i>	EV	24	nedetek.	nedetek.	nedetek.	37	nedetek.
7 (ředěný vzorek 1)	Infekce viry GI traktu	EV	19	nedetek.	31	nedetek.	31	nedetek.
		HPeV	26	nedetek.	43	nedetek.	37	nedetek.
		Směs* rotavirů	neměřeno	neměřeno	neměřeno	neměřeno	neměřeno	neměřeno
8 (ředěný vzorek 3)	Infekce dýchacích cest (běžnými respiračními viru)	H3N2*	neměřeno	neměřeno	neměřeno	neměřeno	neměřeno	neměřeno
		RSV B*	neměřeno	neměřeno	neměřeno	neměřeno	neměřeno	neměřeno
9 (ředěný vzorek 5)	BK virus v moči	BKV	nedetek.	17	36	36	36	33
10 (ředěný vzorek 6)	Směs EV, <i>P. aeruginosa</i> a <i>C. albicans</i>	EV	nedetek.	nedetek.	nedetek.	nedetek.	nedetek.	nedetek.
11	Negativní kontrola	voda	neměřeno	neměřeno	neměřeno	neměřeno	neměřeno	neměřeno

**Tabulka 5:** Hodnoty Ct virů v jednotlivých směsích před přípravou knihovny pomocí kitů od ON

\* viry nebyly měřeny u všech připravených vzorků, protože v naší laboratoři nedisponujeme příslušnými protokoly a chemikáliemi. Viry byly detekovány jen v některých vzorcích, a to na Oddělení virologie, FN Motol. (nedetek. = nedetekováno)

Vzorek	Rapid Sequencing kit (po amplifikaci)		PCR Barcoding kit (bude amplifikováno)		Premium whole genome amplification (bude amplifikováno)	
	RNA frakce [ng/μl]	DNA frakce [ng/μl]	RNA frakce [ng/μl]	DNA frakce [ng/μl]	RNA frakce [ng/μl]	DNA frakce [ng/μl]
1	0,9	9,1	<0,0005	<0,0005	<0,0005	<0,0005
2	2,7	2,3	2,3	0,9	2,7	0,8
3	5,4	5,2	<0,0005	<0,0005	<0,0005	<0,0005
4	4,9	5,2	<0,0005	<0,0005	<0,0005	<0,0005
5	0,4	2,2	0,1	<0,0005	0,1	
6	<0,0005	0,7	<0,0005	<0,0005	<0,0005	<0,0005
7 (ředěný vzorek 1)	<0,0005	0,2	<0,0005	<0,0005	<0,0005	<0,0005
8 (ředěný vzorek 3)	0,1	1,5	<0,0005	<0,0005	<0,0005	<0,0005
9 (ředěný vzorek 5)	<0,0005	0,1	<0,0005	<0,0005	<0,0005	<0,0005
10(ředěný vzorek 6)	0,03	0,05	<0,0005	<0,0005	<0,0005	<0,0005
11 (voda)	<0,0005	<0,0005	<0,0005	<0,0005	<0,0005	<0,0005

**Tabulka 6:** Koncentrace celkové DNA v jednotlivých směsích před přípravou knihovny pomocí kitů od ON

Naměřené celkové kvantitativy rozhodně nedosahovaly požadovaných hodnot. Tyto kity jsou však přednostně dělané pro lidskou DNA, PCR fragmenty, viry množené na kultuře a při studiích viromu se běžně s nízkými kvantitativy připravené DNA setkáváme. V této studii jsme testovali praktické využití ON pro vzorky s neznámým obsahem viru, proto jsme použili vzorky s takto nízkými kvantitativy.

#### Ad 9) Smíchání RNA a DNA frakce

Frakce každé směsi byly smíchány dohromady v poměru objemů 1:1. Vzniklo tak 11 směsí, z nichž každá obsahovala RNA i DNA frakci dané směsi.



### 4.3 Metodika příprava knihovny Oxford Nanopore

V této diplomové práci byly testovány tři různé protokoly pro vytvoření knihoven k sekvenování na přístroji MinION. Jednotlivé protokoly se od sebe liší délkou přípravy, délkou sekvenovaných fragmentů a v použití různých způsobů amplifikace (viz. strana 35).

#### 4.3.1 Chemikálie a přístroje

Seznam příslušenství potřebných navíc k přípravě knihovny ON.

##### Komerční kity:

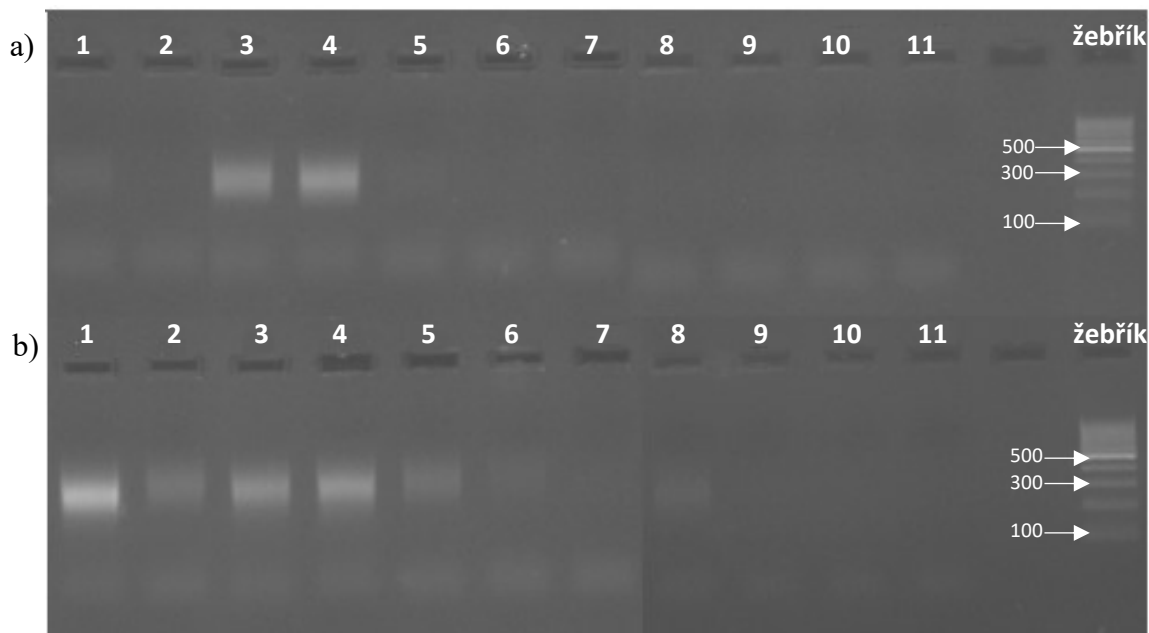
Název	Výrobce
LongAmp Hot Start Taq 2x Master Mix	NewEngland Biolabs, USA
NEB Blunt/TA Ligase Maseter Mix	NewEngland Biolabs, USA
NEBNext FFPE Repair Mix	NewEngland Biolabs, USA
NEBNext Quick Ligation Module	NewEngland Biolabs, USA
NEBNext Ultra II End repair/dA-tailoing Module	NewEngland Biolabs, USA
T7 Endonuclease I	NewEngland Biolabs, USA
Flow Cell Priming Kit (EXP-FLP002)	Oxford Nanopore, Velká Británie
Ligation Sequencing Kit (SQK-LSK109)	Oxford Nanopore, Velká Británie
Native Barcoding Expansion 1-12(EXP-NBD104)	Oxford Nanopore, Velká Británie
PCR Barcoding Kit (SQK-PBK004)	Oxford Nanopore, Velká Británie
Rapid Barcoding Sequencing Kit (SQK-RBK004)	Oxford Nanopore, Velká Británie
REPLI-g Midi Kit	Qiagen, USA

##### Přístroje:

Název	Výrobce
Flow cell R9.4.1	Oxford Nanopore, Velká Británie
MinION Mk1B	Oxford Nanopore, Velká Británie

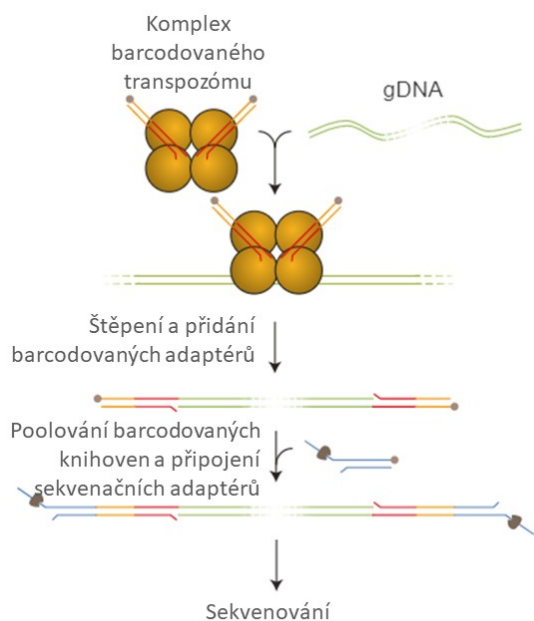
#### 4.3.2 Rapid Sequencing kit

Pro přípravu „Rapid kitem“ máme připraveny náhodné fragmenty všech nukleových kyselin nacházejících se ve vzorku (vzorcích). Tyto fragmenty jsou na koncích opatřeny sekvencemi „tagů“, pomocí kterých byly amplifikovány. Po odmytí velmi krátkých fragmentů pomocí Ampure magnetických kuliček máme fragmenty dlouhé cca 200-500 bp (**obrázek 16**), o kvantitě 5-100x nižší, než je předepsaná (**tabulka 6**).



**Obrázek 16:** Gelová elektroforéza produktů po amplifikaci 30 cykly před přípravou knihovny „Rapid kitem“, a) RNA frakce, b) DNA frakce, čísla na gelu odpovídají číslům vzorků, jako velikostní marker byl použit 100bp DNA žebřík

Obecný princip přípravy knihovny je shrnut na **obrázku 17** a protokol vysvětlen:



- genomická DNA (gDNA) je fragmentována traspozázou a přitom jsou k fragmentům připojeny barcodes, které umožňují rozlišovat jednotlivé vzorky v poolu
- vzorky byly smíchány do jednoho poolu ve stejném objemovém poměru
- nakonec jsou připojeny sekvenční adaptéry, přes které je DNA napojena k motorovému proteinu na povrchu póru ve „flowcell“, který při sekvenování umožňuje průchod vlákna nanoporem

**Obrázek 17:** Schéma přípravy knihovny „Rapid kitem“

### Fragmentace a připojení barcodů

K 7  $\mu$ l amplifikované a přečištěné směsi bylo připipetováno 2,5  $\mu$ l fragmentačního mixu, který obsahoval i barcodes pro označení a následnou identifikaci vzorku. Následovala inkubace v termocykleru s programem:

30°C ... 1 min

80°C ... 1 min

### Poolování vzorků

Po inkubaci byly vzorky zchlazeny na ledu a poolovány – od každého vzorku přišlo do poolu 7  $\mu$ l.

### Přečištění pomocí Ampure XP

Následně byl „pool“ vzorků přečištěn kuličkami AMPure XP v poměru 1:1 shodně s postupem na straně 40. Peleta byla resuspendována v 14  $\mu$ l vody mimo magnet a inkubována při pokojové teplotě 2 minuty. Po inkubaci byl pool vrácen na magnet a do nové zkumavky bylo odebráno 12  $\mu$ l supernatantu pro další kroky.

### Připojení adaptérů

K 10  $\mu$ l připravené knihovny byl nakonec přidán 1  $\mu$ l „Rapid adapteru“ a spolu se tak inkubovali 5 minut v pokojové teplotě.

### Příprava „flowcellly“:

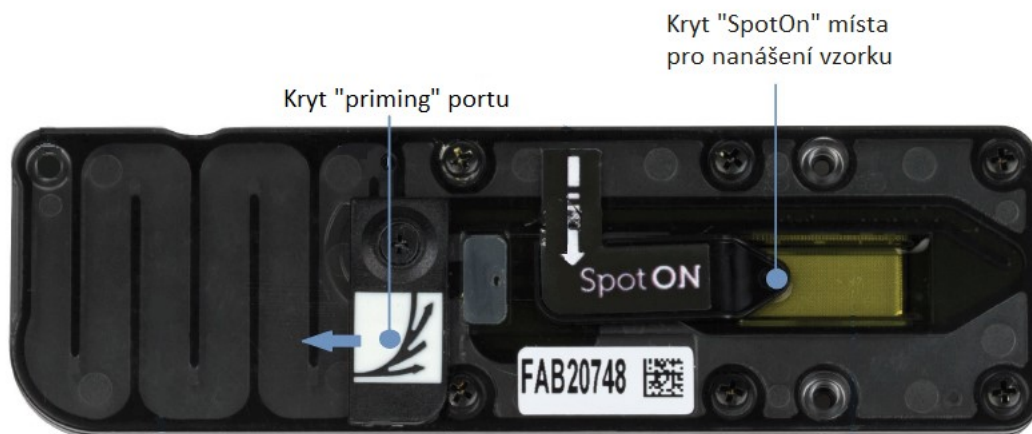
Flowcella R9.4.1 (**obrázek 18**) byla vložena do sekvenátoru Minion Mk 1B. Z „priming“ portu bylo odsáto 200  $\mu$ l pufu tak, aby se do „flowcellly“ nedostala žádná bublina. Poté bylo do něj napipetováno 800  $\mu$ l „priming“ mixu, který byl připraven z 30  $\mu$ l „Flush tetheru“ a jedné alikvóty „flush“ pufu. Po 5 minutách bylo ještě do „priming“ portu napipetováno ještě 200  $\mu$ l „priming“ mixu.

V nové zkumavce byla připravena knihovna, která obsahovala:

- 34  $\mu$ l sekvenačního pufu
- 25,5  $\mu$ l nanášecích kuliček
- 4,5  $\mu$ l vody
- 11  $\mu$ l DNA knihovny

75  $\mu$ l knihovny pak mohlo být po kapkách nanášeno na „flowcellu“ v otevřeném „SpotOn“ místě. Poté mohl být „SpotOn“ a „priming“ port uzavřen a nakonec i víko sekvenátoru (**obrázek 18**).

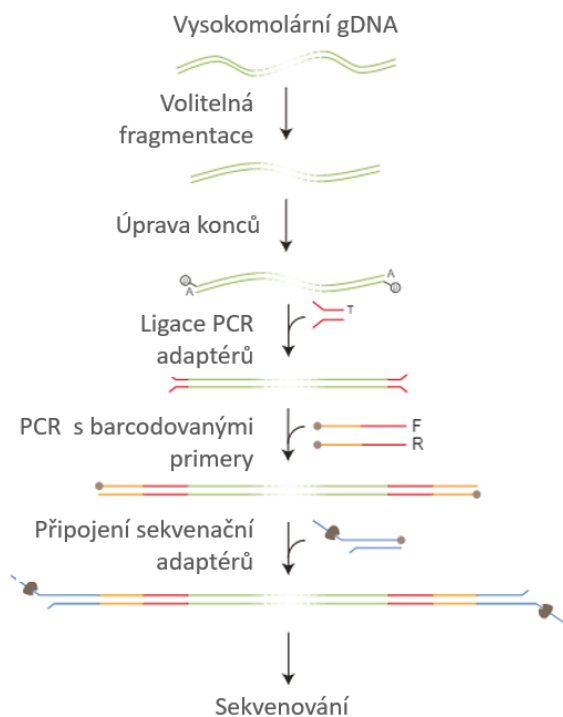
Po připojení sekvenátoru k počítači byla spuštěna sekvenace a probíhala 48 hodin.



**Obrázek 18:** Flowcella do sekvenátoru MinION s vyznačenými místy pro kryt, pod kterým se nachází „priming“ port a kryt, který chrání „SpotON“ místo pro nanášení vzorku (upraveno podle „Oxford Nanopore Technologies“ b.r.)

### 4.3.3 PCR Barcoding kit

Většina vzorků, které byly připraveny pro testování přípravy knihovny pomocí „PCR kitu“ měla koncentraci neměřitelnou pomocí kitu High Sensitivity na přístroji Biolanalyzer (**tabulka 6**). Toto je u směsí virových nukleových kyselin ve vzorku běžné. Namnožení (amplifikace) nukleových kyselin je teprve součástí přípravy knihovny. Vzorky byly fragmentovány na fragmenty dlouhé přibližně 1500 bází a přečištěny pomocí Ampure XP. Následný princip přípravy knihovny je shrnut na **obrázku 19** a popsán vedle v textu.



**Obrázek 19:** Schéma přípravy knihovny „PCR kitem“

- fragmentace pomocí Covarisu na fragmenty o délce 1500 bází (popsáno již v kapitole 4.1.2, strana 29)
- úprava konců po fragmentaci – připojení dAMP (z angl. deoxyadenosine monophosphate) na 3' konec vlákna DNA a fosfátu na 5' konec vlákna, aby mohly být připojeny adaptéry
- ligace adaptérů obsahujících místo pro vazbu primerů
- amplifikace vzorku a přečištění (pro amplifikaci jsou použity primery, které obsahují sekvenci barcodu k rozlišení vzorku)

### Úprava konců DNA a připojení dAMP a přečištění

50 µl fragmentované DNA každé směsi bylo smícháno se 7 µl Ultra II End-prep reakčním pufrům a 3 µl Ultra II End-prep Enzym mixem a společně byly inkubovány v termocykleru programem:

20°C ... 5 min

65°C ... 5 min

Následně byly vzorky přečištěny kuličkami AMPure XP v poměru 1:1 a eluovány ve 20µl vody. Odebráno bylo 16 µl vzorku.

### Ligace adaptérů a přečištění

K 15 µl DNA s upravenými konci bylo přidáno 25 µl Blunt/TA Ligase Master mixu a 10 µl „Barcode adapter“, společně inkubováno 10 minut při pokojové teplotě.

Vzorky byly dále přečištěny pomocí Ampure XP v poměru vzorek : Ampure 1:0,7 k odmytí fragmentů kratších než 200 bp.

Před dalším krokem byla u vzorků změřena koncentrace DNA ds DNA High Sensitivity kitem na Qubitu. U všech vzorků kromě vzorku 2, u kterého byla naměřená koncentrace 0,792 ng/µl, byla koncentrace neměřitelná.

### Amplifikace a připojení barcodů

Pro amplifikaci byla použita chemie LongAmp Hot Start Taq 2x Master Mix a primery označené barcody. Koncentrace požadovaná pro tuto reakci byla 0,2 ng/µl, proto musel být vzorek 2 naředěn (k 12,6 µl vzorku bylo přidáno 11,4 µl vody). Ostatní vzorky nebyly ředěny. Reakční směs pro jeden vzorek obsahovala:

- 25 µl LongAmp Hot Start Taq 2x Master Mixu
- 1 µl primerů
- 24 µl DNA

Reakční směs byla vložena do termocykleru s následujícím programem:

95 °C ... 3 min

95 °C ... 15 sec

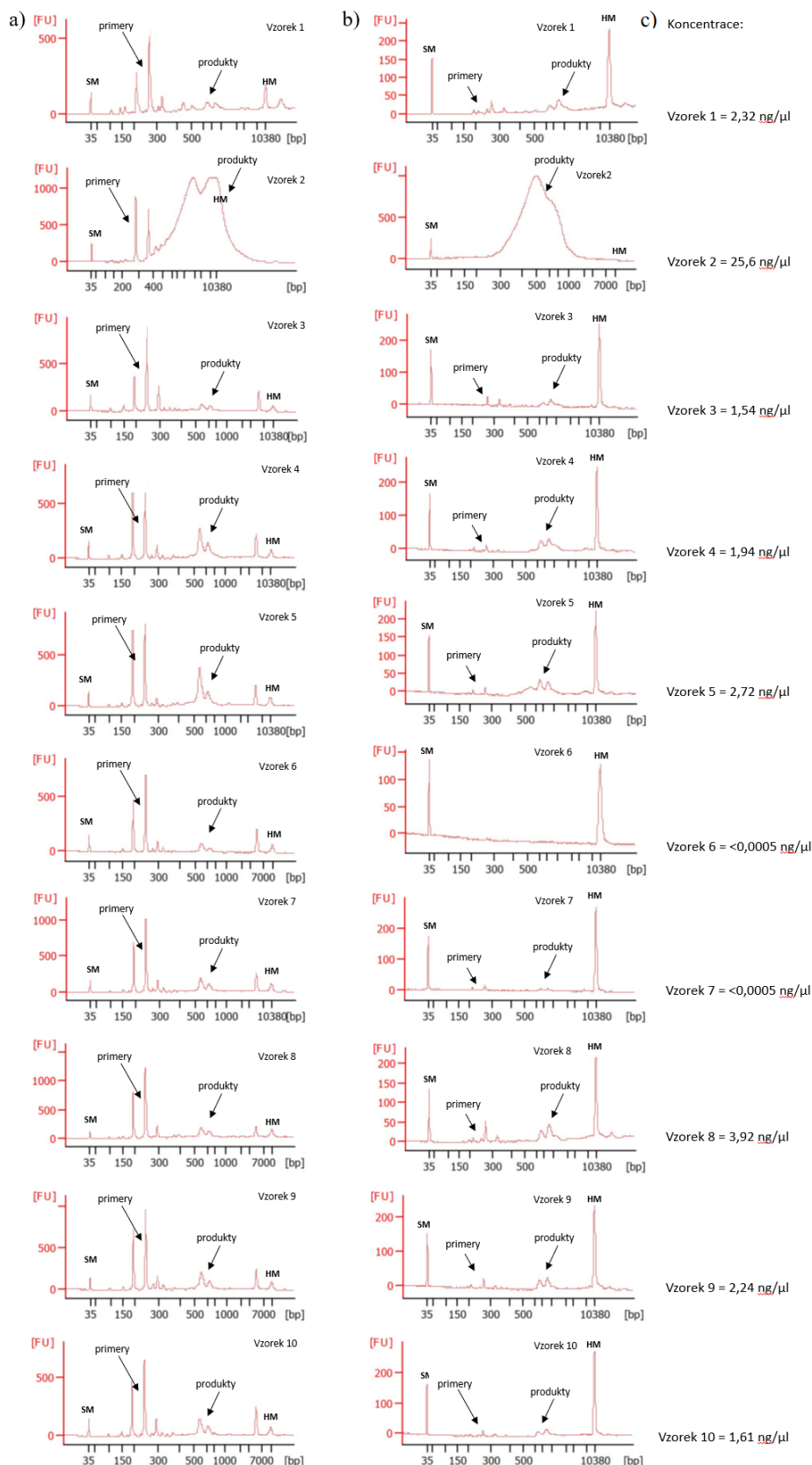
56 °C ... 15 sec      30 cyklů

65 °C ... 50 sec

65 °C ... 6 min

4°C ... ∞

Po amplifikaci byla zkontrolována délka a koncentrace fragmentů na přístroji Agilent 2100 Bioanalyzer Systém s čipem a chemií High Sensitivity DNA (**obrázek 20a**).



**Obrázek 20:** a) Elektroforeogramy z přístroje Bioanalyzer po amplifikaci vzorků 30 cykly, b) Elektroforeogramy z přístroje Bioanalyzer po amplifikaci vzorků 30 cykly a přečištění, c) koncentrace DNA amplifikovaných vzorků na přístroji Qubit, pomocí kitu dsDNA High Sensitivity, po přečištění  
Šípkami jsou vyznačeny primery z reakce a naše produkty, SM = spodní marker (součástí kitu), HM = horní marker (součástí kitu)

Vzorky byly následně přečištěny poměrem vzorek : Ampure 1:0,6. DNA byla eluována ve 13  $\mu$ l vody 2 minuty při pokojové teplotě a poté bylo odebráno 10  $\mu$ l eluátu do nových zkumavek. Vzorky byly následně kontrolovány na přístroji Bioanalyzér pomocí High sensitivity kitu (**obrázek 20b**). Na obrázku lze pozorovat úspěšné odmytí primerů s barcody. Následně byla měřena koncentrace pomocí přístroje Qubit a výsledné koncentrace jsou vyneseny na **obrázku 20c**.

Ačkoli jsme očekávali fragmenty delší než 1500 bazí (požadované délka fragmentace Covaris), produkty jsou kratší. Toto se týká všech vzorků, kromě vzorku 2 s vysokou kvantitou. Vysvětlením může být právě tato kvantita, která pravděpodobně hraje roli také při výsledku fragmentace.

#### Poolování vzorků

Předtím, než byly vzorky poolovány, musel být naředěn vzorek 2, kvůli jeho vysoké koncentraci 1417 fmol/ $\mu$ l. Požadovaná koncentrace poolu byla 5-10 fmol/ $\mu$ l.

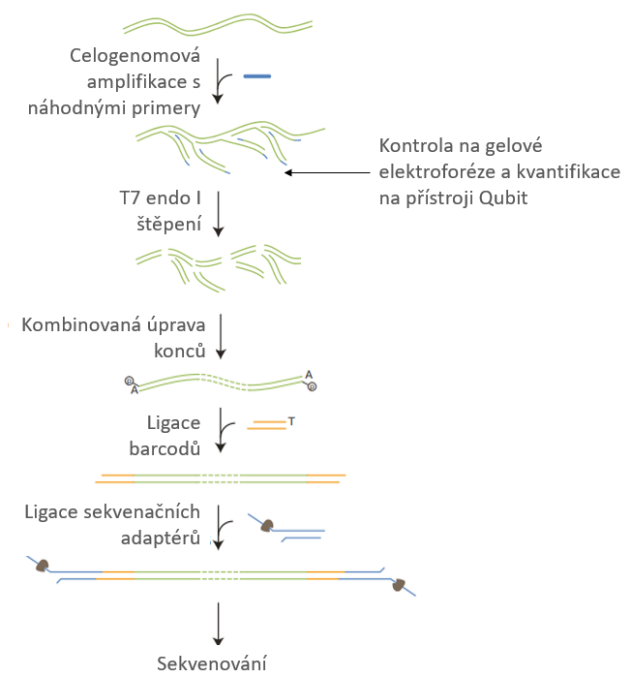
Do poolu pak byly přidány 2  $\mu$ l od každého vzorku. Nakonec byla ještě přeměřena koncentrace celého poolu pomocí přístroje Qubit a dsDNA High Sensitivity kitu. Výsledná koncentrace byla 0,374 ng/ $\mu$ l.

Následovalo připojení adaptérů, příprava „flowcelly“ a samotné sekvenování, které probíhalo stejným způsobem jako u „Rapid kitu“ (strana 45).



#### 4.3.4 Premium whole genome amplification

Vzorky pro přípravu pomocí „WGS kitu“ byly připraveny pomocí náhodných primerů a přečištěny tak, aby došlo k co nejnižším ztrátám (poměr vzorek : Ampure 1:1,8). Tento kit byl vybrán proto, že umí připravit k sekvenaci nukleové kyseliny velmi dlouhé (až 100 kb). Amplifikace vzorku je součástí postupu. Doporučená vstupní koncentrace DNA ve vzorku před izotermální amplifikací byla 2 pg/μl. Proto musel být vzorek 2 o koncentraci 1,84 ng/μl a vzorek 5 o koncentraci 620 pg/μl naředěn.



Obrázek 21: Schéma přípravy knihovny „WGS kitem“

#### Celogenomová izotermální amplifikace

Pro celogenomovou izotermální amplifikaci byla použita chemie REPLI-g Midi kit (Qiagen, USA). Předtím, než byla DNA amplifikována, tak byla inkubována s naředěným DLB pufrem („denturing lysis buffer“) – 9 μl DLB pufru bylo ředěno ve 32 μl vody. V nové zkumavce pak bylo smícháno 5 μl DNA a 5 μl ředěného DLB pufru a spolu se inkubovaly 3 minuty v pokojové teplotě. Po inkubaci byla reakce zastavena 10 μl ředěným „Stop pufrem“, který vznikl smícháním 12 μl „Stop solution“ a 68 μl vody.

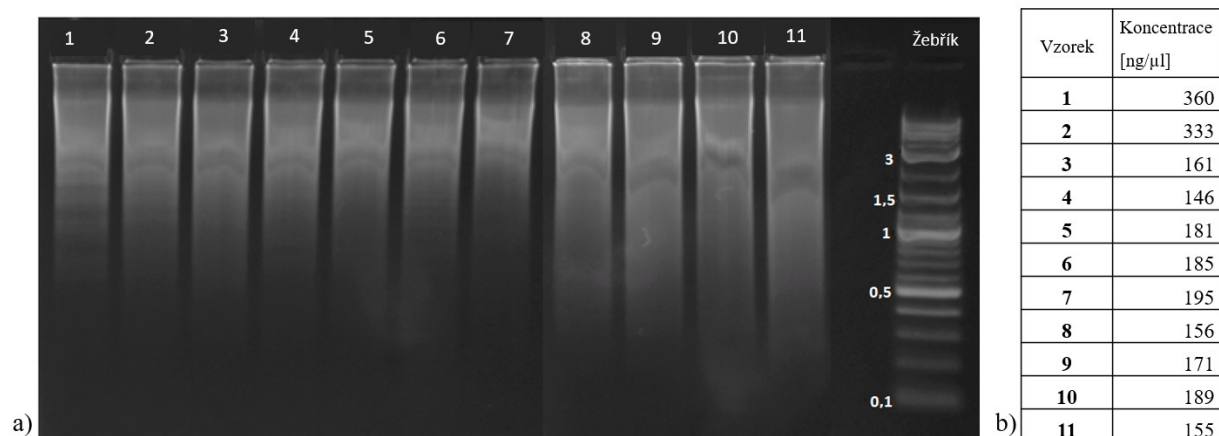
Poté byl připraven polymerázový mastermix smícháním 29 μl REPLI-g Midi reakčního pufru a 1 μl REPLI-g Midi DNA polymerázy. Mastermix o objemu 30 μl byl přidán k DNA reakci a promíchán pipetováním. Amplifikace probíhala v termocykleru s programem:

- celogenomová izotermální amplifikace metodou „multiple displacement amplification (MDA) s náhodnými primery
- kontrola koncentrace a délek DNA
- fragmentace T7 endonukleázou I
- úprava konců po fragmentaci – připojení dAMP na 3' konec vlákna DNA a fosfátu na 5' konec vlákna
- připojení barcodů a následně ligace sekvenačních adaptérů

30 °C ... 16 hod

65 °C ... 3 min

Pro zjištění, zda jsme připravili dostatečné množství DNA (a také pro zjištění délky), byla použita gelová elektroforéza (**Obrázek 22a**). Na 1% agarózový gel byla nanášena amplifikovaná DNA a elektroforéza probíhala s napětím 70 V/cm po dobu 95 minut.



**Obrázek 22:** a) Gelová elektroforéza po izotermální amplifikaci, jako velikostní marker byl použit 1kb DNA žebřík, čísla na gelu odpovídají číslům vzorků, b) Koncentrace DNA měřená po izotermální amplifikaci a přečištění

### Přečištění pomocí Ampure XP

Následně byly vzorky přečištěny kuličkami AMPure XP v poměru 1:1,8. Postup probíhal shodně s postupem na straně 40. Vzorek byl eluován do 100 μl vody a odebráno bylo 95 μl.

Do další reakce bylo zapotřebí celkem 1,5 μg DNA ve 25,5 μl vody (0,06 μg/μl), proto byla u vzorků přeměřena koncentrace DNA na Qubitu s pomocí dsDNA Broad Range kitu a poté naředěna vodou na požadovanou koncentraci.

### Fragmentace T7 endonukleázou I

Ke 25,5 μl DNA ve vodě byl napipetován mix, který obsahoval reagenty:

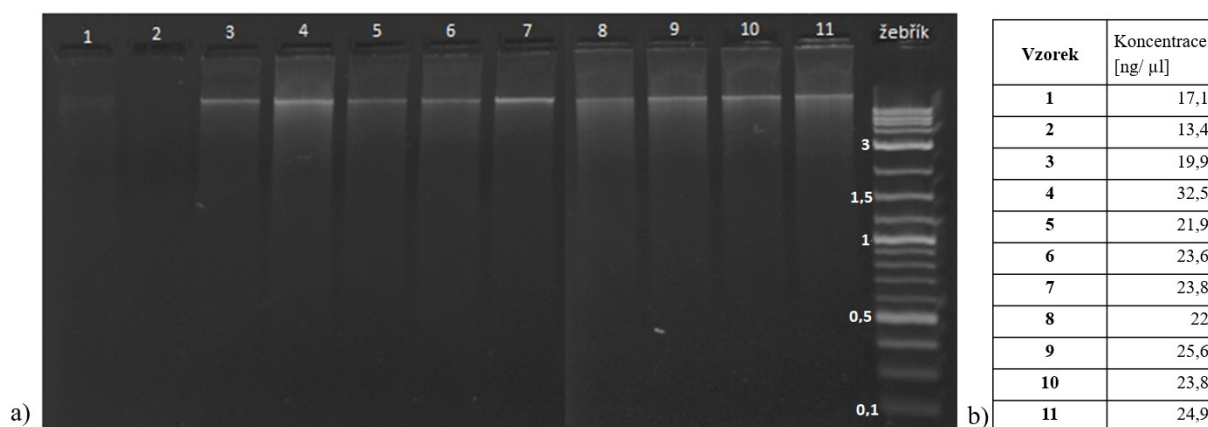
- 3 μl NEBuffer 2
- 1 μl T7 Endonuclease I

a poté následovala inkubace v termocykleru při 37 °C 15 minut. Po inkubaci byla reakce inhibována přidáním 1,5 μl 0,25M EDTA přímo do zkumavek s DNA.

### Přečištění pomocí Ampure XP

Před přečištěním bylo ke vzorků přidáno 20,5  $\mu\text{l}$  vody, aby byl objem vzorků 50  $\mu\text{l}$ . Přečištění probíhalo stejným postupem jako je uvedeno na straně 40, s poměrem vzorek : Ampure 1:0,7 a konečná eluce kuliček byla do 55  $\mu\text{l}$  vody a poté bylo odsáto 53  $\mu\text{l}$  eluátu do nových zkumavek.

Poté byla zkontrolována koncentrace DNA na přístroji Qubit kitem ds DNA High Sensitivity (**obrázek 23b**) a pomocí gelové elektroforézy délka DNA (1% agarózový gel, napětí 70 V/cm, 100 minut, **obrázek 23a**).



**Obrázek 23:** a) Gelová elektroforéza po reakci s T7 endonukleázou I a po přečištění, jako velikostní marker byl použit 1kb DNA žebřík, čísla na gelu odpovídají číslům vzorků, b) Koncentrace DNA měřená po reakci s T7 endonukleázou I a po přečištění

### Úprava konců DNA a připojení dAMP

Požadované množství DNA v dalším kroku bylo 1  $\mu\text{g}$  v 48  $\mu\text{l}$  vody (koncentrace 0,02  $\mu\text{g}/\mu\text{l}$ ). Vzorky byly naředěny vodou.

K naředěným vzorků pak mohla být přidána reakční směs obsahující reagentie NEBNext FFPE DNA Repair Mix a NEBNext End repair / dA-tailing Module, mix pro jeden vzorek:

- 3,5  $\mu\text{l}$  NEBNext FFPE DNA Repair Buffer
- 2  $\mu\text{l}$  NEBNext FFPE DNA Repair Mix
- 3,5  $\mu\text{l}$  Ultra II End-prep reaction buffer
- 3  $\mu\text{l}$  Ultra II End-prep enzyme mix.

Vše bylo promícháno pipetováním a poté probíhala inkubace v termocykleru s programem:

20 °C ... 5 min

65 °C ... 5 min

### Přečištění pomocí Ampure XP

Následovalo přečištění vzorků Ampurem XP, poměr mezi vzorkem a Ampure byl 1:1. Konečná eluce byla do 30  $\mu$ l vody a po inkubaci bylo odebráno 28  $\mu$ l do nových zkumavek.

Po přečištění bylo zapotřebí zkontrolovat koncentraci DNA, protože pro ligaci barcodů bylo požadované množství DNA 500 ng v 22,5  $\mu$ l vody (22,2 ng/ $\mu$ l). Koncentrace DNA byla měřena na Qubitů pomocí kitu dsDNA High Sensitivity a poté byly vzorky naředěny.

### Ligace barcodů

Naředěná DNA mohla být poté smíchána s 25  $\mu$ l Blunt/TA Ligase Master Mixem a 2,5  $\mu$ l Native Barcode. Pro každý vzorek byl použit unikátní barcode. Po smíchání byly vzorky společně inkubovány 10 minut při pokojové teplotě.

Po ligaci byly vzorky opět přečištěny (poměr vzorku a Ampure 1:1, eluce do 30  $\mu$ l vody a poté odebráno 28  $\mu$ l do nových zkumavek) a byla změřena koncentrace DNA na přístroji Qubit, kitem dsDNA High Sensitivity.

### Poolování vzorků

V poolu bylo smícháno ekvimolární množství každého vzorku s vlastním barcode. Celkem bylo v poolu 700 ng DNA a objem byl doplněn vodou do 65  $\mu$ l.

### Připojení adaptérů

Adaptéry byly připojeny reagenциemi NEBNext Quick Ligation Module v reakční směsi:

- 5  $\mu$ l Adapter Mix II
- 20  $\mu$ l NEBNext Quick Ligation Reaction Buffer (5x)
- 10  $\mu$ l Quick T4 DNA Ligase
- 65  $\mu$ l poolu

Reakční směs byla 10 minut inkubována v pokojové teplotě.

### Přečištění pomocí Ampure XP

Nakonec byl pool přečištěn kuličkami AMPure XP v poměru 1:0,5 (pool : Ampure). Pool s kuličkami byl inkubován 5 minut při pokojové teplotě, poté byly dány na magnet a po vyčerení byl odpipetován supernatant. Na magnetu pak byla peleta dvakrát promyta 250  $\mu$ l

Long Fragment pufrem, abychom získaly fragmenty o délce 3 kb a více. Po odstranění pufu a vysušení byla peleta resuspendována a inkubována v termocykleru při 37 °C 10 minut v 18 µl elučního pufu mimo magnet. Po inkubaci byl pool vrácen na magnet a do nové zkumavky bylo odebráno 15 µl supernatantu.

Před sekvenováním byla ještě přeměřena koncentrace poolu na Qubitu pomocí dsDNA High Sensitivity kitu. Výsledná koncentrace poolu byla 11,7 ng/µl

#### Příprava „flowcelly“:

Příprava „flowcelly“ probíhala stejným způsobem jako u „Rapid kitu“ (strana 45).

V nové zkumavce byla připravena knihovna, která obsahovala:

- 37,5 µl sekvenačního pufu
- 25,5 µl nanášecích kuliček
- 12 µl DNA knihovny

75 µl knihovny pak mohlo být po kapkách nanášeno na flowcellu v otevřeném „SpotOn“ místě. Poté mohl být „SpotOn“ a „priming“ port uzavřen a nakonec i víko sekvenátoru.

Po připojení sekvenátoru k počítači mohla být sekvenace spuštěna a probíhala 48 hodin.

## 4.4 Bioinformatická analýza

### 4.4.1 „Basecalling“ a „barcoding“

„Basecalling“ je proces, při kterém dochází k převedení elektrického signálu vytvářenému při procházení vlákna DNA/RNA nanoporem na sekvenci bází daného vlákna. „Basecalling“ může být prováděn různými softwary. V této práci byl využit nástroj „Guppy“, který je integrován přímo v softwaru „MinKNOW“, určeného také k ovládní sekvenátoru a spouštění sekvenace. Po převedení zápisu na sekvence následuje tzv. „barcoding“, který také umožňuje rozdělení sekvencí podle vzorků pomocí barcodů.

### 4.4.2 Identifikace sekvencí

Nejprve jsme chtěli zjistit, jaký je obsah výsledných sekvencí: kolik sekvencí viru najdeme? Jaký bude obsah kontaminujících sekvencí? Vytváří se při sekvenování artefakty a co je jejich obsahem a co je jejich? Co je obsahem dlouhých čtení u „WGS kitu“, která nemohou být virová? K tomu byly použity tři různé postupy:

**VIPIE** (z anglického „Virome Pipeline Extraction“):

nástroj sestavený v Laboratoři molekulární genetiky k identifikaci sekvencí u studií viromu. Tento nástroj umí také identifikovat jiné, kontaminující sekvence. VIPIE využívá k identifikaci celé virové sekvence, získané z databáze NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)). Výhodou je, že pokud je to možné, umí identifikovat sekvenci až do nejnižší taxonomické úrovně, až k nalezení „accession number“ jeho nejbližší referenční sekvence (Lin et al. 2017).

Byl primárně vyvinut k analýze sekvencí z přístroje MiSeq (Illumina), které se velmi liší od námi vyprodukovaných sekvencí: sekvence z MiSequ jsou kratší (2x250 bp) a mají mnohem lepší kvalitu. Proto byl tento nástroj použit u naší studie jen pro nástin výsledků a k identifikaci alespoň trochu blízkých referenčních sekvencí virů. Kvůli horší kvalitě sekvencí z ON jsme očekávali, že VIPIE identifikuje jen málo virových či jiných čtení.

**MG-RAST** (z anglického „Metagenomics Rapid Annotation using Subsystem Technology“):

je veřejně přístupný software umožňující analyzovat metagenomická data po sekvenování (Meyer et al. 2008). Identifikaci sekvencí provádí pomocí několika databází což představuje nespornou výhodu oproti jiným nástrojům:

- proteinové: The SEED (Overbeek et al. 2005), GenBank (Benson et al. 2013), RefSeq (Pruitt, Tatusova, a Maglott 2007), IMG/M (Markowitz et al. 2008), UniProt (Magrane

a Consortium 2011), eggNOG (Jensen et al. 2008), KEGG (Kanehisa 2002), PATRIC (Snyder et al. 2007)

- ribosomální RNA: greengenes (DeSantis et al. 2006) , SILVA (Pruesse et al. 2007), RDP (Cole et al. 2003)

Nevýhodou tohoto nástroje je, že je schopen určit organismus, ze kterého sekvence pochází pouze do úrovně rodu.

### **Software Geneious:**

Pro práci s tímto softwarem jsme využili toho, že víme, které viry byly ve vzorku smíchány a sekvenovány. Také zhruba víme, jaké nejčastější kontaminující sekvence lze očekávat. Proto jsme si nejprve sestavili set referenčních sekvencí, na které jsme nechali namapovat všechna čtení ve vzorku. Tímto způsobem jsme tzv. „vytáhli“ jen ta čtení, která pocházela z daného viru. Výhodou tohoto postupu oproti předchozím dvěma nástrojům je to, že si lze prohlédnout, jak vypadají namapovaná čtení. Zejména, pokud se jedná jen o nepravidelně pokrytý genom, nebo o mapování artefaktů.

Při práci se softwarem Geneious jsme postupovali takto:

- 1) Nejprve jsme se chtěli podívat na **obsah virových čtení v sekvencích**. Díky znalosti, které viry se nachází ve vzorku, jsme sestavili set referenčních sekvencí. Byly to náhodně vybrané referenční sekvence virů se shodnou taxonomií jako sekvenované viry, které pocházely z databáze GenBank ([www.ncbi.nlm.nih.gov/genbank/](http://www.ncbi.nlm.nih.gov/genbank/)). Zařazeny byly také nejbližší referenční sekvence identifikované nástrojem VIPIE. Pro každý virus jsme měli k dispozici cca 10 referencí.
  - a) Každá směs virů a kontaminujících sekvencí byla připravena třemi různými způsoby: „Rapid kitem“, „PCR kitem“ a „WGS kitem“. Všechny virová čtení ze všech tří způsobů přípravy byly namapovány na všechny dostupné referenční sekvence. Poté jsme **vybrali jednu referenci**, na kterou se namapoval největší počet čtení ze všech tří různých příprav vzorku (**tabulka 7**). Pokud nebyla takováto reference k dispozici (u některých vzorků připravených různými kity byly identifikovány různé nejbližší referenční sekvence), byla vybrána reference, na kterou se namapovalo nejvíce čtení alespoň v jedné z příprav vzorku. Nejčastěji se jednalo o vzorek připravený „Rapid kitem“, kde byl identifikován největší počet virových čtení. Tuto referenci

považujeme za nejbližší a jelikož se jedná stále o stejný virus připravený různě, měla by fungovat pro všechny tři vzorky. Výběr jen jedné reference pro všechny tři přípravy vzorku byl za tím účelem, abychom mohli výsledky mezi sebou porovnat.

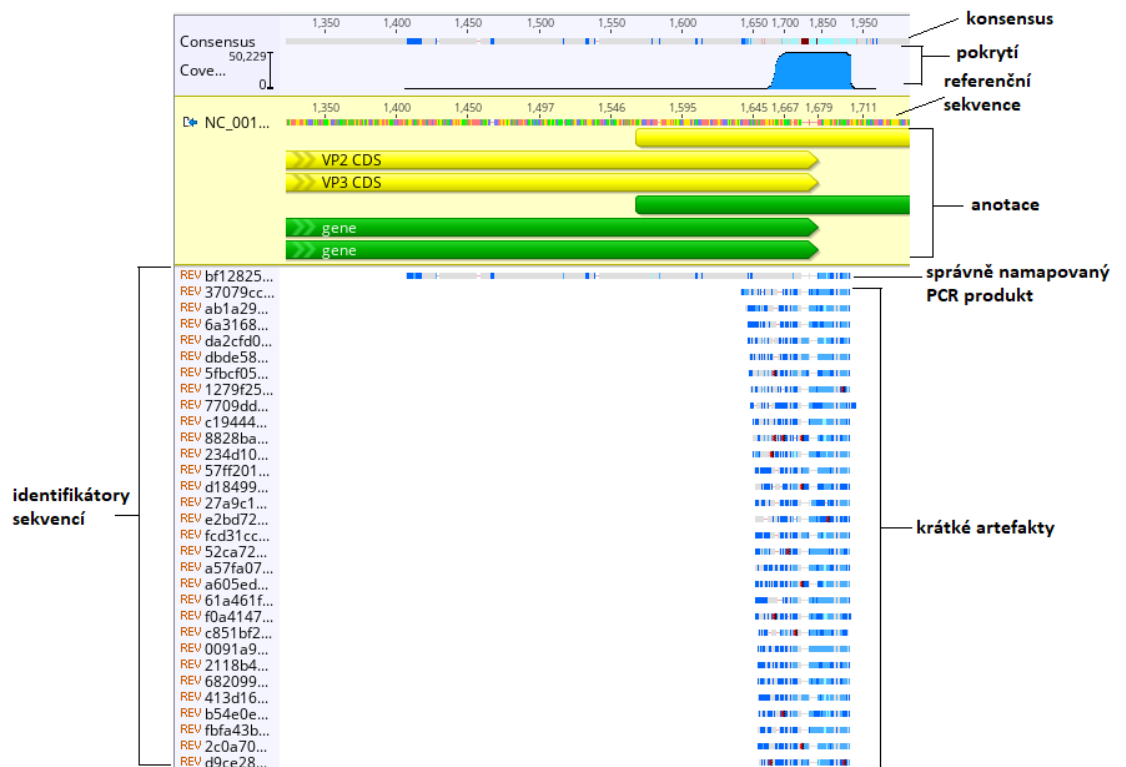
		Reference s největším počtem namapovaných čtení	Počet namapovaných čtení		
			Rapid kit	PCR kit	WGS kit
Vzorek 1	EV (E30)	MF101537	712	17	2
	HPeV-1	GQ183023	10	0	0
	Rotavirus A, kmeny G1-4 a P1A[8]	KF729669	1	0	0
Vzorek 2	HSV-1	GU734771	17	1	1
Vzorek 3	H3N2	MN493833	236	1	0
	RSV B	KP317952	2	0	0
Vzorek 4	MeV	MN893225	3095	249	0
Vzorek 5	BKV	AB211369	285	8	28
Vzorek 6	EV(E30)	MF101537	0	0	0
	<i>C. albicans</i>	NC 0320(89.1-96.1)	17	0	2
	<i>P. aeruginosa</i>	NC 002516	7732	5	77
Vzorek 7	E30	MF101537	58	0	0
	HPeV-1	GQ183023	0	0	0
	Rotavirus A, typ G1-4 a P1A[8]	KF729669	0	0	0
Vzorek 8	H3N2	MN493833	1	0	0
	RSV B	KP317952	0	0	0
Vzorek 9	BKV	AB211369	14	0	0
Vzorek 10	E30	MF101537	0	0	0
	<i>C. albicans</i>	NC 0320(89.1-96.1)	61	0	1
	<i>P. aeruginosa</i>	NC 002516	1193	0	0

**Tabulka 7:** Porovnání počtu nalezených čtení mezi jednotlivými vzorky a jejich způsobem přípravy před sekvenováním

- b) Po prohlídce výsledků mapování v softwaru Geneious bylo zjištěno, že u „PCR kitu“ a „WGS kitu“ existují mezi namapovanými sekvencemi také **bloky velmi krátkých fragmentů** do cca 50 nukleotidů (**obrázek 24**). Tyto sekvence se neshodovaly se sekvencí žádného barcodu, ani adaptéru a jejich původ se nám nepodařilo dohledat. Nicméně, ač se jednalo o cca stejnou sekvenci, mapovala se na více míst a vůbec se neshodovala s vybranou referencí. Zjistili jsme, že důvodem jejich mapování je nastavení opakování mapování 5x („Iterate up to 5 times“). Toto se běžně používá k nalezení co největšího počtu sekvencí. Námí vygenerovaná sekvence však již obsahovala tento fragment a díky opakovanému mapování jsme z dat „vychytili“ i další tyto fragmenty. Muselo tedy dojít k **úpravě mapování pro vzorky**, připravené



oběma zmíněnými kity. Artefaktů u jsme se zbavili trimováním konců sekvencí o 25 nukleotidů.



**Obrázek 24:** Ukázka mapování krátkých čtení bez použití zkracování, které se neshodují s referencí BKV u „PCR kitu“

- c) Dále jsme chtěli provést kontrolu, zda jsme opravdu „vychytali“ největší počet čtení daného viru. Proto jsme místo referenční sekvence z databáze NCBI použili jako referenci konsensuální sekvenci (výslednou) ze vzorku připraveným kitem, kde bylo identifikováno nejvíce čtení daného viru. Téměř všude to byly vzorky připravené „Rapid kitem“. **Počet čtení namapovaných na referenci a konsensus z „Rapid kitu“ byl srovnatelný.**
- d) Posledním problémem zjištěným během mapování byly výsledky mapování u vzorků, připravených „WGS kitem“. Většina čtení připravených tímto kitem byla delší než 10 kb, což u některých virů přesahuje délku celého genomu. Nejednalo se však jen o sekvence lidské či bakteriální, ale také **sekvence virové, které se několikanásobně opakovaly**. Tato čtení se namapovala jedním z opakování virové sekvence na referenci a další opakování bylo označeno jako nekomplementární sekvence.

2) Dále jsme zjišťovali, jaké je množství a obsah ostatních (kontaminujících) sekvencí. K tomu byla použita čtení, které se nenamapovaly k virovým referencím. Nejprve probíhalo mapování k referenci lidského genomu a poté byla další zbylá čtení mapována k referencím bakteriálních genomů. Ty byly vybrány na základě výsledků z nástroje MG-RAST a také podle typu vzorku (při sekvenování vzorku stolice byly vybrány bakterie, které jsou běžnou součástí střevní flóry). Nelze mapovat všechny sekvence ve vzorku na všechny možné typy bakteriálních genomů běžně se ve vzorku vyskytující. Mapování bylo použito zejména proto, abychom odhalili skryté artefakty aj. Bakteriální a lidské referenční sekvence jsou shrnuty v **tabulce 8**. Reference bakteriálních genomů byly vybírány podle výsledků z analýzy v nástroji MG-RASTu, pro každý vzorek byly vybrány 1-3 nejvíce zastoupené bakterie.

- a) Také při hledání lidských čtení jsme museli vyzkoušet více parametrů mapování. Co se týče trimmování 3' a 5' konců o 25 nukleotidů, tak bylo ponecháno stejné nastavení, jako při hledání virových sekvencí (u vzorků „Rapid kitu“ bez zkracování, vzorky „PCR kitu“ a „WGS kitu“ se zkracováním). Dalšími důležitými parametry, kde byly provedeny změny, bylo opakování a senzitivita mapování. Čím nižší senzitivita je, tím je mapování rychlejší a namapuje se méně čtení, ale s větší přesností. Nejlepšího výsledku mapování bylo dosaženo při středně nízké senzitivitě a bez opakování (**tabulka 8**). Pokud jsme zvýšili opakování mapování, došlo opět k mapování kratších, nesprávných sekvencí, které se neshodovali s referencí. tak se zase mapovaly kratší nesprávné sekvence, které se neshodovali s referencí.

Výsledky byly zpracovány do grafů a tabulek v programu Microsoft Excel.

Vzorek		Reference lidského genomu	Počet namapovaných čtení na lidský genom			Bakterie použité při hledání ostatních sekvencí	Reference bakteriálních genomů	Počet namapovaných čtení na bakteriální genomy		
			Rapid kit	PCR kit	WGS kit			Rapid kit	PCR kit	WGS kit
1	EV (E30)	NC_000001-9 NC_000010-24	24263	4472	4487	<i>Bacteroides vulgatus</i>	NC_009614	411	48	0
	HPeV-1					<i>Escherichia coli</i>	NZ_CP028592			
	Rotavirus, kmeny: G1-G4, P1A[8]									
2	HSV-1		91616	6904	9101	<i>Escherichia coli</i>	NZ_CP028592	0	5	2
3	H3N2		68777	7111	13516	<i>Haemophilus influenzae</i>	NZ_CP009610	34	11	19
	RSV B									
4	MeV		36736	33161	3396	<i>Escherichia coli</i>	NZ_CP028592	0	17	33
5	BKV		1939	1243	6808	<i>Helicobacter cinaedi</i>	AP012280	5	10	14
						<i>Mycobacterium tuberculosis</i>	NC_000962			
						<i>Escherichia coli</i>	NZ_CP028592			
6	EV (E30)	1939	430	5653			0	0	0	
	<i>P. aeruginosa</i>									
	<i>C. albicans</i>									

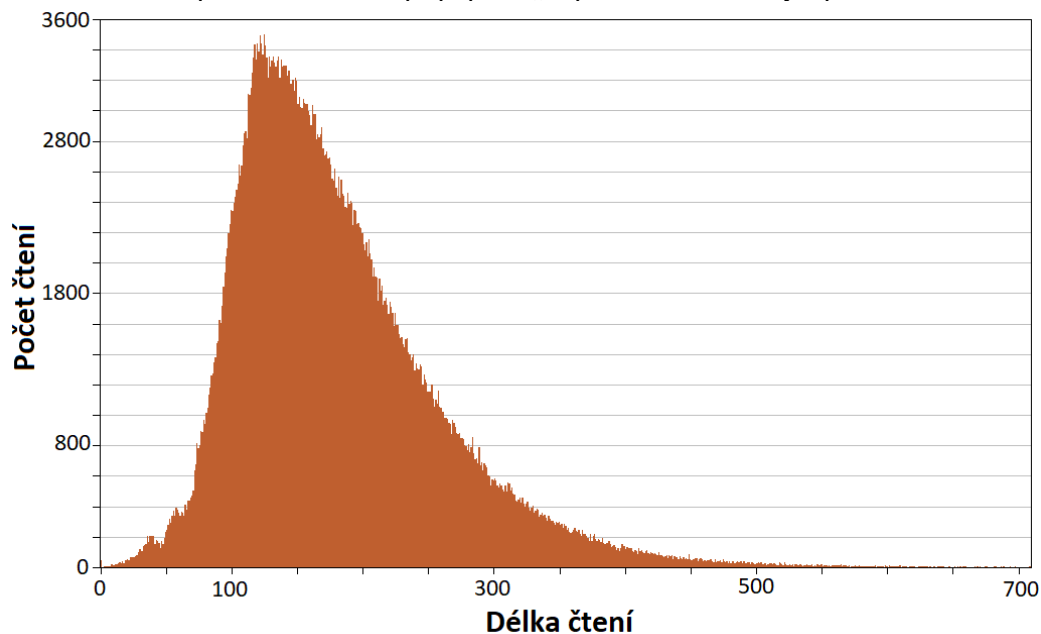
**Tabulka 8:** Počet namapovaných čtení k referenci lidského genomu a referencím bakteriálních genomů

## 5. Výsledky

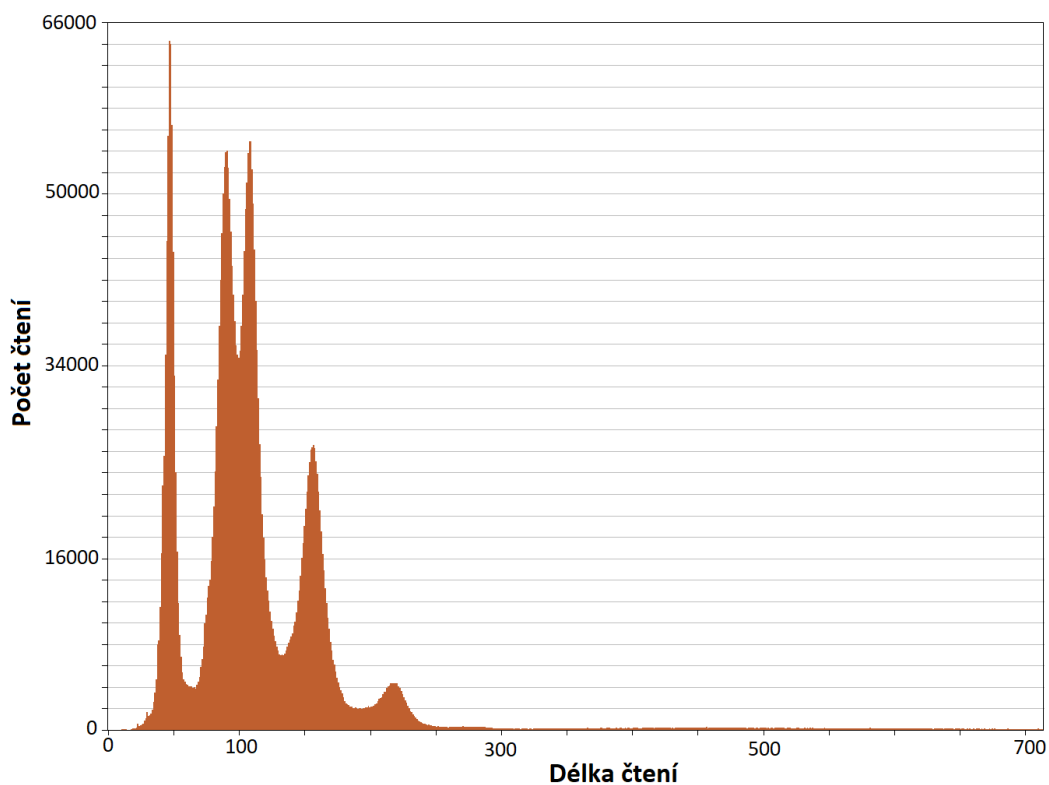
### 5.1 Celkové hodnocení sekvenace

#### 5.1.1 Počty a délky čtení pro tři různé přípravy knihovny – analyzováno softwarem Geneious

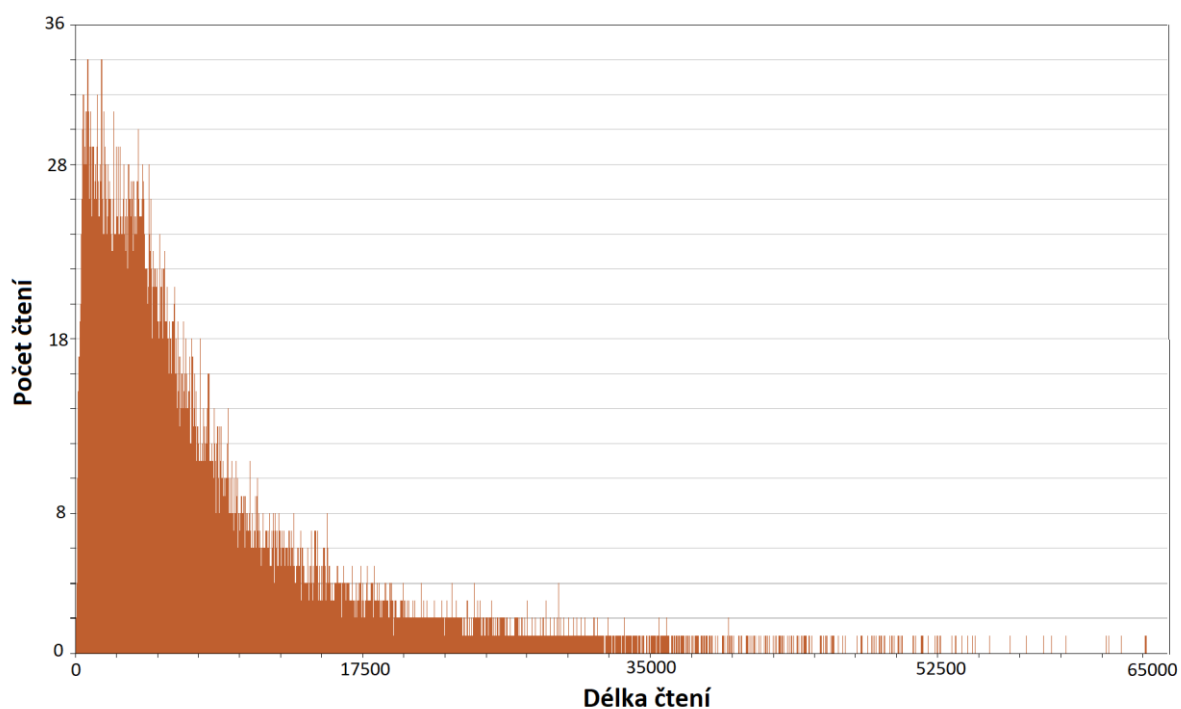
Obrázek 25: Graf počtů a délek čtení po přípravě „Rapid kitem“ s celkovým počtem 504713 sekvencí



Obrázek 26: Graf počtů a délek čtení po přípravě „PCR kitem“ s celkovým počtem 3355064 sekvencí



**Obrázek 27:** Graf počtů a délek čtení po přípravě „WGS kitem“ s celkovým počtem 135862 sekvencí



**Tabulka 9:** Celkový počet osekvenovaných bází ve jednotlivých vzorcích a součet všech bází podle způsobu přípravy

Počet osekvenovaných bází v jednotlivých vzorcích			
vzorek	Rapid kit	PCR kit	WGS kit
1	13 301 268	94 483 490	76 230 557
2	6 562 274	5 725 859	79 392 154
3	18 908 665	31 329 313	100 915 524
4	17 765 571	65 005 272	71 991 564
5	6 562 274	37 225 000	89 716 957
6	3 439 453	3 452 038	76 648 683
7	1 658 398	11 114 237	71 720 667
8	2 915 568	79 174 151	52 775 190
9	430 104	33 347 261	52 693 338
10	1 448 938	21 368 162	54 434 536
11	170 804	10 691 607	29 348 763
<b>Celkem</b>	<b>76 717 809</b>	<b>438 587 205</b>	<b>841 955 913</b>

## 5.1.2 Celkové hodnocení identifikace sekvencí nástrojem MG-RAST

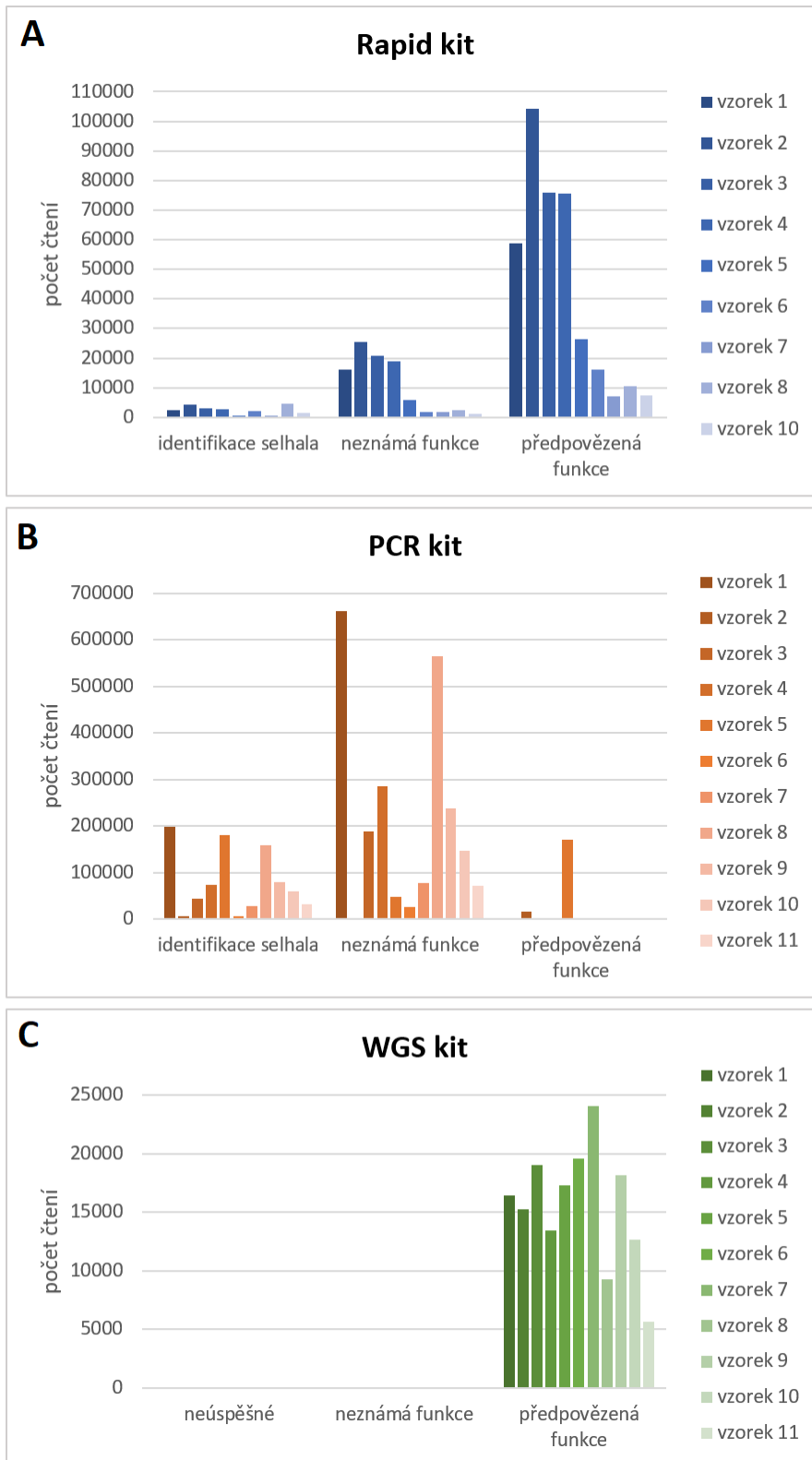
### Výsledky z hodnocení identifikace čtení u každého způsobu zpracování vzorků softwarem MG-RAST

Sekvence byly identifikovány na základě předpovězené funkce po překladu do proteinové sekvence. Na ose x jsou čtení ve vzorcích rozděleny do skupin: „identifikace selhala“ – z nějakého důvodu software nebyl schopen identifikaci dokončit, identifikováno jako „neznámá funkce“ – identifikace proběhla do konce, ale nebyla nalezena podobná sekvence v žádné z databází, identifikováno na základě „předpovězené funkce“ – v databázích byla nalezena podobnost. V grafu pro vzorky připravené „Rapid kitem“ je pouze 9 vzorků, protože ve vzorku 9 a 11 nebylo dostatečné množství sekvenovaných bp (1000000 bp) pro analyzování sekvencí v MG-RASTu.

**Obrázek 28:** A – Klasifikace sekvencí po přípravě protokolem „Rapid kit“. Celkový počet sekvencí s předpovězenou funkcí je 381462, s neznámou funkcí 94376 a u 21708 sekvencí identifikace selhala.

B – Klasifikace sekvencí po přípravě protokolem „PCR kit“. Celkový počet sekvencí s předpovězenou funkcí je 187785 a kromě vzorku 6 se podařilo identifikovat nějaké čtení v každém vzorku (v rozmezí 5-169959 čtení ve vzorku), s neznámou funkcí 2233475 a u 831684 sekvencí identifikace selhala.

C – Klasifikace sekvencí po přípravě protokolem „WGS kit“, u všech sekvencí byla předpovězena funkce



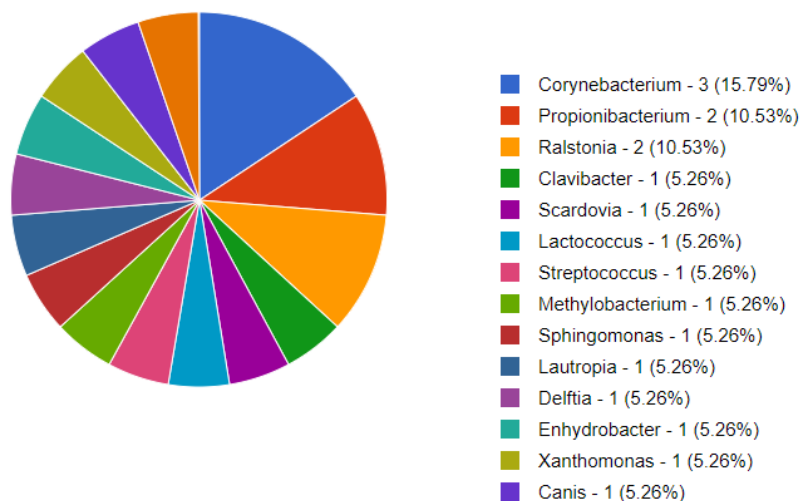
## 5.2 Obsah sekvencí v negativních kontrolách

### Identifikace sekvencí v negativních kontrolách pomocí softwaru MG-RAST

Analyzovány byly pouze negativní kontroly z protokolů „PCR kit“ a „WGS kit“, protože sekvenováním negativní kontroly protokolu „Rapid kit“ nebylo osekvenováno dostatečné množství bp (1000000 bp).

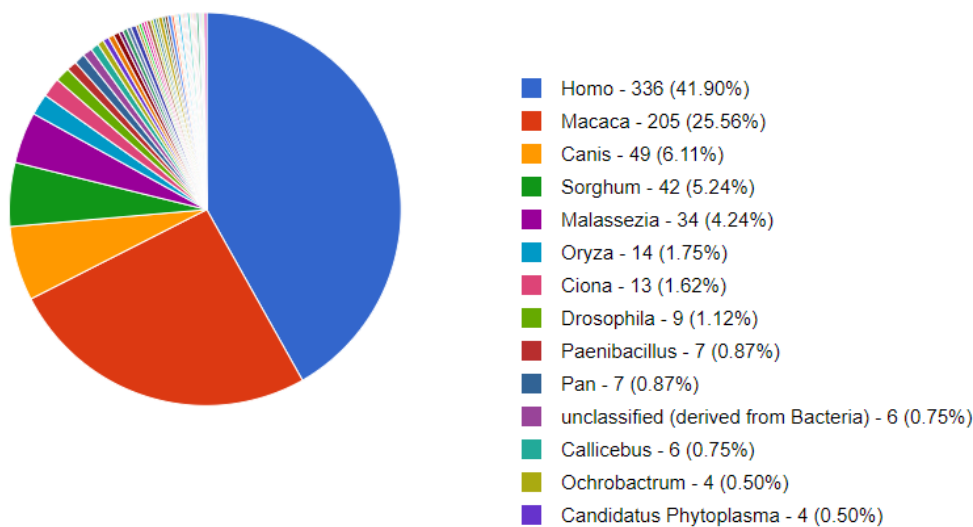
### Obrázek 29: Identifikace sekvencí v negativní kontrole z protokolu „PCR kit“

Identifikace je zapsána jako první rodové jméno z latinského názvu, počet nalezených čtení a počet čtení přepočtený na procenta. Většina sekvencí byla identifikována jako sekvence bakteriálního původu.



### Obrázek 30: Identifikace sekvencí v negativní kontrole z protokolu „WGS kit“

Identifikace je zapsána jako první slovo z latinského názvu, počet nalezených čtení a počet čtení přepočtený na procenta. Většina sekvencí byla identifikována jako sekvence lidského či opičího původu, další sekvence byly identifikovány jako eukaryota (rostliny i živočichové) i bakterie





### **5.3 Identifikace sekvencí v jednotlivých vzorcích – pomocí softwaru Geneious a MG-RAST**

Čtení získaná sekvenováním vzorků byla analyzována v softwarech Geneious a MG-RAST. V grafech je vyjádřen vztah mezi dekadickým logaritmem počtu kopií daného organismu ve vzorku (osa x) a počtu nalezených čtení daného organismu, nebo skupiny organismů (kontaminující sekvence bakteriálního původu), která jsou přepočítána na 1000 nt genomu daného organismu (osa y). V některých případech musela být místo počtu kopií uvedena hodnota Ct, protože daným virům (kmeny rotaviru, H3N2, RSV B a MeV) nebyla v Laboratoři molekulární genetiky k dispozici PCR „assay“ ani kalibrační křivka, hodnoty Ct byly získány z jiného pracoviště. V každém grafu jsou vyneseny výsledky získané v obou softwarech, které jsou od sebe odlišeny barevně. Tvar dále rozlišuje mezi jednotlivými organismy (viry či kontaminujícími sekvencemi lidského či bakteriálního původu).

## Vzorek 1 – Enterovirus, parechovirus a rotavirus ve vzorku stolice

### Obrázek 31: Obsah sekvencí ve vzorku 1

E30: mezi virovými čtení byl E30 nejvíce zastoupený, což odpovídá i původnímu počtu kopií viru ve vzorku („Rapid kit“ 712 čtení – Geneious, 228 čtení – MG-RAST, „PCR kit“ 17 čtení Geneious, 7 čtení MG-RAST, „WGS kit“ 2 čtení Geneious i MG-RAST)

HPeV-1: osekvenovaných čtení bylo výrazně méně, než kolik bylo původně ve vzorku

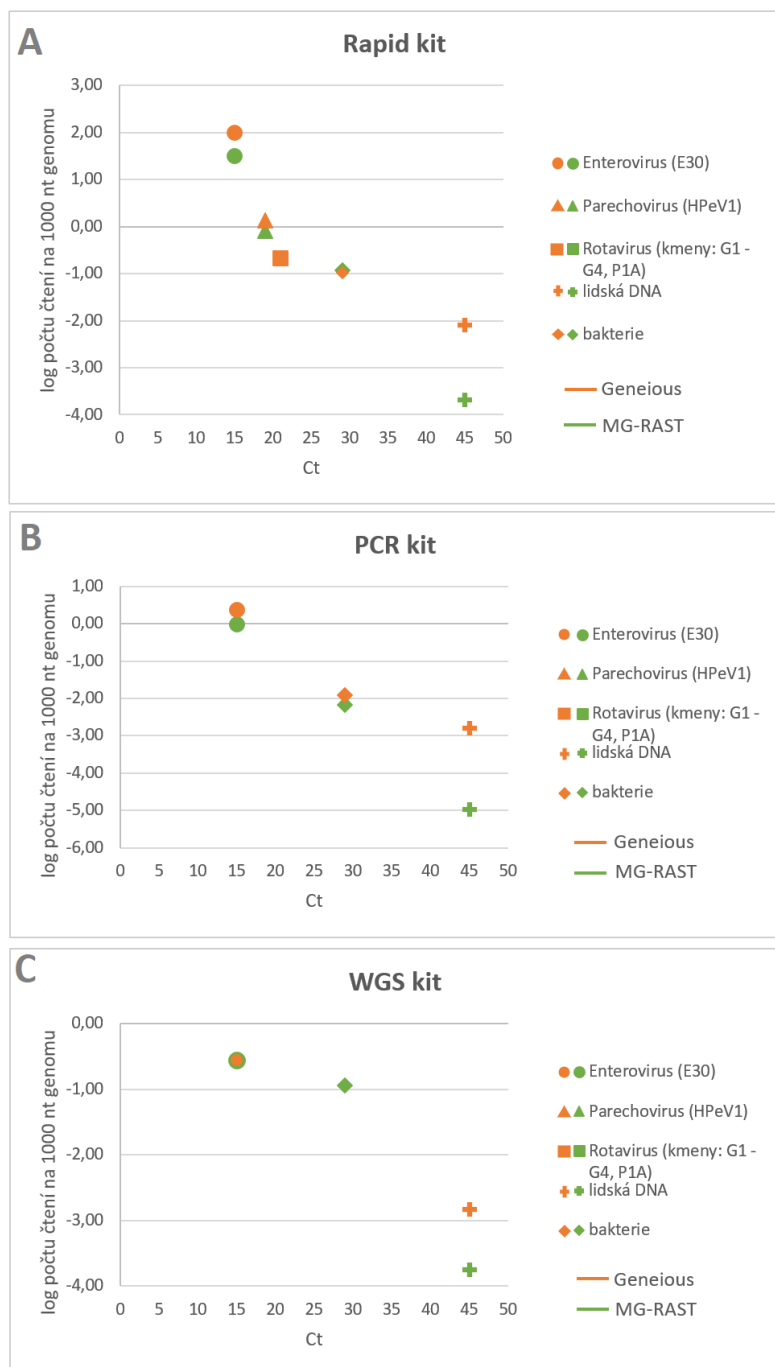
(„Rapid kit“ 10 čtení – Geneious, 6 čtení – MG-RAST)

Rotavirus (kmeny: G1-G4, P1A): osekvenovaných čtení bylo jen stopové množství („Rapid kit“ 4 čtení – Geneious)

Kontaminující sekvence: všechny tři způsoby přípravy odhalily větší podíl bakteriálních nukleových kyselin než lidských.

Analýzy softwaru Geneious a MG-RAST jsou srovnatelné až na identifikaci lidské DNA, kdy v Geneiosu bylo nalezeno mnohem více sekvencí.

Zdaleka nejvíce identifikovaných čtení se našlo přípravou pomocí „Rapid kitu“.



## Vzorek 2 – HSV-1 ve vzorku aspirátu

### Obrázek 32: Obsah sekvencí ve vzorku 2

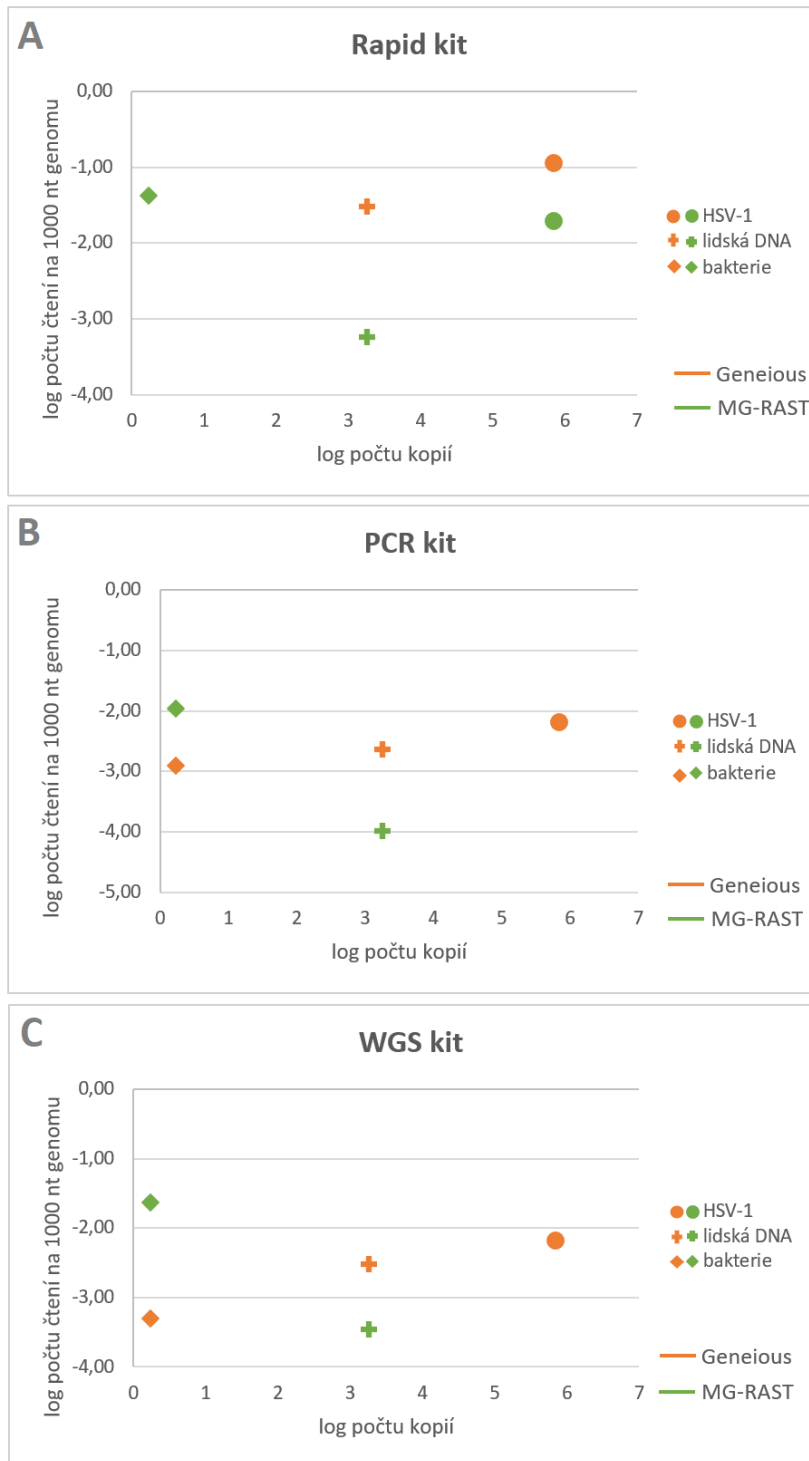
**HSV-1:** největší počet sekvencí bylo identifikováno jako HSV-1 přípravou „Rapid kit“, přípravou „PCR kit“ a „WGS kit“ bylo identifikováno HSV-1 s druhým největším počtem sekvencí („Rapid kit“ 17 čtení – Geneious, 3 čtení – MG-RAST, „PCR kit“ 1 čtení – Geneious, „WGS kit“ 1 čtení – Geneious)

**Kontaminující sekvence:** všechny tři způsoby přípravy odhalily větší podíl bakteriálních čtení než lidských.

Přípravou „PCR kit“ a „WGS kit“ byl identifikován největší počet sekvencí jako bakteriální.

Analýza softwaru Geneious a MG-RAST se lišila. Virových a lidských čtení bylo nejvíce identifikováno pomocí Geneious a bakteriálních čtení pomocí MG-RAST.

Zdaleka nejvíce identifikovaných čtení se našlo přípravou pomocí „Rapid kitu“.



### Vzorek 3 – simulace infekce dýchacích cest virem chřipky typu A a respiračním syncytiálním virem B

**Obrázek 33: Obsah sekvencí ve vzorku 3**

**H3N2:** největší počet sekvencí bylo identifikováno jako H3N2 přípravou „Rapid kit“ a „PCR kit“ („Rapid kit“ 236 čtení – Geneious, 194 čtení – MG-RAST, „PCR kit“ 1 čtení – Geneious, 4 čtení – MG-RAST)

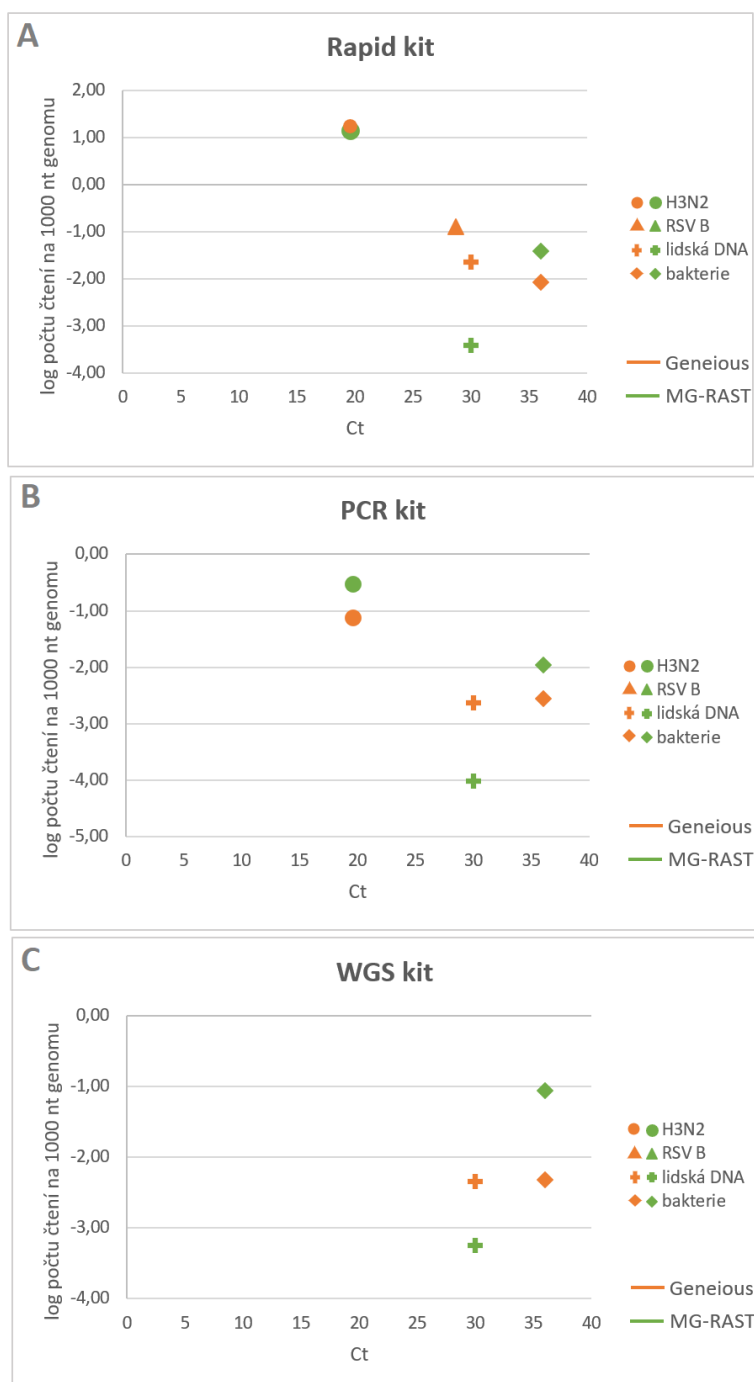
**RSV B:** s druhým největším počtem sekvencí byl identifikován RSV B přípravou „Rapid kit“ (2 čtení – Geneious)

**Kontaminující sekvence:** všechny tři způsoby přípravy odhalily větší podíl bakteriálních čtení než lidských.

Přípravou „WGS kit“ byl identifikován největší počet sekvencí jako bakteriální.

Analýza softwaru Geneious a MG-RAST se shodovala při identifikaci virových čtení. Bakteriálních čtení bylo více identifikováno pomocí MG-RAST a lidských čtení pomocí Geneious.

Zdaleka nejvíce identifikovaných čtení se našlo přípravou pomocí „Rapid kitu“.



## Vzorek 4 – simulace infekce v moči virem spalniček

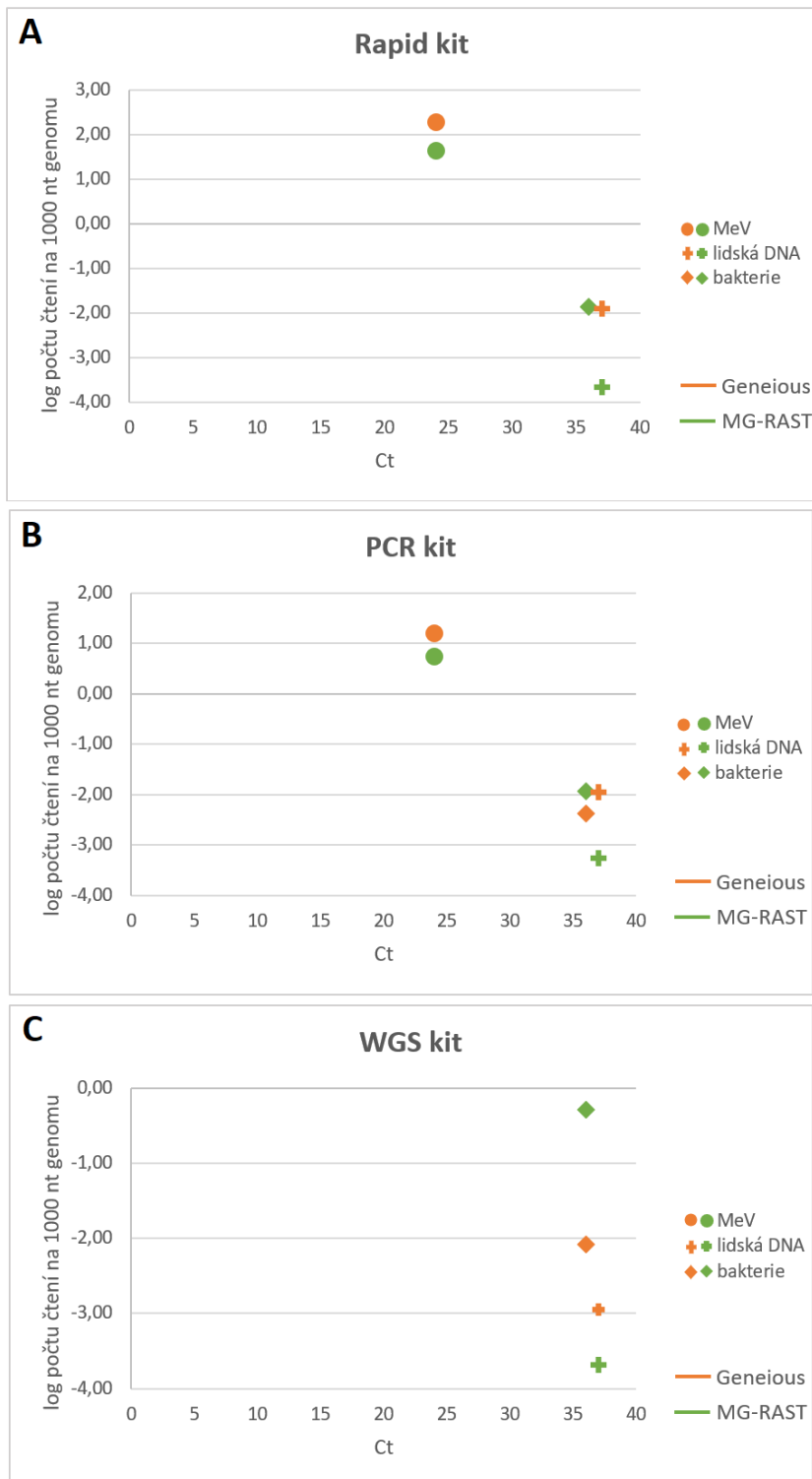
### Obrázek 34: Obsah sekvencí ve vzorku 4

MeV: největší počet sekvencí bylo identifikováno jako MeV v přípravou „Rapid kit“ a „PCR kit“ („Rapid kit“ 3095 čtení – Geneious, 706 čtení – MG-RAST, „PCR kit“ 248 čtení – Geneious, 88 čtení – MG-RAST)

Kontaminující sekvence: u přípravy „WGS kitem“ bylo nejvíce sekvencí identifikováno jako bakterie. V ostatních přípravách byl počet identifikovaných čtení jako lidské nebo bakteriální srovnatelný.

Analýza softwaru Geneious a MG-RAST se lišila. Virových a lidských čtení bylo nejvíce identifikováno pomocí Geneious a bakteriálních čtení pomocí MG-RAST.

Zdaleka nejvíce identifikovaných čtení se našlo přípravou pomocí „Rapid kitu“.



## Vzorek 5 – simulace infekce v moči BK virem

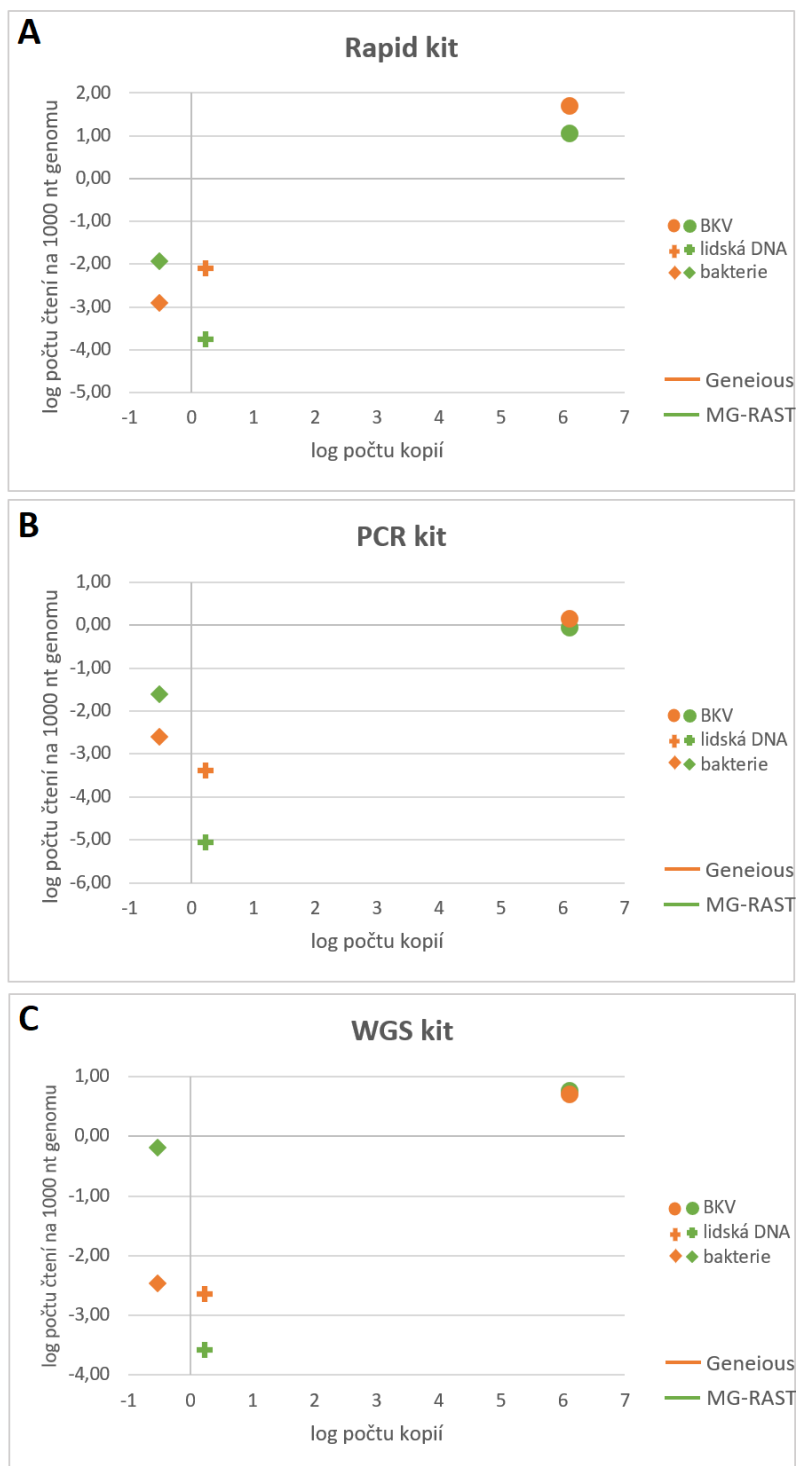
### Obrázek 35: Obsah sekvencí ve vzorku 5

**BKV:** největší počet sekvencí bylo identifikováno jako BKV ve všech třech kitech („Rapid kit“ 285 čtení – Geneious, 63 čtení – MG-RAST, „PCR kit“ 8 čtení – Geneious, 5 čtení – MG-RAST, „WGS kit“ 28 čtení – Geneious, 33 čtení – MG-RAST)

**Kontaminující sekvence:** všechny tři způsoby přípravy odhalily větší podíl bakteriálních nukleových kyselin než lidských.

Analýza softwaru Geneious a MG-RAST se shodovala při identifikaci virových čtení. Bakteriálních čtení bylo více identifikováno pomocí MG-RAST a lidských čtení pomocí Geneious.

Nejvíce virových a lidských čtení se našlo přípravou pomocí „Rapid kitu“, u bakteriálních čtení pomocí „WGS kitu“.



## Vzorek 6 – testování sekvenování různých typů genomů

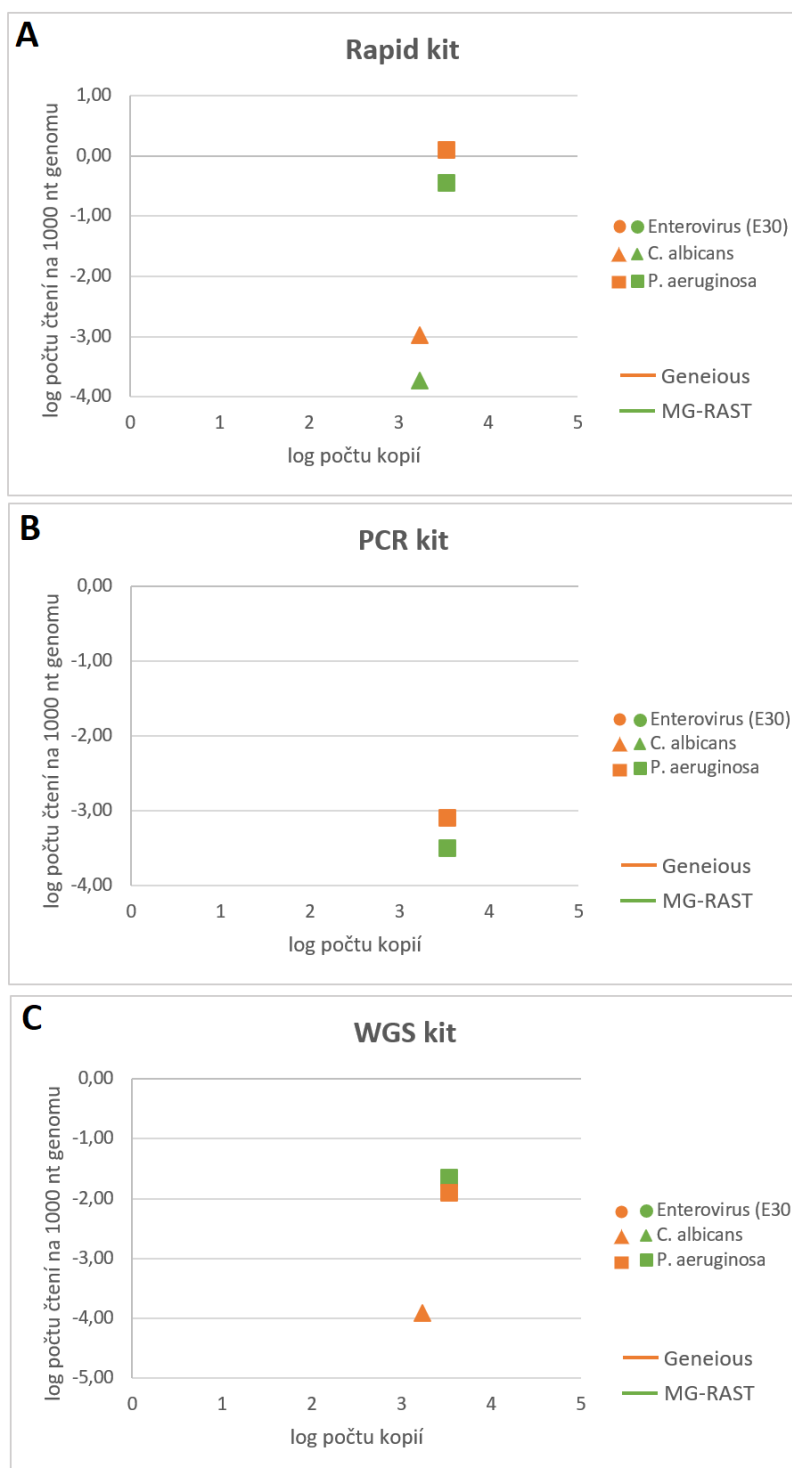
### Obrázek 36: Obsah sekvencí ve vzorku 6

*C. albicans*: bylo identifikováno méně čtení, než kolik bylo původně ve vzorku („Rapid kit“ 17 čtení – Geneious, 3 čtení – MG-RAST, „WGS kit“ 2 čtení Geneious)

*P. aeruginosa*: mezi identifikovanými čteními byla nejvíce zastoupená, což odpovídá i původnímu počtu kopií bakterie ve vzorku („Rapid kit“ 7732 čtení – Geneious, 2194 čtení – MG-RAST, „PCR kit“ 5 čtení Geneious, 2 čtení MG-RAST, „WGS kit“ 77 čtení – Geneious, 140 čtení – MG-RAST)

Analýza softwaru Geneious a MG-RAST se většinou shodovala, až na „WGS kit“, kde MG-RAST nenalezl žádná čtení *C. albicans*.

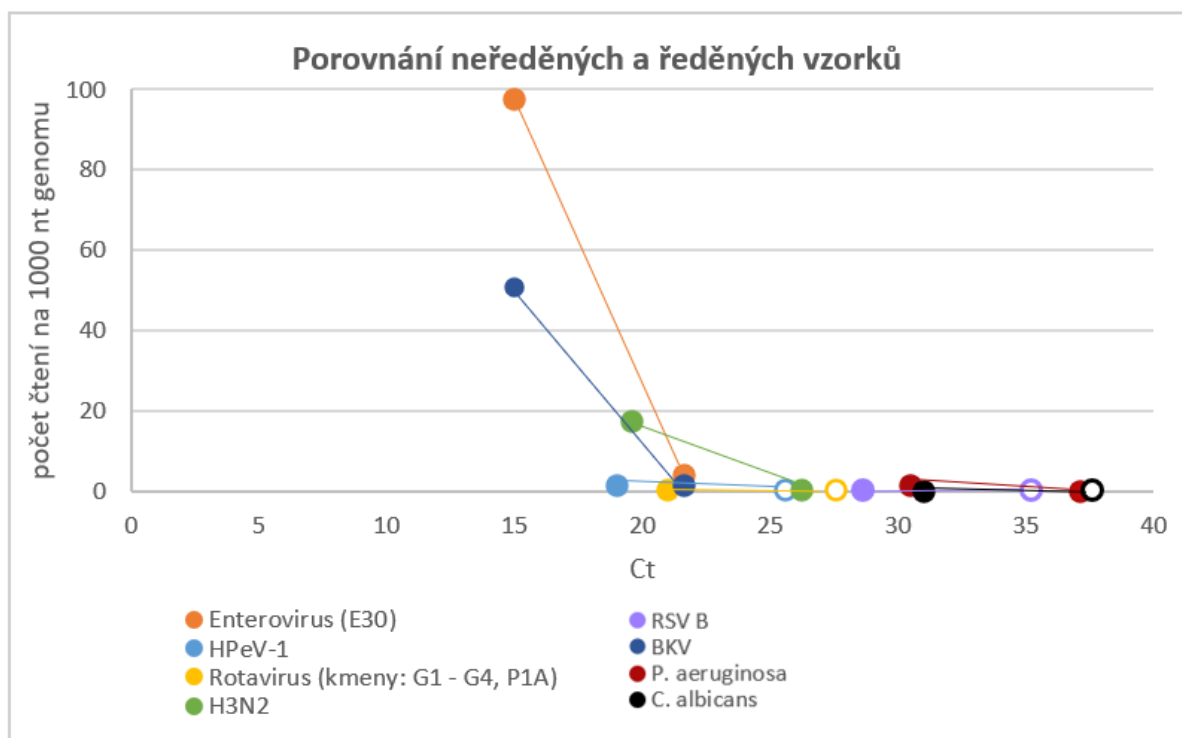
Zdaleka nejvíce identifikovaných čtení se našlo přípravou pomocí „Rapid kitu“.



## 5.4 Porovnání počtů identifikovaných čtení v neřaděných a řaděných vzorcích

Čtení získaná sekvenováním vzorků byla analyzována softwarem Geneious. V grafu je vyjádřen vztah mezi hodnotou prahového cyklu daného organismu ve vzorku (osa x) a počtem nalezených čtení organismu přepočítaná na 1000 nt genomu (osa y). Čtení se podařilo identifikovat pouze ve vzorcích připravených „Rapid kitem“.

**Obrázek 37:** Graf porovnání neřaděných a řaděných vzorků. Počet identifikovaných čtení se u řaděných vzorků pohybuje v rozmezí 0-30 čtení (E30 – 30 čtení, H3N2 – 4 čtení, BKV – 8 čtení, *P. aeruginosa* – 1 čtení, HPeV-1, kmeny rotaviru, RSV B a *C. albicans* – 0 čtení), řaděné vzorky, u kterých nebylo nalezeno žádné čtení jsou v grafu vyznačeny barevnými kroužky s bílou výplní





## 5.5 Identifikace dlouhých a krátkých čtení pomocí softwaru Geneious

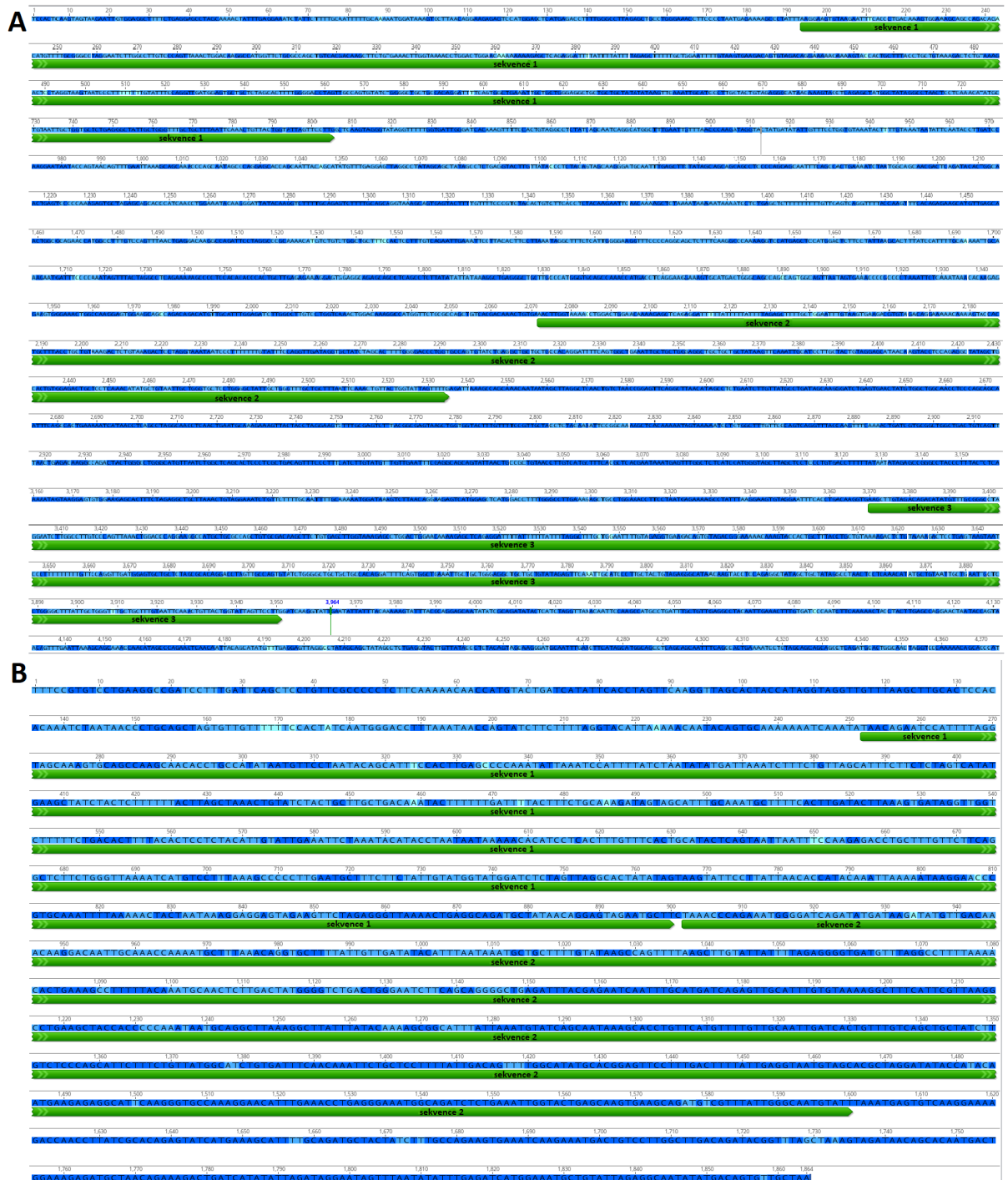
### 5.5.1 Identifikace řetězcích se sekvencí v jednom čtení ve vzorcích připravených „WGS kitem“

Při důkladnější analýze dlouhých čteních z „WGS kitu“ v softwaru Geneious bylo objeveno, že dochází k řetězní kratších sekvencí za sebou v jednom dlouhém čtení.

**Obrázek 28: Opakování sekvencí**

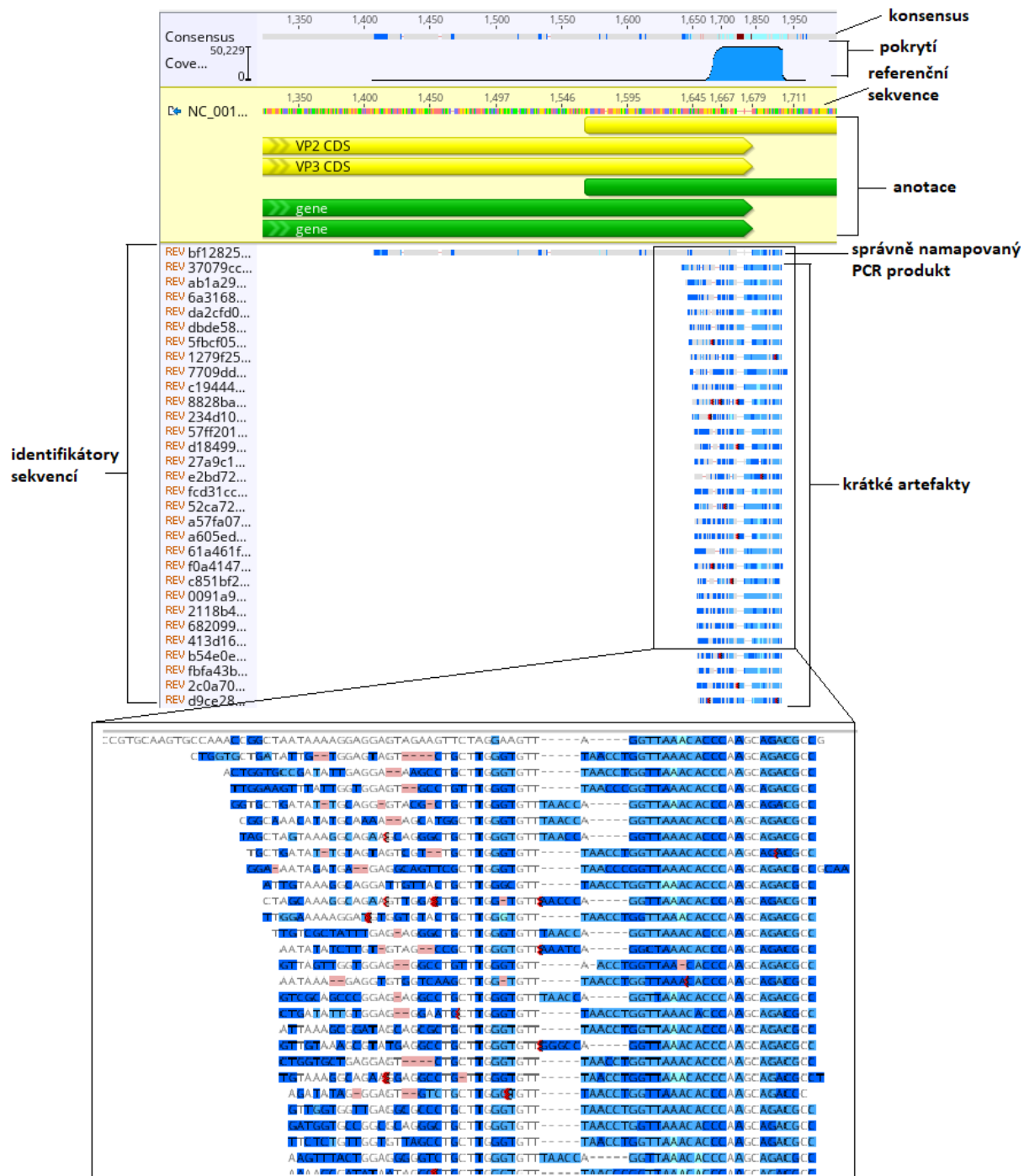
**A** – ve čtení dlouhém 4370 nt byly nalezeny tři stejné sekvence, které se opakovaly (zeleně označené): první opakování o délce 611 nt, druhé opakování o délce 463 nukleotidů a třetí opakování o délce 581 nt. Mezi prvním a druhým opakováním byla mezer 1267 nt a mezi druhým a třetím opakováním byla mezer 835 nt.

**B** – ve čtení dlouhém 1864 nt byly nalezeny dvě opakující se sekvence (zeleně označené): první opakování o délce 648 nt a druhé opakování o délce 698 nt. Mezi sekvencemi byla jednonukleotidová mezer.



## 5.5.2 Identifikace krátkých artefaktů ve vzorcích připravených „PCR kitem“

**Obrázek 29:** Krátké artefakty o délce přibližně 50 nt byly mapovány na správně namapovaný PCR produkt k referenční sekvenci, ale nebyly komplementární s referenční sekvencí



## 6. Diskuze

Tato práce shrnuje výsledky testování různých typů příprav vzorků k sekvenování nové generace pomocí techniky Oxford Nanopore. Účelem bylo otestovat, zda by se tento sekvenátor mohl využít při diagnostice v klinické virologii pro případy, kdy konvenčními metodami nebyl rozpoznán původce infekce. Protože se jedná o neznámého potenciálního patogena bylo nutné použít metagenomické sekvenování. Připravili jsme pět vzorků simulujících infekci ve střevě, respirační a infekci herpesvirem, virem spalniček a BK virem. Vzorky obsahovaly různý poměr virů a kontaminujících sekvencí. Naposled byly testovány vzorky, které byly naředěny vodou, abychom zjistili sensitivitu metody.

K testování byly vybrány tři různé přípravy vzorků. Rapid Barcoding Sequencing kit byl zvolen pro jeho rychlou a jednoduchou přípravu a zároveň byl nejpodobnější metodě, která je u nás v laboratoři zavedená (Kramná et al. 2015). Dále byl vybrán PCR Barcoding kit, který produkuje delší čtení (průměrně 4-5 kb dlouhá) a před přípravou knihovny kitem od ON není amplifikace, při které by mohly vzniknout chyby v amplikonech. Jako poslední byl použit Premium whole genome amplification protokol, kterým lze získat čtení dlouhá několik desítek kilobází, mohlo by se jím tak získat celý virový genom pouze z jednoho čtení.

Nejvíce čtení bylo získáno přípravou „PCR kit“ – 3355064 čtení s průměrnou délkou 100-200 nt, dále bylo získáno 504713 čtení „Rapid kitem“ s průměrnou délkou 100-250 nt a „WGS kitem“ bylo sekvenováno 135862 čtení o průměrné délce 500-7500 nt. Ačkoliv jsme „PCR kitem“ získali nejvíce sekvencí, u čtvrtiny z nich nástrojem MG-RAST selhala identifikace sekvencí a více než polovina sekvencí byla funkce neznámá. Oproti tomu u „Rapid kitu“ se nástroji MG-RAST podařilo předpovědět funkci už 75 % sekvencí. U „WGS kitu“ bylo dokonce 100 % sekvencí s předpovězenou funkcí, ale virové sekvence byly identifikovány pouze ve třech vzorcích a u ostatních vzorků převažovaly lidské a bakteriální sekvence, které byly pravděpodobně přednostně sekvenovány. Při analýze sekvencí připravených „PCR kitem“ v softwaru Geneious bylo zjištěno, že obsahuje velké množství krátkých fragmentů (cca 50 nt dlouhých) a objevují se v každém vzorku. Tyto fragmenty byly objeveny při opakovaném mapování (iteracích), kde se fragmenty ve velkém množství mapovaly na konec správně namapované čtení k referenci, ale samotné fragmenty nemají s referencí nic společného. Fragmenty byly testovány na shodu se barcodey a adaptéry „PCR kitu“, ale shoda nalezena nebyla. I při testování na shodu pomocí databáze NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)) se tato krátká čtení nepodařilo identifikovat. Dalším problémem, který jsme identifikovali během mapování

bylo, že u vzorků připravených „WGS kitem“, kde se nám podařilo získat dlouhá čtení, bylo zjištěno, že jde o kratší fragmenty dlouhé přibližně 600 nt, které se opakovaně řetězí za sebou a mezi kterými je různý počet nukleotidů. Nejlepších výsledků bylo dosaženo přípravou „Rapid kitem“, kterým bylo identifikováno nejvíce virových sekvencí a některé virové genomy byly rovnoměrně pokryty celé.

Co se týče shody délek sekvencí kontrolovaných před sekvenováním a po, u „Rapid kitu“ byly amplikony před přípravou knihovny „Rapid kitem“ dlouhé 300-500 bp. Sekvenováním jsme získali čtení s průměrnou délkou 250 nt, což bylo zřejmě způsobeno enzymatickou fragmentací, která je součástí přípravy tohoto kitu. U „PCR kitu“ při kontrole délek PCR produktů na Bioanalyzeru byly vidět kromě amplikonů v délkách 500-1000 bp i menší vrcholy v oblastech 150-300, ale průměrná délka sekvenovaných čtení byla 200 nt. Je tedy možné, že kratší čtení vznikla nedostatečným osekvenováním amplikonu (sekvenace čtení byla z nějakého důvodu zastavena). Při přípravě „WGS kitem“ po testování vzorků na gelové elektroforéze bylo vidět, že po izotermální amplifikaci a následné reakci s T7 endonukleázou I vznikly produkty o stejné délce 10 kbp. Z toho se dá usoudit, že k řetězení sekvencí došlo už při tomto kroku, protože k testování byly použity i viry s kratší délkou genomů.

V negativních kontrolách byly identifikovány v malém množství sekvence hlavně bakteriálního („PCR kit“) a lidského/opičího („WGS kit“) původu. To je ale běžné, každý kit obsahuje nějaké kontaminace (Salter et al. 2014; Laurence, Hatzis, a Brash 2014). Ty mohly mít podíl na vyšší obsah lidské DNA ve výsledcích u „WGS kitu“, ale u jiných kitů ke zkreslování výsledků nedocházelo.

Nejlepších výsledků bylo dosaženo přípravou „Rapid kit“, kterým bylo nejen identifikováno nejvíce virových čtení, ale také byly nejvíce rovnoměrně pokryté virové genomy a bylo nalezeno všech osm sekvenovaných virů. Největšího pokrytí bylo dosaženo u vzorku 4 (MeV ve vzorku moči), kde jsme získali 95 % genomu viru spalniček s pokrytím alespoň 5 čtení. U vzorku 1 (E30, HPeV-1 a kmeny rotaviru ve vzorku stolice) bylo získáno s pokrytím >5 čtení 90 % genomu echoviru 30 a ve vzorku 5 (BKV ve vzorku moči) přes 70 % genomu BK viru s minimálním pokrytím 3 čtení. U vzorků připravených ostatními dvěma kity byla pokrytí spíše sporadická. Sekvenováním „PCR kitem“ bylo nalezeno šest virů a „WGS kitem“ pouze tři viry z původních osmi sekvenovaných virů.

Celková kvantita nukleových kyselin v našich vzorcích nesplňovala předepsanou vstupní kvantitu jednotlivých kitů. Tento fakt je častým problémem při metagenomickém sekvenování

virů z původních vzorků a celkově použití sekvenování ve virologii znesnadňuje narozdíl např. od sekvenování lidské DNA, které je ve vzorku hodně a neobsahuje tolik kontaminací (Kramná et al. 2015).

Výtěžnost virových sekvencí byla ovlivněna přítomností kontaminujících sekvencí. U vzorku 2 (HSV-1 v aspirátu) bylo nalezeno nejvíce kontaminujících sekvencí před i po sekvenování, a to nejspíš proto, že nukleové kyseliny byly izolovány přímo z lidského aspirátu. Obdobně tomu bylo i u vzorku 3 (H3N2 a RSV B ve vzorku aspirátu. Ve vzorku 1 bylo nalezeno zase velké množství bakteriálních sekvencí, které pocházely především ze vzorku stolice. I přesto, že v původním vzorku 1 nebyla detekována téměř žádná lidská DNA, tak byla osekvenována ve větším množství hlavně „Rapid kitem“, zřejmě díky pomnožení DNA z buněčných kultur, na kterých byl kultivován enterovirus E30. Nejvíce sekvencí a nejlepšího pokrytí genomu bylo dosaženo u viru spalniček a BK viru, které byly ve vzorku moči, ve kterém bylo pouze stopové množství kontaminujících nukleových kyselin. U 100x ředěných vzorků se podařilo detekovat pouze tři viry z šesti (E30, H3N2 a BKV) nejspíš proto, že kvantita virů v ředěných vzorcích už byla příliš nízká a je možné, že tak došlo přednostně k namnožení kontaminujících sekvencí.

Lewandowski et al. (2019) chtěli metagenomickým sekvenováním na MinIONu detekovat virus chřipky přímo z respiračního vzorku. Množství virových sekvencí bylo negativně ovlivněno zvýšeným množstvím hostitelských a bakteriálních čtení. RNA byla po izolaci ze vzorků náhodně amplifikována a knihovna byla připravena stejným způsobem, jako v našem případě u „WGS kitu“ jen bez izotermální amplifikace. Ve vzorcích byl chřipkový virus také kvantifikován pomocí qPCR. U 27 vzorků s hodnotami Ct <30 se virus podařilo detekovat, ale pouze u šesti z 13 vzorků s Ct >30 byly získány sekvence viru. Tyto výsledky představují o mnoho řádů lepší sensitivitu, v této práci jsme identifikovali pouze jeden virus ze čtyř s hodnotami Ct > 25.

Ve studii Greninger et al. (2015) se potýkali s vysokým procentem kontaminujících sekvencí. Pomocí metagenomického NGS se snažili detekovat RNA viry ve vzorku krve od čtyř pacientů. RNA byla izolována a následně náhodně amplifikována. Knihovna byla připravena stejným způsobem, jako v našem případě u „WGS kitu“ bez izotermální amplifikace. Z celkových počtů identifikovaných čtení tvořily 78-96 % sekvence lidského původu a virový výtěžek byl v rozmezí 0,6-6,5 % z identifikovaných čtení. Ve studii Mostafa et al. (2020) chtěli pomocí metagenomického NGS sekvenovat virus u pacientů s onemocněním

COVID-19. U osmi vzorků pozitivních na virus SARS-CoV-2 s hodnotami Ct v rozmezí 21-36,6, virus nebyl detekován a autoři konstatují, že je možné, že by virová čtení mohla být maskována lidskými.

Pro identifikaci neznámého patogenu v klinickém vzorku byla v našem testování příprava „Rapid kitem“ neúspěšnější, a to jak jednoduchostí metody, tak výsledky, které nejvíce odpovídaly původnímu složení vzorku. Podařilo se identifikovat všech osm virů s největším počtem čtení a k tomu tři viry ve vzorcích s nízkou kvantitou. Příprava kitem „PCR“ se pro identifikaci neznámého potenciálního virového patogenu nehodí, zřejmě kvůli nedostatečné vstupní kvantitě DNA těchto vzorků. Ani příprava vzorků kitem „WGS“ se pro naše účely neosvědčila, stále ale považujeme tento způsob přípravy jako velmi přínosný pro možnost sekvenovat dlouhá čtení. Cílový virus by však musel být izolován a pomnožen na buněčné kultuře, nebo by jeho kvantita ve vzorku musela být vysoká a obsah kontaminujících sekvencí nízký, abychom získali požadovaná dlouhá čtení.

## **7. Souhrn**

V této práci bylo sekvenováno několik vzorků, které simulovaly infekční vzorek s různým obsahem virových sekvencí, k otestování tří různých příprav vzorků s použitím tří různých kitů od firmy Oxford Nanopore, s cílem vybrat ten nejvhodnější pro identifikaci patogenu v klinickém vzorku s neznámou infekcí. Také byla testována senzitivita použitých metod na řadě infekčních vzorcích. Součástí práce bylo také testování různých nástrojů pro analýzu dat. Nejlepších výsledků bylo dosaženo přípravou vzorku pomocí Rapid Barcoding Sequencing kitu.

## 8. Seznam použité literatury

- Ashikawa, Sae, Norihito Tarumoto, Kazuo Imai, Jun Sakai, Masahiro Kodana, Toru Kawamura, Kenji Ikebuchi, et al. 2018. „Rapid Identification of Pathogens from Positive Blood Culture Bottles with the MinION Nanopore Sequencer". *Journal of Medical Microbiology* 67 (11): 1589–95. <https://doi.org/10.1099/jmm.0.000855>.
- Benson, Dennis A., Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell, a Eric W. Sayers. 2013. „GenBank". *Nucleic Acids Research* 41 (Database issue): D36–42. <https://doi.org/10.1093/nar/gks1195>.
- Brinkmann, Annika, Sophie-Luisa Ulm, Steven Uddin, Sophie Förster, Dominique Seifert, Rainer Oehme, Merle Corty, Lars Schaade, Janine Michel, a Andreas Nitsche. 2021. „AmpliCoV: Rapid Whole-Genome Sequencing Using Multiplex PCR Amplification and Real-Time Oxford Nanopore MinION Sequencing Enables Rapid Variant Identification of SARS-CoV-2". *Frontiers in Microbiology* 12 (červenec): 651151. <https://doi.org/10.3389/fmicb.2021.651151>.
- Buermans, H.P.J., a J.T. den Dunnen. 2014. „Next Generation Sequencing Technology: Advances and Applications". *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1842 (10): 1932–41. <https://doi.org/10.1016/j.bbadis.2014.06.015>.
- Cinek, O, E Witso, S Jeansson, T Rasmussen, P Drevinek, T Wetlesen, J Vavrínek, B Grinde, a K Ronningen. 2006. „Longitudinal Observation of Enterovirus and Adenovirus in Stool Samples from Norwegian Infants with the Highest Genetic Risk of Type 1 Diabetes". *Journal of Clinical Virology* 35 (1): 33–40. <https://doi.org/10.1016/j.jcv.2005.03.007>.
- Cole, J. R., B. Chai, T. L. Marsh, R. J. Farris, Q. Wang, S. A. Kulam, S. Chandra, et al. 2003. „The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy". *Nucleic Acids Research* 31 (1): 442–43.
- „Decoding DNA with a pocket-sized sequencer | www.scienceinschool.org". b.r. <https://www.scienceinschool.org/content/decoding-dna-pocket-sized-sequencer>.
- DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, a G. L. Andersen. 2006. „Greengenes, a Chimera-Checked 16S RRNA Gene Database and Workbench Compatible with ARB". *Applied and Environmental Microbiology* 72 (7): 5069–72. <https://doi.org/10.1128/AEM.03006-05>.
- Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, et al. 2009. „Real-Time DNA Sequencing from Single Polymerase Molecules". *Science* 323 (5910): 133–38. <https://doi.org/10.1126/science.1162986>.
- Foret, Frantisek, Karel Klepárník, a Marek Minárik. 2017. „Nucleic Acids | Chromatographic and Electrophoretic Methods ☆". In *Reference Module in Chemistry, Molecular Sciences and Chemical Engineering*, B9780124095472142000. Elsevier. <https://doi.org/10.1016/B978-0-12-409547-2.14225-8>.
- „Geneious | Bioinformatics Software for Sequence Data Analysis". b.r. Geneious. <https://www.geneious.com/>.
- Greninger, Alexander L., Samia N. Naccache, Scot Federman, Guixia Yu, Placide Mbala, Vanessa Bres, Doug Stryke, et al. 2015. „Rapid Metagenomic Identification of Viral Pathogens in Clinical Samples by Real-Time Nanopore Sequencing Analysis". *Genome Medicine* 7 (1): 99. <https://doi.org/10.1186/s13073-015-0220-9>.



- „Illumina | Sequencing". b.r. <https://www.illumina.com/>.
- Jain, Miten, Hugh E. Olsen, Benedict Paten, a Mark Akeson. 2016. „The Oxford Nanopore MinION: Delivery of Nanopore Sequencing to the Genomics Community". *Genome Biology* 17 (1): 239. <https://doi.org/10.1186/s13059-016-1103-0>.
- Jensen, Lars Juhl, Philippe Julien, Michael Kuhn, Christian von Mering, Jean Muller, Tobias Doerks, a Peer Bork. 2008. „EggNOG: Automated Construction and Annotation of Orthologous Groups of Genes". *Nucleic Acids Research* 36 (Database issue): D250-254. <https://doi.org/10.1093/nar/gkm796>.
- Kanehisa, Minoru. 2002. „The KEGG Database". *Novartis Foundation Symposium* 247: 91–101; discussion 101-103, 119–28, 244–52.
- Kennedy, Peter G. E., Phenix-Lan Quan, a W. Ian Lipkin. 2017. „Viral Encephalitis of Unknown Cause: Current Perspective and Recent Advances". *Viruses* 9 (6): 138. <https://doi.org/10.3390/v9060138>.
- Kramná, Lenka, Kateřina Kolářová, Sami Oikarinen, Juha-Pekka Pursiheimo, Jorma Ilonen, Olli Simell, Mikael Knip, Riitta Veijola, Heikki Hyöty, a Ondrej Cinek. 2015. „Gut Virome Sequencing in Children With Early Islet Autoimmunity: Figure 1". *Diabetes Care* 38 (5): 930–33. <https://doi.org/10.2337/dc14-2490>.
- Laurence, Martin, Christos Hatzis, a Douglas E. Brash. 2014. „Common Contaminants in Next-Generation Sequencing That Hinder Discovery of Low-Abundance Microbes". Editoval Tom Gilbert. *PLoS ONE* 9 (5): e97876. <https://doi.org/10.1371/journal.pone.0097876>.
- Laver, T., J. Harrison, P.A. O'Neill, K. Moore, A. Farbos, K. Paszkiewicz, a D.J. Studholme. 2015. „Assessing the Performance of the Oxford Nanopore Technologies MinION". *Biomolecular Detection and Quantification* 3 (březen): 1–8. <https://doi.org/10.1016/j.bdq.2015.02.001>.
- Lewandowski, Kuiama, Yifei Xu, Steven T. Pullan, Sheila F. Lumley, Dona Foster, Nicholas Sanderson, Alison Vaughan, et al. 2019. „Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical Respiratory Samples". *Journal of Clinical Microbiology* 58 (1): e00963-19. <https://doi.org/10.1128/JCM.00963-19>.
- Lin, Jake, Lenka Kramna, Reija Autio, Heikki Hyöty, Matti Nykter, a Ondrej Cinek. 2017. „Vipie: web pipeline for parallel characterization of viral populations from multiple NGS samples". *BMC Genomics* 18 (1): 378. <https://doi.org/10.1186/s12864-017-3721-7>.
- Loit, Kaire, Kalev Adamson, Mohammad Bahram, Rasmus Puusepp, Sten Anslan, Riinu Kiiker, Rein Drenkhan, a Leho Tedersoo. 2019. „Relative Performance of MinION (Oxford Nanopore Technologies) versus Sequel (Pacific Biosciences) Third-Generation Sequencing Instruments in Identification of Agricultural and Forest Fungal Pathogens". *Applied and Environmental Microbiology* 85 (21): e01368-19. <https://doi.org/10.1128/AEM.01368-19>.
- Lu, Hengyun, Francesca Giordano, a Zemin Ning. 2016. „Oxford Nanopore MinION Sequencing and Genome Assembly". *Genomics, Proteomics & Bioinformatics* 14 (5): 265–79. <https://doi.org/10.1016/j.gpb.2016.05.004>.
- Magrane, Michele, a UniProt Consortium. 2011. „UniProt Knowledgebase: a hub of integrated protein data". *Database* 2011 (bar009). <https://doi.org/10.1093/database/bar009>.

- Margulies, Marcel, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bembien, Jan Berka, et al. 2005. „Genome Sequencing in Open Microfabricated High Density Picoliter Reactors". *Nature* 437 (7057): 376–80. <https://doi.org/10.1038/nature03959>.
- Markowitz, Victor M., Natalia N. Ivanova, Ernest Szeto, Krishna Palaniappan, Ken Chu, Daniel Dalevi, I-Min A. Chen, et al. 2008. „IMG/M: a data management and analysis system for metagenomes". *Nucleic Acids Research* 36 (Database issue): D534–38. <https://doi.org/10.1093/nar/gkm869>.
- Metzker, Michael L. 2010. „Sequencing Technologies — the next Generation". *Nature Reviews Genetics* 11 (1): 31–46. <https://doi.org/10.1038/nrg2626>.
- Meyer, F., D. Paarmann, M. D’Souza, R. Olson, EM Glass, M. Kubal, T. Paczian, et al. 2008. „The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes". *BMC Bioinformatics* 9 (1): 386. <https://doi.org/10.1186/1471-2105-9-386>.
- Mostafa, Heba H., John A. Fissel, Brian Fanelli, Yehudit Bergman, Victoria Gniazdowski, Manoj Dadlani, Karen C. Carroll, Rita R. Colwell, a Patricia J. Simner. 2020. „Metagenomic Next-Generation Sequencing of Nasopharyngeal Specimens Collected from Confirmed and Suspect COVID-19 Patients". *mBio* 11 (6): e01969-20. <https://doi.org/10.1128/mBio.01969-20>.
- „National Center for Biotechnology Information". b.r. <https://www.ncbi.nlm.nih.gov/>.
- Overbeek, Ross, Tadhg Begley, Ralph M. Butler, Jomuna V. Choudhuri, Han-Yu Chuang, Matthew Cohoon, Valérie de Crécy-Lagard, et al. 2005. „The Subsystems Approach to Genome Annotation and Its Use in the Project to Annotate 1000 Genomes". *Nucleic Acids Research* 33 (17): 5691–5702. <https://doi.org/10.1093/nar/gki866>.
- „Oxford Nanopore Technologies". b.r. <https://nanoporetech.com/>.
- Palacios, Gustavo, Julian Druce, Lei Du, Thomas Tran, Chris Birch, Thomas Briese, Sean Conlan, et al. 2008. „A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases". *The New England Journal of Medicine* 358 (10): 991–98. <https://doi.org/10.1056/NEJMoa073785>.
- Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, a F. O. Glockner. 2007. „SILVA: A Comprehensive Online Resource for Quality Checked and Aligned Ribosomal RNA Sequence Data Compatible with ARB". *Nucleic Acids Research* 35 (21): 7188–96. <https://doi.org/10.1093/nar/gkm864>.
- Pruitt, Kim D., Tatiana Tatusova, a Donna R. Maglott. 2007. „NCBI Reference Sequences (RefSeq): A Curated Non-Redundant Sequence Database of Genomes, Transcripts and Proteins". *Nucleic Acids Research* 35 (Database issue): D61–65. <https://doi.org/10.1093/nar/gkl842>.
- Quick, Joshua, Nathan D Grubaugh, Steven T Pullan, Ingra M Claro, Andrew D Smith, Karthik Gangavarapu, Glenn Oliveira, et al. 2017a. „Multiplex PCR Method for MinION and Illumina Sequencing of Zika and Other Virus Genomes Directly from Clinical Samples". *Nature Protocols* 12 (6): 1261–76. <https://doi.org/10.1038/nprot.2017.066>.
- . 2017b. „Multiplex PCR Method for MinION and Illumina Sequencing of Zika and Other Virus Genomes Directly from Clinical Samples". *Nature Protocols* 12 (6): 1261–76. <https://doi.org/10.1038/nprot.2017.066>.
- Quick, Joshua, Nicholas J. Loman, Sophie Duraffour, Jared T. Simpson, Ettore Severi, Lauren Cowley, Joseph Akoi Bore, et al. 2016. „Real-time, portable genome sequencing for Ebola surveillance". *Nature* 530 (7589): 228–32. <https://doi.org/10.1038/nature16996>.

- Rang, Franka J., Wigard P. Kloosterman, a Jeroen de Ridder. 2018. „From Squiggle to Basepair: Computational Approaches for Improving Nanopore Sequencing Read Accuracy". *Genome Biology* 19 (1): 90. <https://doi.org/10.1186/s13059-018-1462-9>.
- „RDP Taxonomy - Sequence Analysis Tools". b.r. <http://rdp.cme.msu.edu/>.
- „Roche - Doing now what patients need next". b.r. <https://www.roche.com/>.
- Salter, Susannah J., Michael J. Cox, Elena M. Turek, Szymon T. Calus, William O. Cookson, Miriam F. Moffatt, Paul Turner, Julian Parkhill, Nicholas J. Loman, a Alan W. Walker. 2014. „Reagent and laboratory contamination can critically impact sequence-based microbiome analyses". *BMC Biology* 12 (1): 87. <https://doi.org/10.1186/s12915-014-0087-z>.
- Shahi, Shailesh K., Samantha N. Freedman, a Ashutosh K. Mangalam. 2017. „Gut Microbiome in Multiple Sclerosis: The Players Involved and the Roles They Play". *Gut Microbes* 8 (6): 607–15. <https://doi.org/10.1080/19490976.2017.1349041>.
- Snyder, E. E., N. Kampanya, J. Lu, E. K. Nordberg, H. R. Karur, M. Shukla, J. Soneja, et al. 2007. „PATRIC: The VBI PathoSystems Resource Integration Center". *Nucleic Acids Research* 35 (Database): D401–6. <https://doi.org/10.1093/nar/gkl858>.
- Stoler, Nicholas, a Anton Nekrutenko. 2021. „Sequencing error profiles of Illumina sequencing instruments". *NAR Genomics and Bioinformatics* 3 (1). <https://doi.org/10.1093/nargab/lqab019>.
- Towner, Jonathan S., Tara K. Sealy, Marina L. Khristova, César G. Albariño, Sean Conlan, Serena A. Reeder, Phenix-Lan Quan, et al. 2008. „Newly Discovered Ebola Virus Associated with Hemorrhagic Fever Outbreak in Uganda". *PLOS Pathogens* 4 (11): e1000212. <https://doi.org/10.1371/journal.ppat.1000212>.
- Turnbaugh, Peter J., Ruth E. Ley, Micah Hamady, Claire M. Fraser-Liggett, Rob Knight, a Jeffrey I. Gordon. 2007. „The Human Microbiome Project". *Nature* 449 (7164): 804–10. <https://doi.org/10.1038/nature06244>.
- Venter, J Craig, Mark D Adams, Eugene W Myers, Peter W Li, Richard J Mural, Granger G Sutton, Hamilton O Smith, et al. 2001. „The Sequence of the Human Genome". *THE HUMAN GENOME* 291: 49.
- Wang, Jing, Nicole Elizabeth Moore, Yi-Mo Deng, David A. Eccles, a Richard J. Hall. 2015. „MinION Nanopore Sequencing of an Influenza Genome". *Frontiers in Microbiology* 0. <https://doi.org/10.3389/fmicb.2015.00766>.
- Wawina-Bokalanga, Tony, Bert Vanmechelen, Joan Martí-Carreras, Valentijn Vergote, Kurt Vermeire, Jean-Jacques Muyembe-Tamfum, Steve Ahuka-Mundeke, a Piet Maes. 2019. „Complete Genome Sequence of a New Ebola Virus Strain Isolated during the 2017 Likati Outbreak in the Democratic Republic of the Congo". *Microbiology Resource Announcements* 8 (20): e00360-19. <https://doi.org/10.1128/MRA.00360-19>.
- Wheeler, David A., Maithreyan Srinivasan, Michael Egholm, Yufeng Shen, Lei Chen, Amy McGuire, Wen He, et al. 2008. „The Complete Genome of an Individual by Massively Parallel DNA Sequencing". *Nature* 452 (7189): 872–76. <https://doi.org/10.1038/nature06884>.
- Wilson, Michael R., Hannah A. Sample, Kelsey C. Zorn, Shaun Arevalo, Guixia Yu, John Neuhaus, Scot Federman, et al. 2019. „Clinical Metagenomic Sequencing for Diagnosis of Meningitis

and Encephalitis". *New England Journal of Medicine* 380 (24): 2327–40. <https://doi.org/10.1056/NEJMoa1803396>.

Wu, Fan, Su Zhao, Bin Yu, Yan-Mei Chen, Wen Wang, Zhi-Gang Song, Yi Hu, et al. 2020. „A New Coronavirus Associated with Human Respiratory Disease in China". *Nature* 579 (7798): 265–69. <https://doi.org/10.1038/s41586-020-2008-3>.

Zanella, M.-C., L. Lenggenhager, J. Schrenzel, S. Cordey, a L. Kaiser. 2019. „High-Throughput Sequencing for the Aetiologic Identification of Viral Encephalitis, Meningoencephalitis, and Meningitis. A Narrative Review and Clinical Appraisal". *Clinical Microbiology and Infection* 25 (4): 422–30. <https://doi.org/10.1016/j.cmi.2018.12.022>.

Zhang, Jun, Rod Chiodini, Ahmed Badr, a Genfa Zhang. 2011. „The impact of next-generation sequencing on genomics". *Journal of genetics and genomics = Yi chuan xue bao* 38 (3): 95–109. <https://doi.org/10.1016/j.jgg.2011.02.003>.

Zhu, Na, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, et al. 2020. „A Novel Coronavirus from Patients with Pneumonia in China, 2019". *New England Journal of Medicine* 382 (8): 727–33. <https://doi.org/10.1056/NEJMoa2001017>.