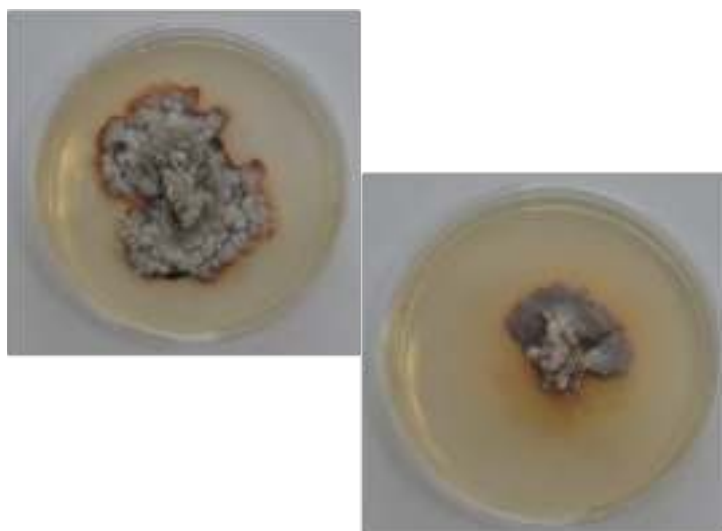




Zemědělská
fakulta
Faculty
of Agriculture

Jihočeská univerzita
v Českých Budějovicích
University of South Bohemia
in České Budějovice

Metodika identifikace, determinace a přenosu mykovirů u hub rodu *Armillaria*



Autoři:

**Mgr. Tomáš Tonka Ph.D., Ing. Lucie Walterová, Ing. Mgr. Ondřej Hejna, Ph.D.,
prof. Ing. Vladislav Čurn, Ph.D.**

České Budějovice, 2021

Jihočeská univerzita v Českých Budějovicích Zemědělská fakulta

**Metodika byla vypracovaná jako výstup projektu NAZV QK1920412
“Mykoviry jako součást potenciálních biopreparátů v ochraně smrkových porostů proti
václavkám“**

**Mgr. Tomáš Tonka, Ph.D.
Ing. Lucie Walterová
Ing. Mgr. Ondřej Hejna, Ph.D.
prof. Ing. Vladislav Čurn, Ph.D.**

České Budějovice, září 2021

Vypracováno za podpory projektu NAZV QK1920412 "Mykoviry jako součást potenciálních biopreparátů v ochraně smrkových porostů proti václavkám".

Poděkování: Veškeré zpracování sekvenčních dat probíhalo na počítačích Metacentrum pořízených z projektu "e-Infrastruktura CZ" (e-INFRA CZ ID:90140) MSMT CR.

Autoři: Mgr. Tomáš Tonka, Ph.D., Ing. Lucie Walterová, Ing. Mgr. Ondřej Hejna, Ph.D., prof. Ing. Vladislav Čurn, Ph.D.

Oponenti:

Ing. Vlasta Knorová, MZE, odbor hospodářské úpravy a ochrany lesů
Těšnov 65/17, 110 00 Praha, vlasta.knorova@mze.cz

Ing. Pavel Hobza, LČR, s.p., Lesní závod Boubín, polesí Klet, Náměstí 7, 38203 Křemže,
pavel.hobza@lesycr.cz

Obrázek na titulní straně: vlevo neinfikovaná kultura václavky smrkové, vpravo izolát václavky smrkové infikovaný ambi-like mykovirem

Text: ©2021 Tonka T., Walterová L., Hejna O., Čurn V.

Foto: ©2021 Tonka T.

Grafická úprava: ©2021 Tonka T.

Vydáno bez jazykové úpravy

ISBN 978-80-7394-891-7



Obsah

1. Cíl metodiky	5
2. Vlastní popis metodiky	7
3. Příprava biologického materiálu	9
3.1. Odběr vzorků	9
4. Detekce virů v hostitelských václavkách	10
4.1. Biologický materiál	10
4.2. Izolace dsRNA	10
4.3. Izolace celkové RNA	12
4.3.1. Izolace celkové RNA pomocí SPLIT RNA Extraction Kit	12
4.3.2. Extrakce RNA pomocí činidla Trizol	13
4.4. Kontrola kvality vyizolované celkové RNA	15
4.5. Stanovení koncentrace izolované RNA	15
5. NGS sekvenování	17
5.1. Příprava knihoven	19
6. Zpracování sekvenáčnických dat – bioinformatická analýza RNAseq sekvencí ..	20
6.1. Kontrola kvality sekvenovaných dat	20
6.2. Odstranění adaptorů (trimming) z hrubých sekvencí	22
6.3. Mapování sekvencí na referenční genom	23
6.4. <i>De novo</i> assembly nenamapovaných sekvencí	26
6.5. Analýza nově sestavených sekvencí	28
7. Vyšetření hub na přítomnost mykovirů	31
7.1. Navrhování primerů	31
7.2. RT izolované celkové RNA	32
7.3. PCR detekce ambi-like virů	33
8. Testování přenosu mykovirů v laboratoři	34
8.1. Hybridizace kmenů václavek	34
9. Srovnání novosti postupů	36
10. Popis uplatnění metodiky	36
11. Ekonomické aspekty	37
12. Seznam použité literatury	38

1. Cíl metodiky

Houby rodu václavka (*Armillaria*) způsobují ročně vysoké ztráty na smrkových porostech. Stromy, oslabené suchem, jsou náchylnější k napadení houbovými i hmyzími škůdci. Je předpoklad, že v souvislosti se změnami klimatu, jako jsou rostoucí teploty a zvyšující se deficit srážek, bude docházet k většímu infekčnímu tlaku václavek na smrkové porosty a václavky se tak mohou stát daleko významnějším problémem v lesním hospodaření, než jsou dnes. Ochrana proti václavkám a obecně proti houbovým patogenům není v praktickém lesním managementu účinná a v zásadě není proti václavkám rodu *Armillaria* ani možná. Proto vyvstává otázka, jestli vůbec a jaká je možná účinná ochrana v rámci lesního managementu proti václavkám. Na základě požadavků státní správy je realizován výzkumný projekt, jehož jedním z cílů bylo najít a identifikovat mykoviry u václavek r. *Armillaria* pomocí metod Next Generation Sequencing (NGS), které by se daly v budoucnu využít v biologické ochraně proti těmto patogenním houbám.

Václavky r. *Armillaria* jsou rozšířeny po celém světě, kde způsobují značné hospodářské a ekonomické škody na hostitelských rostlinách (Coetzee a kol., 2018; Guillaumin a kol., 1993). V ČR bylo popsáno 7 druhů václavek r. *Armillaria*, které škodí zejména v lesním hospodářství, ze kterých je václavka smrková (*Armillaria ostoyae*) dominantním patogenem ve smrkových porostech (Antonín a kol., 2009). Stupeň rozšíření a související ekonomické ztráty jsou nesnadno vyčíslitelné vzhledem k tomu, že evidence poškozeného dřeva se vykazuje pouze při identifikaci myceliem poškozeného stromu. Daleko větší efekt na zdravotní stav stromů má ale podzemní mycelium, které oslabuje hostitelské rostliny čerpáním živin z napadených stromů. Tyto dopady nejsou jasně viditelné a tudíž ekonomické dopady takto napadených stromů se odhadují obtížně.

Hlavním ekonomickým dopadem václavek je poškození jinými biotickými nebo abiotickými stresory oslabených stromů a znehodnocení pokácené dřevní hmoty. Václavky škodí ale i v lesních školkách na semenáčcích, takže hned od počátku pěstování jsou smrky pod tlakem václavek r. *Armillaria*.

Opatření na ochranu proti václavkám nejsou účinná, protože nebyla doposud nalezena účinná metoda proti těmto patogenním houbám. Jednou z potenciálních metod v ochraně smrkových porostů proti václavkám je hledání biologických prostředků ochrany, které jsou součástí půdního ekosystému václavka - hostitel a mohou regulovat její výskyt nebo životní funkce. Jedním z těchto potenciálních bioagens mohou být mykoviry.

Jedním z dílčích cílů řešení projektu QK1920412 "Mykoviry jako součást potenciálních biopreparátů v ochraně smrkových porostů proti václavkám", v rámci kterého byla metodika vypracována, je pomocí metod NGS identifikovat potenciální mykoviry, které by se daly použít v biologické ochraně proti václavkám.

Cílem metodiky je podat ucelený soubor analytických postupů pro detekci virů uvnitř hostitelských organizmů a možnost izolace nových virů a jejich využití jako případných prostředků biologické ochrany v lesním managementu a ochraně smrkových a jiných porostů proti dřevokazným a patogenním houbám.

Většina mykovirů jsou viry ze skupiny dsRNA virů (viry s genomem dvouvláknové RNA), které se dají identifikovat standardními molekulárními metodami (Ghabrial a kol., 2015; Ghabrial a Suzuki, 2009). V souvislosti s rychlým rozvojem a dostupností technik celogenomového sekvenování (NGS) a rozšiřováním úrovně znalostí o této skupině virů a zpřesňováním

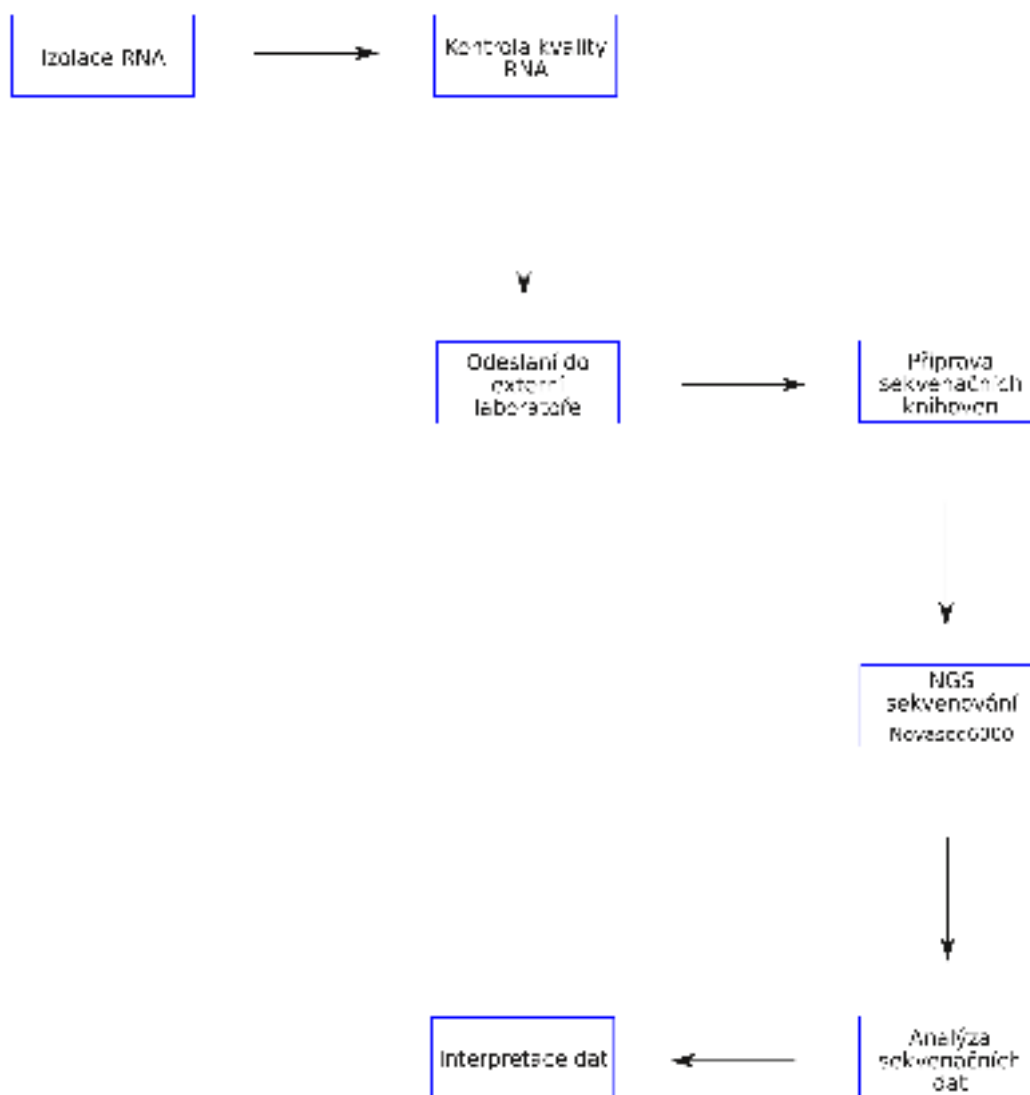
genetických databází se ukazuje, že značná část mykovirů patří do skupiny ssRNA virů, které se dají efektivně najít pouze metodami NGS (Zhang a kol., 2018).

V současné době s rozvojem technik NGS dochází k explozivnímu nárůstu popisů nových virů u hub. Většina těchto virů patří do skupiny RNA virů. Mykoviry, jak se tato skupina souhrnně označuje, byly identifikovány u všech hlavních skupin hub, včetně Chytridiomycota, Ascomycota, Deuteromycota a Basidiomycota (Herrero a kol., 2012; Son a kol., 2015). První mykovirus byl popsán u žampionů (*Agaricus bisporus*) již v 60. letech minulého století (Hollings 1962). Mykoviry byly popsány i u několika významných fytopatogenních hub, např. *Rhizoctonia solani*, *Cryphonectria parasitica*, *Sclerotinia sclerotiorum*, *Botrytis cinerea* (Ghabrial a kol., 2015; Bryner a kol., 2012; Hao a kol., 2018; Boland 1992; Abdoulaye a kol., 2019). Jejich vliv na hostitele je v zásadě neznámý. Většina mykovirů uvnitř hostitele přetrvává v latentní fázi bez viditelných symptomů u hostitelských organizmů (Garcia - Pendrejas a kol., 2019). Na druhou stranu, existují mykoviry, které dramaticky ovlivňují růst hostitele, nebo ovlivňují jeho pohlavní rozmnožování. Ale nejdůležitějším efektem na hostitele je tzv. hypovirulence, neboli schopnost mykoviru redukovat virulenci houbového patogena (Boland 1992; Nuss 2005; Sharma a kol., 2018; Son a kol., 2015; Xie a Jiang, 2014).

Vzhledem k tomu, že bylo popsáno doposud pouze několik mykovirů václavek (Linnakoski a kol., 2021; Rumbou a kol., 2021), a to pouze NGS metodami, nejsou známy žádné vztahy mezi viry a hostitelskými houbami a jejich vliv na patogenitu václavek. Kromě studia těchto vztahů je důležité hledat i nové možnosti biologických prostředků na ochranu proti václavkám a tato metodika přináší podrobný návod na identifikaci nových virů nejenom u václavek, ale i u hub obecně.

2. Vlastní popis metodiky

V současné době v souvislosti se zaváděním a cenovou dostupností nových sekvenačních metod dochází k enormnímu nárůstu nově popsaných a objevených organismů napříč celým systémem živé přírody. Tyto výsledky se nevyhýbají ani nově popsaným virům u hub. V poslední době byla identifikována celá řada virů u fytopatogenních hub (Sutella a kol., 2019; Hillman a kol. 2018; Ghabrail a kol., 2015). Úloha virů v životních cyklech hub ale nebyla doposud podrobněji popsána, takže není zatím úplně jasná role virů a jejich vliv na hostitele. Tato metodika má za cíl podrobně popsat postup vedoucí k identifikaci nových virů u hub a jejich testování v laboratorních podmínkách tak, aby uživatelé metodiky byli schopni popsanými postupy identifikovat virus a otestovat jeho případný vliv na hostitelskou houbu. Metodika se zabývá stopkovýtrosými houbami rodu *Armillaria*, ale uvedené metodické postupy se dají s menšími úpravami, vztahujícími se k životním podmínkám mitosporických hub, použít i k izolaci virových částic u této skupiny hub (Obr. 1).



Obr. 1. Schematické znázornění postupu determinace mykovirů u hub rodu *Armillaria*

Současné poznatky ukazují, že používané diagnostické metody nejsou schopny virové infekce u hub detekovat vzhledem k malému množství informací o vlivu virů na fenotyp hostitelské houby obecně. Nejspolehlivějším detekčním mechanismem je kombinace genetických a molekulárně biologických metod. Díky předchozím výzkumům, založeným na NGS metodách byla identifikována celá řada mykovirů (Rumbou a kol., 2021). Tyto viry patří do skupiny dsRNA virů, které se kromě NGS metod dají identifikovat izolací RNA a separací dsRNA pomocí gelové elektroforézy a ssRNA virů, které se dají identifikovat pouze metodami NGS a zpracováním výstupů sekvenování bioinformatickými metodami. V momentě, kdy takto získáme sekvenci ssRNA virů, dá se již dalšími rutinními postupy molekulárně biologickými identifikovat virus v izolované RNA hostitelského organismu. Uvedenými postupy byla identifikována celá řada houbových virů, včetně několika mykovirů u václavek, vyskytujících se po celém světě (Rusko, Finsko, JAR). S cenovou dostupností metod sekvenování genomu hub budou přibývat i informace o výskytu virů u jednotlivých skupin hub. Je na dalším zkoumání, zejména biologickém, jaká je role virů v hostitelích a jestli je možnost využití těchto virů případně v biologické ochraně nejenom proti václavkám, ale obecně proti dalším houbovým patogenům.

Vzhledem k významu václavek jako parazitů na celé řadě druhů dřevin je potřeba podporovat výzkum nejenom biologický, ale i molekulární, který povede k získávání poznatků o vztahu václavka - virus i na molekulární, biochemické a genetické úrovni.

Tato problematika RNA virů a jejich vlivu na hostitele není u hub doposud zcela uspokojivě vysvětlena. Dá říci, že stojíme na počátku výzkumu houbových virů a jejich vlivu na hostitele. Vzhledem k rostoucímu spektru houbových patogenů a ohrožení nejenom dřevin, ale celé řady rostlin, může nabývat na významu výzkum houbových virů nejen z hlediska identifikace nových virových kmenů/druhů, ale i výzkum interakcí mezi virem a hostitelem. Hledání nových kmenů virů jako potencionální biologické agens v ochraně proti houbovým patogenům se jeví jako jedna z cest v budoucí ochraně rostlin, v našem případě smrkových porostů proti václavkám a dalším patogenním houbám.

3. Příprava biologického materiálu

Vzorky na izolaci nukleových kyselin připravujeme v zásadě třemi způsoby. Nukleové kyseliny můžeme extrahovat z plodnic, mycelia nebo rhizomorf. Pro dlouhodobější uchování materiálu pro izolaci nukleových kyselin jsou nejvhodnější části plodnic. Na druhou stranu i z hlediska dalších pokusů a zkoumání vztahů mykovirus - hostitel je nutné mít živý materiál jak viru, tak hostitele. Proto je důležité izolovat živé kultury václavek jak ze spor z plodnic, tak z rhizomorf. Podrobnější metodika kultivace václavek byla popsána v předchozí metodice (Čurn a kol., 2019).

3.1. Odběr vzorků

Izolaci virů z václavek je možné provádět jak z fixovaného materiálu, tak ze živých hub. ssRNA viry se identifikují z celkové RNA hostitelského organismu. Fixáž tkáně pro další analýzy se provádí v RNAlateru (např. ThermoFisher), kde fixovanou tkáň uchováváme v mrazáku při $-20\text{ }^{\circ}\text{C}$. Z takto fixované tkáně se izoluje celková RNA a dsRNA, které se dále zpracovávají k dalšímu analytickému zkoumání.

Z hlediska identifikace nových virů a testování jejich vlivu na hostitele je daleko výhodnější uchovávat vzorky a kultury václavek v laboratoři pro dlouhodobé zkoumání. Z plodnic se mikrobiologickou klíčkou přenesou zralé spory na ME agar v Petriho misce, rozetřou se po povrchu a dají se do termostatu temperovaného na $25\text{ }^{\circ}\text{C}$. Po vyklíčení spor se nově vytvořené mycélium přenesou na nové misky s agarem.

Vzhledem k tomu, že plodnice václavek jsou vázané svým výskytem pouze na krátkou část roku, je výhodnější - pokud známe ohnisko výskytu václavek - nasbírat rhizomorfy z půdy a z nich izolovat DNA či zakládat nové kultury *in vitro*. Od paty napadeného stromu vybereme část rhizomorfy václavky a s trochou půdy ji dáme do plastového zipovacího sáčku, který uzavřeme. Takto uchovávané rhizomorfy, pokud nevyschnou, vydrží v chladícím boxu nebo lednici několik týdnů beze ztráty schopnosti tvorby mycélia.

Rhizomorfy vyndáme ze sáčku na filtrační papír, omyjeme je ve sterilní destilované H_2O . Následně rhizomorfu skalpelem rozdělíme na několik částí. Záleží na délce a tloušťce rhizomorfy, ideální je délka 5 – 10 cm.

Ve sterilním boxu si připravíme tři Petriho misky: jednu se 70% etanolem, druhou s roztokem 3% SAVA a jednu s čistou sterilní destilovanou vodou. Kousek rhizomorfy namočíme na 30 sekund do 70% etanolu, pak přendáme rhizomorfu na 10 minut do roztoku SAVA a nakonec jí důkladně propláchneme ve sterilní destilované vodě. Na kousku sterilního filtračního papíru odřízneme konce rhizomorfy - cca 1 cm, a rhizomorfu pokládáme na ME agar v Petriho miskách. Misky dáme do termostatu ($25\text{ }^{\circ}\text{C}$) a po několika dnech nově narostlé mycelium rozpasážujeme na nové misky s ME agarem. Pracujeme se sterilními nástroji (pinzety, sklápel) ve sterilním boxu, které mezi každým krokem opalujeme nad plynovým kahanem.

Takto nově kultivované kultury slouží nejen jako studijní materiál k identifikaci nových virů, ale zároveň jako zásobárna pro pokusy virus-hostitel.

4. Detekce virů v hostitelských václavkách

Přítomnost mykovirů v hostitelských organizmech se dá detekovat několika způsoby. dsRNA mykoviry se identifikují izolací dsRNA z mycelia kultivovaných václavek, ssRNA mykoviry hledáme NGS analýzou izolované celkové RNA z mycelia a následným bioinformatickým zpracováním získaných sekvencí příslušnými postupy, které mají odhalit přítomnost virových sekvencí v genomu hostitelských hub.

4. 1. Biologický materiál

Kultivace hub probíhá podle standardních protokolů, pospaných dřívě (Čurn a kol., 2019). Celkovou RNA i dsRNA izolujeme z čerstvého materiálu nebo z materiálu fixovaného (RNAlater). dsRNA se může izolovat i z lyofilizovaného mycelia nebo houbové tkáně.

4. 2. Izolace dsRNA

Při tomto postupu izolujeme dsRNA z biomasy hostitele. Metoda využívá afinity dsRNA k celulóze, kdy z celkového vyizolovaného objemu nukleových kyselin se navázáním na celulózu oddělí dsRNA. Tato metoda slouží k izolaci dsRNA virů, ale není vhodná pro detekci ssRNA virů. dsRNA se neizoluje pomocí komerčních kitů, protože základem izolace je navázání dsRNA na celulózu. Izolace se provádí podle původního protokolu izolace dsRNA (Morris a Dodds, 1979), ale modifikovaného pro potřeby izolací viru z hub.

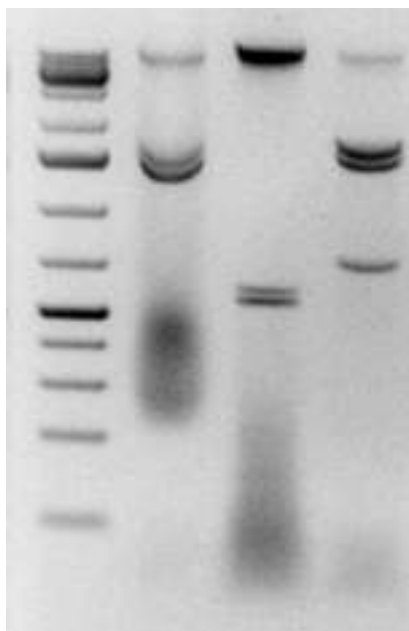
Homogenizace houbového mycelia se provádí v porcelánové třecí misce. Vzorek se může homogenizovat v čerstvém stavu, stejně jako zmražený. Naváží se mycelium - 0,1 - 10 mg - a tloučkem vzorek zhomogenizujeme.

- 1 K zhomogenizovanému vzorku se postupně přidá 1ml 2xSTE pufru, směs se důkladně promíchá, přidá se 400 µl chloroformu, směs se opět promíchá, přidá se 600 µl fenolu a po promíchání se přidá nakonec 100 µl 10% SDS (viz Čurn a kol., 2019).
- 2 Takto připravenou směs přelijeme z třecí misky do mikrozkušavky a necháme směs 1 hodinu točit na rotátoru při 4 °C.
- 3 Poté se mikrocentrifugační zkumavky se směsí stočí 15 minut při 15 000 ot. ve 4 °C v centrifuze.
- 4 Supernatant se po centrifugaci přendá do nových mikrozkušavek, k supernatantu se přidá 0,1 g celulózy a na 1ml supernatantu se přidá 160 µl ledového 96% ethanolu (z mrazáku -20 °C). Pro správnou izolaci je naprosto nutné mít výslednou koncentraci ethanolu 16%.
- 5 Tato směs se následně nechá 30 minut točit na rotátoru při 4 °C.
- 6 Po 30ti minutách se mikrozkušavky stočí 3 minuty při 15000 ot. ve 4 °C
- 7 Po centrifugaci odstraníme supernatant.
- 8 Pelet ve zkumavce promyjeme 1 ml roztoku 1 x STE pufru v 16% ethanolu. Stočíme na centrifuze při 15000 ot. 3 minuty při 4 °C. Odstraníme supernatant.
- 9 Krok 8 opakujeme 5x.
- 10 Po posledním promytí odstraníme supernatant a peletovanou celulózu s navázanou dsRNA důkladně vysušíme v termobloku, temperovaném na 55 °C. Doba sušení celulózy je cca 5 - 10 minut. Mikrozkušavky s celulózu necháváme v termobloku otevřené. Smyslem je nechat odpařit zbytek ethanolu, který by mohl ovlivnit rozpustnost nukleových kyselin.

- 11 Po vysušení k celulóze přidáme 500 μ l 1 x STE pufru a inkubujeme v termobloku při 65 $^{\circ}$ C 5 minut.
- 12 Poté zkumavky stočíme 3 minuty ve 4 $^{\circ}$ C při 15000 ot. Supernatant slijeme do nových mikrozkuvek o objemu 2 ml.
- 13 K supernatantu přidáme 1500 μ l 96% ethanolu (4 $^{\circ}$ C) a necháme 20 minut inkubovat v mrazáku při -20 $^{\circ}$ C. Zde je možné izolaci přerušit.
- 14 Zkumavky vyndáme z mrazáku, stočíme 15 minut ve 4 $^{\circ}$ C při 15000 ot.
- 15 Důkladně slijeme supernatant
- 16 Peletovanou dsRNA na dně zkumavky rozpustíme ve 30 μ l sterilní vody. 15 μ l rozpuštěné dsRNA se dává na agarózový gel (viz obr.).

STE pufr 2x 0,2 mol/l NaCl; 100 mmol/l Tris; 2 mmol/l EDTA

STE pufr 1x + 15% ethylalkohol (4 ml STE 10x + 6 ml etylalkoholu 99,8% pro UV spektroskopii + 30 ml dEPC vody



Obr. 2. Gel s dsRNA mykoviry. První sloupec je velikostní marker, pak tři vzorky s dsRNA o velikosti 10000 bp, 5500 bp a s dvěma dsRNA o velikosti 10000 bp a 6000 bp.

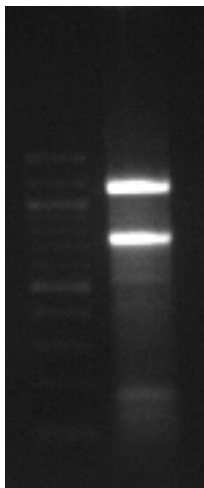
Po vizualizaci na agarózovém gelu se pruhy s dsRNA z gelu vyříznou, komerčními kity se vyizolují z agaru (např. ZymoClean Gel RNA Recovery Kit, Zymo Research) podle protokolu výrobce. S takto vyizolovanou RNA byla provedena reverzní transkripce dle protokolu výrobce (např. ImProm-II™ Reverse Transcription System, Promega, Madison, USA) a vzniklá cDNA byla zaslána do specializované laboratoře na sekvenování

4. 3. Izolace celkové RNA

Pro detekci ssRNA virů je nutné vyizolovat celkovou RNA z hostitele a takto získanou RNA sekvenovat příslušnými metodami. Celkovou RNA je možné izolovat pomocí mikroextrakčních metod (komerčně dostupnými kity) nebo pomocí standardních metod izolace, kdy se získá větší množství celkové RNA. Metody izolace pomocí kitů jsou založeny na principu extrakce RNA pomocí mikrokolonek, kdy se na matrix uvnitř kolonky naváže RNA z lyzované tkáně.

RNA je obecně méně stabilní než DNA, včetně větší náchylnosti k teplotní degradaci. Důležitou součástí prostředí jsou všudypřítomné RNAzy (ribonuklázy), enzymy degradující RNA. I malé stopové množství RNAz může degradovat celou RNA. Zdrojem RNAz je např. kůže, prach, vzorek ze kterého RNA izolujeme, chemikálie potřebné k izolaci atd. Kontaminace při extrakci vede k degradaci vzorku, který se projeví při elektroforetické separaci při kontrole kvality izolované RNA. Nedegradovaná, správně vyizolovaná RNA se na gelu jeví jako dva ostré proužky (horní 28S rRNA, dolní 18S rRNA, viz obr.). Po degradaci je na gelu místo proužků vidět např. neostrá šmouha a takto degradovaná RNA není k dalšímu použití vhodná.

Obr. 3. Celková RNA na agarozovém gelu. Horní pruh odpovídá 28S rRNA, spodní pruh 18S rRNA.



Proto izolace RNA musí probíhat ve sterilním, čistém prostředí, vzhledem k tomu, že RNA je náchylná na rychlou degradaci a poškození. Je nutné pracovat přesně, pečlivě a rychle. Minimalizovat dobu otevření nádob, zkumavek nebo boxů na špičky. Je třeba měnit

často ochranné rukavice a nedotýkat se vlasů, kůže a všech nesterilních předmětů. Po celou dobu izolace se používají pouze DNA a RNA free plasty, špičky s filtrem a čisté sterilní pipety. Před samotnou izolací RNA je dobré UV lampou vysterilovat místo izolace a sterilizovat všechny plasty a nástroje, potřebné k extrakci celkové RNA.

4. 3. 1. Izolace celkové RNA pomocí SPLIT RNA Extraction Kit (Lexogen)

- 1 Před izolací se nechají zkumavky Phase Lock Gel 30 minut temperovat při laboratorní teplotě. Pro každý vzorek použijeme jednu zkumavku.
- 2 Poté zkumavku Phase Lock Gel 1 minutu centrifugujeme při 12000 ot. v 18 °C, aby se gel usadil na dně zkumavky.
- 3 Vzorek mycelia (10 - 100 mg) v tekutém dusíku zhomogenizujeme v třecí misce a rozpustíme ve 400 µl studeného (4 °C) izolačního pufru (IB).

- 4 Homogenizovaný vzorek v izolačním pufru přepipetujeme do zkumavky Phase Lock Gel.
- 5 Přidáme 400 µl roztoku fenolu pH 4,3 (RNAzol) a promícháme 5x převrácením zkumavky.
- 6 Přidáme 150 µl pufru (AB) a opatrně promícháme pipetováním.
- 7 Přidáme 200 µl chloroformu a promícháme převrácením zkumavky 15 vteřin (Nevortexovat!)
- 8 Zkumavky necháme inkubovat při pokojové teplotě 2 minuty.
- 9 Poté se vzorky centrifugují 2 minuty při 18 °C a 12000 ot.
- 10 Horní fázi nad gelem ve zkumavce přeneseme dekantováním do nové 2 ml zkumavky. UPOZORNĚNÍ: nepřenášíme pipetováním, aby nedošlo k přenosu gelu z Phase Lock Gel zkumavky.
- 11 Určíme objem přenesené vodné fáze, který se může lišit v závislosti na objemu vzorku a objemu homogenizace a extrakce. Přidáme isopropanol v množství 1,75 násobku objemu přeneseného supernatantu.
- 12 Směs promícháme vortexováním po dobu 10ti vteřin.
- 13 Do sběrné zkumavky z kitu dáme purifikační kolonku a přeneseme do ní maximálně 800 µl roztoku vodné fáze s isopropanolem.
- 14 Zkumavku centrifugujeme 20 vteřin při 12000 ot. při 18 °C, odstraníme obsah sběrné zkumavky.
- 15 Krok 14 opakujeme do spotřebování roztoku vodné fáze s isopropanolem ve 2 µl zkumavce.
- 16 Na kolonku přidáme 500 µl promývacího pufru (WB) a centrifugujeme 20 vteřin při 12000 ot. při 18 °C. Tento krok 2x opakujeme (celkem tedy promýváme kolonku WB pufrem 3x).
- 17 Po promytí vzorky centrifugujeme 1 minutu při 12000 ot. při 18 °C pro úplné odstranění zbytků promývacího pufru.
- 18 Přenesem purifikační kolonku do nové 1,5 ml zkumavky.
- 19 Eluční pufr (EB) předehejeme v termobloku nebo vodní lázni po dobu 5 minut na 70 °C.
- 20 Přidáme 10 - 50 µl předeheřátého elučního pufru (EB) do kolonky a inkubujeme 1 minutu při laboratorní teplotě.
- 21 Zkumavky centrifugujeme 1 minutu při 12000 ot. při 18 °C.
- 22 Odebereme kolonku a k extrahované celkové RNA přidáme 1 µl inhibitoru Rnáz.
- 23 V tomto okamžiku je celková RNA vyčištěna a připravena ke kontrole kvality a k dalšímu použití.

4. 3. 2. Extrakce RNA pomocí činidla Trizol

Trizol je chemické činidlo (nebo roztok) používané v izolaci nukleových kyselin posledních 40 let. Trizol je komerční název tohoto roztoku, a existuje celá řada extrakčních činidel s podobným názvem (značky dle různých výrobců) a složením, ale založená na směsi, obsahující guanidin thiokyanát a fenol. Správný název metody izolace RNA by měl být guanidin thiokyanátová fenol - chloroformová extrakce. Použitím Trizolu je možné dosáhnout vyššího výnosu extrahované RNA, vyšší čistoty a stability při skladování. Nevýhodou jsou zdravotní rizika spojená s manipulací s fenolem a chloroformem a také poměrně časově náročná doba centrifugace.

Při práci s Trizolem dodržujeme bezpečnostní pravidla. Je dobré používat dvoje nitrilové rukavice a chránit se tak proti přímému kontaktu s chemickou látkou, včetně nádob a

nástrojů, které s ní přišly do styku. Samozřejmě práce se zdravím poškozujícími sloučeninami, obsahujícími fenol, chloroform, thiokyanát apod., je práce v dobře odvětrané místnosti nebo v digestoři.

Výsledný izolát tvoří buď samotná celková RNA, nebo je v něm zastoupena jak RNA, tak i DNA a proteiny. Všechny tyto složky lze získat i z malého množství tkáně. RNA je izolována neporušená, kontaminace hotového izolátu DNA a bílkovinou je minimální.

Na izolaci používáme biologický materiál v čistém stavu, čerstvý nebo hluboko zamražený (-80 °C). Homogenizace se dá alternativně provádět mechanicky pomocí mlýnků s vysokou frekvencí. V takovém případě se používají mikrozkušavky s kulatým dnem o objemu 2 ml se safe - lock uzávěrem. Před homogenizací se do zkumavky přidávají kuličky, které pomáhají rozbít homogenizovanou tkáň. Kuličky skleněné necháme ve vzorku, kuličky kovové se po sterilizaci dají použít opakovaně.

- 1 Vzorek mycelia (0,1 - 10 mg) vložíme do mikrocentrifugační zkumavky o objemu 1,5 ml. Do zkumavky přidáme 1 ml TRI Reagent předehřátý na 60 °C a celá suspenze se homogenizuje tyčinkou. V některých případech je vhodné vzorek homogenizovat v 500 µl TRI reagentu a zbývajících 500 µl přidat až po homogenizaci vzorku. V takovém případě je nutné vzorek jemně tyčinkou promíchat. Homogenizovaný vzorek inkubujeme 5 minut při pokojové teplotě.
- 2 Po homogenizaci vzorek centrifugujeme při 3000 ot 20 minut při teplotě 4 °C.
- 3 Po centrifugaci odebereme supernatant do nové zkumavky a přidáme 200 µl chloroformu. Vzorek 15 s vortexujeme a 5 min. inkubujeme při laboratorní teplotě. Ve zkumavce se oddělí dvě fáze, jedna růžová naspodu zkumavky, druhá čirá navrchu růžové vrstvy. Pak vzorek centrifugujeme ve 4 °C 15 min. a 12000 ot.
- 4 Do nové zkumavky odebereme opatrně čistou vodnou vrchní fázi supernatantu a opakujeme krok 3.
- 5 Vodnou fázi z horní části převedeme do nové zkumavky a přidáme 500 µl ledového propanolu a 500 µl 1,2 M NaCl. Promícháme a necháme inkubovat při laboratorní teplotě 10 minut. Pak vzorek centrifugujeme 10 minut při 12 000 ot ve 4 °C.
- 6 Odstraníme supernatant a RNA pelet promyjeme přidáním 1 ml 75 % ethanolu. Vzorek promícháme na vortexu a centrifugujeme 5 minut při 8 500 ot. ve 4 °C.
- 7 Odstraníme supernatant a opakujeme krok 6. V této fázi izolace mohou být vzorky RNA pelet v etanolu skladovány nejméně týden v lednici, několik měsíců v mrazáku při -20 °C.
- 8 Po centrifugaci a pečlivém odstranění veškerého supernatantu necháme RNA pelet 5 - 10 min sušit. Je potřeba odstranit stopové zbytky ethanolu, který by mohl ovlivnit následnou kvalitu nukleové kyseliny. Pelet RNA nesmí být sušen delší dobu, protože se zvyšující se dobou sušení klesá rozpustnost RNA.
- 9 K peletu RNA přidáme 20 µl RNA-free vody, případně 1 µl inhibitoru RNAz, a dáme inkubovat na 10 minut do termobloku nastaveného na 60 °C.
- 10 Extrahovanou celkovou RNA používáme hned k dalším analýzám (kontrola kvality, měření koncentrace, cDNA knihovny, sekvenování) nebo můžeme vzorek uchovávat v mrazáku v -80 °C.

4. 4. Kontrola kvality vyizolované celkové RNA.

Kvalitu RNA kontrolujeme na 1% agarózovém gelu. Jedná se o kontrolu, zda není RNA degradovaná, popř kontaminovaná DNA. Kontrola probíhá vizualizací gelu v UV světle.

- 1 Připravíme 1,5% ní agarózový gel. Navážíme 1,5 mg agarózy, kterou rozpustíme ve 100 ml TBE pufru. Směs zahříváme v mikrovlnné troubě do úplného rozpuštění agarózy. Roztok musí být čirý. Po ochlazení pod tekoucí vodou přidáme 2 μ l etidium bromidu.
- 2 Gel necháme 30 min. ztuhnout. Barvivo 6X DNA Loading Dye (Thermo Scientific) rozředíme na koncentraci 1x. Takto zředěný loading dye smícháme v poměru 8:2 s vyizolovanou celkovou RNA. Celkový objem 10 μ l ladderu a kontrolované RNA napipetujeme do vytvořených slotů v gelu. Nanese stejného množství ladderu jako kontrolované RNA umožní orientační posouzení množství vyizolované RNA.
- 3 Po nanese vzorků na gel zapneme zdroj napětí a nastavíme např. 90V na 45 min. běhu elektroforézy.
- 4 Po ukončení elektroforézy vyjmeme gel z TBE pufru a položíme ho na vizualizační UV čtecí zařízení a zkontrolujeme kvalitu extrahované RNA (viz obr.).

TBE pufr 1x: 10,8 g Tris, 5,5 g k. borité, 4 ml 0,5 M EDTA na 1l destilované H₂O

4. 5. Určení koncentrace izolované RNA

Koncentrace extrahované celkové RNA je důležitým hlediskem při přípravě sekvenačních knihoven. Proto je před přípravou vzorků pro sekvenování nezbytné zjistit, kromě čistoty, i koncentraci celkové RNA, kterou chceme sekvenovat. Jednou z metod určení koncentrace nukleových kyselin je fluorometrické stanovení koncentrace pomocí fluorescenčních barev. Toto měření koncentrace je přesnější než např. metody měřící koncentraci spektrofotometricky. U této metody jsou specificky se vážící flurochromy jiné na dsDNA, jiné s afinitou k RNA nebo k proteinům.

Koncentrace se měří pomocí přístroje QubitTM Fluorometer (InvitrogenTM, Massachusetts, USA)

- 1 Pro měření koncentrace celkové RNA používáme Qubit[®] RNA BR Assay Kits (Invitrogene).
- 2 Všechny potřebné chemikálie pro měření koncentrace celkové RNA necháme vytemperovat na laboratorní teplotu. Skladují se v lednici při 4 °C.
- 3 Připravíme si pracovní roztok smícháním Quant-iTTM reagentu s Quant-iTTM pufrům v poměru 1:199 μ l v množství 200 μ l pro každý vzorek + 2 vzorky, které slouží jako standardy.
- 4 Připravíme si kalibrační roztoky standardů pro měření ssRNA. Ve speciálních mikrozkušavkách určených pro fluorometrická měření. Pro kalibrační standard 1 smícháme 190 μ l pracovního roztoku s 10 μ l Standardu 1, pro kalibrační roztok 2 smícháme stejné objemy Standardu 2.
- 5 Vzorky s celkovou RNA připravíme k měření smícháním 180 - 199 μ l pracovního roztoku, připraveného v kroku 2, s 1 - 20 μ l vzorku s RNA do finálního objemu 200 μ l v mikrozkušavkách určených pro měření.
- 6 Standardy i vzorky 2-3 vteřiny vortexujeme a necháme 2 minuty inkubovat při laboratorní teplotě.

- 7 Na fluometru Quibit™ si pomocí šipek vybereme v hlavním menu měření ssRNA, které potvrdíme tlačítkem GO.
- 8 Pak v menu zvolíme možnost nové kalibrace - Run new calibration a opět zmáčkneme tlačítko Go.
- 9 Do měřících otvorů vložíme připravený kalibrační roztok standard 1 a potvrdíme tlačítkem GO.
- 10 Následně vložíme roztok standardu 2, potvrdíme tlačítkem GO a kalibrace pro měření ssRNA je hotová.
- 11 Po kalibraci následně vkládáme vzorky s celkovou RNA připravené k měření a po každém měření zmáčkneme tlačítko GO.
- 12 Koncentrace RNA se zobrazuje na displayi fluorometru pár vteřin po změření vzorku. Celková koncentrace se vypočítává dle následujícího vzorce:

$$\text{Koncentrace vzorku} = Q \times \frac{200}{x}$$

kde

Q = Hodnota fluorescence získaná z Quibit™ fluorometru
x = počet mikrolitrů vzorku použitých v měření (viz krok 4)



Obr. 4. Příklad na měření koncentrace nukleových kyselin – Quibit™.

Takto připravený vzorek, tzn. zkontrolovaný na integritu RNA a s určenou přesnou koncentrací, se nechá osekvenovat na specializovaných pracovištích, která disponují NGS sekvenátory. Např. v laboratoři firmy SeqMe. s.r.o. (Dobříš, CR).

5. NGS sekvenování

NGS sekvenování (next generation sequencing) je jednou z nových metod sekvenování nukleových kyselin. Zavedení těchto metod je spojeno s rozvojem biologických metod využívaných v projektu sekvenování lidského genomu na přelomu tisíciletí. V současné době metody sekvenování NGS převládají a vzhledem ke klesající ceně sekvenování vzorků se stávají stále dostupnějšími v biologickém výzkumu.

Základním znakem NGS technik je opakující se paralelní čtení klonálně amplifikovaných klastrů stejných DNA molekul. Klíčový rozdíl od první generace sekvenování (Sangerovo sekvenování) je tedy to, že u sangerových sekvenátorů představuje jeden vzorek soubor stejných molekul DNA, ze kterých sekvenátor získá jedno unikátní čtení. Sleduje tedy linii jeden vzorek se stejnými fragmenty DNA -> jedna enzymatická reakce -> jedna přečtená sekvence DNA. Celá sekvence DNA ve vzorku je přečtena za předpokladu, že nepřekračuje délku zhruba 1000 bází.

Metoda NGS využívá masivní paralelizaci, kdy v rámci jedné enzymatické reakce jsou přečteny až miliardy fragmentů DNA současně. Sekvenovaný vzorek obsahuje směs různých fragmentů DNA a jen část jich je nakonec přečtena. NGS sekvenování tedy probíhá tak, že jeden vzorek s různými fragmenty DNA -> jedna enzymatická reakce -> některé sekvence přečteny. Tato zdánlivá nedokonalost je ale kompenzována obrovským množstvím výstupních dat a náhodností výběru fragmentů k sekvenování. Ve vzorku je každá sekvence DNA zastoupena v mnoha kopiích. Díky náhodnému výběru fragmentů k sekvenování výstupní sekvenační data obsahují stejný poměr sekvencí jako je ve vstupním vzorku. Není tedy nutné přečíst každý jednotlivý fragment ve vzorku. V případě sekvenování RNA je distribuce sekvencí ve vzorku nerovnoměrná a je třeba zajistit dostatečné množství výstupních dat, aby bylo zajištěno ve výstupu i dostatečné množství málo zastoupených sekvencí RNA. Hlavní výhody NGS jsou generování obrovského množství sekvenačních dat za krátkou dobu a výrazně nižší cena za osekvenovanou bázi. Nejběžněji používanou NGS platformou je Illumina.

RNA sekvenování - RNAseq metoda NGS zahrnuje izolaci RNA a tvorbu cDNA sekvenačních knihoven z vyizolované RNA. Tyto knihovny jsou pak sekvenovány na sekvenátoru a obvykle získáme miliony krátkých sekvencí (cca 150 bp) které se následně zpracovávají bioinformatickými metodami.

Příprava knihoven je kritickým krokem a důležitým prvkem pro kvalitu výsledných sekvenačních dat. Při přípravě knihoven se jako standard v procesu tvorby knihoven pro Illumina sekvenování používá klasická PCR reakce, která zaručí amplifikaci většího množství úseků cDNA, které jsou pak efektivněji sekvenovány.

Na naamplifikované úseky se navážou tzv. adaptory, což jsou řetězce nukleotidů o známé sekvenci, které jsou na cDNA vlákno napojeny v obou směrech. Takto označené úseky cDNA jsou připraveny k sekvenování. Adaptory jsou specifické pro každou sekvenační platformu. Jsou nutné pro procesy při sekvenování a označují specificky každý sekvenovaný vzorek.

Celý proces přípravy vzorků (knihoven) pro NGS metodou PCR je celkem jednoduchý a zvládnutelný v běžné laboratoři. Zároveň je možné pro většinu sekvenačních platform použít při přípravě cDNA knihovny kromě adaptorů i tzv. barcoding*. Ke každému úseku DNA s nalogovaným adaptorem se ještě připojí krátká sekvence, která umožňuje rozlišit

desítky až stovky různých vzorků, které se tak mohou v rámci jednoho běhu přístroje sekvenovat naráz (čímž se také snižuje cena sekvenování).

Většina sekvenačních platform není založena na detekci signálu pouze z jednoho fragmentu připravené knihovny. Je nutné prostorově odděleně namnožit každý fragment na takový počet stejných kopií, aby byl signál natolik silný k identifikaci bází. Tento krok je proveden klonální amplifikací následující po přípravě knihovny. Klonální amplifikace už je součástí technologie NGS.

Pomocí PCR (polymerázové řetězové reakce – polymerase chain reaction) se vytvoří klonální klastry složené ze stejných kopií cDNA fragmentů. Celý proces je lokalizován na místo, kde následně probíhá samotná enzymatická reakce sekvenování. Tato část sekvenátorů se označuje jako flow cell, což je v případě platformy Illumina skleněná destička, která v sobě obsahuje několik plochých kanálků, kterými je možné nechat protékat kapaliny. Vnitřní povrch kanálků je hustě pokryt dvěma druhy specifických oligonukleotidových řetězců komplementárních k jednomu z adaptorů připojeného při přípravě knihovny. Následně se nanese připravená knihovna do flow cell. Po úpravách cDNA vlákna jsou ve flow cell pouze klastry shodných fragmentů zhruba s tisíci jednovláčnými molekulami DNA. Následně je přistoupeno k samotné sekvenaci.

Klonální amplifikace fragmentů knihovny je provedena na vnitřním povrchu samotné flow cell, která obsahuje několik kanálků. Po vytvoření klonů jsou přidány sekvenační primery, které nasednou na druhou část adaptorové sekvence. Samotná sekvenace je prováděna prodlužováním jednovláčkových fragmentů v klastru DNA polymerázou. Sekvenování je zaznamenáváno pomocí fluorescenčních sond (různé barvy pro různé báze). V momentě, kdy je získaná požadovaná délka sekvence, probíhá sekvenování jednotlivých vláken. Získané snímky jsou analyzovány na počítači, který pro každý identifikovaný klastr na snímku převede barevné signály do sekvence nukleotidů. Sekvence s nekvalitním čtením a ze spojených klastrů jsou odfiltrovány.

Z výše uvedeného vyplývá, že samotné sekvenování není praktické provádět v běžných laboratořích. Na NGS sekvenování se specializují komerční laboratoře a firmy (např. u nás Seqme s.r.o, Macrogen Korea s laboratořemi v Německu nebo Španělsku, Eurofins Genomics Německo), které jsou schopny připravit knihovny a osekvenovat požadované vzorky ve velmi dobré kvalitě za přijatelnou cenu.

V lokální laboratoři je možné připravit sekvenační knihovny, které se amplifikují a klonují pomocí komerčně dostupných kitů, ale cena za zpracování vzorků a přípravu knihoven a v neposlední řadě i zkušenosti komerčních laboratořích s přípravou knihoven hovoří ve prospěch zadávání zakázek komerční cestou.

5. 1. Příprava knihoven

Po izolaci celkové RNA a ověření kvality a změření koncentrace se vzorky odešlou do komerční laboratoře (Seqme s.r.o., Dobříš). V laboratoři se provede opět kontrola kvality námi dodané RNA, protože pokud by došlo k degradaci a poškození nukleové kyseliny během transportu, nemá smysl připravovat z nekvalitního materiálu knihovny pro sekvenování. Kontrola kvality probíhá pomocí kitu Qubit RNA BR Assay Kit a přístroje Agilent Bioanalyzer RNA 6000 Nano.

Před přípravou knihovny se odstraní ribozomální RNA (RNA deplece) pomocí kitu NEBNext rRNA Depletion Kit (Human/Mouse/Rat), která odstraní rRNA hostitelského

organizmu, což vede k větší přesnosti při hledání neukaryotních genomů v sekvenovaných datech.

Po kontrole kvality se přistoupí k přípravě knihoven. Používají se kity, navržené přímo pro sekvenační platformu Illumina - NEBNext Ultra II Directional RNA Library Prep Kit for Illumina a NEBNext multiplex Oligos for Illumina (Unique Dual Index Primer Pairs). Následně se zkontroluje i kvalita připravených knihoven pomocí kitů Agilent Bioanalyzer 2100 High sensitivity DNA Kit, Invitrogen Collibri Library Quantification Kit a Qubit 1X dsDNA High-Sensitivity Assay Kit. Takto naklonované a zkontrolované vzorky jsou připravené k sekvenování.

Sekvenování probíhá na platformě NovaSeq 6000 System - 2 x 150 bp oběma směry - 3 - 5 a 5 - 3 (Obr.). Výstupem z komerční laboratoře je zpráva o sekvenování (Obr.), kde jsou informace o tvorbě knihovny (viz výše), informace o kvalitě a koncentraci vstupní RNA, informace o sekvenování, počet readů celkový (sekvenci), průměrná délka readů. Výsledkem NGS sekvenování je soubor se sekvencemi z obou směrů sekvenovaných readů, který je dodán zadavateli a který se dále zpracovává bioinformatickými postupy.



Obr. 5. NovaSeq 6000 bioanalyzátor (<https://www.illumina.com>)

Číslo zakázky: 212233515					Číslo zakázky: 212233515									
Sekvenování					Vzorky									
Číslo vzorku	Objem (nl)	Průměrná délka readů (bp)	Kapacita (milión readů)	Reálná (milión readů)	Číslo vzorku	Konzentrace (ng/ml)	Objem (nl)	Konzentrace (ng/ml)	Číslo vzorku	Číslo	Index	Index2	Velikost (bp)	Směr
...

QC samples:

- Qubit 1X dsDNA High-Sensitivity Assay Kit
- Agilent Bioanalyzer 2100 High-Sensitivity DNA Kit

Library preparation:

- NEBNext Ultra II Directional RNA Library Prep Kit for Illumina
- NEBNext Multiplex Oligos for Illumina (Unique Dual Index Primer Pairs)

Library (R1):

- Agilent Bioanalyzer 2100 High-Sensitivity DNA Kit
- Invitrogen Collibri Library Quantification Kit
- Qubit 1X dsDNA High-Sensitivity Assay Kit

Obr. 6. Zpráva o sekvenování jako informační výstup z externí laboratoře (upraveno).

6. Zpracování sekvenačních dat - bioinformatická analýza RNAseq sekvencí

Výstupem ze sekvenátoru, které jsou dále zpracovávány, jsou dva soubory se sekvencemi, jeden pro každý směr čtení. Zpracování dat může být součástí objednávky NGS sekvenování, protože všechny specializované firmy a laboratoře nabízejí jako součást své nabídky i zpracování výstupních dat. V takovém případě je výsledkem soubor se sekvencemi, odpovídajícími záznamům v databázích (viz dále).

Na druhou stranu, pro alespoň se základními počítačovými znalostmi obeznámeného člověka není problém se se zpracováním bioinformatických dat seznámit. Existuje celá řada automatizovaných postupů, kde většina úkonů, spojených se zpracováním dat, probíhá automaticky a konečným výstupem je soubor s výsledky. Na druhou stranu je celá řada skriptů, programů apod., která je k dispozici na internetu a která může pomoci ve zpracování dat i lidem, kteří nemají programátorské zkušenosti. Je ale třeba říci, že analýza dat je jen začátek, protože hlavním důvodem zpracování dat je hledání biologických údajů v počítačových datech.

Zpracování NGS dat je ale velmi náročné na počítačové vybavení. S malými soubory se dá pracovat i na průměrně vybavených stolních počítačích, ale vzhledem k tomu, že výstupem z analyzátorů jsou mnoha gigabytové soubory, na zpracování celého objemu dat je potřebná i počítačová infrastruktura.

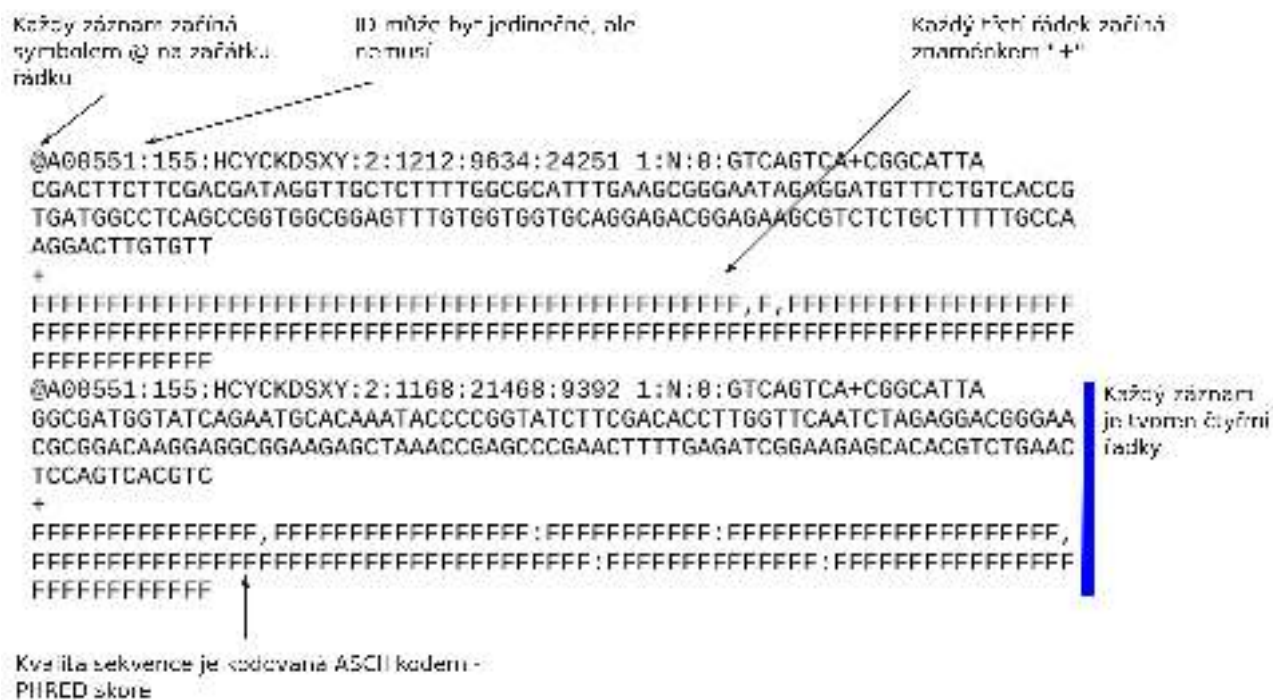
Pro akademickou sféru v ČR je k dispozici platforma MetaCentrum VO (www.metacentrum.cz). Jedná se o počítačovou infrastrukturu, otevřenou všem akademickým pracovníkům a studentům s možností bezplatného využití výpočetní a úložné kapacity a řady aplikačních programů pro zpracování velkého objemu dat (podrobnosti viz stránky Metacentra).

Všechny dále uvedené kroky a postupy se prováděly přes vzdálený přístup na počítačích Metacentrum. Proto budou i příkazy uvedené ve formátu pro počítačovou strukturu Metacentrum. Verze programů a skriptů se mohou lišit od zde popsaných v závislosti na době vydávání nových verzí příslušných nástrojů a programovacích balíčků. Všechny dále uvedené programy a nástroje pro zpracování dat jsou dostupné na platformě Metacentrum a není potřeba je instalovat.

6. 1. Kontrola kvality sekvenovaných dat

NGS technologie umožňuje získání důležitých biologických údajů o genomu celé řady organismů. Přesnost vyhodnocení údajů záleží na hrubých sekvenovaných datech a jejich kvalitě. Je jasné, že během sekvenování dochází k celé řadě chyb, a je proto potřeba většinu těchto chyb odstranit, aby nedocházelo ke zkreslení výstupů při zpracování dat.

Sekvence nukleotidů, dodané po skončení sekvenování, jsou uloženy v souborech formátu FASTQ (.fastq, .fq). Jedná se o textový soubor, ke kterému jsou přidány informace o kvalitě sekvence (obr.). Každý záznam - sekvence - v tomto souboru se skládá ze 4 řádků. První řádek je tzv. FASTA identifikátor, druhá řádka je nukleotidová sekvence. Třetí řádek začíná znaménkem + a znamená konec sekvence. Poslední řádek udává informaci o kvalitě sekvence. Jedná se o tzv. PHRED hodnotu, neboli pravděpodobnost počtu nekvalitních bází.



Obr. 7. Fastq sekvence jako výstup z bioanalyzátoru.

Pro kontrolu kvality NGS sekvencí se použije nástroj FastQC (Andrews 2010, <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Jedná se o software, jehož součástí je i příkaz fastqc (viz dále). Následující sled příkazů provede kontrolu kvality hrubých sekvenovaných dat. Pro jednodušší orientaci jsou sobory s opačným směrem čtení pojmenované R1 a R2.

```

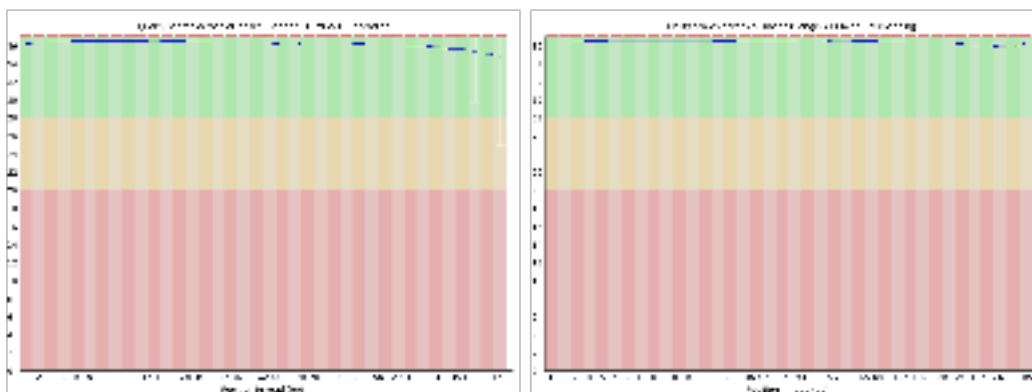
fastqc -t N R1.fq -o výstup
fastqc -t N R2.fq -o výstup

```

kde parametr -t N udává počet procesorů, parametr -o udává adresář, kde budou uloženy výsledky příkazu

Výsledkem jsou dva soubory R1.html a R1.zip pro soubor R1.fq a R2.html a R2.zip pro soubor R2.fq. V komprimovaných zip souborech jsou podrobná data o kvalitě sekvencí a data, ze kterých je tvořen vizuálně přívětivější html soubor. Ten otevřeme např. v prohlížeči Firefox a dostaneme grafický výstup fastqc analýzy. Jedná se o 11 kategorií, které vyhodnocují kvalitu sekvencí ve fastqc souboru. U každé kategorie je symbol, který označuje výsledky kategorie.

Všech 11 kategorií ovlivňuje kvalitu dalšího zpracování dat (mapování sekvencí na referenční genom, kompletování nových kontigů, viz dále), ale nejdůležitějšími hledisky jsou kvalita sekvencí a obsah adaptorů v sekvencích. Některé z kvalitativních parametrů se dají řešit na úrovni přípravy knihovny a během sekvenování, ale následující kroky v analýze sekvencí se zaměřují pouze na odstranění nekvalitních sekvencí a odstranění adaptorů ze sekvencí, a tím ke zvýšení kvality čtení ve fastq souborech.



Obr. 8. Kvalita sekvencí před odstraněním špatných sekvencí příkazem fastqc (vlevo), a po odstranění nekvalitních sekvencí.

Dobré sekvenční jsou v zelené oblasti grafu. Všechny báze mají průměrně kvalitu větší než 30 a proto můžeme mluvit o kvalitně osekvenovaných sekvencích. Jinými slovy PHRED skóre kvalitních sekvencí je nad 30.

Obsah adaptorů je v neupravených sekvencích značný, (viz obr. 8) a proto musíme přistoupit k dalšímu kroku zpracování dat.

6. 2. Odstranění adaptorů (trimming) z hrubých sekvencí

Adaptory jsou krátké řetězce nukleotidů o známé sekvenci, a jsou specifické pro každou sekvenční platformu. Jsou nutné pro procesy při sekvenování, ale je nutné je odstranit při následné analýze sekvenčních dat. Společně s adaptory se ze souborů hrubých sekvencí odstraní i méně kvalitní sekvenční, které by mohly následně negativně ovlivnit výsledky analýzy.

Pro odstranění adaptorů a nekvalitních sekvencí se používá celá řada programů (Trimmomatic, TrimGalore, Cutadapt, Adapterremoval, atd.). Většina těchto programů zná adaptorové sekvence používané na nejběžnějších sekvenčních platformách, ale tyto sekvence se dají zadat přímo do příkazu ať už ve formě sekvence nebo ve formě souboru se sekvencemi adaptorů. Pokud nevíme sekvenci adaptorů a dalších sekvencí, které se přidávaly během tvorby sekvenčních knihoven (barcodes, indexy), vyžádáme si u sekvenční laboratoře tyto informace.

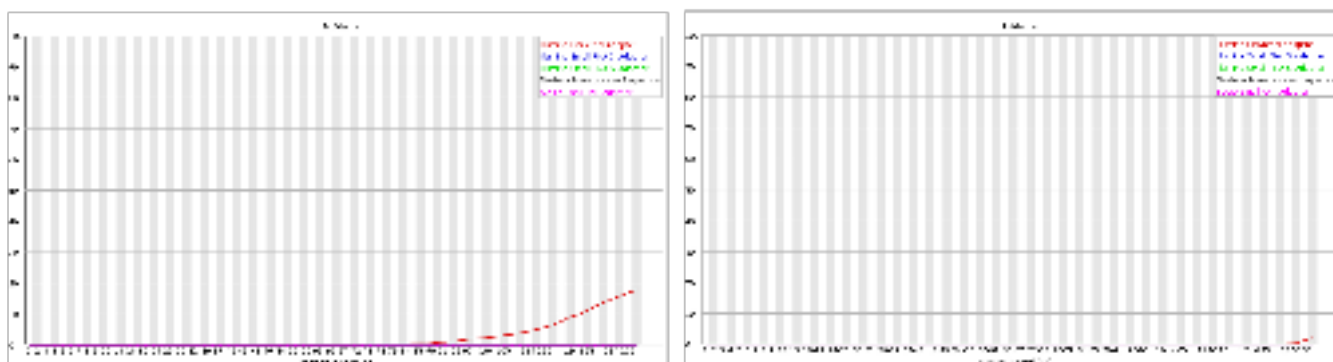
Sekvence adaptorů pro analyzátoři Illumina z laboratoře Seqme jsou pro soubor R1 AGATCGGAAGAGCACACGTCTGAACTCCAGTCA a pro soubor R2 AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT.

Příkaz k odstranění adaptorů je následující (v případě použití programu cutadapt – Martin 2012, <https://cutadapt.readthedocs.io/en/stable>)

```
cutadapt --cores N --quality-cutoff 30 --trim-n --minimum-length 50 -a
AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
-o R1.trim.fq -p R2.trim.fq R1.fq R2.fq
```

Parametr `--cores` udává počet procesorů, parametr `--quality-cutoff 30` udává příkazu odstranit sekvence s kvalitou nižší než PHRED skóre 30, `--trimm-n` odstraňuje konce s neznámými bázemi (NNNNN), `---minimum-length 50` odstraňuje sekvence kratší, než 50 bp, parametr `-a` je sekvence R1 adaptoru, `-A` je sekvence adaptoru R2, parametr `-o` udává název souboru se sekvencemi bez adaptorů a bez sekvencí s nízkou kvalitou čtení (R1.trim.fq pro R1 soubor, R2.trim.fq pro R2 soubor), R1.fq a R2.fq jsou vstupní soubory.

Následuje opět kontrola kvality kontrol sekvencí (viz výše), kde se ověří obsah adaptorů v sekvencích a kvalita sekvencí, když byly odstraněny méně kvalitní sekvence. Vstupními soubory příkazu jsou soubory R1.trim.fq a R2.trim.fq. Výsledky by v kvalitě sekvencí měly odpovídat sekvencím na obr. . Kontrola odstranění adaptoru je na obr. , kde je vidět, že adaptory byly ze sekvencí odstraněny.



Obr. 9. Obsah adaptorů před použitím příkazu cutadapt (vlevo) a po odstranění adaptorů (vpravo).

6. 3. Mapování sekvencí na referenční genom

Vzhledem k tomu, že i po odstranění nekvalitních sekvencí je v souboru celá řada sekvencí, která nemá souvislost s genomem virů, který hledáme, slouží tento krok v analýze sekvenačních dat k odstranění sekvencí, které jsou primárně houbové. Použijeme genom houby václavky smrkové (*Armillaria ostoyae*), který je v databázi GenBank (např. ftp.ncbi.nlm.nih.gov/genomes/genbank/fungi/Armillaria_ostoyae/latest_assembly_versions/GCA_900157425.1_version_2/ z 9.3.2020). Václavka smrková, stejně jako několik dalších druhů václavek, má osekvenovaný genom se známým počtem genů a proto je praktické odstranit ze sekvenovaných dat data, která patří ke genomu houby. Nejenom, že se tím zpřesní následné analýzy, ale také zrychlí, protože odstraněním sekvenovaných houbových dat dojde i ke zmenšení počtu sekvencí a tím se zmenší i vstupní soubory pro následné analýzy.

K dispozici je několik programů, které mapují získané čtení na referenční sekvence (bwa, hisat2 např.) a program STAR (Dobin a kol., 2013, <https://github.com/alexdobin/STAR>), který je použit pro namapování sekvenovaných dat ke genomu *A. ostoyae*. Ten pracuje ve dvou krocích, kdy se nejprve vytvoří indexový soubor genomu referenčního organismu (v. smrková) a následně se k tomuto genomu namapují sekvence z námi analyzovaných souborů.

Pro referenční genom můžeme použít genomovou sekvenci GCA_900157425.1_version_2_genomic.fna.gz, a GTF anotační soubor genů GCA_900157425.1_version_2_genomic.gtf.gz.

Nejprve vytvoříme indexovaný soubor následujícím příkazem:

```
STAR --runThreadN N --runMode genomeGenerate --genomeDir
star_genome --genomeFastaFiles
GCA_900157425.1_version_2_genomic.fna.gz --sjdbGTFfile
GCA_900157425.1_version_2_genomic.gtf.gz. --sjdbOverhang
149
```

Parametr `--runThreadN` udává počet procesorů, `--runMode genomeGenerate` udává příkazu tvorbu indexu, `--genomeDir` je adresář, kde bude index soubor umístěn, `--genomeFastaFiles` specifikuje cestu k souboru s referenčním genomem ve formátu fasta, `--sjdbGTFfile` specifikuje cestu s popisem genomu *A. ostoyae* ve formátu GTF, `--sjdbOverhang` určuje délku sekvencí, ideálně by měla být o jednu bázi menší, než je délka sekvencí v sekvenačním výstupu ze sekvenátoru (v souborech R1.fq a R2.fq)

Výsledkem uvedeného příkazu je vygenerovaný index soubor, ke kterému se mapují soubory s odstraněnými adaptory R1.trim.fq a R2.trim.fq.

Mapování sekvencí k referenčnímu genomu zavedeme následujícím příkazem:

```
STAR --runThreadN N --outSAMattributes All --outSAMtype BAM
Unsorted SortedByCoordinate --outReadsUnmapped Fastx --
quantMode TranscriptomeSAM GeneCounts --limitBAMsortRAM
250000000000 --genomeDir star_genome --readFilesIn
R1.trim.fq R2.trim.fq
```

`--outSAMattributes All` do výstupu celkového jsou zařazeny všechny dostupné atributy
`--outSAMtype BAM Unsorted SortedByCoordinate` formát výstupního souboru - BAM neseřazený, seřazený podle koordinát
`--outReadsUnmapped Fastx` výstup nenamapovaných sekvencí v oddělených fasta souborech, tzn. Že pro každý směr čtení bude vytvořen jeden soubor s nenamapovanými, nebo částečně namapovanými sekvencemi,
`--quantMode TranscriptomeSAM GeneCounts`
`--limitBAMsortRAM 250000000000` maximum operační paměti pro seřazení BAM souboru
`--genomeDir` ukazuje cestu k indexovanému referenčnímu genomu `star_genome`
`--readFilesIn` specifikuje vstupní soubory R1.trim.fq a R2.trim.fq

Výsledkem alignovacích programů je obvykle soubor ve formátu SAM (Sequence alignment Map), v zásadě textový soubor se seznamem genů a namapovaných sekvencí a nebo, jako v našem případě BAM soubor (Binary alignment Map), což je v zásadě zkomprimovaný SAM soubor.

6. 4. *De novo* seskupování (assembly) nenamapovaných sekvencí

De novo znamená latinsky od začátku. V tomto slova smyslu se jedná o skládání sekvencí bez referenčního genomu, jinými slovy bez toho, aniž bychom znali jak sekvence vypadají. Skládání a tvorba nových sekvencí je finálním krokem bioinformatické analýzy.

Programy, které seskupují sekvence dohromady, se nazývají assembly. Výsledkem je fasta soubor se sekvencemi různé velikosti které označujeme jako blízké sekvence - kontigy (contigs - contiguous sequences) nebo jako scaffolds (česky možná konstrukce). Scaffolds jsou části genomu rekonstruované z kontigů a obsahují mezery.

Programů (assemblerů), které dávají dohromady sekvence genomu, je celá řada, pracují s různými algoritmy. Dále jsou popsány příkazy dvou assemblerů, které se používají nejčastěji. Je také potřeba říci, že pokud máme stejný soubor s nenamapovanými sekvencemi, a použijeme různé assembly, může se stát, že se některé finální sekvence budou lišit.

První assemblerem, který je dnes široce používán při bioinformatické analýze, je program Trinity (Grabherr a kol., 2011, <https://github.com/trinityrnaseq/trinityrnaseq/wiki>). Assembler Trinity spustíme následujícím příkazem:

```
Trinity --seqType fq --left R1_unmap.fq --right R2_unmap.fq
--CPU N --max_memory 480G --min_contig_length 500
```

kde parametr `--seqType` znamená, v jakém formátu jsou vstupní soubory, parametry `--left` a `--right` jsou parametry pro soubory se sekvencemi podle směru jejich čtení, `--CPU` počet procesorů, `--max-memory` je maximální možná paměť pro běh příkazu a `--min_contig_length` je minimální délka složených výsledných sekvencí.

Výsledkem běhu příkazu je soubor Trinity.fasta, který obsahuje většinou několik tisíc různě dlouhých nově sestavených sekvencí:

```
>trinity_DN11763_c0_g2_i1
GACATATACGTCCATGACCCATATCCCGAAATAATTCCCTTCTCCAACCTTCGCACCCCACTTCAAC
CTACGACTACACGATCCTCTCGAGTCTGCCATACTGCTAGATATAGCCCGTGGTATTCCAATTCT
TCCATATACGATGTTTTCGAAGCCAGAGTGCATCCTGGTCGAAAGAGATACGTCTTTTACACGGA
GTAAAGAGGGAGGCTGCCGTTAACTGGGATGTATTCTTCGAGGAAGGGTGCCAGCATCTATCCATT
GCAGACGGTGTGGATCCCAATCCCAGATGGCAAGACGCATTCATCATCAGCTAAAGAAATTCGA
GAGAGTATCCGTGGTCGTCGGTTCTTTCTCGGGCAACCCATAAGAACATGAGCAGGAGTACACGA
CGAGACTTCGAGAGACAAACAGGGCATTCAATTGGAGGGGATTCCCATCTTCGGACAAGATAATTGG
GGTGCGCATTACCACAAGACGGGGAAGAACTAGGTGGTTCCAGCGAGATGAGACAGAAGTGGTAT
CACCATGGTGCAAAGCCAGGACCTACTTCGCAATGGGGGGCGAGGCGTACGAGGCTTGTCGATTC
CTACAGGACTTCTTCACCGATATAGTAGGCTTCTTCATGCCAACAACCCACAAAACACGGTTACAG
CCAGATCGTCTGTTCTATCATCAAATATGACAAAGAGGACCCCATTTCCGCATTTAC
>trinity_DN11763_c0_g1_i1
GATATGGGTCATGGACGTATATGTCAGGATGTTGTTTCGATCATTTGGCGCAAGGTGGAGAAGGAGT
CGGGTCGTGGCCATGATTGCGCTTGGCAGGATTTGATCAGGAGATTACGTGCTGAAAGCGGGAGTC
GAGAGAGAGTTCCGATAACCGTACGGTATTCGGGGTACACCAGAGAGAAGAGTTCCTTCAGATCGG
TTCGGGTGTGCATCCTCTCCTCTGCTGCTCTTCTTAGAACAACTTGAGAGTGCTCCGGTATTTTCA
GAAGAAGTATCGAGCGAGCAACACACACACCCCGAAGATCATGAACTCCTTAATCTATCTCGGTG
AAACCCCATCCGACCGACAATCGGTATCCG
```

Druhým assemblerem, který se v současné době nejčastěji používá, je program SPAdes (Bankevich a kol., 2012, <https://cab.spbu.ru/software/spades>). Ten spustíme příkazem

```
spades.py -t N -m 480 -1 R1_unmap.fq -2 R2_unmap.fq -o  
spadesRNA --rna
```

kde parametr -t znamená počet procesorů, parametr -m určuje maximálně použitelnou operační paměť pro běh příkazu, parametry -1 a -2 jsou vstupní soubory příkazu s nenamapovanými sekvencemi, parametr -o je parametr pro adresář výstupních souborů a přepínač --rna znamená, že program přistupuje k datům jakoby pocházely ze sekvenování RNA, tzn. že skládá sekvence genů, nikoli celý genom.

Výsledkem běhu spades s parametrem --rna je soubor transcripts.fasta (pokud nepoužijeme přepínač --rna je výsledkem soubor scaffolds.fasta), který, stejně jako v případě Trinity je tvořen několika tisíci sekvencemi různé délky:

```
>NODE_10213_length_501_cov_0.440299  
ATGGGTTATAGCGAGCGGGACGTGGAAGGCATCAGATGGCAATCATCGGTTATGCGCGAC  
GCTGAACTCACCTCAAAGGGTTGGAGCAGGATACGTCCCGGTCTCGAGGAGGCGATCAA  
GAAAACCGAGTTCGGCAAGCCAAAGCGGACCCTAGTACGGCGCTATACGCACGTGCTGAA  
ATTGCGAGTGGCCTCTCCGTA AAAATACGCACCGATTGCGTACACGTCCAACATTGCTACC  
GTGAATACCAAAGGACAGGCTGGTTTTGATGAACACGGGACGGTTCGTGCCAAGATATTTA  
AGAGCGGCCCGGATATGAATTACCCTCCAGCCACGACATCGCCTTTCTCGGCGGAGTCCTG  
AACAATGTTACGCTAAACTTGTACTTCAATACTCGCTAACAAAGGCGGGTTCAGCCCATA  
GGTCACTCCGACAGACGGCAGTCGCCGCCAAGGGAAAGGCTGGCGAAAGTTTCTTACTCT  
ACACGATCATAACCATGCGGAC  
>NODE_10214_length_500_cov_22601.112219  
AATAGATTGTTAAGTGAAAATCAAGGACGTGACAAAATCTTATACAGTTGTGAAGTGAGC  
TTGGTAGTCGCTACTTATAATGATCGTATGTAACAACGCATCAACAGTTTTTAGCTTTGT  
TACGCCGAAAATTTAACGGGTCTTAAACAATCAACCGATATCGTGTTGATATTGAATTTA  
TAAATTTTACCCTTCTATTTAAAAGTTAAACCCCAACCTTCCCCTACTATATCTTTGTA  
AAAGTGGGTAGAAAAGGGTTGTCGTTAACGCTAGGGCTAAATGTTTATTCTTTATCTAGG  
TAGTAGAACATTCAGTCAATCTATTTATAAAATAGTGTTTCATTATTTTTTAGGCAACTG  
AAGAGAGAATGCTGACATGAGTAACGTAAAAGAAGAATATAATTCTTCTCGCCGAAAACG  
AAAGGGTTACAATAAAAGGTTAAATCATTTTGTGTTAATGCGGCCTCTAAGGACTATAAC  
CAAAGTTATGGCTGATGAGT
```

6. 5. Analýza nově sestavených sekvencí

Nově sestavené sekvence - kontigy a scaffolds v příslušných, výše uvedených souborech (viz kap.) - je potřeba srovnat z doposud známými sekvencemi a zjistit, zda se v námi sekvenovaných datech vyskytují nové, doposud neobjevené nebo nepublikované virové sekvence. Tento proces porovnávání sekvencí se nazývá hledání podobných sekvencí (sequence similarity search) a je jedním z důležitých kroků v bioinformatické analýze hledání nových virových sekvencí. Nejpoužívanějším nástrojem je tzv. BLAST (Basic local alignment search tool – <https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST>, Altschul a kol., 1990), který porovnává sekvence s databází již známých sekvencí (Sayers a kol., 2019). Tento program byl vyvinut Národním centrem biologických informací (NCBI - Bethesda, MD), které také spravuje NCBI databázi všech dosposud deponovaných sekvencí, nukleotidových i proteinových. Úkolem BLAST programu je najít co nejpodobnější sekvence našim, nově vytvořeným. Vložíme sekvenci, která nás zajímá, a hledáme v databázích sekvencí identickou nebo podobnou. Vzhledem k tomu, že v poslední době roste počet záznamů v databázích exponenciálním tempem, jsou některé databáze připravené přímo a pouze na virové sekvence, které se dají použít. Nicméně je třeba si dát pozor na stáří a dobu obnovování nových záznamů v těchto databázích.

Je třeba říci, že BLAST hledání v databázích je jedna z časově nejnáročnějších částí bioinformatické analýzy, kdy doba trvání běhu příkazu může být i několik dní, než jsou všechny námi vytvořené sekvence porovnány s databází. Soubory s kontigy obsahují několik tisíc sekvencí. Proto je např. lepší rozdělit soubor na několik menších souborů, např. podle velikosti kontigů, a všechny kontigy kratší než např. 500 (nebo 1000, záleží na tom, co hledáme) bp odstranit ze souboru úplně.

Prvně se zjišťuje, zda-li jsou v nových sekvencích úseky, odpovídající již známým virovým sekvencím. Prvně porovnáme sekvence s databází virových referenčních genomů z NCBI (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/> -> all nucleotides)

```
blastn -num_threads N -query soubor_s_contigy.fasta -db
navez_database -max_target_seqs 1 -evaluate 1e-3 -outfmt '6
qseqid pident length qlen slen qstart qend evaluate bitscore
salltitles sscinames' -out soubor_s_vysledky.txt
```

kde -num_threads je počet procesorů, -query je vstupní soubor s kontigy (v našem případě transcripts.fasta nebo Trinity.fasta), -max_target_seqs je parametr, který udává maximální počet cílových sekvencí podobných vstupním kontigům, -evaluate je parametr udávající jaká je pravděpodobnost že je dotazovaná sekvence přítomná v databázi (expected value), -outfmt je parametr, kterým se určuje podoba výstupního souboru (viz dále) a -out je název výstupního souboru s výsledky hledání. Blastn příkaz znamená hledání v nukleotidových databázích, kdy se porovnávají nukleotidové sekvence s nukleotidovými sekvencemi.

Výstupem je textový soubor s několika sloupci. První sloupec identifikuje kontig, u kterého se našla v databázi shoda (qseqid). Druhý sloupec udává procento podobnosti kontigu s vyhledanou sekvencí (pident). Třetí sloupec udává délku alignmentu obou sekvencí (length). Délka sekvence kontigu udává, jak je dlouhá sekvence (qlen) ve srovnání s délkou referenční sekvence (slen). Začátek (qstart) a konec (qend) alignmentu udává, v jak je dlouhá sekvence, kde se jednotlivé nukleotidy (proteiny v případě blastx, viz dále) překrývají. Evaluate (evaluate) je hodnota, která odpovídá počtu záznamů, které by byly statisticky významně podobné zájmové sekvenci a zároveň měly nízkou pravděpodobnost,

že se natolik podobné záznamy v databázi vyskytují náhodou – čím nižší je tato hodnota, tím lepší je výsledek hledání.

Contig	% podobnosti s referenční sekvencí	Délka alignmentu	Délka sekvence contigu	Délka referenční sekvence	Začátek alignmentu	Konec alignmentu	Poslední očekávaná hodnota	Název organismu v BLAST databázi
NCBI_25_knjg_3037	96,77%	4245	2130	4522	1	4042	0,0	[Parvovirus ambli-like virus 2 strain AT29, complete genome]
NCBI_10_knjg_2221	95,76%	2531	2291	4522	729	2802	0,0	[Parvovirus ambli-like virus 2 strain AT29, complete genome]
NCBI_12_knjg_2043	96,99%	2477	2130	4522	611	3106	0,0	[Parvovirus ambli-like virus 2 strain AT29, complete genome]
NCBI_307_knjg_1423	96,90%	2422	2190	4522	77	1-92	0,0	[Parvovirus ambli-like virus 2 strain AT29, complete genome]

Obr. 12. Pozitivní shoda sekvencí s nukleotidovými sekvencemi z NCBI databáze. Popis sloupců viz text.

Dále můžeme kontigy porovnat s nukleotidovou databází všech organismů, spravovanou NCBI (<https://www.ncbi.nlm.nih.gov/nucleotide>)

```
blastn -num_threads N -query soubor_s_contigy.fasta -db nt
-max_target_seqs 1 -evaluate 1e-5 -outfmt '6 qseqid pident
length qlen slen qstart qend evaluate bitscore salltitles
sscinames' -out soubor_s_vysledky.txt
```

Nakonec porovnáme námi vytvořené kontigy s proteinovou databází (např. nr z NCBI, swissprot z EMBL), v našem případě s databází UniProt (<https://www.uniprot.org/uniprot/?query=taxonomy:10239> pro viry). Pokud není žádná nebo minimální shoda našich kontigů s nukleotidovými záznamy, použijeme srovnání kontigů s proteinovou databází. Databáze UniProt obsahuje referenční sekvence proteinu, genů a popisů, nutných pro studium exprese genů a proteinů.

```
blastx -num_threads N -query soubor_s_contigy.fasta -db
uniprot -max_target_seqs 1 -evaluate 1e-5 -outfmt '6 qseqid
pident length qlen slen qstart qend evaluate bitscore
salltitles sscinames' -out soubor_s_vysledky.txt
```

Contig	% podobnosti s referenční sekvencí	Délka alignmentu	Délka sekvence contigu	Délka referenční sekvence	Začátek alignmentu	Konec alignmentu	Poslední očekávaná hodnota	Název organismu v BLAST databázi
NCBI_1_knjg_1552	34,26%	137	1552	627	1907	2771	5,13e-49	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
NCBI_15_knjg_4571	30,67%	590	4571	590	1820	2667	0,0	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
NCBI_16_knjg_4576	30,20%	717	4576	717	279	2114	0,0	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
NCBI_129_knjg_5881	32,51%	690	5881	690	2616	4222	0,0	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
NCBI_199_knjg_5517	31,21%	471	5517	471	2735	3222	9,12e-51	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
NCBI_209_knjg_5507	29,22%	512	5507	512	2949	3224	4,22e-43	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
NCBI_209_knjg_5507	27,55%	639	5507	732	2949	3242	7,22e-42	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
NCBI_264_knjg_4116	34,83%	691	4116	691	2758	2951	0,0	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
NCBI_308_knjg_4410	38,36%	469	4410	469	2	2519	0,0	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
Contig_069199_01_01_01	62,44%	412	666	767	1181	8	3,4e-48	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
Contig_069199_02_01_01	33,65%	449	727	727	2064	4774	5,32e-52	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
Contig_069199_03_01_01	31,21%	421	727	579	2064	4447	1,12e-50	[Hypovirus 1 strain Yersinia enterocolitica strain 2]
Contig_069199_04_01_01	37,47%	537	2142	537	1591	2122	0,0	[Hypovirus 1 strain Yersinia enterocolitica strain 2]

Obr. 13. Pozitivní shoda nových kontigů s proteinovými sekvencemi ambi – like známých virů z UniProt databáze. Popis sloupců viz text výše.

Těmito kroky a srovnání sekvencí s databázemi nukleotidových a proteinových sekvencí jsme získali záznamy o potenciálních virových sekvencích. Pokud budeme popisovat nové viry, je nutné ale všechny tyto sekvence otestovat a zjistit, jak jsou viry v hostiteli uspořádány, jestli se jedná o provirus, nebo o celý virus a jeho přesnou sekvenci.

Pokud nebudeme chtít nový virus popisovat a bude nám stačit pouhá jeho identifikace, můžeme rovnou přistoupit k testování jeho virulence a přenosu.



Obr. 14. Sekvence nově detekovaného ambi-like viru s vyznačenými kodujícími oblastmi v podobě ORF (open reading frames). Jeden ORF je ve směru 3' -> 5', druhý ORF ve směru 5' -> 3'.

7. Vyšetření biologického materiálu na přítomnost mykovirů

V momentu, kdy máme k dispozici sekvenci viru, získanou bioinformatickou analýzou NGS dat, můžeme začít testovat přítomnost virů v jednotlivých izolátech z biobanky václavek. Postup spočívá v navržení specifických primerů pro sekvenci ambi-like virů, screening přítomnosti ambi-like virů v izolovaných kulturách václavek a testech, vedoucích k přenosu virů z infikovaných na neinfikované houby. K tomuto přenosu dochází pomocí hybridizace hyfových vláken, která jsou infikovaná mykovirem a hyfových vláken bez přítomnosti mykovirů. Zároveň jsou ověřovány kmeny, které budou vhodnými adepty na pokusy s přenosem viru, protože se v nich ambi-like viry nevyskytují.

7.1. Navrhování primerů

Prvním krokem v analýze mykovirů v kmenech hub, u nichž nebylo prováděno NGS sekvenování, je navržení specifických primerů pro část sekvence ambi-like viru. Po PCR analýze je amplifikovaný fragment virové sekvence identifikován vizuálně na agarózovém gelu a poté je osekvenován a získaná sekvence je porovnána pomocí nástroje BLAST s databází sekvencí NCBI (GenBank). K navrhování primerů existuje celá řada programů, nejpoužívanějším je software Primer3 (Untergasser a kol., 2012), který je součástí balíku Geneious.

1. Sekvenci sestavenou z NGS sekvenování nahrajeme do programu Geneious (ve verzi 8.1.9 – Biomatters, Ltd.) a spuštíme program Primer3. Program sám vyhledá nejvhodnější sekvence primerů ve směru 3' -> 5' i ve směru 5' -> 3'.
2. Vybereme sekvence primerů (forward a reverse primerový pár), které odpovídají nejlépe zadaným kritériím – např. navržený fragment byl dlouhý ideálně kolem 500 bp, oba F i R primery měli co nejpodobnější T_m (melting temperature), primery nesmí vytvářet vlásenky a dimery, poměr G-C párů bází by měl být ideálně mezi 40-60 % a primery by se měly primárně navrhovat z konzervované části DNA.
3. Následně vybrané primery porovnáme pomocí BLAST s databází. Zvolené sekvence primerů se nesmí vyskytovat v jiném organismu, který by se mohl vyskytovat společně s námi zkoumaným materiálem. .
4. Sekvence primerů, které jsme vybrali, necháme nasynthetizovat v komerční laboratoři (např. KRD Praha, Macrogen Europe, Holandsko).
5. Příklad navržených primerů pro sekvenci ambi-like virů jsou v tabulce 1

Tab. 1. Navržené sekvence primerů pro ambi-like virus

Název primeru	Sekvence primeru
Ambi F	GCTATGGCTGACTCCTCGTC
Ambi R	ACAGGGCAATCATTGGAGGG

7. 2. Reverzní transkripce izolované celkové RNA

PCR reakce probíhá na templátu DNA, analýza je založena na replikaci DNA a klíčovým enzymem je DNA polymeráza. Proto je nutné izolovanou celkovou RNA (viz kap. 4.3.) přepsat procesem reverzní transkripce na cDNA, která slouží jako templát pro PCR s primery pro ambi-like virus. Existuje celá řada komerčních kitů pro reverzní transkripci. Jedná se v zásadě o jednu reakci kombinující reverzní transkripci a PCR, která probíhá v termocyklerech, stejně jako klasická PCR. Proces reverzní transkripce probíhá podle manuálu výrobce kitu.

Reverzní transkripce se připraví pomocí kitu, např. kitem ImProm-II Reverse Transcription System (Promega, Madison, USA)

1. Na ledu smícháme 0,5 μ l forward primeru, 0,5 μ l reverse primeru, 3 μ l celkové RNA a 1 μ l vody bez nukleáz.
2. Směs necháme denaturovat při 70 °C 5 minut, poté okamžitě přemístíme zpět na led a necháme inkubovat alespoň 5 minut.
3. Připravíme reakční mix pro reverzní transkripci smícháním následujících komponent na ledu.

Voda bez nukleáz	3,7 μ l
ImProm-II 5x reakční pufr	4 μ l
MgCl ₂	4,8 μ l
dNTP mix	1 μ l
Rekombinantní RNasin ribonukleázový inhibitor	0,5 μ l
ImProm-II reverzní transkriptáza	1 μ l
<hr/> Finální objem	<hr/> 15 μ l

4. Celý namíchaný objem 15 μ l reakčního mixu přidáme k denaturované celkové RNA s navrženými ambi-like F i R primery.
5. Vzorek vložíme do termo-bloku a nastavíme teplotní profil reakce následovně

25 °C - 5 minut
42 °C - 60 minut
70 °C - 15 minut

6. Výsledkem reakce je cDNA, kterou můžeme použít pro další analýzy, popř. zamrazit při -80° C do doby použití.

7.3. Klasická PCR reakce a důkaz ambi-like virů

Jako templát klasické PCR slouží cDNA získaná reverzní transkripcí. Klasickou PCR provádíme proto, abychom:

- 1) potvrdili a zjistili, kde v NGS sekvenovaných vzorcích hub se ambi-like virus nachází, protože celková RNA se izoluje jako pool několika vzorků václavek,
- 2) zjistili, jestli se ambi-like viry nenacházejí také v jiných, NGS nesekvenovaných, izolátech a vzorcích mycelia václavek.

Pro PCR použijeme navržené primery specifické pro námi identifikované ambi-like viry (viz výše).

Příprava reakční směsi:

H ₂ O	4 µl
Ambi F primer	1 µl
Ambi R primer	1 µl
PPP Master Mix (Top-Bio, Vestec, ČR)	10 µl
cDNA	4 µl
<hr/>	
Finální objem	20 µl

Mikrozkumavky s PCR směsí vložíme do termocykleru a spustíme reakci o následujícím teplotním profilu:

	Počáteční denaturace:	94 °C 2 minuty
	Denaturace	94 °C 1 minuta
25 cyklů:	Annealing	60 °C 1 minuta (dle T _m primerů)
	Elongace	72 °C 2 minuty
	Konečná elongace:	72 °C 5 minut

Po dokončení PCR vizualizujeme amplifikované fragmenty na agarózovém gelu a amplifikované PCR produkty necháme osekvenovat v externí laboratoři s ambi-like primery za účelem potvrzení specifčnosti analýzy. Výsledné sekvence porovnáme s databází sekvencí NCBI (podrobnější postup viz. Čurn a kol., 2019).

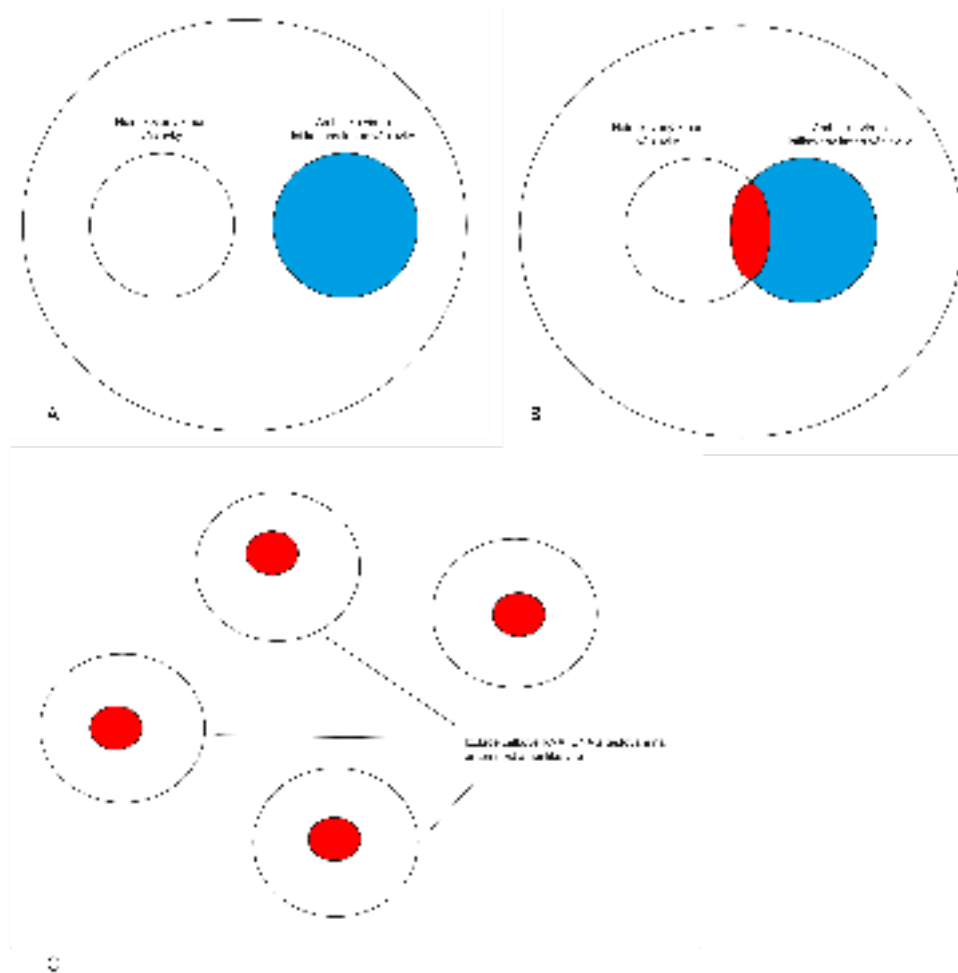
8. Testování přenosu mykovirů v laboratoři

Vzhledem k tomu, že se mykoviry považují za potenciální prostředky biologické ochrany proti celé řadě fytopatogenních hub, je důležité testovat přenositelnost mykovirů mezi jedinci populací fytopatogenních druhů hub. Doposud byly popsány přenosy mykovirů sporami nebo splynutím hyf anastomózou (Nuss 2005; Zhao a kol. 2020; Yu a kol., 2013). Z výše uvedeného vyplývá, že přenos mykovirů mezi jedinci je možný a to může mít i potenciální dopad na biologickou ochranu proti václavkám.

8.1. Hybridizace kmenů václavek

Přenos mykovirů splýváním hyf je proces, kdy dochází k výměně genetických informací mezi jedinci a možnému přenosu informací z jednoho jedince na druhého. Testy jsou postavené na prorůstání hyfových vláken infikovaných mykovirem a neinfikovaných mykovirem kmenů václavek na jedné agarové misce. Popsaný postup je graficky znázorněn na obrázku 15.

1. Připravíme si misky s ME agarem
2. Na misku s agarem dáme na jednu polovinu inokulum mycelia kmene václavky, která není infikována ambi-like virem. Na druhou polovinu agarů v misce dáme kmen václavky, který je ambi-like virem infikovaný (obr. 15A.)
3. Jako kontrolu nainokulujeme každý kmen václavky zvlášť na samostatnou misku s agarem.
4. Misky umístíme do termostatu a necháme kultivovat (viz Čurn a kol., 2019).
5. Po cca 3 týdnech misky z termostatu vyndáme a část nově narostlého mycelia na styku obou kmenů odebereme a nainokulujeme na nové misky s agarem. Opět dáme kultivovat do termostatu (obr. 15C).
6. Pasážování provádíme po třech týdnech. Nukleové kyseliny ale můžeme izolovat již po 2 týdnech, kdy odeberem mycelium a vyizolujeme celkovou RNA a/nebo DNA. Extrahované kyseliny zpracujeme příslušnými postupy (viz předešlé kapitoly, Čurn a kol., 2019). Nukleové kyseliny extrahujeme i z kontrolních vzorků.
7. V případě, že PCR reakcí s příslušnými primery získáme amplikony mykovirů v původně neinfikovaném kmeni a Sangerovo sekvenování potvrdí přítomnost sekvencí mykoviru v původně neinfekčním kmeni, dosáhli jsme hyfálního přenosu mykoviru.



Obr. 15, Schematické znázornění postupu testování přenosu virů.

9. Srovnání novosti postupů

V současné době není v ČR k dispozici dostupná metodika, zabývající se touto problematikou. Dosud dostupné informace jsou dílčí, velké množství informací týkajících se metod NGS je dostupné v celé řadě vědeckých publikací a monografiích, zabývajících se problematikou detekce virových nákaz u hub, ale nikde dostupné v takto podrobné a komplexní podobě.

Předkládaná metodika zahrnuje popis všech metodických a analytických postupů od sběru a kultivace hub, přes izolaci nukleových kyselin a jejich zpracování pro NGS analýzu až po podrobný popis bioinformatického zpracování NGS dat a popis detekce nových virů.

Novost postupu spočívá v kombinaci metod molekulárně biologických a bioinformatických použitelných k detekci virů a jejich identifikaci u václavek, které nebyly doposud v ČR pro diagnostiku virů václavek použity.

10. Popis uplatnění metodiky

Metodika "Metodika identifikace a determinace mykovirů u hub rodu *Armillaria*" v první části popisuje teoretický úvod do problematiky. V praktické části jsou uvedeny postupy potřebné pro nakládání s biologickými vzorky, izolaci nukleových kyselin, přípravu vzorků pro NGS sekvenování a vyhodnocení sekvenačních dat pomocí bioinformatických nástrojů pro detekci viru v houbových hostitelích.

V metodice jsou shrnuty soubory optimalizovaných postupů a návodů pro molekulární a bioinformatickou analýzu sekvencí nových virů, na jejichž základě lze dle uvedených postupů provádět analýzy houbových hostitelů s cílem optimální detekce přítomnosti RNA virových elementů. Výstupem analýzy je detekce potenciálních biologických agens, použitelných v rámci integrované, popř. biologické ochrany proti václavkám, resp. houbám obecně.

Uživatelé metodiky jsou výzkumná lesnická pracoviště, která mohou využít přednosti optimalizovaného postupu detekce nových potenciálních biologicky aktivních virových elementů u patogenních hub. Metodiku mohou využít v rámci vědecké činnosti i vysoké a střední školy se zaměřením na ochranu lesa. Metodika se může uplatnit v celé řadě výzkumných a vysokoškolských organizací vzhledem k obecnému postupu při detekci nových organismů metodami NGS.

11. Ekonomické aspekty

Předpokládané přínosy metodiky v rovině ekonomické nelze přesně vyčíslit. V případě úspěšné identifikace viru, který bude zásadně ovlivňovat vztah václavka - smrk ve prospěch smrku, by byly ekonomické přínosy vyčísleny do výše několika stovek tisíc korun ve formě vyšších výnosů z produkčních ploch a kvalitní, výhodně obchodovatelné dřevní hmoty. Identifikace viru vhodného k biologické ochraně smrků proti václvkám by také přinesla finanční prospěch ve formě užitečných vzorů a patentových přihlášek.

Dalšími přínosy předkládané metodiky jsou rozšíření spektra technik a metodických postupů používaných v diagnostických laboratořích, rozšíření portfolia technik a služeb prováděných v laboratoři a ekonomický přínos související s těmito službami a v rovině metodické a vzdělávací. Metodika bude využita ve výzkumných laboratořích, na vysokých a středních školách.

12. Seznam použité literatury

- Abdoulaye A.H., Foda M.F., Kotta-Loizou I. 2019: Viruses Infecting the Plant Pathogenic Fungus *Rhizoctonia solani*. *Viruses* 11(12): 1113. doi:10.3390/v11121113
- Altschul S. F., Gish W., Miller W., Myers E.W., Lipman D. J. 1990: Basic local alignment search tool. *J Mol Biol.* 215 (3): 403-10. doi: 10.1016/S0022-2836(05)80360-2.
- Andrews S. 2010: FastQC: AQuality Control Tool for High Throughput Sequence Data. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Antonín V., Tomšovský M., Sedlák P., Májek T., Jankovský L. 2009: Morphological and molecular characterization of the *Armillaria cepistipes*—*A. gallica* complex in the Czech Republic and Slovakia. *Mycol. Prog.* 8: 259–271. doi: 10.1007/s11557-009-0597-1.
- Boland G. J. 1992: Hypovirulence and double-stranded RNA in *Sclerotinia sclerotiorum*. *Can. J. Plant Pathol.* 14, 10–17. doi: 10.1080/07060669209500900
- Bryner S. F., Rigling D., Brunner P. C. 2012: Invasion history and demographic pattern of *Cryphonectria hypovirus 1* across European populations of the chestnut blight fungus. *Ecol. Evol.* 2, 3227–3241. doi: 10.1002/ece3.429
- Coetzee M.P.A., Wingfield B.D., Wingfield M.J. 2018: *Armillaria* Root-Rot Pathogens: Species Boundaries and Global Distribution. *Pathogens* 24;7(4):83. doi: 10.3390/pathogens7040083.
- Čurn V., Tonka T., Křížová L., Jozová E. 2019: Metodika izolace DNA a analýzy molekulárních markerů u hub. ZF České Budějovice. ISBN: 978-80-7394-781-1
- Dobin A., Davies C. A., Schlesinger F. et al. 2013: STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* Vol 29 (1), Pages 15–21. doi: 10.1093/bioinformatics/bts635
- García-Pedrajas M.D., Cañizares M.C., Sarmiento-Villamil J.L., Jacquat A.G., Dambolena J.S. 2019: Mycoviruses in Biological Control: From Basic Research to Field Implementation. *Phytopathology.* Nov. 109(11): 1828-1839. doi: 10.1094/PHYTO-05-19-0166-RVW
- Ghabrial S. A., Caston, J. R., Jiang, D., Nibert, M. L. & Suzuki, N. 2015: 50-plus years of fungal viruses. *Virology* 479–480, 356–368. doi: 10.1016/j.virol.2015.02.034
- Grabherr M.G., Haas B.J., Yassour M., et al. 2011: Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29 (7): 644-652. doi:10.1038/nbt.1883
- Guillaumin J.-J., et al. 1993: Geographical distribution and ecology of the *Armillaria* species in western Europe. *Eur. J. Plant Pathol.* 23: 321 - 341.
- Hao F., Wu M. Li, G. 2018: Molecular characterization and geographic distribution of a mymovirus in the population of *Botrytis cinerea*. *Viruses* 10:432. doi: 10.3390/v10080432

- Herrero N., Dueñas E., Quesada-Moraga E., Zabalgoceazcoa I. 2012: Prevalence and diversity of viruses in the entomopathogenic fungus *Beauveria bassiana*. *Appl. Environ. Microbiol.* 78, 8523–8530. doi: 10.1128/AEM.01954-12
- Hillman B. I., Aulia A. & Suzuki N. 2018: Viruses of plant-interacting fungi. *Adv. Virus Res.* 100, 99–116.
- Hollings M. 1962. Viruses associated with a die-back disease of cultivated mushroom. *Nature* 196, 962–965. doi: 10.1038/196962a0
- Linnakoski R., Sutela S., Coetzee M.P.A. et al. 2021: *Armillaria* root rot fungi host single-stranded RNA viruses. *Sci. Rep.* Apr 1;11(1): 7336. doi: 10.1038/s41598-021-86343-7.
- Martin M. 2011: Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, v. 17, n. 1, p. pp. 10-12, doi:10.14806/ej.17.1.200.
- Morris T. J., Dodds J. A. 1979: Isolation and analysis of double-stranded RNA from virus-infected plant and fungal tissue. *Phytopathology* 69, 854–858. doi: 10.1094/Phyto-69-854
- Nuss D. 2005: Hypovirulence: Mycoviruses at the fungal–plant interface. *Nat. Rev. Microbiol.* 3, 632–642. doi: 10.1038/nrmicro1206
- Robinson J. T., Thorvaldsdóttir H., Winckler W., et al. 2011: Integrative genomics viewer. *Nat Biotechnol.* 29 (1): 24-26. doi:10.1038/nbt.1754
- Rumbou A., Vainio E.J., Büttner C. 2021: Towards the Forest Virome: High-Throughput Sequencing Drastically Expands Our Understanding on Virosphere in Temperate Forest Ecosystems. *Microorganisms.* 9 (8) :1730. doi:10.3390/microorganisms9081730
- Sayers E. W., Agarwala R., Bolton E. E., et al. 2019: Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* Jan 8; 47 (D1): D23-D28. doi: 10.1093/nar/gky1069.
- Sharma M., Guleria S., Singh K., Chauhan A., Kulshrestha S. 2018: Mycovirus associated hypovirulence, a potential method for biological control of *Fusarium* species. *Virus disease.* 29(2): 134-140. doi:10.1007/s13337-018-0438-4
- Son M., Yu J., Kim K.-H. 2015: Five Questions about Mycoviruses. *PLoS Pathog* 11(11): e1005172. doi: 10.1371/journal.ppat.1005172
- Sutela, S. Poimala A. & Vainio E. J. 2019: Viruses of fungi and oomycetes in the soil environment. *FEMS Microbiol. Ecol.* **95**, 119. doi: 10.1093/femsec/fiz119.
- Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M., Rozen S.G. 2012: Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 2012 Aug 1;40(15):e115.
- Xie J., Jiang D. 2014: New insights into mycoviruses and exploration for the biological control of crop fungal diseases. *Annu. Rev. Phytopathol.* 52: 45-68. doi: 10.1146/annurev-phyto-102313-050222.

Yu, X., Li, B., Fu, Y., Xie, J., Cheng, J., Ghabrial, S. A., et al. 2013“ Extracellular transmission of a DNA mycovirus and its use as a natural fungicide. Proc. Natl. Acad. Sci. U.S.A. 110, 1452-1457. doi: 10.1073/pnas.1213755110

Zhang Y., Guo Li, Ma L. J. 2018: A Computational Protocol to Analyze Metatranscriptomic Data Capturing Fungal - Host Interactions. Wenbo Ma and Thomas Wolpert (eds.), Plant Pathogenic Fungi and Oomycetes: Methods and Protocols, Methods in Molecular Biology, vol. 1848, doi: 10.1007/978-1-4939-8724-5_15,

Zhao Y., Zhang Y., Wan X., She Y., Li M., Xi H., Xie J., Wen C. 2020: A Novel Ourmia-Like Mycovirus Confers Hypovirulence-Associated Traits on *Fusarium oxysporum*. Front Microbiol. 2020 Dec 9;11:569869. doi: 10.3389/fmicb.2020.569869. PMID: 33362731; PMCID: PMC7756082.