

**Univerzita Palackého v Olomouci**

**Diplomová práce**

**Olomouc 2020**

**Hana Novotná**

**Univerzita Palackého v Olomouci**  
**Přírodovědecká fakulta**  
**Katedra buněčné biologie a genetiky**



# **Sekvenování nové generace jako nástroj identifikace nových rostlinných virů**

**Diplomová práce**

**Bc. Hana Novotná**

Studijní program: Biologie

Studijní obor: Molekulární a buněčná biologie

Forma studia: Prezenční

**Olomouc 2020**

**Vedoucí práce: Mgr. Dana Šafářová, Ph.D.**

## **Bibliografické údaje**

Jméno a příjmení autora: Hana Novotná  
Název práce: Sekvenování nové generace jako nástroj identifikace nových rostlinných virů  
Typ práce: diplomová  
Pracoviště: Katedra buněčné biologie a genetiky  
Vedoucí práce: Mgr. Dana Šafářová, Ph.D.  
Rok obhajoby práce: 2020  
Klíčová slova: *Nucleorhabdovirus*, bez černý, Illumina, viry ovocných dřevin, bioinformatická analýza  
Počet stran: 52  
Počet příloh: 2  
Jazyk: český

## SOUHRN

Diplomová práce se věnuje charakterizaci nového nukleorhabdoviru infikujícího bez černý (Sambucus nigra) pomocí sekvenování nové generace a Sangerova sekvenování.

Teoretická část popisuje metody vhodné pro identifikaci nových druhů virů se zaměřením na viry ovocných dřevin. Největší pozornost je věnována sekvenování nové generace. Je zde popsán princip sekvenování syntézou firmy Illumina, výhody a nevýhody využití různých typů templátových molekul, celkové RNA a DNA, siRNA, dsRNA a nukleových kyselin asociovaných s virionem. Dále se práce zabývá bioinformatickým zpracováním dat ze sekvenování nové generace, včetně trimování sekvencí, *de novo* skládání i mapování ´readů´ na referenční sekvenci, a identifikace získaných kontigů. V teoretické části jsou také popsány viry z čeledi *Rhabdoviridae* infikující rostliny, příznaky, které způsobují, morfologie jejich virionů a jejich molekulární charakteristika.

Standardními technikami molekulární biologie a Sangerovým sekvenováním byla získána téměř úplná sekvence genomu viru z rodu *Nucleorhabdovirus*, který byl identifikován na bezu černém pomocí sekvenování nové generace. V jeho genomu bylo identifikováno šest otevřených čtecích rámců s organizací typickou pro viry z rodu *Nucleorhabdovirus* a konzervativní mezigenové sekvence. Byla provedena analýza identity celogenomové sekvence, pláštěvého proteinu, ´movement´ proteinu a RNA-dependentní RNA polymerázy a jejich genů u tohoto nového izolátu, ostatních virů z rodu *Nucleorhabdovirus* a typových zástupců dalších rodů z čeledi *Rhabdoviridae* infikujících rostliny. S nejpodobnějším virem má nový izolát ve všech analyzovaných nukleotidových sekvencích identitu lehce přesahující 50 %, a tak těsně nesplňuje kritérium pro uznání jako samostatný druh v rámci rodu *Nucleorhabdovirus* na základě sekvenční rozdílnosti. Lze jej ale považovat za samostatný druh na základě hostitelské specifity. Byla provedena fylogenetické analýza pro zjištění příbuznosti tohoto nového druhu s ostatními viry z rodu *Nucleorhabdovirus*.

## **Bibliographical identification**

Author's first name and surname: Hana Novotná

Title: High-throughput sequencing as a tool in identification of new plant viruses

Type of thesis: master

Department: Department of cell biology and genetics

Supervisor: Mgr. Dana Šafářová, Ph.D.

The year of presentation: 2020

Keywords: *Nucleorhabdovirus*, black elderberry, Illumina, fruit tree viruses, bioinformatic analysis

Number of pages: 52

Number of appendices: 2

Language: Czech

## SUMMARY

This thesis focuses on characterization of a novel nucleorhabdovirus found in black elderberry (*Sambucus nigra*) identified by high-throughput sequencing.

Theoretical part describes methods that enable discovering of novel virus species. It focuses on discovering viruses of fruit trees and shrubs. The biggest attention is given to high-throughput sequencing. It describes principles of Illumina sequencing by synthesis, advantages and disadvantages of using different template molecules, total RNA or DNA, siRNA, dsRNA or virion associated nucleic acid. Further, bioinformatic data processing is described including data trimming, *de novo* assembly and mapping reads against the reference sequence, and identification of obtained contigs. Theoretical part also focuses on characterization of viruses from family *Rhabdoviridae* infecting plants, symptoms they cause, virion morphology and their molecular characterization.

Nearly complete genome sequence of novel virus from genus *Nucleorhabdovirus* identified on black elderberry by high-throughput sequencing was obtained by standard molecular technics and Sanger sequencing. Its genome comprises six open reading frames with typical organization of nucleorhabdoviruses and conservative intergenic sequences. Identities of genome sequence, nucleocapsid protein, movement protein and RNA-dependent RNA polymerase, and their genes were analyzed in novel isolate, other nucleorhabdoviruses and type species of other genera of plant rhabdoviruses. The most similar virus to the novel isolate shares the identity slightly exceeding 50 % which means it doesn't fulfill species demarcation criteria based on sequence distance in genus *Nucleorhabdovirus*, but it can be considered a distinct species based on different ecology. The phylogenetic analysis was made to find out the congeniality of the new species to other viruses from genus *Nucleorhabdovirus*.

Prohlášení:

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně za pomoci vedoucí práce Mgr. Dany Šafářové, Ph.D. a uvedených literárních zdrojů.

V Olomouci dne: .....

Podpis: .....

Děkuji Mgr. Daně Šafářové, Ph.D. za odborné vedení, trpělivost, vstřícnost a čas, který mi věnovala během zpracovávání diplomové práce. Dále děkuji paní Janě Veselské za pomoc a rady při práci v laboratoři. Diplomová práce byla vypracována s podporou grantu IGA\_2020\_006.



# OBSAH

<b>1</b>	<b>Úvod.....</b>	<b>1</b>
<b>2</b>	<b>Cíle práce .....</b>	<b>2</b>
<b>3</b>	<b>Literární přehled .....</b>	<b>3</b>
3.1	Virózy ovocných dřevin .....	3
3.2	Metody pro identifikaci nových virů .....	3
3.3	Sekvenování nové generace.....	5
3.4	Identifikace nových virů rostlin pomocí sekvenování nové generace.....	6
3.4.1	Metodika přípravy vzorku pro NGS.....	8
3.4.2	Zpracování dat.....	8
3.4.3	Verifikace sekvence a doplnění mezer .....	10
3.5	Sekvenování nové generace a rutinní diagnostika virů rostlin .....	10
3.6	Charakteristika čeledi <i>Rhabdoviridae</i> .....	11
3.6.1	Viry rodu <i>Nucleorhabdovirus</i> .....	13
<b>4</b>	<b>Materiál a metody .....</b>	<b>15</b>
4.1	Biologický materiál .....	15
4.2	Použité chemikálie, soupravy a roztoky .....	15
4.2.1	Použité chemikálie .....	15
4.2.2	Použité soupravy .....	15
4.2.3	Použité roztoky a jejich příprava.....	16
4.3	Použité přístroje .....	16
4.4	Použité experimentální a vyhodnocovací postupy .....	17
4.4.1	NGS sekvenování .....	17
4.4.2	Analýza NGS dat.....	17
4.4.3	Izolace RNA .....	18
4.4.4	Reverzní transkripce.....	18

4.4.5	PCR amplifikace .....	19
4.4.6	RACE PCR.....	21
4.4.7	Elektroforetická separace .....	21
4.4.8	Izolace PCR produktu z gelu.....	22
4.4.9	Klonování PCR produktu .....	22
4.4.10	Sangerovo sekvenování.....	24
4.4.11	Bioinformatická analýza .....	24
<b>5</b>	<b>Výsledky .....</b>	<b>26</b>
<b>6</b>	<b>Diskuze .....</b>	<b>39</b>
<b>7</b>	<b>Závěr.....</b>	<b>44</b>
<b>8</b>	<b>Literatura.....</b>	<b>45</b>
<b>9</b>	<b>Přílohy .....</b>	<b>53</b>

## SEZNAM ZKRATEK

AaNv	Alfalfa-associated nukleorhabdovirus
ApRVA	Apple rootstock virus A
BCaRV-1	Black currant-associated rhabdovirus
BFTV-1	Bird's-foot trefoil nucleorhabdovirus
DYVV	<i>Datura yellow vein nucleorhabdovirus</i>
EMDV	<i>Eggplant mottled dwarf nucleorhabdovirus</i>
G	glykoprotein
GSPNuV	Green Sichuan pepper nucleorhabdovirus
ICTV	International Committee on Taxonomy of Viruses, Mezinárodní výbor pro klasifikaci virů
L	RNA dependentní RNA polymeráza virů z čeledi <i>Rhabdoviridae</i>
LBVV	<i>Lettuce big-vein virus</i>
LNyV	<i>Lettuce necrotic yellows cytorhabdovirus</i>
M	matrix protein
MFSV	<i>Maize fine streak nucleorhabdovirus</i>
MIMV	<i>Maize Iranian mosaic nucleorhabdovirus</i>
MMaV	Morogoro maize-associated virus
MMV	<i>Maize mosaic nucleorhabdovirus</i>
N	nukleokapsid protein
NGS	'next generation sequencing', sekvenování nové generace
OFV	<i>Orchid fleck dichorhavirus</i>
ORF	'open reading frame', otevřený čtecí rámec
P	fosfoprotein
P3	'movement protein' virů rostlin z čeledi <i>Rhabdoviridae</i>
PhCMoV	Physostegia chlorotic mottle virus
PPV	<i>Plum pox virus</i>
PYDV	<i>Potato yellow dwarf nucleorhabdovirus</i>
RT	reverzní transkripce
RYSV	<i>Rice yellow stunt nucleorhabdovirus</i>
SYNV	<i>Sonchus yellow net nukleorhabdovirus</i>
SYVV	<i>Sowthistle yellow vein nucleorhabdovirus</i>
TaVCV	<i>Taro vein chlorosis nucleorhabdovirus</i>
VANA	'virion-associated nucleic acid', nukleová kyselina asociovaná s virionem
WYSV	<i>Wheat yellow striate virus</i>

## SEZNAM OBRÁZKŮ

Obrázek 1: Schéma genomu nového nukleorhabdoviru B152 s lokalizovanými ORF

Obrázek 2: Fylogenetický strom vytvořený na základě celogenomové sekvence virů z rodu *Nucleorhabdovirus*

Obrázek 3: Fylogenetický strom vytvořený na základě aminokyselinové sekvence proteinu L virů z rodu *Nucleorhabdovirus* a zástupců rodů z čeledi *Rhabdoviridae*, kteří infikují rostliny

Obrázek 4: Fylogenetický strom vytvořený na základě aminokyselinové sekvence proteinu N virů z rodu *Nucleorhabdovirus* a zástupců rodů z čeledi *Rhabdoviridae*, kteří infikují rostliny

Obrázek 5: Fylogenetický strom vytvořený na základě aminokyselinové sekvence proteinu P3 virů z rodu *Nucleorhabdovirus*

## SEZNAM TABULEK

- Tabulka 1: Složení reakčních směsí pro reverzní transkripci
- Tabulka 2: Složení reakční pro PCR s MyTaq DNA polymerázou
- Tabulka 3: Složení reakční pro PCR s MyFi DNA polymerázou
- Tabulka 4: Seznam použitých primerů
- Tabulka 5: Teplotní průběh PCR reakce s primery RBD 430 F/908 R a RBD 1395 F/2058 R
- Tabulka 6: Teplotní průběh PCR reakce s primery RBD 699 F/1598 R
- Tabulka 7: Teplotní průběh PCR reakce s primery RBD 1569 F/3178 R, 3007 F/4684 R a 8333 F/9963 R
- Tabulka 8: Teplotní průběh PCR reakce s primery RBD 4408 F/8258 R
- Tabulka 9: Teplotní průběh PCR reakce s primery RBD 5276 F/7918 R, 9939 F/12480 R
- Tabulka 10: Složení 'Master Mix' směsi pro RACE PCR
- Tabulka 11: Složení reakční směsi pro RACE PCR
- Tabulka 12: Teplotní průběh PCR amplifikace konců genomu
- Tabulka 13: Složení ligační směsi pro klonování PCR produktu
- Tabulka 14: Složení reakční směsi pro PCR amplifikaci pomocí MyTaq DNA polymerázy
- Tabulka 15: Složení reakční směsi pro PCR amplifikaci pomocí MyFi DNA polymerázy
- Tabulka 16: Teplotní průběh PCR reakce
- Tabulka 17: Teplotní průběh sekvenačního značení
- Tabulka 18: Konzervovaná mezigenová sekvence a její funkce
- Tabulka 19: Procentuální identita genomických nukleotidových sekvencí B152 a dalších zástupců rodu *Nucleorhabdovirus*
- Tabulka 20: Procentuální identita nukleotidové sekvence genu pro L protein a aminokyselinové sekvence tohoto proteinu u virů z rodu *Nucleorhabdovirus* a typových zástupců rodů z čeledi *Rhabdoviridae*, kteří infikují rostliny
- Tabulka 21: Procentuální identita nukleotidové sekvence genu pro N protein a aminokyselinové sekvence tohoto proteinu u virů z rodu *Nucleorhabdovirus* a typových zástupců rodů z čeledi *Rhabdoviridae*, kteří infikují rostliny
- Tabulka 22: Procentuální identita nukleotidové sekvence genu pro P3 protein a aminokyselinové sekvence tohoto proteinu u virů z rodu *Nucleorhabdovirus*
- Tabulka 23: Procentuální identita všech nukleotidových sekvencí genů a aminokyselinových sekvencí proteinů viru *Datura yellow vein virus* s izolátem B152

# 1 ÚVOD

Víceleté rostliny, jako jsou ovocné dřeviny, jsou náchylné k akumulaci virových infekcí. Některé viry způsobují pěstitelům velké ztráty úrody, jiné virové infekce jsou latentní, ale v kombinaci s infekcí jiným virem mohou působit synergicky. Kromě vektorů se viry v ovocnářství přenášejí především díky vegetativnímu množení a roubování. Je tedy velmi důležité, aby množitelský materiál byl viruprostý. Ke kontrole množitelského materiálu se používají především techniky zjišťující přítomnost konkrétního viru, jakou jsou ELISA a RT-PCR. Aby mohly být tyto techniky uplatněny, je třeba znát virom daného druhu plodiny. Sekvenování nové generace umožňuje detekci všech nukleových kyselin virového původu v testované rostlině bez ohledu na předchozí znalost jejího viromu. Je tedy účinným nástrojem pro detekci, objevování a sekvenování nových druhů virů.

## **2 CÍLE PRÁCE**

- 1) Shrnutí aktuálních nejvýznamnějších pokroků v identifikaci nových virů ovocných dřevin.
- 2) Analýza NGS dat získaných sekvenováním siRNA a celkové RNA izolované z bezu černého s využitím Illumina platformy.
- 3) Molekulárně genetická charakteristika detekovaných nových nebo nově popsaných virů pomocí standardních technik RT-PCR, Sangerova sekvenování a bioinformatické analýzy.

## 3 LITERÁRNÍ PŘEHLED

### 3.1 Virózy ovocných dřevin

Viry jsou významnými patogeny ovocných dřevin. Existuje velké množství virů, které tyto dřeviny infikují. V průběhu let se v nich také mohou kumulovat různé druhy virů, a tak dochází ke vzniku směsných infekcí. Zatím bylo u ovocných dřevin detekováno 163 virů patřících do 45 rodů z 23 čeledí, ale další viry jsou stále objevovány. Více než 70 % těchto virů má genom tvořený jednovláknovou RNA s funkcí mRNA (+ssRNA). Řada virových infekcí je latentní, ale některé viry způsobují závažné příznaky a s tím spojené velké finanční ztráty pro pěstitele. Mezi příznaky virových infekcí na listech ovocných dřevin patří žloutnutí, chlorotické mozaiky, chlorotické lemování žilek, nekrotické tečky a pruhy, tvorba výrůstků a jiné deformace listů. Plody napadených dřevin mohou být malé, deformované a mít jinou chuť. Podobně jako u listů, i u plodů virózy způsobují změnu barvy, barevné tečky a pruhy. Viry způsobují také deformace kmene a větví, zakrslost, předčasné opadávání listů případně i odumření dřeviny (Umer a kol., 2019). Jedním z hospodářsky nejvýznamnějších virů ovocných dřevin je *Plum pox virus* (PPV) z čeledi *Potyviriidae*, který infikuje stromy z rodu *Prunus*. PPV způsobuje měknutí a předčasný opad plodů. Plody infikovaných stromů mají nižší cukernatost, nižší zastoupení fenolických látek, flavonoidů a antioxidantů, naopak obsahují větší množství organických kyselin (Usenic a Marn, 2017; Milošević a kol., 2019). Viry ovocných dřevin bývají často přenášeny vegetativním množením, roubováním a hmyzími vektory (Umer a kol., 2019). Viry jsou v rámci rostliny nerovnoměrně rozloženy a jejich titr v průběhu roku kolísá, což znesnadňuje detekci a identifikaci virů. Obzvláště v létě je titr virů nízký (Dal Zotto a kol., 1999). Jednotlivé populace virů v různých částech rostliny se pak vyvíjí nezávisle na sobě (Jridi a kol., 2006).

### 3.2 Metody pro identifikaci nových virů

V rutinních diagnostických laboratořích jsou detekovány konkrétní viry pomocí protilátek, specifických primerů a sond. Aby tomu tak mohlo být, musí být viry charakterizované, musí být známá alespoň část sekvence jejich genomu nebo jejich sérologické vlastnosti. Je zapotřebí, aby diagnostické laboratoře znaly virom rostlin, které na přítomnost virů testují. Následující techniky umožňují identifikovat nové dosud necharakterizované viry.

Biologické indexování je metoda založená na přenesení virové infekce na indikátorovou rostlinu a vyhodnocování jejích příznaků. Přenos se provádí pomocí mechanické inokulace na listy indikátorové byliny nebo roubováním na dřevinu (Al Rwahnih a kol., 2015). Jedná se



o metodu náročnou na čas i práci a je v podstatě nemožné její pomocí určit konkrétní druh viru. Výsledky je třeba potvrdit a upřesnit pomocí dalších metod (Rott a kol., 2017).

Další metodou schopnou zjistit přítomnost viru bez předchozí znalosti viromu testované rostliny je transmisní elektronová mikroskopie. Viriony ale mohou být přehlédnuty, pokud nejsou ve vzorku v dostatečném titru (van Lent a kol., 1980). Pomocí elektronové mikroskopie lze zjistit tvar a velikost virionů, jejich intracelulární lokalizaci a charakterizovat cytopatické změny způsobené virem. Totožnost viru lze obvykle určit jen na úrovni čeledi (Kwon a kol., 2018; Popov a kol., 2019).

ELISA a PCR jsou technikami rutinně používanými k diagnostice virových infekcí. Tyto metody jsou zaměřené na detekci konkrétního viru nebo skupiny příbuzných virů. Pro přípravu antiséra pro metodu ELISA je třeba nejdříve izolovat viriony nebo proteiny ze sledovaného viru. V případě PCR je třeba znát alespoň část sekvence genomu daného viru, aby bylo možné navrhnout primery. ELISA ani PCR proto nejsou vhodnými metodami pro cílenou identifikaci nových virů (Boonham a kol., 2014). Přesto vedla standartní PCR technika s degenerovanými primery k identifikaci nového viru, kdy byl při detekci získán produkt o dvojnásobné velikosti oproti velikosti očekávané. Produkt byl osekvenován a získaná sekvence byla podobná sekvenci genu pro polymerázu několika virů z rodu *Carlavirus*. Podařilo se tak objevit *Potato virus H*.

Extrakcí dsRNA ze vzorku lze získat genomovou RNA většiny ds- i ssRNA virů vyskytujících se v daném pletivu bez ohledu na předchozí znalost viromu testované rostliny. Viry jsou následně identifikovány pomocí gelové elektroforézy a analýzy restričních fragmentů (Morris a Dodds, 1979). DsRNA z virů, které se nepovedou určit nebo ještě nebyly charakterizovány, je možné přepsat do komplementární DNA (cDNA), zaklonovat a osekvenovat (Benthack a kol., 2005).

‘Rolling circle amplification’ je levná metoda vhodná pro hromadnou detekci virů s genomem tvořeným kruhovou DNA. Je tedy možné objevit i nový virus s tímto typem genomu. Pro určení viru, případně pro charakterizaci nově objeveného viru, se používá restriční štěpení nebo Sangerovo sekvenování DNA amplifikované touto metodou. Pro Sangerovo sekvenování je DNA třeba linearizovat. Velkou nevýhodou metody ‘rolling circle amplification’ je její omezení na DNA viry s kruhovým genomem (Haible a kol., 2006), kterých je mezi viry infikující ovocné dřeviny malé množství (Umer a kol., 2019).

Microarray mohou sloužit k hromadné detekci virů, o kterých je známo, že mohou infikovat testovaného hostitele. Kromě sond specifických pro konkrétní druh viru mohou microarray nést i sondy s vysoce konzervativní sekvencí. S nimi mohou hybridizovat i nukleové kyseliny dosud neznámých virů, které mají sekvenci podobnou té, proti které byla sonda s konzervativní sekvencí navržena. Tyto sondy by měly být schopné zachytit všechny viry z jedné čeledi. Microarray jsou vhodné pro detekci virů u rostlin, které může infikovat velké množství různých virů, jako je např. réva vinná (Engel a kol., 2010).

Přelomovou technologií pro hromadnou detekci a objevování nových virů u rostlin se stalo sekvenování nové generace (NGS). NGS nevyžaduje předchozí znalost viromu testované rostliny, je schopné detekovat RNA i DNA viry, a to s velmi nízkým titrem, z asymptomatických rostlin, s různým tkáňovým tropismem a místem replikace. V případě objevení nového viru pomocí NGS není třeba totožnost viru upřesňovat pomocí dalších metod, a přitom je získána alespoň částečná sekvence genomu (Kreuze a kol., 2009). První publikace popisující objev nového viru v rostlině pomocí NGS byly zveřejněny roku 2009. Jednalo se o objev viru *Grapevine Syrah virus-1* z rodu *Marafivirus* na révě vinné (Al Rwahnih a kol., 2009), viru *Gayfeather mild mottle virus* z rodu *Cucumovirus* infikující šuškardu klasnatou (*Liatris spicata*) (Adams a kol., 2009) a dvou virů infikujících povijnici batátovou (*Ipomoea batatas*), viru Sweet potato symptomless virus 1 z rodu *Mastrevirus* a viru z rodu *Badnavirus* (Kreuze a kol., 2009).

### 3.3 Sekvenování nové generace

Sekvenování nové generace (NGS) je novým přístupem k sekvenování nukleových kyselin, který umožňuje získat neznámé sekvence. V porovnání s klasickým sekvenováním Sangerovou metodou je sekvenování NGS levnější na počet párů bází. Finančně se tedy vyplatí, pokud potřebujeme získat velké množství sekvenčních dat (Tucker a kol., 2009). Existují různé přístupy NGS, prvním z nich bylo pyrosekvenování vyvinuté firmou 454 Life Sciences (později odkoupené firmou Roche), které začalo být komerčně dostupné v roce 2005. Jedná se o sekvenování syntézou. DNA je náhodně naštěpena a na její fragmenty jsou ligovány adaptéry komplementární k oligonukleotidům imobilizovaným na kuličkách. Aby mohlo sekvenování proběhnout, musí být na kuličku hybridizován pouze jeden fragment DNA. Následuje amplifikace fragmentů DNA na kuličkách pomocí emulzní PCR. Směs kuliček obohacená o ty, které nesou fragmenty DNA, je nanášena na destičku s jamkami. K těmto kuličkám jsou přidány ještě menší kuličky nesoucí imobilizované enzymy ATP-sulfurylázu a luciferázu. Při samotném sekvenování jsou na 'flow cell' cyklicky nanášeny dNTP. Pokud je nukleotid

inkorporován, uvolní se pyrofosfát, který je substrátem pro ATP-sulfurylázou, jejímž produktem je ATP. ATP je zapotřebí pro reakci katalyzovanou luciferázou, která emituje světlo. Záblesky světla z jednotlivých jamek jsou snímány a zaznamenány. Před nanesením dalšího typu dNTP je na 'flow cell' přidána apyráza, která zbývající nukleotidy rozloží. (Margulies a kol., 2005). Pyrosekvenování následovaly další platformy založené na různých mechanismech (Tucker a kol., 2009). Nejpoužívanějšími jsou platformy firmy Illumina (Nakazato a kol., 2013). Mechanismus je založený na sekvenování syntézou. Nejdříve je provedena příprava knihovny, která zahrnuje ligaci specifických adaptérů. Následně jsou fragmenty DNA z knihovny nanášeny na 'flow cell', kde se hybridizují na oligonukleotidy komplementární k adaptérům. Ty slouží jako primery pro můstkovou PCR amplifikaci. Jejím výsledkem je klastr molekul o stejné sekvenci. Samotné sekvenování probíhá za syntézy komplementárního řetězce pomocí fluorescenčně značených deoxynukleotidtrifosfátů. Fluorescenční značka slouží zároveň jako reverzibilní terminátor syntézy a po každém navázání nukleotidu je enzymaticky odštěpena, což umožní navázání dalšího nukleotidu. Fluorescence je průběžně snímána a zaznamenávána (Bentley a kol., 2008).

### **3.4 Identifikace nových virů rostlin pomocí sekvenování nové generace**

Pomocí NGS je možné získat sekvence ze všech nukleových kyselin ze vzorku včetně genomů virů (Bag a kol., 2015). Pro sekvenování lze zvolit různé typy vzorků, celkovou RNA, dsRNA, siRNA, případně pro identifikaci virů rostlin výjimečně používanou celkovou DNA nebo 'virion-associated nucleic acid' (VANA).

Nejjednodušším přístupem je sekvenování celkové RNA. Tu je možné z rostliny vyizolovat pomocí komerčních kitů. Nevýhodou tohoto přístupu je, že většina 'readů' je získána z hostitelské rostliny, což znemožňuje identifikaci virů, které se v rostlině vyskytují s nízkým titrem (Maliogka a kol., 2018). Tato nevýhoda může být překonána odstraněním ribozomální RNA z hostitelské rostliny (Villamor a kol., 2016). Pomocí tohoto přístupu byl objeven např. virus *Nectarine virus M* z rodu *Marafivirus* na nektarince (*Prunus persica* cv. *nectarina*) a virus z rodu *Nucleorhabdovirus* na černém rybízu (*Ribes nigrum*) prozatímně pojmenovaný Black currant-associated rhabdovirus (Villamor a kol., 2016; Wu a kol., 2018).

Další možností je sekvenování dsRNA. Výhodou tohoto přístupu je velké zastoupení virové nukleové kyseliny ve vzorku (Maliogka a kol., 2018). dsRNA tvoří genom některých virů nebo je meziproduktem replikace virů s genomem tvořeným ssRNA. Tento přístup je tedy zaměřen na RNA viry, přesto se touto metodou podařilo identifikovat i DNA viry. Spekuluje se, že by

molekuly dsRNA mohly vznikat hybridizací transkriptů z obou vláken dsDNA virů, případně ssDNA virů s 'ambisense' genomem (Rott a kol., 2017). Vzorek se získává z rostliny pomocí fenol-chloroformové extrakce celkové DNA a RNA následované 'long polymer cellulose matrix' chromatografií, která vzorek obohatí o dsRNA (Maliogka a kol., 2018). Pomocí tohoto přístupu byl objeven např. *Nectarine stem pitting-associated virus* z rodu *Luteovirus* na nektarince (*Prunus persica* cv. *nectarina*), a několik navržených virů na bezu černém (*Sambucus nigra*), Elderberry aureusvirus 1 z rodu *Aureusvirus*, Sambucus virus S z rodu *Bromovirus* a pět příbuzných druhů carlavirů prozatímne pojmenovaných Elderberry virus A-E (Bag a kol., 2015; Šafářová a kol., 2018a; 2018b). Rott a kol. (2017) při porovnávání identifikace virů pomocí NGS s konvenčními detekčními metodami objevili nový virus s genomem tvořeným ssDNA z čeledi *Geminiviridae*.

Dalším vhodným přístupem je sekvenování malých RNA. Tato frakce je obohacena o siRNA virového původu. Molekuly siRNA vznikají při RNA interferenci. Jedná se o mechanismus, kterým se rostlina brání virové infekci tím, že štěpí RNA virového původu na malé molekuly (21-24 nukleotidů) (Maliogka a kol., 2018). Největší zastoupení RNA molekul odvozených od virů je mezi molekulami o velikosti 22 nukleotidů. Sekvenování malých RNA umožňuje identifikaci RNA i DNA virů, a to i v případě že se v rostlině vyskytují s velmi nízkým titrem (Kreuze a kol., 2009). Tímto přístupem byly objeveny např. *Mulberry mosaic dwarf-associated virus*, ssDNA virus z čeledi *Geminiviridae*, izolovaný z morušovníku bílého (*Morus alba*) a *Citrus leprosis virus cytoplasmic type 2*, (+)ssRNA virus z rodu *Cilevirus*, izolovaný z pomerančovníku (*Citrus sinensis*) (Roy a kol., 2013; Ma a kol., 2015). Nevýhodou sekvenování malých molekul RNA je, že se ze získaných krátkých 'readů' obtížně skládá celý genom *de novo* (Maliogka a kol., 2018). Může být výhodné tento přístup nakombinovat se sekvenováním celkové RNA, z kterého jsou získány delší 'ready' (Cao a kol., 2019).

Sekvenování celkové DNA vyizolované z rostliny není vhodným přístupem pro identifikaci nových virů na ovocných dřevinách, protože umožňuje pouze detekci virů s genomem tvořeným DNA (Maliogka a kol., 2018), které tvoří jen malou část virů infikujících ovocné dřeviny (Umer a kol., 2019). Sekvenováním celkové DNA byl objeven nový virus na citronovníku (*Citrus limon*) postiženém chorobou Citrus chlorotic dwarf disease. Jedná se o virus z čeledi *Geminiviridae* a byl prozatímne pojmenován Citrus chlorotic dwarf-associated virus. Při studii bylo využito i sekvenování malých RNA (Loconsole a kol., 2012).

Sekvenování nukleových kyselin asociovaných s virionem (VANA) je přístupem aplikovaným při studii virů živočichů a člověka nebo virů v environmentálních vzorcích. U rostlin se

téměř nepoužívá. Jedná se o přístup umožňující sekvenování RNA i DNA nacházející se uvnitř kapsidy (Reavy a kol., 2015; Roossinck a kol., 2015). Izolace VANA je složitým procesem, který zahrnuje centrifugaci homogenátu testovaného pletiva, filtraci, digesci volných nukleových kyselin pomocí DNázy a RNázy, a samotnou extrakci nukleových kyselin z izolovaných virionů pomocí kitu. Tímto přístupem zkombinovaným se sekvenováním malých RNA byl objeven virus *Sugarcane white streak virus* z rodu *Mastrevirus* na cukrové třtině (*Saccharum officinarum*) (Candresse a kol., 2014). Dále byla pomocí sekvenování VANA získána sekvence izolátu *Plum pox virus* z planě rostoucí slivoně (*Prunus domestica*) (Sheveleva a kol., 2013).

### 3.4.1 Metodika přípravy vzorku pro NGS

Extrakci nukleové kyseliny z rostlinných pletiv následuje další zpracování vzorku. Příprava RNA vzorků pro sekvenování syntézou od firmy Illumina zahrnuje fragmentaci molekul v případě použití celkové RNA nebo dsRNA nebo selekci velikostní frakce pomocí polyakrylamidové gelové elektroforézy v případě sekvenování malých RNA. RNA je následně přepsána do cDNA reverzní transkripcí s random primery, v dalším kroku je dosyntetizováno druhé vlákno cDNA. Na získané dsDNA jsou vytvořeny tupé konce. Přesahy na 3' konci jsou odstraněny exonukleázou a přesahy na 5' konci jsou dosyntetizovány polymerázou. Na 3' konce je následně přidáno po jednu nukleotidu 'A'. V dalším kroku jsou na fragmenty DNA ligovány adaptéry, které mají na svém 3' konci jednonukleotidový přesah tvořený nukleotidem 'T'. Tento mechanismus zajišťuje, aby se při ligaci adaptérů náhodně nespojovaly fragmenty DNA ze vzorku mezi sebou, a nevznikaly tak chimérické molekuly. Molekuly DNA, které mají naligované adaptéry z obou stran, jsou namnoženy pomocí PCR se specifickými primery, které nasedají na konce adaptérů. Adaptéry umožňují, aby fragmenty DNA mohly později hybridizovat s oligonukleotidy na 'flow cell'. Získanou DNA je třeba přesně kvantifikovat pomocí kvantitativní PCR a zjistit, jestli má vzorek dostatečnou čistotu a optimální velikost fragmentů DNA (Kreuze a kol., 2009; TruSeq RNA sample preparation v2 guide, 2014). Následuje můstková PCR na 'flow cell' a samotné sekvenování, jejichž mechanismus byl popsán dříve v kapitole 3.3 (Bentley a kol., 2008).

### 3.4.2 Zpracování dat

Výstupem ze sekvenování jsou 'ready' ve formátu FASTQ, který kromě sekvence bází zahrnuje také tzv. Q-skóre vyjadřující kvalitu sekvence. Q-skóre odráží pravděpodobnost, že daná báze byla zařazena chybně. Pro vztah pravděpodobnosti chyby (p) a Q-skóre platí, že:

$$p = 10^{\frac{-Q}{10}}$$

Použití 'readů' s nízkým Q-skóre (menším, než 20, spolehlivost 99 %) může znemožnit skládání kontigů, nebo vést ke skládání nespolehlivých kontigů (Del Fabbro a kol., 2013; Blawid a kol., 2017). Při tzv. trimování sekvencí jsou odstraněna data o nižší kvalitě, než je nastavený 'threshold' a také adaptéry použité při sekvenování. Nicméně je třeba trimování používat s opatrností, aby nedocházelo k velké ztrátě dat (Blawid a kol., 2017).

Trimované 'ready' mohou být potom skládány *de novo* do kontigů na základě překryvů mezi nimi, aniž by byly předem známy jejich sekvence. Nejvíce používanými algoritmy pro *de novo* skládání genomů virů jsou Velvet a SPAdes. Algoritmus Velvet byl vyvinut pro skládání krátkých 'readů' a SPAdes pro skládání 'readů' nerovnoměrně rozložených v rámci genomu. SPAdes je navíc schopen skládat dlouhé kontigy (Blawid a kol., 2017). *De novo* skládání může být problematické, pokud je virus variabilní, nebo pokud jednu rostlinu infikuje více různých izolátů jednoho druhu viru. Tento problém je nejvýraznější, pokud se jedná o skládání krátkých 'readů' (Maliogka a kol., 2018).

Alternativou k *de novo* skládání je mapování 'readů' na referenční sekvenci. Aby bylo možné tento přístup použít, je potřeba mít představu, které viry rostlinu infikují a mít přístupné referenční sekvence genomů těchto virů nebo jim blízké příbuzných (De Andrade a Vaslin, 2014). Tento přístup se používá také k odstranění 'readů' z hostitele. Odstranění části dat snižuje výpočetní náročnost následného *de novo* skládání a s tím ušetří i čas. Nicméně je třeba dát pozor, aby společně s 'ready' z hostitele nebyly odstraněny i virové 'ready'. Na referenční sekvenci lze mapovat i poskládané kontigy. Mapování se používá také ke zjištění diverzity populace studovaného viru v rostlině tím, že se 'ready' mapují na finální poskládanou sekvenci genomu (Blawid a kol., 2017).

Získané kontigy neznámého původu jsou identifikovány prohledáváním virové databáze algoritmem BLAST. Pro tento účel přichází v úvahu varianty algoritmu BLASTn, BLASTx a tBLASTx. BLASTn je schopen najít shodu jen u velmi podobných sekvencí, zatímco BLASTx a tBLASTx objeví shodu na úrovni proteinu i u taxonomicky vzdálenějších taxonů virů, než je druh. Do genomů rostlin mohou být intergované části genomů virů, které mohou být mylně interpretovány jako infekční virus (Zheng a kol., 2017). Kontigy, které jsou podobné sekvencemi téhož druhu, případně rodu virů, jsou dále zpracovány společně. Kontigy jsou prodlužovány opakovaným namapováváním trimovaných 'readů'. Prodloužené kontigy pak mohou být složeny stejným způsobem, jako se skládají 'ready' *de novo* (Blawid a kol., 2017).

Když není možné složit z kontigů celý genom, je možné namapovat je na referenční sekvenci příbuzného viru, a zjistit tak jejich pořadí a přibližnou vzdálenost mezi nimi (Villamor a kol., 2016).

### **3.4.3 Verifikace sekvence a doplnění mezer**

K potvrzení správnosti sekvence získané pomocí NGS a k doplnění děr mezi kontigy se využívá sekvenování Sangerovou metodou. Vzorčky pro sekvenování jsou amplifikovány pomocí RT-PCR se specifickými primery navrženými na základě NGS dat. PCR produkty jsou zaklonovány a osekvenovány (Villamor a kol., 2016). Konce genomu jsou amplifikovány pomocí 5' a 3' RACE PCR (Frohman a kol., 1988).

## **3.5 Sekvenování nové generace a rutinní diagnostika virů rostlin**

Kromě objevování nových druhů virů má NGS potenciál pro využití i v rutinní diagnostice virů rostlin. NGS může teoreticky detekovat všechny viry přítomné v testované rostlině a popsat i jejich variabilitu na sekvenční úrovni. NGS tak dává informaci o celkovém zdravotním stavu rostliny. Tato schopnost je ale závislá na kompletnosti virových databází, se kterými jsou data porovnávána (Rott a kol., 2017; Maree a kol., 2018). Naopak PCR se specifickými primery nemusí detekovat všechny izoláty viru, pro jehož detekci byly primery navrženy, pokud je jejich genom variabilní v oblasti, na kterou primery nasedají (Al Rwahnih a kol., 2015) NGS může být i finančně výhodnější než soubor klasických metod pro detekci konkrétních druhů virů, pokud jich testovanou rostlinu může infikovat velké množství. Dále je NGS rychlejší než biologické indexování dřevin, které zabere minimálně 2 roky času, zatímco NGS a analýza dat trvá 1-2 měsíce (Maree a kol., 2018). Schopnost detekovat všechny viry bez ohledu na to, jaký typ nukleové kyseliny tvoří jejich genom, má sekvenování celkové RNA ochuzené o ribozomální RNA hostitele a sekvenování malých RNA. *De novo* skládáním 'readů' ze sekvenování celkové RNA se daří získat delší kontigy pokrývající větší část genomu viru, než z 'readů' získaných sekvenováním malých RNA (Visser a kol., 2016). Sekvenování celkové RNA je schopné detekovat RNA viry i s velmi nízkým titrem a umožňuje získání větší části sekvence jejich genomu oproti sekvenování malých RNA. Naopak sekvenování malých RNA dává lepší výsledky pro DNA viry a viroidy (Pecman a kol., 2017).

Při zpracování NGS dat je třeba dát pozor na falešně pozitivní výsledky. Získaná NGS data obsahují kromě sekvencí virů infikujících testovanou rostlinu také sekvence bakteriofágů a dalších virů parazitujících na organismech asociovaných s testovanou rostlinou (Rott a kol., 2017). Při zpracování vzorku pro NGS sekvenování může dojít k jeho kontaminaci

a při analýze dat může dojít k chybě. Proto by přítomnost viru detekovaného pomocí NGS měla být potvrzena některou z klasických metod, nejčastěji (RT)-PCR se specifickými primery. Speciální pozornost je třeba věnovat DNA virům. Fragmenty jejich genomu mohou být integrovány do genomu rostliny a detekovány pomocí NGS, přestože v buňkách testované rostliny nejsou přítomny infekční částice těchto virů. Nejčastěji se do genomu hostitele integrují viry z čeledi *Caulimoviridae* (Massart a kol., 2017).

Při zavedení NGS do rutinní diagnostiky by mohl nastat problém se srovnatelností různých přístupů. Bylo by potřeba standardizovat protokol a zavést systém s negativní kontrolou a vnitřní pozitivní kontrolou (Maree a kol., 2018).

Využití NGS v rostlinné virologii má vliv na diagnostiku, aniž by v tomto oboru bylo přímo používáno. Diagnostické laboratoře a rostlinolékařská správa musí reagovat na existenci nově objevených virů, u nichž je známá sekvence jejich genomu, ale ne jejich biologické vlastnosti. Informace o příznacích, které virus způsobuje, o jeho hostitelském okruhu, rozšíření, způsobu šíření a vektorech, a hospodářském dopadu na plodiny jsou zásadní pro zhodnocení jeho nebezpečnosti a zavedení opatření, která zabrání jeho dalšímu šíření. Zjištění těchto informací je zásadním úkolem navazujícím na objevení viru pomocí NGS. Někjaká provizorní opatření je ale třeba udělat rychle, i bez znalosti těchto informací. Předběžný odhad vlastností a nebezpečnosti viru lze provést na základě příznaků na rostlině, na které byl virus identifikován, a to i s ohledem na to, jaké další viry tuto rostlinu infikovaly, dále také podle vlastností příbuzných virů (Massart a kol., 2017). U některých plodin je nedosažitelným požadavkem, aby byly zcela prosté všech virů, proto je třeba nakládat s opatřeními týkajícími se nově objevených virů opatrně (Maree a kol., 2018).

Sekvence genomu nových druhů virů, získané pomocí NGS, umožňují rychlé navržení specifických primerů pro jejich detekci (RT)-PCR (Massart a kol., 2017). Podobně umožňují NGS data získat přehled o variabilitě studovaného viru a navrhnout primery nasedající na konzervativní oblasti genomu, jejichž pomocí bude možné detekovat širší spektrum izolátů studovaného viru (Marais a kol., 2014).

### **3.6 Charakteristika čeledi *Rhabdoviridae***

Čeď *Rhabdoviridae* je velkou a ekologicky velmi diverzifikovanou skupinou virů, patřící do řádu *Mononegavirales*. Její zástupci infikují bezobratlé živočichy, obratlovce a rostliny. Taxonomie virů v této čeledi se vyvíjí velmi rychle, což je způsobeno z velké části tím, že bylo v poslední době objeveno mnoho nových virů pomocí NGS (Dietzgen a kol., 2017). V současné



době se do této čeledě řadí 18 rodů (Amarasinghe a kol., 2018), přičemž většina z nich zahrnuje viry infikující živočichy. Mezi rhabdoviry patří např. i virus vztekliny. Většina virů z této čeledi je přenášena členovci. Nicméně, viry z rodů *Lyssavirus* a *Novirhabdovirus* si vyvinuly strategii přenosu mezi obratlovci bez pomoci vektora, viry z rodu *Sigmavirus* jsou přenášeny mezi generacemi u octomilek (Dietzgen a kol., 2017) a *Lettuce big-vein virus* (rod *Varicosavirus*) je šířen v půdě zoosporami houby *Olpidium brassicaceae* (Roggero a kol., 2000).

Viriony virů z čeledi *Rhabdoviridae* mají tvar podobný projektilu, některé jsou baciliformní, a až na výjimky jsou obalené. Jejich genom je tvořen ssRNA s negativní polaritou o velikosti 11-16 kb (Dietzgen a kol., 2017). Genom všech virů v této čeledi obsahuje 5 čtecích rámců (ORF) kodujících základní proteiny, jsou to plášťový protein (N), fosfoprotein asociovaný s polymerázou (P), matrix protein (M), glykoprotein (G) a RNA dependentní RNA polymeráza (L). Geny pro tyto proteiny jsou v genomu uspořádány v tomto pořadí ve směru od 3' konce k 5' konci. V genomu některých rhabdovirů jsou i další doplňkové otevřené čtecí rámce (ORF) kódující různé proteiny, některé z nich jsou charakteristické pro daný rod virů (Walker a kol., 2015).

Části genomu virů z čeledi *Rhabdoviridae* mohou být integrované do genomu hostitelské rostliny nebo členovce. V genomu rostlin byly nalezeny sekvence podobné genu pro N protein. V genomu členovců byly nalezeny i další geny virů z této čeledi (Chiba a kol., 2011; Fort a kol., 2012).

Z této čeledi infikují rostliny viry ze 4 rodů, jedná se o rody *Cytorhabdovirus*, *Nucleorhabdovirus*, *Dichoravirus* a *Varicosavirus* (Dietzgen a kol., 2017). Závažnost příznaků, které tyto viry u rostlin vyvolávají se výrazně liší. Některé způsobují bezpříznakové infekce, jiné mohou vézt i k úhynu rostliny. Nejčastěji způsobují zakrslost, žloutnutí žilnatiny a chlorotické mozaiky, případně lokální nekrózy. Rozšíření těchto virů je omezené na rozšíření vektora, který je přenáší. Podobně je tomu s okruhem hostitelů virů z čeledi *Rhabdoviridae*, které infikují rostliny. Přirozeně je jejich okruh hostitelů úzký a souvisí s tím, kterými rostlinami se vektor živí. Nicméně, experimentálně lze tyto viry přenést na větší množství druhů rostlin (Jackson a kol., 2005). Genom virů z rodů *Cytorhabdovirus* a *Nucleorhabdovirus* je stejně jako u rhabdovirů infikujících živočichy tvořen jednou molekulou (-)ssRNA, zatímco u virů z rodů *Dichoravirus* a *Varicosavirus* se genom skládá ze dvou molekul RNA. Viry z rodů *Cytorhabdovirus*, *Varicosavirus* a rhabdoviry infikující živočichy se replikují v cytoplasmě, zatímco u virů z rodů *Nucleorhabdovirus* a *Dichoravirus* dochází k replikaci v jádře hostitelské buňky (Dietzgen a kol., 2017).

Zástupcem rodu *Varicosavirus* je jen jeden uznávaný druh, *Lettuce big-vein virus* (LBVV). Vytváří neobalené virové částice tvaru projektilu (ViralZone, 2020). Jeho RNA1 o velikosti cca 6,8 kb nese jeden dlouhý ORF, který kóduje L protein (Sasaya a kol., 2002). RNA2 o velikosti cca 6,1 kb nese dalších 5 ORF (Sasaya a kol., 2004). Fylogenetická analýza podle proteinu L naznačuje, že rod *Varicosavirus* je příbuznější ostatním rhabdovirům infikujícím rostliny než těm, které infikují živočichy, nicméně je fylogeneticky vzdálenější, než jsou ostatní rhabdoviry rostlin mezi sebou (Sasaya a kol., 2002).

Viry z rodu *Dichoravirus* vytváří baciliformní viriony, případně viriony ve tvaru projektilu, které jsou neobalené. Jejich RNA1 o velikosti cca 6,4 kb nese 5 ORF a RNA2 o velikosti cca 6,0 kb nese jeden ORF kódující L protein. Geny a jejich pořadí jsou shodné s těmi u rodu *Nucleorhabdovirus* (Dietzgen a kol., 2014). I podle fylogenetických analýz aminokyselinové sekvence proteinu L tento rod zapadá mezi nukleorhabdoviry (Cao a kol., 2019; Gaafar a kol., 2019), nicméně se liší svými biologickými vlastnostmi jako jsou příznaky a vektorů, vlastnostmi virionu a segmentovaným genomem (Dietzgen a kol., 2014).

### 3.6.1 Viry rodu *Nucleorhabdovirus*

Viry z rodu *Nucleorhabdovirus* jsou přenášeny hmyzem, konkrétně křísy, křísky a mšicemi. Jejich viriony mají tvar projektilu. V jejich genomu se vedle pěti základních ORF rhabdovirů nachází další ORF označený P3, který kóduje 'movement protein'. Nachází se mezi genem pro fosfoprotein a matrix protein. Geny jsou tedy v genomu v pořadí 3'-N-P-P3-M-G-L-5' (Dietzgen a kol., 2017). Typovým zástupcem rodu *Nucleorhabdovirus* je *Potato yellow dwarf virus* (PYDV). Kromě ORFs typických pro tento rod, nese sedmý ORF s neznámou funkcí, nacházející se mezi geny pro plášťový protein a fosfoprotein. Jednotlivé geny jsou od sebe odděleny konzervovanými oblastmi, jejichž součástí je signál pro začátek transkripce. Tato konzervovaná sekvence se nachází i před prvním ORF (Bandyopadhyay a kol., 2010).

V současnosti patří do rodu *Nucleorhabdovirus* deset uznávaných druhů virů, jsou to *Potato yellow dwarf nucleorhabdovirus* (PYDV), *Datura yellow vein nucleorhabdovirus* (DYVV), *Eggplant mottled dwarf nucleorhabdovirus* (EMDV), *Maize Iranian mosaic nucleorhabdovirus* (MIMV), *Maize fine streak nucleorhabdovirus* (MFSV), *Maize mosaic nucleorhabdovirus* (MMV), *Rice yellow stunt nucleorhabdovirus* (RYSV), *Sonchus yellow net nucleorhabdovirus* (SYNV), *Sowthistle yellow vein nucleorhabdovirus* (SYVV) a *Taro vein chlorosis nucleorhabdovirus* (TaVVCV) (Amarasinghe a kol., 2018). Nicméně bylo v poslední době popsáno velké množství dalších potenciálních zástupců rodu *Nucleorhabdovirus*, kteří doposud

nebyli oficiálně uznáni jako druh. Jedná se o viry Black currant-associated rhabdovirus (BCaRV-1) (Wu a kol., 2018), Apple rootstock virus A (ApRVA) (Baek a kol., 2019), Morogoro maize-associated virus (MMaV) (Read a kol., 2019), Green Sichuan pepper nucleorhabdovirus (GSPNuV) (Cao a kol., 2019), Bird's-foot trefoil nucleorhabdovirus (BFTV-1) (Debat a Bajerman, 2019) a Alfalfa-associated nucleorhabdovirus (AaNv), který má ve svém genomu mezi geny pro M a G sedmý ORF o neznámé funkci (Gaafar a kol., 2019).

## 4 MATERIÁL A METODY

### 4.1 Biologický materiál

Pomocí NGS byl analyzován vzorek listu bezu černého (*Sambucus nigra*) odebraný v roce 2016 označený B15. Jednalo se o vzorek ze Slatinek (střední Morava) s příznaky virové infekce. Analyzovaný keř měl chlorotické skvrny na listech, zakrslé listy a menší množství plodů, které měly nerovnoměrné rozložení. Kromě infekce zde popsaným novým nukleorhabdovirem byl tento keř infikován i nepovirem *Cherry leaf roll virus* (CLRV), carlaviry Elderberry virus A a B, bromovirem *Sambucus virus S* a aureusvirem Elderberry aureusvirus 1 (Šafářová a kol., 2018a). Vzorek B152 pochází ze stejného keře, byl odebrán v roce 2017.

### 4.2 Použité chemikálie, soupravy a roztoky

#### 4.2.1 Použité chemikálie

Agaróza (AMRESCO, kat.č. 0710-500g)

Deionizovaná voda

dEPC voda

dNTP Mix 10 mmol·l<sup>-1</sup> each (Fermentas, kat.č. R0192)

99,6% ethanol (Lach-Ner, kat.č. 20025-U99-M1000-1)

Ethidium bromid, 5 mg·ml<sup>-1</sup> ve vodě, (Biotium, kat.č. 40042)

Gel Red Nucleic Acid Stain (Biotium, kat.č. 41003)

GeneRuler™ 100bp DNA Ladder (Fermentas, kat.č. SM0241)

GeneRuler™ 100bp Plus DNA Ladder (ThermoFisher Scientific, kat.č. SM323)

Glycerol (Lachema, kat.č. 40294)

FastRuler High range DNA Ladder (ThermoFisher Scientific, kat.č. SM1123)

Hoechst 33258, pentahydrate (Life Technologies, kat.č. H1398)

Isopropanol (Lach-Ner, kat.č. 20037-AT0)

2-Mercaptoethanol (Sigma-Aldrich, kat.č. 63700)

Primery

Random primers (20 pmol·μl<sup>-1</sup>) (Promega, kat.č. C1181)

Rnasin ribonuclease inhibitor (40 U·μl<sup>-1</sup>) (Promega, kat.č. N2111)

#### 4.2.2 Použité soupravy

BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems™, kat.č. 4337455)

BioScript Reverse Transcriptase (Bioline, kat.č. BIO-27036)

QIAquick gel extraction kit (250) (Quiagen, kat.č. 28706)  
FMyFi DNA polymeráza (Bioline, kat.č. BIO-21118)  
MyTaq Red DNA polymeráza (Bioline, kat.č. BIO-21110)  
NucleoSpin RNA Plant kit (Macherey-Nagel, kat.č. 740949.250)  
pGEM-T Vector system (Promega, kat.č. A3600)  
SMARTer® RACE 5'/3' Kit (Takara Bio, kat. č. 634858)

### 4.2.3 Použité roztoky a jejich příprava

50x TAE (242 g Tris base; 57,1 ml ledové kyseliny octové; 100 ml EDTA  $c = 0,5 \text{ mol}\cdot\text{l}^{-1}$ ,  
doplněno na 1 l deionizovanou vodou)  
0,1% bromfenolová modř v 30% glycerolu  
10x TNE (12,1 g Tris.Cl  $c = 0,1 \text{ mol}\cdot\text{l}^{-1}$  pH 7,4; 3,7 g EDTA  $c = 0,01 \text{ mol}\cdot\text{l}^{-1}$ ; 58,4 g NaCl  
 $c = 1 \text{ mol}\cdot\text{l}^{-1}$ )  
Ampicilin ( $100 \mu\text{g}\cdot\mu\text{l}^{-1}$ , sterilizace filtrací)  
IPTG (isopropyl- $\beta$ -D-galaktopyranosid) (400 mg IPTG rozpustit ve 2 ml deionizované vody,  
sterilizovat filtrací)  
LB médium, pevné (10 g kasein hydrolyzátu; 5 g kvasničného autolyzátu; 5 g peptonu; 10 g  
NaCl; 18 g agaru; 1 l vody; sterilizovat autoklávováním 25 min)  
LB médium, tekuté (10 g kasein hydrolyzátu; 5 g kvasničného autolyzátu; 5 g peptonu; 10 g  
NaCl; 1 l vody; sterilizovat autoklávováním 25 min)  
SOC médium (0,5% kvasničný extrakt; 2% trypton; 10 mM NaCl, 2,5 mM KCl; 10 mM MgCl<sub>2</sub>;  
10 mM MgSO<sub>4</sub>; 20 mM glukóza; sterilizovat autoklávováním, glukózu přidat až dodatečně)  
X-Gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galaktopyranosid) (20 mg X-Gal rozpustit v 1 ml  
dimethylformamidu)

### 4.3 Použité přístroje

Autokláv (Microjet)  
Centrifuga Biofuge pico (Heraeus)  
Elektroforetická komora HU10 MINI (Fisherbrand)  
Fluorimetr DynaQuant DQ 200 (Amersham Biosciences)  
Homogenizátor FastPrep 24 (MP Biomedicals)  
Laminární box Telstar (PV-100)  
Minicentrifuga PRISM™ mini (Labnet International, Inc.)  
Spektrofotometr NanoDrop 1000 (Thermo Scientific)

Suchá lázeň - DryBlock (Labnet International, Inc.)  
Suchá lázeň Mixing block, MB-102 (Bioer)  
Termostat – SM 30 control, (Edmund Bühler GmbH)  
Termostatová rotační třepačka - Bio RS – 24 (Biosan)  
Thermocycler T100™ Thermal Cycler (Bio-Rad)  
Thermocycler – T personal (Biometra)  
Transiluminátor G : BOX (SynGene) - dokumentační systém GeneSnap  
Transiluminátor UVT – 14L (Herolab)  
Váhy AND EK 200G (Schoellar Pharma Prague, Ltd.)  
Vortex MS2 Minishaker (IKA)  
Výrobník ledu F100 Compact (Icematic)  
Zdroj stejnosměrného proudu MS major science, MP-250V (BioTech)  
Zdroj stejnosměrného proudu Power Station 300 (Labnet International, Inc.)

## **4.4 Použité experimentální a vyhodnocovací postupy**

### **4.4.1 NGS sekvenování**

Pro NGS sekvenování byl vybrán vzorek B15. Pro sekvenování byla provedena pracovníky Laboratoře molekulární biologie mikroorganismů izolace dsRNA pomocí afinitní celulozové chromatografie, která byla sekvenována na platformě HiSeq2500 (Illumina) komerční firmou SEQme s.r.o.

### **4.4.2 Analýza NGS dat**

‘Ready’ získané sekvenováním dsRNA byly analyzovány pomocí programu Geneious 9.1.8. Pro skládání kontigů byla použita strategie *de novo*.

Získané ‘ready’ byly trimovány (Q skóre 30). Bylo provedeno *de novo* skládání kontigů za využití assembleru SPAdes (K-mer 35; minimum contig length 95 bp; identita 80 %). Získané kontigy byly anotovány prohledáváním virové databáze na GenBank pomocí algoritmů BLASTx (parametry: maximální e-value  $10^{-5}$ ; defaultní nastavení). Kontigy, které vykazovaly podobnost se sekvencemi známých nukleorhabdovirů, byly vybrány a prodlouženy mapováním ‘readů’ (senzitivity: low-fast; 5 opakování). Správnost prodloužení kontigů s cílem vyloučit chimérické sekvence byla zkontrolována pomocí algoritmu BLASTx. Získané prodloužené a zkontrolované kontigy byly mapovány na referenční sekvenci genomu nukleorhabdoviru *Datura yellow vein virus* (parametry: highest senzitivity - slow). Překrývající se kontigy byly ručně spojeny v delší pomocí programu MEGA 7.0. Pokrytí finálních kontigů a sekvence

ověřené pomocí Sangerova sekvenování bylo stanoveno pomocí mapování 'readů' prostřednictvím algoritmu Geneious mapper (senzitivity: medium-low-fast, 10 opakování).

#### 4.4.3 Izolace RNA

Celková RNA byla izolována ze 70-80 mg listů vzorku pomocí kitu NucleoSpin RNA Plant (Macherey-Nagel) podle návodu výrobce, který byl modifikován. List byl homogenizován ve dvojnásobku doporučeného objemu homogenizačního pufru (700  $\mu$ l pufru RA1 a 7  $\mu$ l  $\beta$ -merkaptoethanolu), následně bylo použito dvojnásobné množství 70% ethanolu (700  $\mu$ l). Koncentrace získané RNA byla stanovena spektrofotometricky. RNA byla dlouhodobě uchovávána při -80 °C.

#### 4.4.4 Reverzní transkripce

Pro přepsání izolované RNA do cDNA bylo využito reverzní transkripce (RT) s použitím Biscript™ MMLV reverzní transkriptázy (Bioline) a random primerů. Vzorky izolované RNA byly zředěny pomocí dEPC vody na koncentraci cca 150  $\mu$ g·ml<sup>-1</sup>.

Byly připraveny dvě reakční směsi, první smícháním dEPC vody, dNTP a random primerů, druhá smícháním dEPC vody, RT-PCR reakčního pufru, inhibitoru ribonukleáz a reverzní transkriptázy (složení viz Tab. 1) vynásobeném počtem vzorků.

K 7  $\mu$ l reakční směsi I byly přidány 3  $\mu$ l vzorků RNA (tj. 450 ng na reakci). Zkumavky s takto připravenou směsí byly přemístěny do thermocycleru, kde byly po dobu 5 min inkubovány při 70 °C. Poté byly okamžitě přemístěny na led. Do zkumavek bylo přidáno 10  $\mu$ l reakční směsi II. Zkumavky s takto připravenou směsí byly inkubovány po dobu 10 min při 25 °C, poté 30 min při 42 °C a 5 min při 85 °C. Získaná cDNA byla uchovávána v -20 °C.

Tabulka 1: Složení reakčních směsí pro reverzní transkripci

Položka	Koncentrace prac. roztoku	Konečná koncentrace	1 test [ $\mu$ l]
<b>Reakční směs I</b>			
voda			5
dNTP	10 mmol·l <sup>-1</sup>	500 $\mu$ mol·l <sup>-1</sup>	1
random primer	20 $\mu$ mol·l <sup>-1</sup>	1 $\mu$ mol·l <sup>-1</sup>	1
<b>Reakční směs II</b>			
voda			4
RT-PCR pufr	5x	1x	4
RNase inhibitor	40 U· $\mu$ l <sup>-1</sup>	40 U	1
BioScript™	10 U· $\mu$ l <sup>-1</sup>	10 U	1

#### 4.4.5 PCR amplifikace

Na základě NGS dat byly navrženy primery tak, aby s jejich pomocí získané PCR produkty pokrývaly celý genom viru, a tak byla ověřena data a aby byly vyplněny mezery ve scaffoldu (seznam použitých primerů viz Tab. 4).

Byla připravena reakční směs pro PCR smícháním pufru, sterilní deionizované vody, forward a reverse primerů a polymerázy v množství odpovídajícímu počtu vzorků, celkový objem reakce byl 25  $\mu\text{l}$ . V závislosti na očekávané velikosti PCR produktu byla provedena amplifikace pomocí MyTaq DNA polymerázy (Bioline) nebo MyFi DNA polymeráza (Bioline) (složení reakčních směsí viz Tab. 2 a 3). Produkty do velikosti 2500 bp byly amplifikovány MyTaq DNA polymerázou a pro očekávané větší produkty byla použita MyFi DNA polymeráza. K 23  $\mu\text{l}$  reakční směsi a byly přidány 2  $\mu\text{l}$  vzorků cDNA. Teplotní průběhy PCR reakcí s jednotlivými páry primerů jsou uvedeny v Tab. 5 a 6.

Tabulka 2: Složení reakční pro PCR s MyTaq DNA polymerázou

<b>Položka</b>	<b>Koncentrace prac. roztoku</b>	<b>Konečná koncentrace</b>	<b>1 test [<math>\mu\text{l}</math>]</b>
<b>pufr</b>	5x	1x	5
<b>voda</b>			17,54
<b>f primer</b>	20 $\mu\text{mol}\cdot\text{l}^{-1}$	0,1 $\mu\text{mol}\cdot\text{l}^{-1}$	0,13
<b>r primer</b>	20 $\mu\text{mol}\cdot\text{l}^{-1}$	0,1 $\mu\text{mol}\cdot\text{l}^{-1}$	0,13
<b>MyTaq DNA pol</b>	5 U· $\mu\text{l}^{-1}$	1 U	0,2

Tabulka 3: Složení reakční pro PCR s MyFi DNA polymerázou

<b>Položka</b>	<b>Koncentrace prac. roztoku</b>	<b>Konečná koncentrace</b>	<b>1 test [<math>\mu\text{l}</math>]</b>
<b>pufr</b>	5x	1x	5
<b>voda</b>			17,24
<b>f primer</b>	20 $\mu\text{mol}\cdot\text{l}^{-1}$	0,1 $\mu\text{mol}\cdot\text{l}^{-1}$	0,13
<b>r primer</b>	20 $\mu\text{mol}\cdot\text{l}^{-1}$	0,1 $\mu\text{mol}\cdot\text{l}^{-1}$	0,13
<b>MyFi DNA pol</b>	2 U· $\mu\text{l}^{-1}$	1 U	0,5



Tabulka 4: Seznam použitých primerů

Název primeru	Velikost PCR produktu	Teplota nasedání primerů	Poznámka
430 F 908 R	479 bp	55 °C	detekční
699 F 1598 R	900 bp	53 °C	
1395 F 2058 R	664 bp	54 °C	
1569 F 3178 R	1609 bp	53 °C	
3007 F 4684 R	1678 bp	59 °C	
4408 F 8258 R	3851 bp	54 °C	
5276 F 7918 R			sekvenační
5874 F 7423 R			sekvenační
8333 F 9963 R	1626 bp	55 °C	
9939 F 12480 R	2541 bp	56 °C	
10465 F 11718 R			sekvenační
12122 R 5' 496 R 3'			RACE PCR
M13 F M13 R	-	55 °C	univerzální

Tabulka 5: Teplotní průběh PCR amplifikace genomických fragmentů

Proces	Teplota	Čas	Počet cyklů
Denaturace	95 °C	2 min	1x
Denaturace	95 °C	30 s	
Nasedání	X °C	30 s	35x
Syntéza	72 °C	30 s	
Dosyntetizování	72 °C	5 min	1x

Teplota nasedání primerů pro jejich jednotlivé páry primerů: 430 F/908 R – 55 °C  
1395 F/2058 R – 54 °C

Tabulka 6: Teplotní průběh PCR reakce s primery RBD 699 F/1598 R

Proces	Teplota	Čas	Počet cyklů
Denaturace	95 °C	3 min	1x
Denaturace	95 °C	1 min	
Nasedání	X °C	1 min	35x
Syntéza	72 °C	1 <sup>a</sup> /1,5 <sup>b</sup> /3 <sup>c</sup> /4 <sup>d</sup> min	
Dosyntetizování	72 °C	5 min	1x

<sup>a</sup> – pro primery RBD 699 F/1598 R; <sup>b</sup> - RBD 1569 F/3178 R, 3007 F/4684 R a 8333 F/9963 R; <sup>c</sup> - RBD 5276 F/7918 R, 9939 F/12480 R; <sup>d</sup> - RBD 4408 F/8258 R

Teploty nasedání primerů viz Tab. 4

#### 4.4.6 RACE PCR

Ověření 5' a 3' konce genomické sekvence bylo provedeno SMARTer RACE 5'/3' kitu (Takara Bio) pomocí podle doporučení výrobce.

#### 4.4.7 Elektroforetická separace

Produkty PCR reakce byly analyzovány pomocí elektroforetické separace v 1% agarózovém gelu s přídavkem 2 µl barviva GelRed (Biotium). Elektroforetická separace probíhala v prostředí TAE pufru při 80 V po dobu cca 40 min. Velikost získaných PCR produktů byla odhadnuta srovnáním se standardy molekulové hmotnosti. V závislosti na očekávané velikosti PCR produktu byl použit GeneRuler 100bp DNA Ladder (Fermentas), GeneRuler 100bp Plus DNA Ladder (ThermoFisher Scientific) anebo FastRuler High Range DNA Ladder (ThermoFisher Scientific). Výsledky elektroforetické separace byly vizualizovány pomocí UV-transiluminátoru (G : BOX, Syngene) a programu GeneSnap.

Tabulka 7: Složení 'Master Mix' směsi pro RACE PCR

Položka	1 test [µl]
PCR voda	15,5
2x SeqAmp pufr	25
SeqAmp DNA polymeráza	1

Tabulka 8: Složení reakční směsi pro RACE PCR

Položka	Koncentrace prac. roztoku	Konečná koncentrace	1 test [µl]
cDNA (pro 5' nebo 3' konec)			2,5
UPM	10x	1x	5
5' nebo 3' specifický primer	10 µmol·l <sup>-1</sup>	0,2 µmol·l <sup>-1</sup>	1
Master Mix			41,5

Tabulka 9: Teplotní průběh PCR amplifikace konců genomu

Proces	Teplota	Doba	Počet cyklů
Denaturace	94 °C	30 s	5x
Annealing a syntéza	72 °C	3 min	
Denaturace	94 °C	30 s	5x
Annealing	70 °C	30 s	
Syntéza	72 °C	3 min	
Denaturace	94 °C	30 s	25x
Annealing	68 °C	30 s	
Syntéza	72 °C	3 min	

#### 4.4.8 Izolace PCR produktu z gelu

Pro potřebu následného sekvenčního značení, případně klonování byl PCR produkt přečištěn izolací z gelu pomocí soupravy QIAquick Gel Extraction kit (Quiagen) podle doporučení výrobce.

Byla provedena elektroforetická separace PCR produktů v 1% agarózovém gelu s přídavkem ethidium bromidu v prostředí TAE pufru při napětí 80 V. PCR produkty byly vyříznuty z gelu skalpelem a přemístěny do zkumavek. Byl k nim přidán QG pufr v poměru 3 : 1. Zkumavky byly inkubovány při 50 °C v suché lázni, dokud se gel zcela nerozpustil. Dále byl přidán isopropanol v poměru 1 : 1. DNA byla navázána na silikátovou fritu kolony za pomoci centrifugace 1 min při 13 000 otáčkách (Centrifuga Biofuge pico, Heraeus). DNA byla promyta QG pufrem a následně PE pufrem. DNA byla z kolony eluována 30 µl sterilní deionizované vody. Její koncentrace v roztoku byla zjištěna pomocí fluorimetru DynaQuant DQ 200 (Amersham Biosciences). Izolovaná DNA byla dlouhodobě uchovávána při -20 °C.

#### 4.4.9 Klonování PCR produktu

Izolované PCR produkty o velikosti větší než 1700 bp byly klonovány v buňkách bakterií *Escherichia coli* (kmen DH5α). Menší PCR produkty byly sekvenovány přímo. Dále byly klonovány PCR produkty, z nichž byly získány nejednoznačné sekvence. Ke klonování PCR produktů byl použit pGEM-T Vector systém (Promega).

Byla připravena ligační směs smícháním pGEM-T ligačního pufru, T4 DNA ligázy, pGEM-T plazmidu a PCR produktu (viz Tab. 13). Směs byla inkubována při 4 °C přes noc. Následující den bylo smícháno 50 µl kompetentních buněk s 5 µl ligační směsí, obsah zkumavky byl inkubován 20 min na ledu. Kompetentní buňky byly transformovány pomocí teplotního šoku při teplotě 42 ° po dobu 45 s v suché lázni. Poté byla směs okamžitě umístěna na led, kde byla inkubována minimálně 2 min. Transformované kompetentní buňky byly kultivovány

v SOC médiu v rotační třepačce při teplotě 37 °C po dobu 1,5 hodiny. Suspenzní kultura namnožených transformovaných bakterií byla zakoncentrována centrifugací a přenesena na povrch tuhého LB média s ampicilinem, X-Galem a IPTG. Transformované buňky na médiu byly kultivovány přes noc při 37 °C.

Jednotlivé kolonie byly dalšího dne přeneseny na nové médium a otestovány na přítomnost PCR produktu pomocí PCR reakce s univerzálními primery M13. V závislosti na očekávané velikosti PCR produktu byla použita k amplifikaci buď MyTaq DNA polymeráza (Bioline) (složení reakční směsi viz Tab. 14), nebo MyFi DNA polymeráza (složení reakční směsi viz Tab. 15). Teplotní průběh testovací PCR reakce je uveden v Tab. 16.

Detekce PCR produktů byla provedena pomocí elektroforetické separace za podmínek popsaných v kapitole 4.4.7. Z bakterií, které nesou požadovaný inzert, byly vytvořeny konzervy. Bakterie byly přeočkovány do tekutého LB média s ampicilinem a kultivovány cca 16 hodin na rotační třepačce při 37 °C. Následně bylo 850 µl bakteriální suspenze smícháno se 150 µl sterilního glycerolu. Konzervy byly uchovávány při -80 °C.

Tabulka 10: Složení ligační směsi pro klonování PCR produktu

Položka	1 reakce [µl]
pGEM-T ligační pufr	5
T4 DNA ligáza	1
pGEM-T plazmid	0,5
PCR produkt	3,5

Tabulka 11: Složení reakční směsi pro PCR amplifikaci pomocí MyTaq DNA polymerázy

Položka	Koncentrace prac. roztoku	Konečná koncentrace	1 test [µl]
pufr	5x	1x	5
voda			19,55
F primer	20 µmol·l <sup>-1</sup>	0,1 µmol·l <sup>-1</sup>	0,13
R primer	20 µmol·l <sup>-1</sup>	0,1 µmol·l <sup>-1</sup>	0,13
MyTaq DNA polymeráza	5 U·µl <sup>-1</sup>	1 U	0,2

Tabulka 12: Složení reakční směsi pro PCR amplifikaci pomocí MyFi DNA polymerázy

Položka	Koncentrace prac. roztoku	Konečná koncentrace	1 test [µl]
pufr	5x	1x	6
voda			22,9
F primer	20 µmol·l <sup>-1</sup>	0,2 µmol·l <sup>-1</sup>	0,3
R primer	20 µmol·l <sup>-1</sup>	0,2 µmol·l <sup>-1</sup>	0,3
MyFi DNA polymeráza	2 U·µl <sup>-1</sup>	1 U	0,5

Tabulka 13: Teplotní průběh PCR reakce

Proces	Teplota	Čas	Počet cyklů
Denaturace	95 °C	5 min	1x
Denaturace	95 °C	1 min	
Nasedání	55 °C	1 min	40x
Syntéza	72 °C	4 min	
Dosyntetizování	72 °C	5 min	1x

#### 4.4.10 Sangerovo sekvenování

Izolované PCR produkty byly označeny pomocí soupravy BigDye Terminator v 3.1 Cycle Sequencing kit (Applied Biosystems) podle doporučení výrobce.

Byla připravena reakční směs smícháním 1 µl pufru, 1 µl Big Dye a 2 µl primeru o koncentraci 1,6 µmol·l<sup>-1</sup>. K reakční směsi bylo přidáno 6 µl DNA. Množství templátové DNA bylo zvoleno v závislosti na velikosti fragmentu (20 ng pro fragment o velikosti do 500 bp; 25 ng - 1000 bp; 30 ng - 1500 bp; 50 ng - 2000 bp a více). Teplotní průběh reakce je uveden v Tab. 17.

Vlastní sekvenování bylo provedeno na genetickém analyzátoru ABI PRISM 3730 na Ústavu experimentální botaniky AV ČR (CR Haná) v Olomouci.

#### 4.4.11 Bioinformatická analýza

Výstupy ze Sangerova sekvenování byly skládány do kontigů pomocí programu Lasergene (DNASTAR). V programu bylo nastaveno trimování 'readů' s nízkou kvalitou, jinak bylo použito výchozí nastavení programu. Výsledky skládání byly manuálně kontrolovány.

Výsledná sekvence, s cílem vyloučit možné artefakty, byla opětovně ověřena porovnáním pomocí algoritmu BLASTN a BLASTX.

V získané sekvenci byly identifikovány potenciální čtecí rámce pomocí programu ORF Finder (NCBI) (minimal ORF length: 600 nt; genetic code: standard; ORF start codon: ATG only; ignore nested ORF). Identifikované potenciální počátky ORF byly ověřeny za pomoci identifikace signálních sekvencí pomocí manuální kontroly a mnohočetného alignmentu. Pozice ORF byla vykreslena do fyzické mapy genomu pomocí programu pDRAW32.

Získaná sekvence byla porovnána se sekvencemi příbuzných virů dostupných v databázi GenBank. Byl proveden 'multiple alignment' pomocí algoritmu Clustal W (defaultní nastavení). Identita sekvencí byla stanovena pomocí p-distance analýzy. Fylogenetická analýza byla provedena s využitím Neighbor-Joining metody s bootstrap opakováním 1000x (byly zvoleny modely: Tamura 3-parametrický model pro nukleotidové sekvence, JTT model pro

aminokyselinové sekvence). Fylogenetické stromy byly vizualizovány pomocí Tree Explorer. Všechny tyto analýzy byly provedeny v programu MEGA 7.0.

Tabulka 14: Teplotní průběh sekvenačního značení

<b>Proces</b>	<b>Teplota</b>	<b>Čas</b>	<b>Počet cyklů</b>
<b>Denaturace</b>	95 °C	1 min	1x
<b>Denaturace</b>	95 °C	10 s	
<b>Annealing</b>	50 °C	5 s	40x
<b>Syntéza</b>	60 °C	4 min	

## 5 VÝSLEDKY

Sekvenováním vzorku B15 pomocí NGS bylo získáno celkem 27 747 697 'readů' (s průměrnou délkou 100 nt). Bylo složeno 4 638 kontigů, přičemž 93 z nich vykazovalo podobnost se sekvencemi některých virů z rodu *Nucleorhabdovirus*. Většina těchto kontigů v BLASTx vykazovala největší podobnost se sekvencemi proteinů viru *Datura yellow vein virus* (DYVV) a to 88 z nich. Se sekvencemi viru *Sonchus yellow net virus* (SYNV) vykazovalo největší podobnost 5 kontigů. Většina těchto kontigů byla podobná sekvenci RNA-dependentní RNA polymerázy (L protein), konkrétně 49 z nich vykazovalo podobnost v tomto proteinu viru DYVV a 4 kontigy s tímto proteinem viru SYNV. S každým ze zbývajících pěti proteinů viru DYVV bylo podobných několik kontigů (seznam kontigů podobných sekvencím nukleorhabdovirů viz Příloha 1). Kontigy, které vykazovaly podobnost s některým z těchto virů, byly prodlouženy pomocí mapování 'readů' a některé kontigy se podařilo spojit jejich mapováním na referenční sekvenci genomu viru DYVV, přičemž bylo získáno 81 konečných kontigů. Nejdelší kontig měl délku 1 160 nt a nejkratší 161 nt, průměrná délka kontigu byla 420 nt. Nejvyšší pokrytí kontigu 'ready' bylo 8,5 x a nejnižší 1,3 x, průměrné pokrytí kontigů bylo 3,7 x (viz Příloha 2). Některé kontigy se při namapování na referenční sekvenci překrývaly, ale nemohly být spojeny, protože se jejich sekvence lišily. Svědčí to o přítomnosti více izolátů viru v jedné rostlině.

Ze vzorku B152 byla pomocí Sangerova sekvenování získána téměř úplná sekvence genomu viru, jejíž součástí je i sekvence 3' konce genomu, kterou se podařilo získat pomocí RACE PCR. Naopak 5' RACE PCR se nezdařila, proto je sekvence na 5' konci v genu pro L a nepřekládané oblasti genomu neúplná. Celková délka sekvence je 12 822 bp.

Při opětovné analýze NGS dat bylo na sekvenci izolátu B152 namapováno 639 'readů', které pokrývaly 93,4 % sekvence. Průměrná délka namapovaných 'readů' byla  $98,3 \pm 5$  nt, průměrné pokrytí referenční sekvence 'ready' bylo 4,7x. 'Ready' nacházející se ve stejné pozici v genomu se od sebe navzájem lišily, což opětovně potvrdilo přítomnost více různých izolátů téhož druhu viru infikujících tutéž rostlinu, a to přinejmenším tří.

Získaná genomická sekvence byla analyzována pomocí algoritmu BLASTn. Tak byla potvrzena příslušnost viru B152 k virům z rodu *Nucleorhabdovirus*. Izolát vykazoval v různých parciálních oblastech (pokrytí sekvence 34–10 %) genomu podobnost s šesti zástupci virů z tohoto rodu, Black currant-associated rhabdovirus izolát Veloy, *Datura yellow vein virus*, *Zhuye pepper nucleorhabdovirus* isolate ZPNu1, *Sonchus yellow net virus*, *Apple rootstock*

virus A, *Alfalfa associated nucleorhabdovirus*. BLASTx analýza potvrdila podobnost (97-75 %) oblastí L proteinu s L proteinem DYVV, ZPNuV, BCaRV-1, ApRVA a AaNv.

V sekvenci potenciálně nového rhabdoviru bylo identifikováno 6 ORF s organizací typickou pro viry z rodu *Nucleorhabdovirus* (viz Obr. 1). První ORF má délku 1374 nt (pozice v genomu 179-1552) a kóduje plášťový protein (N) o délce 457 aa a molekulové hmotnosti 50,6 kDa. ORF2 má délku 1032 nt (pozice v genomu 1633-2664) a kóduje fosfoprotein (P) o délce 343 aa a molekulové hmotnosti 37,9 kDa. ORF3 o délce 978 nt (pozice v genomu 2778-3755) kóduje protein označený P3, u něž se předpokládá funkce 'movement proteinu'. P3 protein má velikost 325 aa a molekulovou hmotnost 37,1 kDa. Jako ORF4 byla algoritmem Open Reading Frame Finder identifikována sekvence o délce 1086 nt s pozicí v genomu 3670-4755, nicméně začátek této sekvence zasahuje do předchozího ORF, což se u virů z rodu *Nucleorhabdovirus* nestává. Další iniciační kodony se nachází v pozicích 3685, 3817 a 3922. S ohledem na velikost čtvrtého ORF, kódujícího matrix protein (M), u ostatních nukleorhabdovirů se dá předpokládat, že tento ORF začíná až na posledním z nich. Nachází se tedy v pozici 3922-4755 a má délku 801 nt. Kóduje matrix protein o velikosti 266 aa a molekulové hmotnosti 30,8 kDa. ORF5 má délku 1884 nt (pozice v genomu 4909-6792) a kóduje glykoprotein (G) o velikosti 627 aa a molekulové hmotnosti 70,5 kDa. Posledním, a také největším otevřeným čtecím rámcem je ORF6. Vzhledem k tomu, že se nepodařilo získat sekvenci z 5'-konce genomu, je tento ORF neúplný. Z ORF6 byla získána parciální sekvence 5850 nt (pozice v genomu 6973-12822), tento ORF kóduje RNA-dependentní RNA polymerázu (L), která je tvořena více než 1950 aa a její molekulová hmotnost přesahuje 221,5 kDa.

Mezi ORFs se nachází konzervovaná mezigenová sekvence typická pro rhabdoviry o délce 16 nt (ATATAAGAAAACCAA) (viz Tab. 18), byla identifikována v pozicích 1557-1573, 2691-2706, 3825-3840, 4835-4850 a 6875-6890. Totožnou sekvenci jen o 2 nt kratší než mezigenová sekvence najdeme před začátkem prvního ORF (pozice v genomu 111-124). Tato sekvence zahrnuje signál k polyadenylaci o délce 9 nt, mezigenový mezerník, který je tvořen 2 nt a část signálu pro start transkripce. Před signálem pro polyadenylaci mRNA se nachází 3 nt- sekvence o neznámé funkci, která je shodná u všech mezigenových sekvencí nového nukleorhabdoviru, ale nevyskytuje se u konzervativní sekvence nacházející se před prvním ORF.

Umístění této konzervované mezigenové sekvence v pozici 3825-3840 potvrzuje, že ORF4 nemůže začínat na dřívějším iniciačním kodonu než tom v pozici 3922, jinak by tato sekvence zasahovala do následujícího čtvrtého ORF.





***Nucleorhabdovirus B152***  
**12822 bp**

Obrázek 1: Schéma genomu nového nukleorhabdovirusu B152 s lokalizovanými ORF

N – plášťový protein, P – fosfoprotein, P3 – předpokládaný ‘movement protein’,  
M – matrix protein, G – glykoprotein, L – RNA-dependentní RNA polymeráza (large protein)

Tabulka 15: Konzervovaná mezigenová sekvence a její funkce

	neznámá funkce	polyA signál	mezerník	start transkripce
sekvence v genomu	ATA	TAAGAAAAA	CC	AAMHV*
sekvence v mRNA	UAU	AUUCUUUUU	GG	UUKDB*

\*M – adenin nebo cytozin, H – adenin, cytosin nebo thymin, V – adenin, thymin nebo guanin, K – uracil nebo guanin, D – adenin, guanin nebo uracil, B – cytosin, guanin nebo uracil

P-distance analýza (v programu MEGA 7.0) nově detekovaného rhabdovirusu infikujícího bez černý prokázala jeho podobnost s jinými zástupci rodu *Nucleorhabdovirus* infikujícími různé rostlinné hostitele. Analýzou celé získané sekvence bylo zjištěno, že izolát B152 vykazuje největší 57% podobnost s virem DYVV (NC\_028231.1) izolovaným z thungberie křídlaté (*Thungberia alata*). Dalšími dvěma nejpodobnějšími viry jsou BCaRV-1 (MF543022.1), jediný izolát viru získaný z rybízu černého (*Ribes nigrum*), s 50 % identitou, a SYNIV (NC\_001615.3), izolát z experimentálního hostitele *Nicotiana edwardsonii*, se 49 % identitou. Sekvence ostatních zástupců tohoto rodu uznaných ICTV nebo navržených vykazovaly 34–43 % identitu s izolátem B152 (viz Tab. 19). S typovým zástupcem rodu *Cytorhabdovirus*, virem LNYV (NC\_007642.1), izolát B152 sdílí 34 % celogenomové sekvence, což je identita srovnatelná s viry z rodu *Nucleorhabdovirus* fylogeneticky vzdálenějšími vůči izolátu B152. Analýza identity nukleotidové sekvence celého genomu s typovými zástupci rodů *Dichorhavirus* a *Varicosavirus*, jejichž genom je segmentovaný, nebyla pro zjednodušení provedena.

Obdobné podobnosti byly detekovány i při analýze jednotlivých čtecích rámců a proteinů, které kódují. Nejvyšší identitu v proteinu L vykazuje izolát B152 s virem DYVV, konkrétně 52 % aa (57 % nt), následuje BCaRV-1 s 49% aa (54% nt) identitou a SYNIV s 45% aa (52% nt) identitou. Sekvence aminokyselin v proteinu L ostatních uznaných nebo navržených zástupců rodu *Nucleorhabdovirus* vykazují s izolátem B152 identitu 24-41 % a sekvence ORF kódujícího tento protein 38-49% identitu (viz Tab. 20). Izolát B152 vykazoval jen 20% (aa)

identitu v proteinu L (37% nt) s typovým zástupcem rodu *Cytorhabdovirus*, virem LNYV, podobně je tomu i s virem LBVV (NC\_011558.1), jediným zástupcem rodu *Varicosavirus*. Identita jeho proteinu L s izolátem B152 je 19 % aa (37 % nt). Typový zástupce rodu *Dichorhavirus*, OFV (LC222630.1), izolát Cym07 vykazuje v proteinu L s izolátem B152 identitu 28 % aa (41 % nt).

Zjištěná identita sekvence plášťového proteinu (N) izolátu B152 s viry z rodu *Nucleorhabdovirus* se pohybuje ve velkém rozsahu. Izolát vykazuje nejvyšší, 60% identitu pro aa- a 63% identitu pro nt- sekvenci s plášťovým proteinem viru DYVV. Podobně i s virem SYNIV vykazuje izolát B152 v tomto proteinu vysokou, 50% aa (55% nt) podobnost, s virem BCaRV-1 má v proteinu N identitu 49 % aa (55 % nt). Sekvence N proteinu ostatních uznaných nebo navržených zástupců rodu *Nucleorhabdovirus* vykazuje s izolátem B152 identitu 17-45 % aa a sekvence ORF kódujícího tento protein 36-52% nt identitu (viz Tab. 21). Izolát B152 vykazoval pouze 18% (aa) identitu v proteinu N (35% nt) s typovým zástupcem rodu *Cytorhabdovirus*, virem LNYV (NC\_007642.1). Podobně je tomu i s identitou tohoto proteinu nového izolátu s virem OFV, typovým zástupcem rodu *Dichorhavirus*, která činí 17 % aa (36 % nt), což značí malou podobnost, nicméně podobnost některých zástupců rodu *Nucleorhabdovirus* v tomto proteinu není větší.

Protein P3 vykazuje nejvyšší variabilitu z analyzovaných proteinů. Až na výjimky je identita v tomto proteinu mezi všemi nukleorhabdoviry velmi nízká. Stejně jako tomu bylo u analýzy identity celého genomu i proteinů L a N, jsou i v proteinu P3 nejpodobnější izolátu B152 viry DYVV, BCaRV-1 a SYNIV, přičemž virus DYVV vykazoval 44% aa (53% nt) identitu, BCaRV-1 28% aa (42% nt) identitu a SYNIV 25% aa (41% nt) identitu. Sekvence P3 proteinu ostatních uznaných nebo navržených zástupců rodu *Nucleorhabdovirus* vykazuje s izolátem B152 jen identitu 6-20 % a sekvence ORF kódujícího tento protein 29-36 % identitu (viz Tab. 22). Virus OFV, typový zástupce rodu *Dichorhavirus*, má v proteinu P3 s izolátem B152 identitu 11 % aa (33 % nt), je tedy v tomto proteinu popisovanému izolátu podobnější než některé nukleorhabdoviry.

Identita izolátu B152 a jemu nejpodobnějšímu viru, DYVV, ve všech genech lehce přesahuje 50 %. V genu pro P činí 53 %, v genu pro M 54 % a v genu pro G dokonce 59 %. Podobnost proteinů je o něco nižší než u genů, které je kódují. Podobnost proteinu P mezi izolátem B152 a DYVV je jen 35 %, proteinu M 40 % a proteinu G 49 % (Tab. 23). Až na několik výjimek jsou mezi nukleorhabdoviry identity v proteinech (aa) nižší než identity v genech, které je kódují (nt), což naznačuje velké množství nesynonymních mutací u nového viru i ostatních

zástupců rodu *Nucleorhabdovirus*. Pro splnění kritéria pro uznání izolátu jako nového virového druhu na základě rozdílnosti v sekvenci je zapotřebí, aby izolát nepřesáhl 50% identitu v nukleotidové sekvenci kódujících oblastí. Tuto podmínku izolát těsně nespĺňuje, může být ale považován za nový druh na základě rozdílné ekologie.

Podle fylogenetické analýzy celého genomu vytváří nový nukleorhabdovirus B152 samostatnou statisticky průkaznou větev (Obr. 2). Nejpřibuznější tomuto viru je virus DYVV. Tyto dva viry vytváří klastr společně s viry BCaRV-1, SYNV, ApRVA, AaNv a virem MFSV, který je nejméně příbuzný ostatním virům z tohoto klastru. Ostatní viry z rodu *Nucleorhabdovirus* tvoří samostatný klastr. Virus LNYV z rodu *Cytorhabdovirus* vytváří samostatnou větev lišící se od všech doposud známých nukleorhabdovirů. Tato zjištění potvrdila i analýza jednotlivých ORF. V rámci fylogenetické analýzy sekvence proteinu L byly oproti analýze celého genomu přidány i sekvence typových zástupců rodů rhabdovirů se segmentovaným genomem, OFV z rodu *Dichorhavirus* a LBVV z rodu *Varicosavirus*. Do téhož klastru jako nový nukleorhabdovirus B152 se vyvětvil i virus OFV, přestože patří do samostatného rodu *Dichorhavirus*. Nicméně, podobně jako v případě viru MFSV není umístění v tomto klastru statisticky průkazné. Varicosavirus LBVV, stejně jako cytorhabdovirus LNYV vytváří samostatnou větev mimo klastr tvořený nukleorhabdoviry a dichorhaviry (Obr. 3). Rozdíly ve větvení fylogenetických stromů vytvořených na základě sekvence celého genomu a proteinu L jsou jen v rámci dvou klastrů virů z rodu *Nucleorhabdovirus* a jsou zanedbatelné. Výsledky fylogenetické analýzy proteinu N potvrzují předpoklady vytvořené na základě analýz celého genomu a proteinu L (Obr. 4) Analýza aminokyselinové sekvence proteinu P3 má nižší statistickou průkaznost než předchozí analýzy, což souvisí s vysokou variabilitou v tomto proteinu. Přesto potvrzuje předpoklady předchozích analýz (Obr. 5). Nukleotidové sekvence ORF vykazovaly obdobné větvení.

Tabulka 16: Procentuální identita genomických sekvencí B152 a dalších zástupců rodu *Nucleorhabdovirus*

	celý genom	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	<b>B152</b>															
2	<b>DYVV_(NC_028231.1)</b>	<b>0,57</b>														
3	<b>BCaRV-1_(MF543022.1)</b>	<b>0,50</b>	0,51													
4	<b>SYNV_(NC_001615.3)</b>	<b>0,49</b>	0,49	0,55												
5	<b>ApRVA_(MH778545.1)</b>	<b>0,43</b>	0,44	0,43	0,43											
6	<b>AaNv_(MG948563)</b>	<b>0,43</b>	0,42	0,42	0,42	0,42										
7	<b>MFSV_(NC_005974.1)</b>	<b>0,37</b>	0,38	0,37	0,37	0,37	0,37									
8	<b>TaVCV_(NC_006942.1)</b>	<b>0,37</b>	0,37	0,36	0,36	0,35	0,36	0,36								
9	<b>MMaV_(MK112501.2)</b>	<b>0,37</b>	0,36	0,36	0,36	0,35	0,36	0,36	0,66							
10	<b>PYDV_(KY549567.1)</b>	<b>0,37</b>	0,36	0,36	0,36	0,35	0,35	0,35	0,41	0,42						
11	<b>MMV_(MK828539.1)</b>	<b>0,36</b>	0,36	0,36	0,36	0,35	0,36	0,36	0,59	0,60	0,40					
12	<b>MIMV_(DQ186554.1)</b>	<b>0,36</b>	0,36	0,35	0,36	0,35	0,35	0,36	0,58	0,60	0,41	0,59				
13	<b>EMDV_(KC905081.1)</b>	<b>0,36</b>	0,36	0,35	0,35	0,35	0,36	0,35	0,41	0,42	0,54	0,41	0,41			
14	<b>PhCMoV_(KX636164.1)</b>	<b>0,35</b>	0,36	0,36	0,36	0,35	0,36	0,36	0,41	0,42	0,54	0,41	0,41	0,72		
15	<b>RYSV_(NC_003746.1)</b>	<b>0,34</b>	0,34	0,34	0,34	0,34	0,34	0,34	0,38	0,38	0,39	0,38	0,38	0,38	0,39	
16	<b>LNYV_(NC_007642.1)</b>	<b>0,34</b>	0,34	0,34	0,34	0,33	0,34	0,35	0,36	0,36	0,35	0,35	0,35	0,34	0,35	0,34

B152 – nový nukleorhabdovirus; DYVV – *Datura yellow vein virus*; BcaRV-1 – Black currant-associated rhabdovirus; SYNv – *Sonchus yellow net virus*; ApRVA – Apple rootstock virus A; AaNv – Alfalfa-associated nucleorhabdovirus; MFSV – *Maize fine streak virus*; TaVCV – *Taro vein chlorosis virus*; MmaV – Morogoro maize-associated virus; PYDV – *Potato yellow dwarf virus*; MMV – *Maize mosaic virus*; MIMV – *Maize Iranian mosaic virus*; EMDV – *Eggplant mottled dwarf virus*; PhCMoV – *Physostegia chlorotic mottle virus*; RYSV – *Rice yellow stunt virus*; LNYV – *Lettuce necrotic yellows virus (Cytorhabdovirus)*

Tabulka 17: Procentuální identita nukleotidové sekvence genu pro L protein a aminokyselinové sekvence tohoto proteinu u virů z rodu *Nucleorhabdovirus* a typových zástupců rodů z čeledi *Rhabdoviridae*, kteří infikují rostliny

	L	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	<b>B152</b>		<b>0,52</b>	<b>0,49</b>	<b>0,45</b>	<b>0,41</b>	<b>0,38</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,24</b>	<b>0,25</b>	<b>0,24</b>	<b>0,25</b>	<b>0,25</b>	<b>0,24</b>	<b>0,28</b>	<b>0,20</b>	<b>0,19</b>
2	<b>DYVV_(NC_028231.1)</b>	<b>0,57</b>		0,49	0,44	0,40	0,38	0,25	0,24	0,25	0,24	0,24	0,24	0,24	0,24	0,25	0,27	0,20	0,18
3	<b>BCaRV-1_(MF543022.1)</b>	<b>0,54</b>	0,55		0,45	0,41	0,37	0,25	0,22	0,23	0,24	0,23	0,24	0,24	0,24	0,24	0,26	0,19	0,18
4	<b>SYNV_(NC_001615.3)</b>	<b>0,52</b>	0,51	0,52		0,40	0,37	0,25	0,23	0,24	0,24	0,23	0,24	0,24	0,25	0,24	0,26	0,20	0,18
5	<b>ApRVA_(MH778545.1)</b>	<b>0,49</b>	0,50	0,50	0,49		0,38	0,25	0,24	0,24	0,24	0,24	0,23	0,23	0,23	0,24	0,27	0,20	0,18
6	<b>AaNv_(MG948563.1)</b>	<b>0,49</b>	0,48	0,48	0,48	0,49		0,26	0,23	0,23	0,25	0,23	0,24	0,25	0,25	0,23	0,27	0,20	0,18
7	<b>MFSV_(NC_005974.1)</b>	<b>0,41</b>	0,40	0,41	0,40	0,42	0,42		0,24	0,25	0,24	0,25	0,24	0,26	0,26	0,24	0,26	0,20	0,19
8	<b>TaVcV_(NC_006942.1)</b>	<b>0,40</b>	0,39	0,39	0,39	0,38	0,39	0,39		0,74	0,31	0,65	0,63	0,32	0,32	0,28	0,26	0,21	0,18
9	<b>MMaV_(MK112501.2)</b>	<b>0,40</b>	0,39	0,39	0,39	0,38	0,38	0,39	0,67		0,30	0,65	0,65	0,32	0,31	0,28	0,26	0,20	0,18
10	<b>PYDV_(KY549567.1)</b>	<b>0,39</b>	0,40	0,38	0,39	0,39	0,39	0,39	0,43	0,44		0,31	0,30	0,53	0,53	0,34	0,26	0,20	0,17
11	<b>MMV_(MK828539.1)</b>	<b>0,39</b>	0,39	0,40	0,39	0,38	0,39	0,40	0,62	0,63	0,43		0,64	0,32	0,31	0,29	0,26	0,21	0,19
12	<b>MIMV_(DQ186554.1)</b>	<b>0,40</b>	0,39	0,39	0,40	0,38	0,39	0,39	0,61	0,63	0,44	0,63		0,32	0,32	0,28	0,26	0,20	0,17
13	<b>EMDV_(KC905081.1)</b>	<b>0,39</b>	0,39	0,38	0,39	0,39	0,39	0,39	0,44	0,44	0,57	0,45	0,44		0,84	0,32	0,25	0,19	0,18
14	<b>PhCMoV_(KX636164.1)</b>	<b>0,38</b>	0,39	0,38	0,40	0,39	0,40	0,39	0,44	0,44	0,57	0,44	0,44	0,74		0,32	0,25	0,20	0,18
15	<b>RYSV_(NC_003746.1)</b>	<b>0,39</b>	0,38	0,39	0,38	0,39	0,38	0,39	0,43	0,44	0,48	0,43	0,44	0,46	0,48		0,25	0,21	0,18
16	<b>OFV_(LC222630.1)</b>	<b>0,41</b>	0,40	0,40	0,40	0,41	0,41	0,40	0,40	0,40	0,41	0,40	0,41	0,41	0,40	0,40		0,21	0,19
17	<b>LNYV_(NC_007642.1)</b>	<b>0,37</b>	0,37	0,37	0,36	0,37	0,38	0,37	0,36	0,36	0,36	0,36	0,36	0,36	0,36	0,36	0,36	0,38	0,20
18	<b>LBVV_(NC_011558.1)</b>	<b>0,37</b>	0,37	0,37	0,37	0,36	0,38	0,36	0,36	0,35	0,35	0,36	0,36	0,34	0,35	0,35	0,37	0,36	

Levý dolní roh – identita v nukleotidové sekvenci; Pravý horní roh – identita v aminokyselinové sekvenci; B152 – nový nukleorhabdovirus; DYVV – *Datura yellow vein virus*; BcaRV-1 – Black currant-associated rhabdovirus; SYNV – *Sonchus yellow net virus*; ApRVA – Apple rootstock virus A; AaNv – Alfalfa-associated nucleorhabdovirus; MFSV – *Maize fine streak virus*; TaVcV – *Taro vein chlorosis virus*; MmaV – Morogoro maize-associated virus; PYDV – *Potato yellow dwarf virus*; MMV – *Maize mosaic virus*; MIMV – *Maize Iranian mosaic virus*; EMDV – *Eggplant mottled dwarf virus*; PhCMoV – *Physostegia chlorotic mottle virus*; RYSV – *Rice yellow stunt virus*; OFV – *Orchid fleck virus (Dichorhavirus)*; LNYV – *Lettuce necrotic yellows virus (Cytorhabdovirus)*; LBVV – *Lettuce big-vein virus (Varicosavirus)*

Tabulka 18: Procentuální identita nukleotidové sekvence genu pro N protein a aminokyselinové sekvence tohoto proteinu u virů z rodu *Nucleorhabdovirus* a typových zástupců rodů z čeledi *Rhabdoviridae*, kteří infikují rostliny

	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	<b>B152</b>		<b>0,60</b>	<b>0,49</b>	<b>0,50</b>	<b>0,45</b>	<b>0,38</b>	<b>0,20</b>	<b>0,20</b>	<b>0,21</b>	<b>0,19</b>	<b>0,21</b>	<b>0,20</b>	<b>0,17</b>	<b>0,18</b>	<b>0,18</b>	<b>0,17</b>	<b>0,18</b>
2	<b>DYVV_(NC_028231.1)</b>	<b>0,63</b>		0,49	0,50	0,46	0,39	0,22	0,19	0,20	0,20	0,20	0,19	0,19	0,19	0,19	0,18	0,17
3	<b>BCaRV-1_(MF543022.1)</b>	<b>0,55</b>	0,55		0,48	0,43	0,37	0,20	0,19	0,22	0,19	0,20	0,20	0,18	0,19	0,19	0,19	0,17
4	<b>SYNV_(NC_001615.3)</b>	<b>0,55</b>	0,55	0,54		0,41	0,38	0,20	0,16	0,19	0,19	0,19	0,17	0,16	0,17	0,19	0,17	0,17
5	<b>ApRVA_(MH778545.1)</b>	<b>0,52</b>	0,52	0,54	0,48		0,42	0,19	0,20	0,21	0,17	0,22	0,20	0,18	0,20	0,18	0,17	0,18
6	<b>AaNv_(MG948563.1)</b>	<b>0,50</b>	0,48	0,48	0,48	0,50		0,21	0,20	0,21	0,17	0,20	0,19	0,17	0,18	0,21	0,17	0,18
7	<b>MFSV_(NC_005974.1)</b>	<b>0,38</b>	0,38	0,39	0,39	0,40	0,39		0,20	0,23	0,21	0,21	0,21	0,18	0,20	0,21	0,20	0,14
8	<b>TaVCV_(NC_006942.1)</b>	<b>0,39</b>	0,39	0,40	0,38	0,38	0,39	0,37		0,75	0,21	0,64	0,56	0,21	0,21	0,22	0,19	0,19
9	<b>MMaV_(MK112501.2)</b>	<b>0,37</b>	0,39	0,39	0,38	0,38	0,39	0,38	0,72		0,22	0,68	0,57	0,22	0,24	0,23	0,20	0,19
10	<b>PYDV_(KY549567.1)</b>	<b>0,37</b>	0,36	0,36	0,36	0,34	0,36	0,36	0,39	0,40		0,21	0,20	0,46	0,47	0,21	0,18	0,14
11	<b>MMV_(MK828539.1)</b>	<b>0,36</b>	0,37	0,38	0,38	0,38	0,37	0,37	0,62	0,63	0,39		0,60	0,23	0,22	0,24	0,19	0,19
12	<b>MIMV_(DQ186554.1)</b>	<b>0,38</b>	0,38	0,37	0,37	0,38	0,39	0,38	0,60	0,61	0,39	0,61		0,20	0,20	0,22	0,18	0,18
13	<b>EMDV_(KC905081.1)</b>	<b>0,37</b>	0,36	0,37	0,36	0,36	0,37	0,37	0,40	0,41	0,54	0,41	0,41		0,83	0,22	0,17	0,14
14	<b>PhCMoV_(KX636164.1)</b>	<b>0,36</b>	0,37	0,37	0,37	0,36	0,37	0,38	0,40	0,41	0,54	0,40	0,40	0,73		0,23	0,18	0,14
15	<b>RYSV_(NC_003746.1)</b>	<b>0,36</b>	0,39	0,36	0,38	0,36	0,39	0,37	0,39	0,40	0,39	0,41	0,41	0,40	0,40		0,18	0,13
16	<b>OFV_(LC222629)</b>	<b>0,36</b>	0,36	0,35	0,37	0,35	0,38	0,35	0,37	0,36	0,35	0,37	0,36	0,35	0,35	0,36		0,16
17	<b>LNyV_(NC_007642.1)</b>	<b>0,35</b>	0,35	0,36	0,36	0,35	0,35	0,35	0,34	0,35	0,33	0,36	0,34	0,36	0,35	0,33	0,35	

Levý dolní roh – identita v nukleotidové sekvenci; Pravý horní roh – identita v aminokyselinové sekvenci; B152 – nový nukleorhabdovirus; DYVV – *Datura yellow vein virus*; BcaRV-1 – Black currant-associated rhabdovirus; SYNV – *Sonchus yellow net virus*; ApRVA – Apple rootstock virus A; AaNv – Alfalfa-associated nucleorhabdovirus; MFSV – *Maize fine streak virus*; TaVCV – *Taro vein chlorosis virus*; MMaV – Morogoro maize-associated virus; PYDV – *Potato yellow dwarf virus*; MMV – *Maize mosaic virus*; MIMV – *Maize Iranian mosaic virus*; EMDV – *Eggplant mottled dwarf virus*; PhCMoV – *Physostegia chlorotic mottle virus*; RYSV – *Rice yellow stunt virus*; OFV – *Orchid fleck virus (Dichorhavirus)*; LNyV – *Lettuce necrotic yellows virus (Cytorhabdovirus)*

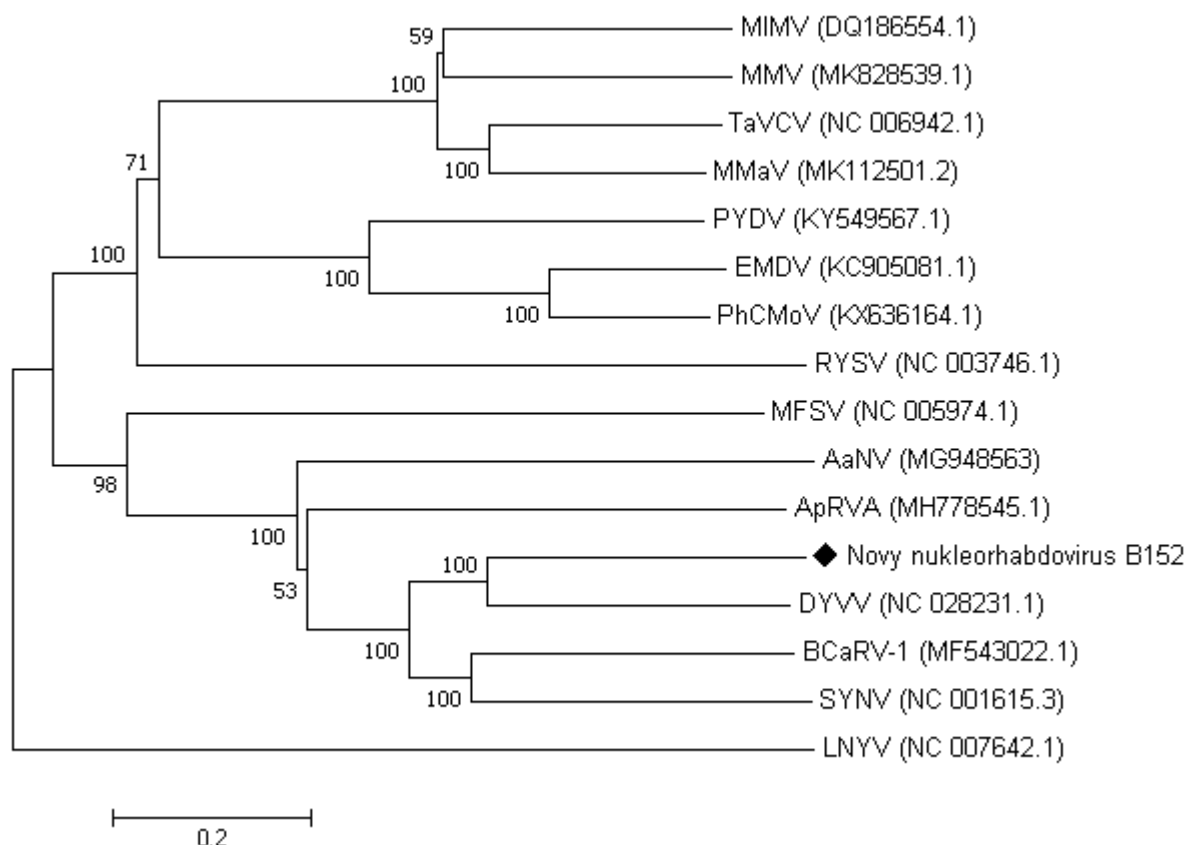
Tabulka 19: Procentuální identita nukleotidové sekvence genu pro P3 protein a aminokyselinové sekvence tohoto proteinu u virů z rodu *Nucleorhabdovirus*

	P3	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	<b>B152</b>		<b>0,44</b>	<b>0,28</b>	<b>0,25</b>	<b>0,20</b>	<b>0,18</b>	<b>0,09</b>	<b>0,10</b>	<b>0,13</b>	<b>0,09</b>	<b>0,06</b>	<b>0,12</b>	<b>0,11</b>	<b>0,11</b>
2	<b>DYVV_(NC_028231.1)</b>	<b>0,53</b>		0,26	0,28	0,20	0,21	0,09	0,12	0,13	0,07	0,09	0,12	0,12	0,12
3	<b>BCaRV-1_(MF543022.1)</b>	<b>0,42</b>	0,45		0,26	0,23	0,15	0,07	0,09	0,11	0,05	0,06	0,11	0,11	0,14
4	<b>SYNV_(NC_001615.3)</b>	<b>0,41</b>	0,42	0,45		0,19	0,19	0,08	0,07	0,14	0,07	0,06	0,12	0,11	0,14
5	<b>ApRVA_(MH778545.1)</b>	<b>0,36</b>	0,38	0,42	0,37		0,19	0,07	0,06	0,11	0,04	0,05	0,11	0,11	0,10
6	<b>AaNv_(MG948563.1)</b>	<b>0,36</b>	0,38	0,34	0,37	0,35		0,10	0,10	0,12	0,10	0,09	0,14	0,13	0,12
7	<b>TaVcV_(NC_006942.1)</b>	<b>0,29</b>	0,33	0,29	0,28	0,30	0,30		0,63	0,10	0,44	0,41	0,13	0,14	0,08
8	<b>MMaV_(MK112501.2)</b>	<b>0,29</b>	0,32	0,30	0,30	0,29	0,30	0,65		0,08	0,46	0,39	0,11	0,12	0,09
9	<b>PYDV_(KY549567.1)</b>	<b>0,36</b>	0,35	0,29	0,31	0,30	0,29	0,31	0,33		0,11	0,13	0,46	0,46	0,12
10	<b>MMV_(MK828539.1)</b>	<b>0,30</b>	0,31	0,30	0,30	0,30	0,29	0,54	0,57	0,33		0,42	0,13	0,12	0,10
11	<b>MIMV_(DQ186554.1)</b>	<b>0,31</b>	0,33	0,31	0,30	0,29	0,32	0,51	0,52	0,35	0,53		0,10	0,11	0,09
12	<b>EMDV_(KC905081.1)</b>	<b>0,35</b>	0,37	0,31	0,34	0,32	0,29	0,32	0,33	0,55	0,34	0,34		0,86	0,10
13	<b>PhCMoV_(KX636164.1)</b>	<b>0,35</b>	0,37	0,31	0,33	0,32	0,29	0,31	0,32	0,54	0,31	0,35	0,75		0,11
14	<b>OFV_(LC222629.1)</b>	<b>0,33</b>	0,31	0,34	0,32	0,31	0,32	0,29	0,32	0,31	0,33	0,31	0,33	0,31	

Levý dolní roh – identita v nukleotidové sekvenci; Pravý horní roh – identita v aminokyselinové sekvenci; B152 – nový nukleorhabdovirus; DYVV – *Datura yellow vein virus*; BcaRV-1 – Black currant-associated rhabdovirus; SYNV – *Sonchus yellow net virus*; ApRVA – Apple rootstock virus A; AaNv – Alfalfa-associated nucleorhabdovirus; TaVcV – *Taro vein chlorosis virus*; MMaV – Morogoro maize-associated virus; PYDV – *Potato yellow dwarf virus*; MMV – *Maize mosaic virus*; MIMV – *Maize Iranian mosaic virus*; EMDV – *Eggplant mottled dwarf virus*; PhCMoV – *Physostegia chlorotic mottle virus*; OFV – *Orchid fleck virus (Dichorhavirus)*

Tabulka 20: Procentuální identita všech nukleotidových sekvencí genů a aminokyselinových sekvencí proteinů viru *Datura yellow vein virus* s izolátem B152

ORF	nt	aa
<b>P</b>	0,53	0,35
<b>M</b>	0,54	0,40
<b>G</b>	0,59	0,49

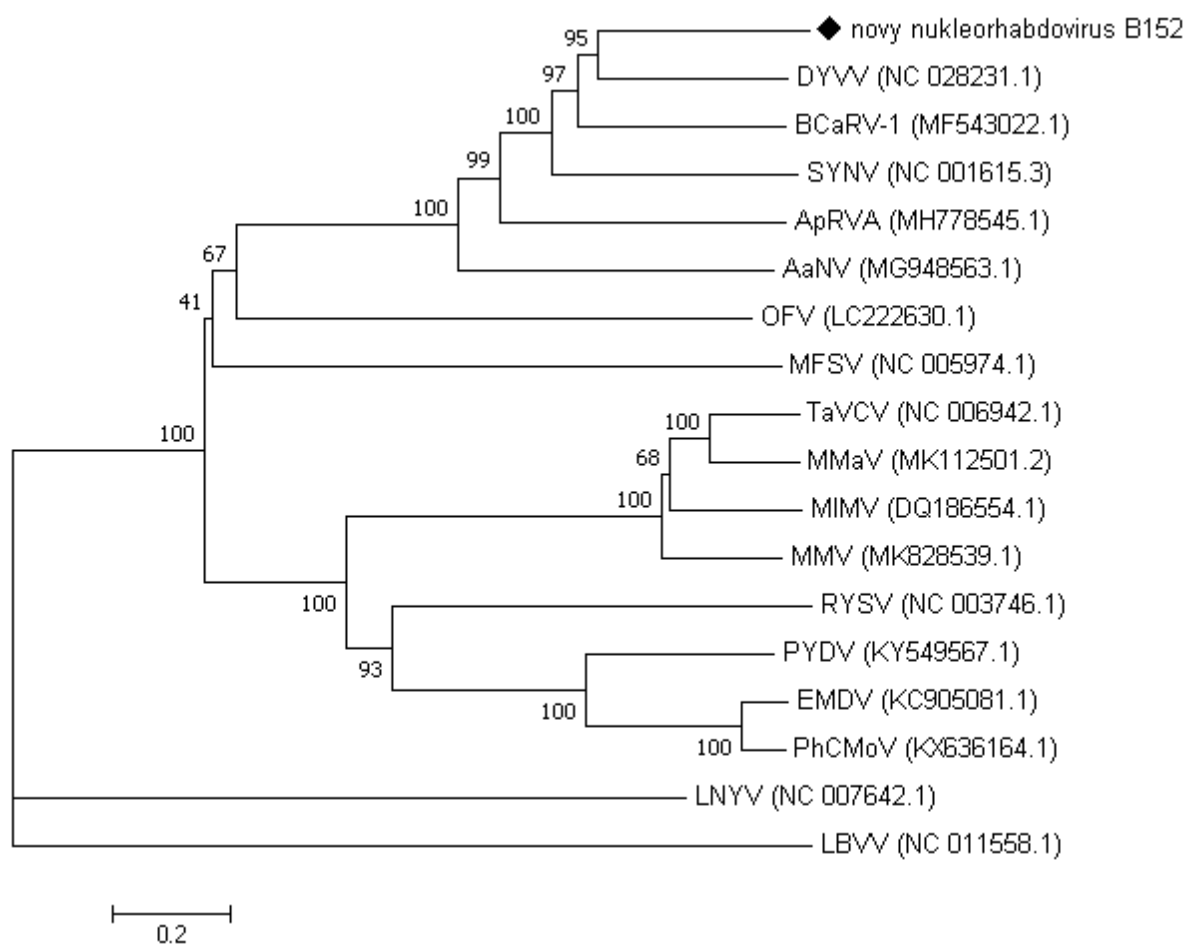


Obrázek 2: Fylogenetický strom vytvořený na základě celogenomové sekvence virů z rodu *Nucleorhabdovirus*

Jednotlivé viry jsou označeny zkratkou a přístupovým číslem databáze GenBank (Acc.No.). Měřítko vyjadřuje 0,2 substituce na pozici.

DYVV – *Datura yellow vein virus*; BCaRV-1 – Black currant-associated rhabdovirus; SYNV – *Sonchus yellow net virus*; ApRVA – Apple rootstock virus A; AaNv – Alfalfa-associated nucleorhabdovirus; MFSV – *Maize fine streak virus*; TaVCV – *Taro vein chlorosis virus*; MMaV – Morogoro maize-associated virus; PYDV - *Potato yellow dwarf virus*; MMV - *Maize mosaic virus*; MIMV - *Maize Iranian mosaic virus*; EMDV - *Eggplant mottled dwarf virus*; PhCMoV - *Physostegia chlorotic mottle virus*; RYSV - *Rice yellow stunt virus*; LNYV – *Lettuce necrotic yellows virus (Cytorhabdovirus)*

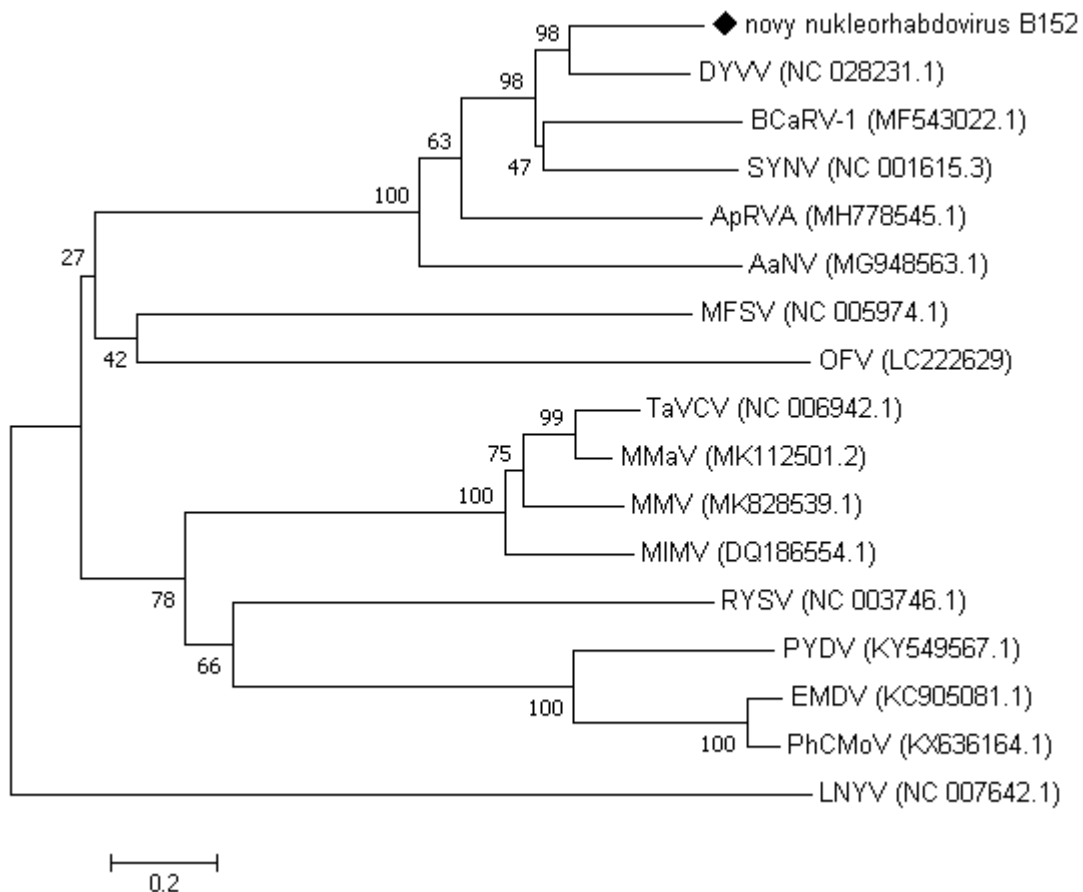




Obrázek 3: Fylogenetický strom vytvořený na základě aminokyselinové sekvence proteinu L virů z rodu *Nucleorhabdovirus* a zástupců rodů z čeledi *Rhabdoviridae*, kteří infikují rostliny

Jednotlivé viry jsou označeny zkratkou a přístupovým číslem databáze GenBank (Acc.No.). Měřítko vyjadřuje 0,2 substituce na pozici.

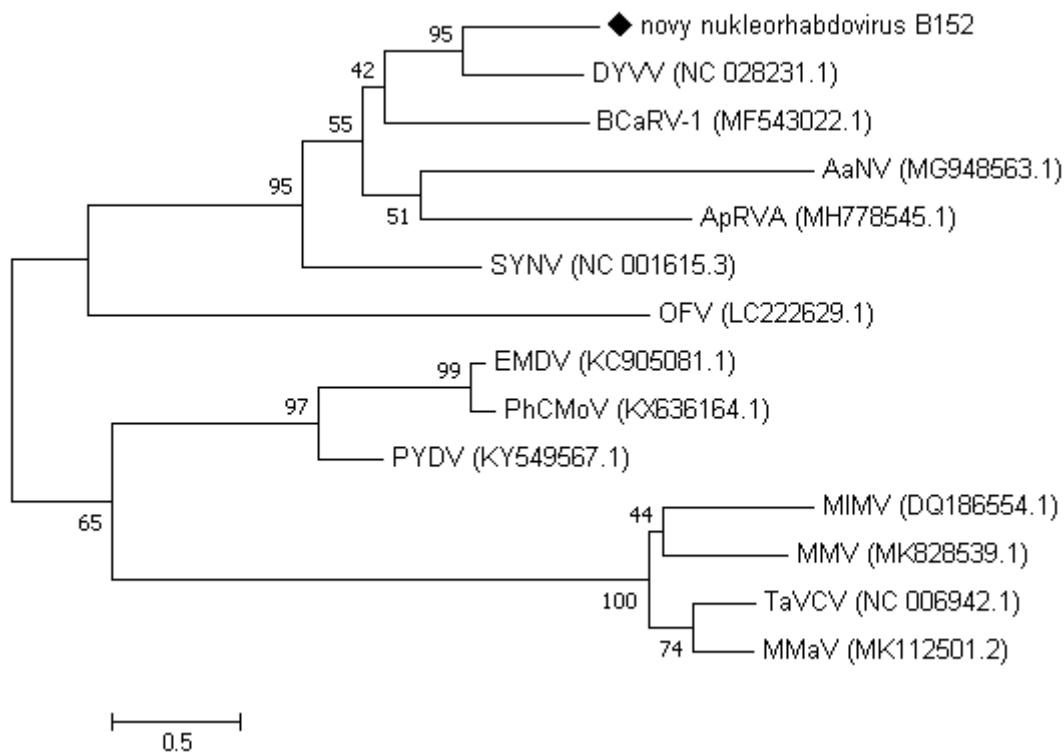
DYVV – *Datura yellow vein virus*; BCaRV-1 – Black currant-associated rhabdovirus; SYNv – *Sonchus yellow net virus*; ApRVA – Apple rootstock virus A; AaNv – Alfalfa-associated nucleorhabdovirus; MFSV – *Maize fine streak virus*; TaVCV – *Taro vein chlorosis virus*; MMaV – Morogoro maize-associated virus; PYDV - *Potato yellow dwarf virus*; MMV - *Maize mosaic virus*; MIMV - *Maize Iranian mosaic virus*; EMDV - *Eggplant mottled dwarf virus*; PhCMoV - *Physostegia chlorotic mottle virus*; RYSV - *Rice yellow stunt virus*; OFV - *Orchid fleck virus (Dichorhavirus)*; LNYV - *Lettuce necrotic yellows virus (Cytorhabdovirus)*; LBVV – *Lettuce big-vein virus (Varicosavirus)*



Obrázek 4: Fylogenetický strom vytvořený na základě aminokyselinové sekvence proteinu N virů z rodu *Nucleorhabdovirus* a zástupců rodů z čeledi *Rhabdoviridae*, kteří infikují rostliny

Jednotlivé viry jsou označeny zkratkou a přístupovým číslem databáze GenBank (Acc.No.). Měřítko vyjadřuje 0,2 substituce na pozici.

DYVV – *Datura yellow vein virus*; BCaRV-1 – Black currant-associated rhabdovirus; SYNV – *Sonchus yellow net virus*; ApRVA – Apple rootstock virus A; AaNv – Alfalfa-associated nucleorhabdovirus; MFSV – *Maize fine streak virus*; TaVCV – *Taro vein chlorosis virus*; MMaV – Morogoro maize-associated virus; PYDV - *Potato yellow dwarf virus*; MMV - *Maize mosaic virus*; MIMV - *Maize Iranian mosaic virus*; EMDV - *Eggplant mottled dwarf virus*; PhCMoV - *Physostegia chlorotic mottle virus*; RYSV - *Rice yellow stunt virus*; OFV - *Orchid fleck virus (Dichorhavirus)*; LNYV - *Lettuce necrotic yellows virus (Cytorhabdovirus)*



Obrázek 5: Fylogenetický strom vytvořený na základě aminokyselinové sekvence proteinu P3 virů z rodu *Nucleorhabdovirus*

Jednotlivé viry jsou označeny zkratkou a přístupovým číslem databáze GenBank (Acc.No.). Měřítko vyjadřuje 0,5 substituce na pozici.

DYVV – *Datura yellow vein virus*; BCaRV-1 – Black currant-associated rhabdovirus; SYNV – *Sonchus yellow net virus*; ApRVA – Apple rootstock virus A; AaNv – Alfalfa-associated nucleorhabdovirus; TaVVCV – *Taro vein chlorosis virus*; MMaV – Morogoro maize-associated virus; PYDV - *Potato yellow dwarf virus*; MMV - *Maize mosaic virus*; MIMV - *Maize Iranian mosaic virus*; EMDV - *Eggplant mottled dwarf virus*; PhCMoV - *Physostegia chlorotic mottle virus*; OFV - *Orchid fleck virus (Dichorhavirus)*

## 6 DISKUZE

Všechny viry, o kterých bylo dosud známo, že infikují bez černý (*Sambucus nigra*), mají genom tvořený (+)ssRNA. Nový nukleorhabdovirus B152 detekovaný na této rostlině pomocí NGS je prvním známým virem, který infikuje bez černý a má genom tvořený (-)ssRNA.

Aby bylo možné nový izolát popsáný v této diplomové práci navrhnout jako nový virový druh náležící do rodu *Nucleorhabdovirus*, musí splňovat některá z následujících kritérií definovaných International Committee on Taxonomy of Viruses (ICTV): mít méně než 50% identitu v nukleotidové sekvenci příbuzných genů, obsazovat jinou ekologickou niku (jedná se o okruh hostitelů a vektora) a být rozlišitelný od ostatních virů z tohoto rodu sérologickými testy či hybridizací nukleových kyselin. Druhy virů v rámci tohoto rodu se nejčastěji liší svým hostitelským okruhem a vektory (ICTV, 2020). S virem DYVV má izolát B152 nejvyšší identitu v celogenomové sekvenci i ve všech studovaných ORF, která u všech těchto sekvencí lehce přesahuje tuto hranici.

Nový virus se liší od ostatních virů z rodu *Nucleorhabdovirus* svou ekologií. Žádný z doposud popsáných virů z tohoto rodu neinfikuje bez, nebo jinou jemu blízkou příbuznou rostlinu. U viru DYVV, sekvenčně nejpodobnějšího virového druhu nukleorhabdoviru B152, jsou známí dva hostitelé, Durman obecný (*Datura stramonium*) z čeledi lilkovité (*Solanaceae*) a Thungberie křídlatá (*Thunbergia alata*) z čeledi paznehtíkovité (*Acanthaceae*). Tento virus byl detekován pouze v Austrálii (Thomas a Dietzgen, 1991; Dietzgen a kol., 1992). Durman a thungberie nejsou příbuzné ani na úrovni řádu, nicméně bez černý je vůči nim ještě fylogeneticky vzdálenější, než jsou tyto dvě rostliny vůči sobě navzájem (Angiosperm Phylogeny Webside, 2020). Nukleorhabdovirus B152 izolovaný v ČR je tedy možné považovat za nový druh na základě rozdílného okruhu hostitelů.

Virus SYNIV infikuje mléč zelinný (*Sonchus oleraceus*), dvouzubec chlupatý (*Bidens pilosa*) a salát (*Lactuca sativa*) z čeledi hvězdicovité (*Asteraceae*) na Floridě (Christie a kol., 1974; Falk a kol., 1986). Virus PYDV infikuje rostliny z čeledi lilkovité (*Solanaceae*), především lilek brambor (*Solanum tuberosum*). Zřídka může infikovat i rostliny jiných čeledí. Tento virus se vyskytuje pouze v Severní Americe (EPPO, 2020a). Virus EMDV infikuje lilek vejcoplodý (*Solanum melongena*), lilek brambor, lilek černý (*Solanum nigrum*), papriku setou (*Capsicum annuum*) a tabák (*Nicotiana tabacum*) z čeledi lilkovitých, okurku (*Cucumis sativum*) a meloun (*Cucumis melo*) z čeledi tykvovitých (*Cucurbitaceae*), zřídka může infikovat i rostliny jiných čeledí. Tento virus je rozšířen v Evropě, severní Africe a na Blízkém východě (EPPO, 2020b).

Virus MMV infikuje kukuřici (*Zea mays*) z čeledi lipnicovitých (*Poaceae*) v jižních oblastech Severní Ameriky a v Jižní Americe (Amar a kol., 2005). Virus MIMV může kromě kukuřice infikovat i pšenici (*Triticum aestivum*), ječmen (*Hordeum vulgare*) a možná i další lipnicovité. Je rozšířen v Íránu (Izadpanah a kol., 1983). Virus MFSV infikuje kukuřici ve Spojených státech (Redinbaugh a kol. 2002) Virus RYSV infikuje rýži (*Oryza sativa*) z čeledi lipnicovité v Číně, Japonsku a v jihovýchodní Asii (EPPO, 2020c). Virus TaVCV infikuje plodinu kolokázii jedlou (*Colocasia esculenta*) z čeledi áronovité (*Araceae*) na ostrovech v Tichém oceánu (Revill a kol., 2005).

Přestože není známo, že by jiný virus z rodu *Nucleorhabdovirus* bez infikoval, nevybočuje bez jako hostitel z „řady“. Nukleorhabdoviry infikují různé rostliny patřící k vyšším dvouděložným a jednoděložným. I když většina jejich hostitelů jsou byliny nebo obiloviny, dřeviny se v omezené míře mezi hostiteli také objevují. Příkladem dřevinného hostitele je rybíz černý (*Ribes nigrum*) z čeledi meruzalkovité (*Grossulariaceae*) (Wu a kol., 2018).

Posledním kritériem, které umožňuje taxonomické vyčlenění izolátu jakožto nového virového druhu, nejen jako vzdáleného typu již popsaného druhu, je možnost jeho odlišení od ostatních nukleorhabdovirů pomocí sérologických metod a metod založených na hybridizaci nukleových kyselin. Touto otázkou se ale tato diplomová práce nezabývá.

Uznat izolát popsáný v této práci jako nový druh v rámci rodu *Nucleorhabdovirus* pouze na základě rozdílnosti sekvencí nelze. Nicméně ani další již oficiálně uznané druhy virů z tohoto rodu nesplňují hranici maximální 50% identity (nt) v genových sekvencích. Z analýzy sekvencí vyplynulo, že např. viry MMV, MIMV a TaVCV mají mezi sebou vzájemnou identitu vyšší než 50 % ve všech zkoumaných ORF (geny pro L, N a P3 proteiny), stejně je tomu také u virů PYDV a EMDV. Pro uznání jako nového druhu by tedy měl stačit výrazně odlišný hostitel.

Dalším znakem, který by naznačoval se jedná o nový druh, by byl odlišný vektor. Ten ale není znám ani u nového nukleorhabdoviru B152, ani u viru DYVV (Thomas a Dietzgen, 1991). Nejpodobnější virus nukleorhabdoviru B152, u kterého je znám vektor, je SYNV, který je přenášen mšicí *Aphis coreopsidis* (Christie a kol., 1974). Kdyby byl znám vektor nového viru, bylo by možné také cíleně zkoumat jeho okruh hostitelů, podle toho, kterými dalšími rostlinami se vektor živí. Existuje hypotéza, že jednotlivé evoluční větve v rámci rodu *Nucleorhabdovirus* mají podobného vektora (Bandyopadhyay a kol., 2010). Dá se tedy předpokládat, že vektorem nového nukleorhabdoviru je nějaký druh mšice. Vzhledem k úzkému okruhu hostitelů mezi

zástupci rodu *Nucleorhabdovirus* (Jackson a kol., 2005) je možné i to, že tento nový virus žádné další hostitele nemá.

Nicméně, taxonomie týkající se virů v současnosti zařazených do rodu *Nucleorhabdovirus*, se pravděpodobně v brzké době změní. V srpnu roku 2019 byla navržena nová kompletní reklasifikace nukleorhabdovirů (Freitas-Astúa a kol., 2019). Viry zařazené do rodu *Nucleorhabdovirus* mají být roztříděny do tří rodů *Alpha-*, *Beta-* a *Gammanucleorhabdovirus*. Důvodem je, že současný rod *Nucleorhabdovirus* není monofyletická skupina, viry rodu *Dichorhavirus* jsou součástí klastru tvořeného především nukleorhabdoviry. Do navrženého rodu *Alphanucleorhabdovirus* by měly patřit viry EMDV, MIMV, MMV, PYDV (typový zástupce), RYSV a TaVCV. Dále by do tohoto rodu měly patřit nově uznané druhy MMaV, PhCMoV a Wheat yellow striate virus (WYSV). Do navrženého rodu *Betanucleorhabdovirus* by měly patřit viry DYVV, SYNV (typový zástupce) a SYVV. Dále by do tohoto rodu měly patřit nově uznané druhy *Blackcurrant betanucleorhabdovirus* (původně navržený název Black currant-associated rhabdovirus), *Trefoil betanucleorhabdovirus* (původně navržený název Bird's-foot trefoil nucleorhabdovirus) a *Alfalfa betanucleorhabdovirus* (původně navržený název Alfalfa-associated nucleorhabdovirus). Virus MFSV by měl být jediným zástupcem rodu *Gammanucleorhabdovirus*. Tyto navržené rody odpovídají jednotlivým klastrům ve fylogenetických stromech identifikovaných v této práci. V rámci tohoto rozdělení by měl patřit nový nukleorhabdovirus B152 do rodu *Betanucleorhabdovirus*. Aby byl izolát uznán za nový virový druh podle této nové reklasifikace, musí splňovat některá ze tří kritérií: mít identitu v celogenomové sekvenci s ostatními betanukleorhabdoviry nižší než 75 %, obsazovat jinou ekologickou niku anebo být rozlišitelný od ostatních virů z tohoto rodu sérologickými testy či metodami založenými na hybridizaci nukleových kyselin. Druhé a třetí kritérium je tedy totožné se současnými kritérii pro uznání nového virového druhu v rámci rodu *Nucleorhabdovirus*. Za těchto taxonomických kritérií nový izolát splňuje první podmínku, jeho identita v celogenomové sekvenci s ostatními viry, jejichž sekvence je dostupná v databázi, nepřesahuje, ani se neblíží 75 %. Identita nového izolátu s viry SYVV a BFTV-1 není známá, protože sekvence těchto virů nejsou dostupné v databázi. Na základě známých skutečností, virus splňuje i druhou taxonomickou podmínku.

Není možné posoudit, jestli příznaky na keři, na kterém byl nový nukleorhabdovirus B152 identifikován, nějak souvisí s infekcí tímto virem, protože tento keř byl infikován řadou dalších virů. Z příznaků, které běžně rhabdoviry u rostlin způsobují (Thomas a Dietzgen, 1991;

Jackson a kol., 2005), byly na tomto keři pozorovány žluté skvrny na listech a zakrslost listů. Jedná se ale o obecné, nespecifické příznaky virové infekce.

Získaná téměř úplná sekvence genomu má délku 12 822 nt. Genom nového nukleorhabdoviru B152 nese 6 ORF typických pro viry z rodu *Nucleorhabdovirus* (Jackson a kol., 2005; Dietzgen a kol., 2015). Nenese žádný doplňkový ORF. Sekvence neobsahuje stop kodon ORF pro L protein, takže není známá délka tohoto ORF, a také není známá sekvence nekódující oblasti na 5' konci genomu. Pro její získání je třeba optimalizovat RACE PCR a navrhnout jiný primer. Dá se odhadnout, že skutečná velikost genomu bude lehce přesahovat 13 000 nt, což odpovídá střední velikosti genomu v rámci rodu *Nucleorhabdovirus*, u nějž se velikost genomu virů pohybuje asi od 12 000 nt (12 020 nt u TaVCV) do 14 500 nt (14432 nt u BCaRV-1).

Na základě dílčích analýz se dá předpokládat, že je pozice jednotlivých ORF v genomu nového nukleorhabdoviru B152 identifikována správně. Vypočtené molekulové hmotnosti proteinů, které by takto identifikované ORF kódovaly, jsou velmi blízké molekulovým hmotnostem odpovídajících proteinů fylogeneticky nejbližšího viru DYVV, tak jak je identifikovali Dietzgen a kol. (2015). Kdyby byl použit kterýkoli jiný potenciální iniciační kodon u kteréhokoli genu, rozdíly by byly větší. Vypočtené molekulové hmotnosti jsou prakticky totožné u proteinů M a G, proteiny P a P3 nového nukleorhabdoviru B152 mají o zhruba 0,9 kDa vyšší molekulovou hmotnost oproti těmto proteinům viru DYVV, zatímco protein M má o cca 0,6 kDa menší molekulovou hmotnost. V případě proteinu L není možné tímto způsobem kalkulovat vzhledem k tomu, že neznáme jeho přesnou molekulovou hmotnost.

Konzervativní mezigenová sekvence o velikosti 16 nt byla identifikovaná mezi všemi ORF nového nukleorhabdoviru B152. Tato sekvence se nachází ve zcela nezměněné formě také v genomech virů DYVV, BCaRV-1 a SYNIV (Wu a kol., 2018). Výjimkou je jedna tato sekvence u viru DYVV, která se liší v jednom nukleotidu (Dietzgen a kol., 2015). Velmi podobné regulační sekvence se nachází napříč genomy všech virů z čeledi *Rhabdoviridae*. Podobnosti lze vyzorovat i u dalších virů, jejichž genom je tvořen (-)ssRNA (Jackson a kol., 2005). Tato regulační sekvence zajišťuje polyadenylaci mRNA přepsané z předchozího genu a začátek transkripce genu následujícího. Regulační element pro start transkripce by měl mít délku 5 nt (Dietzgen a kol., 2015), ale konzervativní jsou jen první 2 nt, další nukleotidy se liší i mezi těmito elementy u nového nukleorhabdoviru B152. Před 9 nt- sekvencí považovanou za signál pro polyadenylaci mRNA (Dietzgen a kol., 2015), se nachází 3 nt- sekvence o neznámé funkci, která je shodná u všech mezigenových sekvencí nového nukleorhabdoviru i virů DYVV, BCaRV-1 a SYNIV. Nenachází se ale u této sekvence v blízkosti 3' konce. To,

že se tato zkrácená konzervativní sekvence nachází před prvním ORF, patrně souvisí s transkripcí RNA v blízkosti 3' konce, která nekóduje žádný protein, a je také polyadenylována. Existence takové RNA byla popsána u viru SYNIV (Wagner a kol., 1996).

Výsledky fylogenetické analýzy na základě proteinu L odpovídají dosavadním znalostem o nukleorhabdovirech (Wu a kol., 2018; Baek a kol., 2019; Gaafar a kol., 2019). Rozsáhlejší fylogenetické analýzy dalších proteinů doposud publikovány nebyly, ale dá se předpokládat stejná topologie stromů. Do stejného klastru s novým nukleorhabdovirem B152 patří hned několik nově objevených virů, konkrétně BCaRV-1, ApRVA a AaNV. Dalším novým virem, který by měl patřit do tohoto klastru, je BFTV-1 izolovaný ze štírovníku růžkatého (*Lotus corniculatus*) v Číně, ale nemohl být zahrnut do fylogenetických analýz, protože jeho sekvence momentálně není dostupná v databázi GenBank. (Debat a Bajerman, 2019).

Ve vzorku B152 byla zjištěna přítomnost více izolátů studovaného druhu nukleorhabdoviru, z nichž se zatím nepodařilo získat větší část sekvence. Tento virus byl detekován také u dvou dalších rostlin bezu. Bude následovat práce na získání celých sekvencí dalších izolátů tohoto viru a zjištění vnitrodruhové variability, která se podle částečných sekvencí zdá být poměrně vysoká.



## 7 ZÁVĚR

Pomocí NGS a Sangerova sekvenování byl na bezu černém (*Sambucus nigra*) identifikován nový (-)ssRNA virus z rodu *Nucleorhabdovirus*. V rámci této diplomové práce byla získána téměř úplná sekvence jeho genomu o délce 12 822 nt. V genomu izolátu bylo identifikováno šest otevřených čtecích rámců s organizací typickou pro viry tohoto taxonu. Jedná se o geny kódující plášťový protein (N), fosfoprotein (P), ´movement protein´ (P3), matrix protein (M), glykoprotein (G) a RNA-dependentní RNA polymerázu (L). V oblastech mezi geny se nachází konzervativní sekvence podobné sekvencím u dalších virů z rodu *Nucleorhabdovirus* odpovídající za iniciaci transkripce a polyadenylaci RNA.

Nukleorhabdovirus popsáný v této práci vykazuje nejvyšší sekvenční identitu s virem *Datura yellow vein virus*. V otevřených čtecích rámcích kódujících plášťový protein, ´movement protein´ a RNA-dependentní RNA polymerázu se jeho identita s tímto virem pohybuje v rozmezí 53-63 %. Nový izolát lze uznat jako samostatný virový druh na základě rozdílné ekologie, infikuje hostitele odlišného od hostitelů viru *Datura yellow vein virus* a je rozšířen na jiném světadílu. Fylogenetická analýza prokázala odlišnost nového viru od ostatních virů z rodu *Nucleorhabdovirus*. Nový nukleorhabdovirus je blíže příbuzný s viry *Datura yellow vein virus*, Black currant-associated rhabdovirus, *Sonchus yellow net virus*, Apple rootstock virus A a Alfalfa-associated nucleorhabdovirus, s nimiž vytváří společný klastř v rámci klastřu tvořeného viry rodu *Nucleorhabdovirus*.

## 8 LITERATURA

- Adams I. P., Glover R. H., Monger W. A., Mumford R., Jackeviciene E., Navalinskiene M., Samuitiene M., Boonham N. (2009): Next-generation sequencing and metagenomic analysis: a universal diagnostic tool in plant virology. *Molecular Plant pathology* 10: 537–545.
- Al Rwahnih M., Daubert S., Golino D., Islas C., Rowhani A. (2015): Comparison of next-generation sequencing versus biological indexing for the optimal detection of viral pathogens in grapevine. *Phytopathology* 105: 758-763.
- Al Rwahnih M., Daubert S., Golino D., Rowhani A. (2009): Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. *Virology* 387: 395–401.
- Amarasinghe G. K., Aréchiga Ceballos N. G., Banyard A. C., Basler Ch. F., Bavari S., ..., Kuhn, J. H. (2018): Taxonomy of the order *Mononegavirales*: Update 2018. *Archives of Virology* 163: 2283–2294.
- Ammar E. D., Gomez Luengo R. G., Gordon D. T., Hogenhout S. A. (2005): Characterization of *Maize Iranian mosaic virus* and comparison with Hawaiian and other isolates of *Maize mosaic virus (Rhabdoviridae)*. *Journal of Phytopathology* 153: 129-136.
- Baek D., Lim S., Ju H.-J., Kim H.-R., Lee S.-H., Moon J. S. (2019): The complete genome sequence of apple rootstock virus A, a novel nucleorhabdovirus identified in apple rootstocks. *Archives of Virology* 164: 2641–2644.
- Bag S., Al Rwahnih M., Li A., Gonzalez A., Rowhani A., Uyemoto J. K., Sudarshana M. R. (2015): Detection of a new luteovirus in imported nectarine trees: A case study to propose adoption of metagenomics in post-entry quarantine. *Phytopathology* 105: 840-846.
- Bandyopadhyay A., Kopperud K., Anderson G., Martin K., Goodin M. (2010): An integrated protein localization and interaction map for *Potato yellow dwarf virus*, type species of the genus *Nucleorhabdovirus*. *Virology* 402: 61–71.
- Benthack W., Mielke N., Büttner C., Mühlbach H.-P. (2005): Double-stranded RNA pattern and partial sequence data indicate plant virus infection associated with the ringspot disease of European mountain ash (*Sorbus aucuparia* L.). *Archives of Virology* 150: 37–52.

- Bentley D. R., Balasubramanian S., Swerdlow H. P. a kol. (2008): Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53–59.
- Blawid R., Silva J. M. F., Nagata T. (2017): Discovering and sequencing new plant viral genomes by next-generation sequencing: description of a practical pipeline. *Annals of Applied Biology* 170: 301-314.
- Boonham N., Kreuze J., Winter S., van der Vlugt R., Bergervoet J., Tomlinson J., Mumford R. (2014): Methods in virus diagnostics: From ELISA to next generation sequencing. *Virus Research* 186: 20–31.
- Candresse T., Filloux D., Muhire B., Julian C., Galzi S., Fort G., Bernardo P., Daugrois J.-H., Fernandez E., Martin D. P., Varsani A., Roumagnac P. (2014): Appearances can be deceptive: Revealing a hidden viral infection with deep sequencing in a plant quarantine context. *PLoS ONE* 9: e102945.
- Cao M., Zhang S., Li M., Liu Y., Dong P., Li S., Kuang M., Li R., Zhou Y. (2019): Discovery of four novel viruses associated with flower yellowing disease of green sichuan pepper (*Zanthoxylum armatum*) by virome analysis. *Viruses* 11: 696.
- Dal Zotto A., Nome S. F., Di Rienzo J. A., Docampo D. M. (1999) Fluctuations of *Prunus necrotic ringspot virus* (PNRSV) at various phenological stages in peach cultivars. *Plant Disease* 83: 1055-1057.
- De Andrade R. R. S., Vaslin M. F. S. (2014): SearchSmallRNA: a graphical interface tool for the assemblage of viral genomes using small RNA libraries data. *Virology Journal* 11: 45.
- Debat H. J., Beijerman N. (2019): Novel bird's-foot trefoil RNA viruses provide insights into a clade of legume-associated enamoviruses and rhabdoviruses. *Archives of Virology* 164: 1419-1426.
- Del Fabbro C., Scalabrin S., Morgante M., Giorgi F. M. (2013): An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS ONE* 8: e85024.
- Dietzgen R. G., Gowanlock D. H., Thomas J. E. (1992): Purification and identification of a rhabdovirus from *Thunbergia alata*. *Journal of Phytopathology* 135: 99–106.
- Dietzgen R. G., Innes D. J., Beijerman N. (2015): Complete genome sequence and intracellular protein localization of *Datura yellow vein nucleorhabdovirus*. *Virus Research* 205: 7–11.

- Dietzgen R. G., Kondo H., Goodin M. M., Kurath G., Vasilakis N. (2017): The family Rhabdoviridae: mono- and bipartite negative-sense RNA viruses with diverse genome organization and common evolutionary origins. *Virus Research* 227: 158–170.
- Dietzgen R. G., Kuhn J. H., Clawson A. N., Freitas-Astúa J., Goodin M. M., Kitajima E. W., Kondo H., Wetzel T., Whitfield A. E. (2014): *Dichorhavirus*: a proposed new genus for *Brevipalpus* mite transmitted, nuclear, bacilliform, bipartite, negative-strand RNA plant viruses. *Archives of Virology* 159: 607–619.
- Engel E. A., Escobar P. F., Rojas L. A., Rivera P. A., Fiore N., Valenzuela P. D. T. (2010): A diagnostic oligonucleotide microarray for simultaneous detection of grapevine viruses. *Journal of Virological Methods* 163: 445–451.
- Falk B. W., Purcifull D. E., Christie S. R. (1986): Natural occurrence of *Sonchus yellow net virus* in Florida lettuce. *Plant Disease* 70: 591-593.
- Fort P., Albertini A., Van-Hua A., Berthomieu A., Roche S., Delsuc F., Pasteur N., Capy P., Gaudin Y., Weill M. (2012): Fossil rhabdoviral sequences integrated into arthropod genomes: Ontogeny, evolution, and potential functionality. *Molecular Biology and Evolution* 29: 381–390.
- Freitas-Astúa J., Dietzgen R. G., Walker P. J., Blasdel K. R., Breyta R., a kol. (2019): Split the genus *Nucleorhabdovirus*, creating three new genera (*Alphanucleorhabdovirus*, *Betanucleorhabdovirus* and *Gammanucleorhabdovirus*) comprising sixteen species, including six new species, in the family *Rhabdoviridae*. ICTV taxonomic proposal 2019.031M.
- Frohman M. A., Dush M. K., Martin G. R. (1988): Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences* 85: 8998-9002.
- Gaafar Y. Z. A., Richert-Pöggeler K. R., Maaß Ch., Vetten H.-J., Ziebell H. (2019): Characterisation of a novel nucleorhabdovirus infecting alfalfa (*Medicago sativa*). *Virology Journal* 16: 55.
- Haible D., Kober S., Jeske H. (2006): Rolling circle amplification revolutionizes diagnosis and genomics of geminiviruses. *Journal of Virological Methods* 135: 9–16.
- Ho T., Quito-Avilad D., Keller K. E., Postman J. D., Martin R. R., Tzanetakis I. E. (2016): Evidence of sympatric speciation of elderberry carlaviruses. *Virus Research* 215: 72–75.

- Chiba S., Kondo H., Tani A., Saisho D., Sakamoto W., Kanematsu S., Suzuki N. (2011): Widespread endogenization of genome sequences of non-retroviral RNA viruses into plant genomes. *PLoS Pathogens* 7: e1002146.
- Christie S. R., Christie R. G., Edwardson J. R. (1974): Transmission of a bacilliform virus of sowthistle and *Bidens pilosa*. *Phytopathology* 64: 840-845.
- Izadpanah K., Ahmadi A. A., Parvin S. Jafari S. A. (1983): Transmission, particle size and additional hosts of the rhabdovirus causing maize mosaic in Shiraz, Iran. *Phytopathologische Zeitschrift* 107: 283-288.
- Jackson A. O., Dietzgen R. G., Goodin M. M., Bragg J. N., Deng M. (2005): Biology of plant rhabdoviruses. *Annual Review of Phytopathology* 43: 623-660.
- Jridi Ch., Martin J.-F., Marie-Jeanne V., Labonne G., Blanc S. (2006): Distinct viral populations differentiate and evolve independently in a single perennial host plant. *Journal of Virology* 80: 2349–2357.
- Kreuze J. F., Perez A., Untiveros M., Quispe D., Fuentes S., Barker I., Simon R. (2009): Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388: 1-7.
- Kwon S.-J., Choi G.-S., Choi B., Seo J.-K. (2018): Molecular characterization of an unusual new plant RNA virus reveals an evolutionary link between two different virus families. *PLoS ONE* 13: e0206382.
- Li Y.-Y., Zhang R.-N., Xiang H.-Y., Abouelnasr H., Li D.-W., Yu J.-L., McBeath J. H., Han Ch.-G. (2013): Discovery and characterization of a novel *Carlavirus* infecting potatoes in China. *PLoS ONE* 8: e69255.
- Loconsole G., Saldarelli P., Doddapaneni H., Savino V., Martelli G. P., Saponari M. (2012): Identification of a single-stranded DNA virus associated with citrus chlorotic dwarf disease, a new member in the family *Geminiviridae*. *Virology* 432: 162–172.
- Ma Y., Navarro B., Zhang Z., Lu M., Zhou X., Chi S., Di Serio F., Li S. (2015): Identification and molecular characterization of a novel monopartite geminivirus associated with mulberry mosaic dwarf disease. *Journal of General Virology* 96: 2421–2434.

- Maliogka V. I., Minafra A., Saldarelli P., Ruiz-García A. B., Glasa M., Katis N., Olmos A. (2018): Recent advances on detection and characterization of fruit tree viruses using high-throughput sequencing technologies. *Viruses* 10: 436.
- Marais A., Faure C., Couture C., Bergey B., Gentit P., Candresse T. (2014): Characterization by deep sequencing of divergent *Plum bark necrosis stem pitting-associated virus* (PBNSPaV) isolates and development of a broad-spectrum PBNSPaV detection assay. *Phytopathology* 104: 660-666.
- Maree H. J., Fox A., Al Rwahnih M., Boonham N., Candresse T. (2018): Application of HTS for routine plant virus diagnostics: State of the art and challenges. *Frontiers in Plant Science* 9: 1082.
- Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., Berka J., Braverman M. S., Chen Y.-J., Chen Z., Dewell S. B., Du L., Fierro J. M., Gomes X. V. a kol. (2005): Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437: 376-380.
- Massart S., Candresse T., Gil J., Lacomme Ch., Predajna L., Ravnihar M., Reynard J.-S., Rumbou A., Saldarelli P., Škori D., Vainio E. J., Valkonen J. P. T., Vanderschuren H., Varveri Ch., Wetzel T. (2017): A frame work for the evaluation of biosecurity, commercial, regulatory, and scientific impacts of plant viruses and viroids identified by NGS technologies. *Frontiers in Microbiology* 8: 45.
- Milošević T., Milošević N., Mladenović J., Jevremović D. (2019): Impact of Sharka disease on tree growth, productivity and fruit quality of apricot (*Prunus armeniaca* L.). *Scientia Horticulturae* 244: 270–276.
- 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.
- Nakazato T., Ohta T., Bono H. (2013): Experimental design-based functional mining and characterization of high-throughput sequencing data in the Sequence read archive. *PLoS ONE* 8: e77910.
- Pecman A., Kutnjak D., Gutiérrez-Aguirre I., Adams I., Fox A., Boonham N., Ravnihar M. (2017): Next generation sequencing for detection and discovery of plant viruses and viroids: Comparison of two approaches. *Frontiers in Microbiology* 8: 1998.

- Popov V. L., Tesh R. B., Weaver S. C., Vasilakis N. (2019): Electron microscopy in discovery of novel and emerging viruses from the collection of the World reference center for emerging viruses and arboviruses (WRCEVA). *Viruses* 11: 477.
- Read D. A., Featherston J., Rees D. J. G., Thompson G. D., Roberts R., Flett B. Ch., Mashingaidze K., Pietersen G., Kiula B., Kullaya A., Mbega E. R. (2019): Molecular characterization of Morogoro maize associated virus, a nucleorhabdovirus detected in maize (*Zea mays*) in Tanzania. *Archives of Virology* 164: 1711-1715.
- Reavy B., Swanson M. M., Cock P. J. A, Dawson L., Freitag T. E., Singh B. K., Torrance L. Mushegian A. R., Talianskya M. (2015): Distinct circular single-stranded DNA viruses exist in different soil types. *Applied and Environmental Microbiology* 81: 3934-3945.
- Redinbaugh M. G., Seifers D. L., Meulia T., Abt J. J., Anderson R. J., Styer W. E., Ackerman J., Salomon R., Houghton W., Creamer R., Gordon D. T., Hogenhout S. A. (2002): *Maize fine streak virus*, a new leafhopper-transmitted rhabdovirus. *Phytopathology* 92: 1167-1174.
- Revell P., Trinh X., Dale J., Harding R. (2005): *Taro vein chlorosis virus*: characterization and variability of a new nucleorhabdovirus. *Journal of General Virology* 86: 491–499.
- Roggero P., Ciuffo M., Vaira A. M., Accotto G. P., Masenga V., Milne R. G. (2000): An *Ophiovirus* isolated from lettuce with big-vein symptoms. *Archives of Virology* 145: 2629-2642.
- Roossinck M. J., Martin D. P., Roumagnac P. (2015): Plant virus metagenomics: Advances in virus discovery. *Phytopathology* 105: 716-727.
- Rott M., Xiang Y., Boyes I., Belton M., Saeed H., Kesanakurti P., Hayes S., Lawrence T., Birch C., Bhagwat B., Rast H. (2017): Application of next generation sequencing for diagnostic testing of tree fruit viruses and viroids. *Plant Disease* 101: 1489-1499.
- Roy A., Choudhary N., Guillermo L. M., Shao J., Govindarajulu A., Achor D., Wei G., Picton D. D., Levy L., Nakhla M. K., Hartung J. S., Brlansky R. H. (2013): A novel virus of the genus *Cilevirus* causing symptoms similar to citrus leprosis. *Phytopathology* 103:488-500.
- Rubio M., Martínez-Gómez P., Marais A., Sánchez-Navarro J. A., Pallás V., Candresse T. (2017): Recent advances and prospects in *Prunus* virology. *Annals of Applied Biology* 171: 125–138.

- Sasaya T., Ishikawa K., Koganezawa H. (2002): The nucleotide sequence of RNA1 of Lettuce big-vein virus, genus Varicosavirus, reveals its relation to nonsegmented negative-strand RNA viruses.
- Sasaya T., Kusaba S., Ishikawa K., Koganezawa H. (2004): Nucleotide sequence of RNA2 of *Lettuce big-vein virus* and evidence for a possible transcription termination/initiation strategy similar to that of rhabdoviruses. *Journal of General Virology* 85: 2709-2717.
- Sheveleva A., Kudryavtseva A., Speranskaya A., Belenikin M., Melnikova N., Chirkov S. (2013): Complete genome sequence of a novel *Plum pox virus* strain W isolate determined by 454 pyrosequencing. *Virus Genes* 47: 385–388.
- Šafářová D., Candresse T., Navrátil M. (2018b): Complete genome sequence of a novel bromovirus infecting elderberry (*Sambucus nigra* L.) in the Czech Republic. *Archives of Virology* 163: 567–570.
- Šafářová D., Vavroušková K., Candresse T., Navrátil M. (2018a): Molecular characterization of a novel Aureusvirus infecting elderberry (*Sambucus nigra* L.). *PLoS ONE* 13: e0200506.
- Thomas J. E., Dietzen R. G. (1991): Characterisation of *Datura yellow vein virus*, a newly described rhabdovirus from Australia. *Annals of Applied Biology* 118: 339–349.
- Tucker T., Marra M., Friedman J. M. (2009): Massively parallel sequencing: The next big thing in genetic medicine. *The American Journal of Human Genetics* 85: 142–154.
- Umer M., Liu J., You H., Xu Ch., Dong K., Luo N., Kong L., Li X., Hong N., Wang G., Fan X., Kotta-Loizou I., Xu W. (2019): Genomic, morphological and biological traits of the viruses infecting major fruit trees. *Viruses* 11: 515.
- Usenic V., Marn M. V. (2017): Sugars and organic acids in plum fruit affected by *Plum pox virus*. *Journal of the Science of Food and Agriculture* 97: 2154–2158.
- van Lent J. W. M., Wit A. J., Dijkstra J. (1980): Characterization of a *Carlavirus* in elderberry (*Sambucus* spp.). *Netherlands Journal of Plant Pathology* 86: 117 -134.
- Villamor D. E. V., Mekuria T. A., Pillai S. S., Eastwell K. C. (2016): High-throughput sequencing identifies novel viruses in nectarine: Insights to the etiology of stem-pitting disease. *Phytopathology* 106: 519-527.
- Visser M., Bester R., Burger J. T., Maree H. J. (2016): Next-generation sequencing for virus detection: covering all the bases. *Virology Journal* 13: 85.



- Wagner J. D. O., Choi T.-J., Jackson A. O. (1996): Extraction of nuclei from *Sonchus yellow net rhabdovirus* - Infected plants yields a polymerase that synthesizes viral mRNAs and polyadenylated plus-strand leader RNA. *Journal of Virology* 70: 468–477.
- Walker P. J., Firth C., Widen S. G., Blasdell K. R., Guzman H., Wood T. G., Paradkar P. N., Holmes E. C., Tesh R. B., Vasilakis N. (2015): Evolution of genome size and complexity in the Rhabdoviridae. *PLoS Pathogens* 11: e1004664.
- Wu L. P., Yang T., Liu H. W., Postman J., Li R. (2018): Molecular characterization of novel rhabdovirus infecting blackcurrant identified by high throuput sequencing. *Archives of Virology*: 163: 1363–1366.
- Zheng Y., Gao S., Padmanabhan Ch., Li R., Galvez M., Gutierrez D., Fuentes S., Ling K.-S., Kreuze J., Fei Z. (2017): VirusDetect: An automated pipeline for efficient virus discovery using deep sequencing of small RNAs. *Virology* 500: 130–138.

## **Internetové zdroje**

- Angiosperm Phylogeny Webside [online] [navštíveno 27.3.2020] Dostupné z <http://www.mobot.org/MOBOT/research/APweb/>
- EPPOa. [online] [navštíveno 24.3.2020] Dostupné z <https://gd.eppo.int/taxon/PYDV00/hosts>
- EPPOb. [online] [navštíveno 24.3.2020] Dostupné z <https://gd.eppo.int/taxon/EMDV00/hosts>
- EPPOc. [online] [navštíveno 9.4.2020] <https://gd.eppo.int/taxon/RYSV00/hosts>
- ICTV. [online] [navštíveno 26.3.2020] Dostupné z [https://talk.ictvonline.org/ictv-reports/ictv\\_online\\_report/negative-sense-rna-viruses/mononegavirales/w/rhabdoviridae/797/genus-nucleorhabdovirus](https://talk.ictvonline.org/ictv-reports/ictv_online_report/negative-sense-rna-viruses/mononegavirales/w/rhabdoviridae/797/genus-nucleorhabdovirus)
- TruSeq RNA sample preparation v2 guide. [online] [navštíveno 21. 11. 2019] Dostupné z [https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry\\_documentation/samplepreps\\_truseq/truseq\\_rna/truseq-rna-sample-prep-v2-guide-15026495-f.pdf](https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseq_rna/truseq-rna-sample-prep-v2-guide-15026495-f.pdf)
- Viralzone. [online] [navštíveno 1.2.2020] Dostupné z [https://viralzone.expasy.org/503?outline=all\\_by\\_species](https://viralzone.expasy.org/503?outline=all_by_species)

## 9 PŘÍLOHY

Příloha 1: Kontigy složené *de novo* vykazující identitu se sekvencemi některého z virů rodu *Nucleorhabdovirus*

Name	Identical Sites	Pairwise Identity	Description	E Value	Hit end	Hit start	Query	Query coverage	Query end	Query start
NP_042285.1	55.9%	55.9%	glycoprotein (SYNV)	3.14e-20	442	375	Contig 5646	92.31%	212	9
NP_042286.1	58.2%	58.2%	polymerase (SYNV)	2.80e-17	1138	1084	Contig 11123	97.06%	168	4
NP_042286.1	65.3%	65.3%	polymerase (SYNV)	6.84e-31	1111	1037	Contig 5123	98.25%	226	2
NP_042286.1	65.6%	65.6%	polymerase (SYNV)	1.88e-22	399	339	Contig 8257	95.81%	186	4
NP_042286.1	43.6%	43.6%	polymerase (SYNV)	3.59e-08	216	162	Contig 9510	91.16%	172	8
YP_009176972.1	54.7%	54.7%	nucleocapsid protein (DYVV)	3.64e-14	82	30	Contig 13433	98.76%	159	1
YP_009176972.1	65.2%	65.2%	nucleocapsid protein (DYVV)	6.04e-51	267	156	Contig 2022	98.82%	336	1
YP_009176972.1	53.3%	53.3%	nucleocapsid protein (DYVV)	4.38e-21	127	53	Contig 2396	71.20%	234	10
YP_009176972.1	85.1%	85.1%	nucleocapsid protein (DYVV)	3.88e-39	395	322	Contig 5409	98.67%	224	3
YP_009176972.1	64.3%	64.3%	nucleocapsid protein (DYVV)	1.23e-29	260	191	Contig 6495	100.00%	210	1
YP_009176972.1	43.1%	43.1%	nucleocapsid protein (DYVV)	1.66e-10	140	76	Contig 7373	96.00%	200	9
YP_009176972.1	79.7%	79.7%	nucleocapsid protein (DYVV)	7.89e-62	411	294	Contig 844	68.34%	356	3
YP_009176972.1	85.0%	85.0%	nucleocapsid protein (DYVV)	6.10e-31	312	253	Contig 9411	98.90%	181	2
YP_009176973.1	62.3%	62.3%	phosphoprotein (DYVV)	1.49e-15	54	2	Contig 3612	60.23%	161	3
YP_009176973.1	35.7%	35.7%	phosphoprotein (DYVV)	6.54e-06	269	214	Contig 4744	71.19%	235	68
YP_009176973.1	55.7%	55.7%	phosphoprotein (DYVV)	9.42e-22	75	6	Contig 6305	99.06%	211	2
YP_009176973.1	53.8%	53.8%	phosphoprotein (DYVV)	5.23e-18	72	8	Contig 7570	98.48%	197	3
YP_009176973.1	62.1%	62.1%	phosphoprotein (DYVV)	1.07e-17	59	2	Contig 9084	94.05%	183	10
YP_009176974.1	40.7%	40.7%	putative cell-to-cell movement protein (DYVV)	2.54e-07	201	151	Contig 11625	91.62%	160	8
YP_009176974.1	48.1%	48.1%	putative cell-to-cell movement protein (DYVV)	6.58e-12	292	241	Contig 11653	93.41%	157	2
YP_009176974.1	49.7%	49.7%	putative cell-to-cell movement protein (DYVV)	1.12e-44	201	62	Contig 1240	97.67%	423	4
YP_009176974.1	42.2%	42.2%	putative cell-to-cell movement protein (DYVV)	3.65e-18	311	229	Contig 2721	83.28%	295	47

Name	Identical Sites	Pairwise Identity	Description	E Value	Hit end	Hit start	Query	Query coverage	Query end	Query start
YP_009176974.1	57.1%	57.1%	putative cell-to-cell movement protein (DYVV)	1.04e-30	145	62	Contig 3930	98.82%	255	4
YP_009176974.1	46.3%	46.3%	putative cell-to-cell movement protein (DYVV)	9.80e-20	95	14	Contig 4129	98.40%	248	3
YP_009176974.1	49.2%	49.2%	putative cell-to-cell movement protein (DYVV)	4.28e-16	247	185	Contig 5744	85.91%	219	31
YP_009176974.1	62.7%	62.7%	putative cell-to-cell movement protein (DYVV)	1.65e-26	138	72	Contig 6365	95.26%	204	4
YP_009176974.1	49.2%	49.2%	putative cell-to-cell movement protein (DYVV)	6.56e-16	157	95	Contig 8590	100.00%	189	1
YP_009176975.1	46.8%	46.8%	matrix protein (DYVV)	1.11e-21	181	103	Contig 4595	99.16%	237	1
YP_009176975.1	45.7%	45.7%	matrix protein (DYVV)	1.01e-15	82	13	Contig 5985	97.22%	211	2
YP_009176975.1	50.0%	50.0%	matrix protein (DYVV)	3.97e-16	167	104	Contig 7868	98.46%	194	3
YP_009176975.1	41.9%	41.9%	matrix protein (DYVV)	4.76e-12	82	21	Contig 8170	96.88%	191	6
YP_009176976.1	53.7%	53.7%	glycoprotein (DYVV)	6.43e-15	269	216	Contig 11215	95.86%	165	4
YP_009176976.1	51.0%	51.0%	glycoprotein (DYVV)	9.40e-13	147	97	Contig 11443	91.07%	166	14
YP_009176976.1	54.7%	54.7%	glycoprotein (DYVV)	3.20e-15	285	233	Contig 11569	95.21%	163	5
YP_009176976.1	48.5%	48.5%	glycoprotein (DYVV)	1.30e-28	554	452	Contig 1437	77.06%	309	1
YP_009176976.1	50.8%	50.8%	glycoprotein (DYVV)	9.51e-37	238	121	Contig 1748	97.25%	363	10
YP_009176976.1	28.2%	28.2%	glycoprotein (DYVV)	6.47e-12	597	495	Contig 2242	95.08%	313	5
YP_009176976.1	65.9%	65.9%	glycoprotein (DYVV)	6.00e-39	342	252	Contig 3146	97.50%	278	6
YP_009176976.1	54.9%	54.9%	glycoprotein (DYVV)	1.76e-13	56	6	Contig 5684	69.23%	154	2
YP_009176976.1	65.7%	65.7%	glycoprotein (DYVV)	1.73e-26	421	352	Contig 5885	96.33%	211	2
YP_009176976.1	69.2%	69.2%	glycoprotein (DYVV)	1.05e-24	333	269	Contig 7358	97.50%	199	5
YP_009176976.1	62.5%	62.5%	glycoprotein (DYVV)	2.22e-25	127	64	Contig 7866	98.46%	194	3
YP_009176976.1	54.5%	54.5%	glycoprotein (DYVV)	3.12e-59	174	21	Contig 830	88.34%	522	61
YP_009176976.1	60.0%	60.0%	glycoprotein (DYVV)	2.55e-19	504	445	Contig 9396	98.90%	182	3
YP_009176977.1	55.7%	56.0%	RNA-dependent RNA polymerase (DYVV)	2.69e-57	961	806	Contig 1033	99.57%	467	3

Name	Identical Sites	Pairwise Identity	Description	E Value	Hit end	Hit start	Query	Query coverage	Query end	Query start
YP_009176977.1	82.8%	82.8%	RNA-dependent RNA polymerase (DYVV)	6.93e-28	863	806	Contig 10438	100.00%	174	1
YP_009176977.1	82.1%	82.1%	RNA-dependent RNA polymerase (DYVV)	4.26e-17	1295	1240	Contig 10814	97.67%	170	3
YP_009176977.1	69.6%	69.6%	RNA-dependent RNA polymerase (DYVV)	1.02e-28	799	744	Contig 10869	98.25%	170	3
YP_009176977.1	89.1%	89.1%	RNA-dependent RNA polymerase (DYVV)	1.25e-06	730	676	Contig 11394	98.21%	166	2
YP_009176977.1	63.6%	63.6%	RNA-dependent RNA polymerase (DYVV)	4.65e-68	1249	1107	Contig 1242	99.77%	429	1
YP_009176977.1	38.9%	38.9%	RNA-dependent RNA polymerase (DYVV)	3.80e-27	921	868	Contig 12420	98.78%	162	1
YP_009176977.1	88.9%	88.9%	RNA-dependent RNA polymerase (DYVV)	5.18e-18	745	692	Contig 12565	98.78%	164	3
YP_009176977.1	76.6%	76.6%	RNA-dependent RNA polymerase (DYVV)	9.10e-15	788	648	Contig 1271	99.30%	424	2
YP_009176977.1	55.6%	55.6%	RNA-dependent RNA polymerase (DYVV)	9.65e-22	1373	1320	Contig 12738	99.39%	162	1
YP_009176977.1	33.8%	33.8%	RNA-dependent RNA polymerase (DYVV)	4.49e-68	1707	1583	Contig 1322	91.59%	416	36
YP_009176977.1	93.2%	93.2%	RNA-dependent RNA polymerase (DYVV)	4.78e-15	759	716	Contig 13976	87.42%	150	19
YP_009176977.1	74.6%	74.6%	RNA-dependent RNA polymerase (DYVV)	1.42e-62	1350	1217	Contig 1430	100.00%	402	1
YP_009176977.1	76.9%	76.9%	RNA-dependent RNA polymerase (DYVV)	1.23e-52	793	664	Contig 1513	99.74%	390	1
YP_009176977.1	64.3%	64.3%	RNA-dependent RNA polymerase (DYVV)	6.92e-08	1084	956	Contig 1571	99.48%	381	1
YP_009176977.1	28.1%	28.1%	RNA-dependent RNA polymerase (DYVV)	3.16e-45	1584	1460	Contig 1572	97.91%	379	5

Name	Identical Sites	Pairwise Identity	Description	E Value	Hit end	Hit start	Query	Query coverage	Query end	Query start
YP_009176977.1	62.5%	62.5%	RNA-dependent RNA polymerase (DYVV)	4.10e-52	921	802	Contig 1608	94.74%	360	1
YP_009176977.1	70.1%	70.1%	RNA-dependent RNA polymerase (DYVV)	3.60e-26	1280	1164	Contig 1872	99.43%	352	2
YP_009176977.1	43.9%	43.9%	RNA-dependent RNA polymerase (DYVV)	3.44e-23	282	169	Contig 1888	95.73%	342	7
YP_009176977.1	46.5%	46.5%	RNA-dependent RNA polymerase (DYVV)	1.39e-26	1452	1353	Contig 1937	87.32%	309	7
YP_009176977.1	47.0%	47.0%	RNA-dependent RNA polymerase (DYVV)	3.95e-14	1441	1331	Contig 2072	99.70%	337	2
YP_009176977.1	36.0%	36.0%	RNA-dependent RNA polymerase (DYVV)	6.82e-39	138	53	Contig 2106	77.25%	334	77
YP_009176977.1	60.4%	60.4%	RNA-dependent RNA polymerase (DYVV)	9.64e-29	1074	969	Contig 2291	96.89%	312	1
YP_009176977.1	51.0%	51.0%	RNA-dependent RNA polymerase (DYVV)	1.72e-53	312	211	Contig 2467	100.00%	312	1
YP_009176977.1	86.9%	86.9%	RNA-dependent RNA polymerase (DYVV)	6.10e-12	619	521	Contig 2666	98.02%	299	3
YP_009176977.1	35.3%	35.3%	RNA-dependent RNA polymerase (DYVV)	2.80e-22	1699	1617	Contig 3153	91.07%	265	11
YP_009176977.1	55.0%	55.0%	RNA-dependent RNA polymerase (DYVV)	1.64e-43	356	277	Contig 3210	86.33%	243	4
YP_009176977.1	81.4%	81.4%	RNA-dependent RNA polymerase (DYVV)	1.23e-26	879	794	Contig 3683	98.47%	260	3
YP_009176977.1	58.2%	58.2%	RNA-dependent RNA polymerase (DYVV)	9.39e-18	109	31	Contig 3946	92.94%	253	17
YP_009176977.1	48.8%	48.8%	RNA-dependent RNA polymerase (DYVV)	8.47e-29	1464	1382	Contig 3966	98.03%	254	6
YP_009176977.1	67.1%	67.1%	RNA-dependent RNA polymerase (DYVV)	1.30e-19	442	370	Contig 4319	89.39%	243	25

Name	Identical Sites	Pairwise Identity	Description	E Value	Hit end	Hit start	Query	Query coverage	Query end	Query start
YP_009176977.1	53.4%	53.4%	RNA-dependent RNA polymerase (DYVV)	9.44e-37	1386	1314	Contig 5304	96.90%	224	6
YP_009176977.1	75.7%	75.7%	RNA-dependent RNA polymerase (DYVV)	1.44e-28	1322	1249	Contig 5435	99.11%	222	1
YP_009176977.1	69.0%	69.0%	RNA-dependent RNA polymerase (DYVV)	4.61e-36	1255	1185	Contig 5721	96.82%	215	3
YP_009176977.1	85.7%	85.7%	RNA-dependent RNA polymerase (DYVV)	6.42e-37	653	584	Contig 6191	98.59%	212	3
YP_009176977.1	92.6%	92.6%	RNA-dependent RNA polymerase (DYVV)	7.67e-26	622	555	Contig 6813	99.03%	204	1
YP_009176977.1	67.6%	67.6%	RNA-dependent RNA polymerase (DYVV)	3.48e-15	525	458	Contig 6821	99.03%	205	2
YP_009176977.1	45.6%	45.6%	RNA-dependent RNA polymerase (DYVV)	2.25e-23	236	169	Contig 6912	99.51%	204	1
YP_009176977.1	64.2%	64.2%	RNA-dependent RNA polymerase (DYVV)	9.64e-31	1217	1151	Contig 7272	100.00%	201	1
YP_009176977.1	20.9%	20.9%	RNA-dependent RNA polymerase (DYVV)	8.73e-10	1908	1784	Contig 769	71.27%	474	88
YP_009176977.1	78.5%	78.5%	RNA-dependent RNA polymerase (DYVV)	3.58e-12	706	642	Contig 7820	100.00%	195	1
YP_009176977.1	29.1%	29.6%	RNA-dependent RNA polymerase (DYVV)	9.40e-71	2106	1990	Contig 807	65.98%	388	38
YP_009176977.1	45.8%	45.8%	RNA-dependent RNA polymerase (DYVV)	8.78e-19	1425	1367	Contig 8127	91.71%	190	14
YP_009176977.1	62.7%	62.7%	RNA-dependent RNA polymerase (DYVV)	3.03e-25	510	452	Contig 8243	87.50%	192	25
YP_009176977.1	77.4%	77.4%	RNA-dependent RNA polymerase (DYVV)	2.48e-27	563	502	Contig 8627	98.94%	188	3
YP_009176977.1	64.5%	64.5%	RNA-dependent RNA polymerase (DYVV)	7.49e-62	1255	1087	Contig 863	98.64%	512	6

Name	Identical Sites	Pairwise Identity	Description	E Value	Hit end	Hit start	Query	Query coverage	Query end	Query start
YP_009176977.1	73.3%	73.3%	RNA-dependent RNA polymerase (DYVV)	1.07e-31	419	360	Contig 9452	98.90%	181	2
YP_009176977.1	65.0%	65.0%	RNA-dependent RNA polymerase (DYVV)	1.57e-50	589	433	Contig 968	97.72%	473	3
YP_009176977.1	91.5%	91.5%	RNA-dependent RNA polymerase (DYVV)	3.84e-27	750	692	Contig 9983	99.44%	178	2

Příloha 2: Seznam konečných *de novo* kontigů a jejich pokrytí 'ready'

Name	Sequences	Pairwise Identity	Mean Coverage	Sequence Length
1033.1 a 1571	57	98,50%	6,19	890
1748.1 a 11215 a 7358.1	35	98,10%	3,60	937
2666.1 a 7820.1	19	99,60%	2,73	656
5423 a 5304 a 3966	44	98,90%	6,44	656
5744 a 11653.1	10	98,00%	2,53	355
5885.1 a 1437.1	28	97,50%	3,36	780
6495.1 a 9411	10	97,30%	2,17	410
8257.1 a 968.1 a 6191.1	84	98,60%	7,05	1160
9510 a 2467.1	28	97,50%	5,29	502
Contig_10438	6	100,00%	2,87	174
Contig_10814.1	5	100,00%	1,58	251
Contig_10869.1	11	99,20%	3,99	247
Contig_11123	5	98,20%	2,16	170
Contig_11394.1	6	100,00%	2,80	178
Contig_11443.1	5	99,60%	1,96	202
Contig_11569	4	100,00%	1,79	167
Contig_11625.1	5	99,20%	1,64	244
Contig_1240.1	44	97,40%	5,77	732
Contig_1242	16	99,60%	3,43	430
Contig_12420.2	12	96,80%	2,27	521
Contig_12565.1	4	100,00%	1,27	234
Contig_1271.1	39	98,60%	8,51	438
Contig_12738	6	99,60%	3,04	164
Contig_1322.1	40	97,60%	7,06	538
Contig_13433	4	99,40%	1,84	161
Contig_13976.1	10	99,10%	3,02	286
Contig_1430.2	24	98,10%	3,53	647
Contig_1513	15	99,60%	3,40	391
Contig_1572.1	24	99,10%	5,31	418
Contig_1608.1	21	98,60%	3,75	531
Contig_1872	10	99,70%	2,52	353
Contig_1888.1	28	98,70%	4,74	558
Contig_1937.1	20	100,00%	4,95	380
Contig_2022.1	26	98,10%	6,60	373
Contig_2072.2	55	98,70%	5,46	981
Contig_2106.1	39	97,90%	6,84	551
Contig_2242.1	41	98,00%	4,80	823
Contig_2291.1	20	99,40%	4,73	400
Contig_2396.1	17	97,60%	4,75	334
Contig_2721. 2	27	98,90%	4,28	591
Contig_3146.1	37	97,80%	4,79	735
Contig_3153.1	25	95,30%	4,65	514
Contig_3210.1	10	99,50%	3,13	286
Contig_3612.2	15	99,50%	5,28	264
Contig_3683.2	14	99,00%	3,36	384
Contig_3930	11	99,80%	3,89	255



Name	Sequences	Pairwise Identity	Mean Coverage	Sequence Length
Contig_3946.1	11	99,20%	2,66	374
Contig_4129	12	96,60%	4,34	250
Contig_4319.1	12	99,50%	3,90	279
Contig_4595.1	25	97,40%	3,88	604
Contig_4744.1	36	99,00%	6,07	571
Contig_5123.1	11	100,00%	2,54	390
Contig_5409	7	99,20%	2,65	225
Contig_5646	10	99,80%	4,06	221
Contig_5684.1	9	99,80%	1,94	404
Contig_5721.1	16	97,50%	3,25	457
Contig_5985.1	20	98,10%	3,52	533
Contig_6305.2	27	97,20%	3,87	639
Contig_6365.1	9	99,20%	2,68	296
Contig_6813.1	10	97,70%	2,64	335
Contig_6821.1	6	100,00%	1,74	286
Contig_6912.1	6	100,00%	2,14	220
Contig_7272.1	7	100,00%	2,65	221
Contig_7373.1	6	99,70%	1,88	264
Contig_7570.1	7	99,70%	2,92	198
Contig_769.1	63	97,00%	7,93	772
Contig_7866.1	8	99,60%	2,35	297
Contig_7868.1	11	95,70%	3,64	276
Contig_807.1	53	97,40%	6,43	797
Contig_8127.1	8	99,50%	2,56	269
Contig_8170.1	11	99,10%	3,08	351
Contig_8243	5	100,00%	2,06	192
Contig_830.1	27	97,50%	4,73	546
Contig_844	20	98,60%	3,61	518
Contig_8590	4	99,50%	1,57	189
Contig_8627.2	13	99,60%	3,97	296
Contig_863	38	98,70%	7,14	515
Contig_9084.1	5	100,00%	1,71	233
Contig_9396	6	100,00%	2,71	182
Contig_9452.1	8	99,50%	1,84	373
Contig_9983	4	99,50%	1,68	178