Univerzita Karlova v Praze

Přírodovědecká fakulta

Studijní program:Biologie

Bc. Kateřina Holková

Genom enterovirů z dětské stolice: kombinace next-generation a klasického Sangerova sekvenování

Enterovirus genomes in stool: a combination of the next generation and Sanger sequencing

Diplomová práce

Školitel: doc. MUDr. Ondřej Cinek Ph.D.

Praha 2014

Poděkování

Ráda bych poděkovala doc. MUDr. Ondřeji Cinkovi Ph.D. za trpělivost a odborné vedení této diplomové práce. Rovněž děkuji Mgr. Lence Kramné a celému kolektivu Laboratoře molekulární genetiky Pediatrické kliniky FN Motol za odbornou a morální podporu a příjemnou přátelskou atmosféru při práci a psání této diplomové práce. Velký dík patří mému příteli a rodině, za intenzivní podporu během celého mého studia.

Prohlášení:

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

V Praze dne 2.5.2014

Abstrakt

Cílem této práce je vývoj strategie pro vyhodnocování dat z next-generation sekvenování. Pomocí bioinformatických nástrojů (programu Galaxy, Velvet, Enterovirus genotyping tool) jsme optimalizovali metodu pro zpracování těchto dat. Analyzovali jsme 22 vzorků. Deset z těchto vzorků bylo pěstováno na buněčných kulturách, zbylých dvanáct pochází z reálných vzorků stolic. Všechny vzorky pocházejí od jednotlivců, kteří jsou geneticky predisponovaní k diabetu 1. typu a všechny byly pozitivní na enterovirus. Enteroviry a jejich infekce jsou po dlouhou dobu považovány za vážné kandidáty, kteří mohou být zapojeni do etiologie onemocnění diabetu 1. typu, což je onemocnění končící absolutním deficitem insulínu v důsledku autoimunitní destrukce beta buněk pankreatu. Genetická složka tohoto onemocnění se zdá být poměrně dobře definována (*HLA, INS, CTLA4, PTPN22, CTLA4, IFIH1* a mnoho dalších genů), environmentální část etiologie zůstává nejasná.

Dokázali jsme sestavit 22 genomů de novo, avšak s četnými mezerami mezi jednotlivými kontigy. U prvních devíti vzorků jsme tyto mezery překlenuli pomocí Sangerova sekvenování. Tímto způsobem jsme sestavili 9 celých virových genomů.

Hlavním přínosem této práce je vytvoření univerzálního postupu analýzy dat z next-generation sekvenování, který se již používá pro další analýzu vzorků, jež jsou podrobeny tomuto typu sekvenování. Pomocí tohoto postupu jsme schopni identifikovat viry ze vzorku, aniž bychom je specificky detekovali.

Klíčová slova

Next-generation sekvenování, Sangerovo sekvenování, enteroviry, diabetes 1. typu, bioinformatika

Abstract

This diploma thesis deals with a development of a strategy for data evaluation generated by next-generation sequencing. Using bioinformatics tools such as Galaxy, Velvet and Enterovirus genotyping tool new aproach of data processing was optimized. There were 22 samples analyzed which of 10 were grown on cell culture. Remaining 12 were obtained from real stool samples. All samples were taken from children at the highest genetic risk of type 1 diabetes. All of them were enterovirus positive. Enteroviruses and their following infections have been suspecting to be involved in ehiology of type 1 diabetes for a long time. That's a disease resulting to an absolut insulin deficiency due to autoimmune destruction of pancreatic beta cells. Genetic components seems to be relatively well defined (the HLA, INS, STLA4, PTPN22, CTLA4, IFIH1 and numerous other genes), the environmental part of the etiology remains obscured.

We were able to assemble 22 genomes de novo. However, there were numerous gaps among the particular contigs. For the first nine samples these gaps were complemented by Sanger sequencing. Nine full-length genomes were assempled this way.

The main contribution of this work was to create a universal process of analyzing data from next-generation sequencing. This has already been using for further analysis of the samples which are subjected to this type of sequencing. Using this procedure we are able to identify viruses from a sample without their specific detection..

Key words

Next-generation sequencing, Sanger sequencing, enteroviruses, type 1. diabetes, bioinfomatics

Zkratky

CVA – coxsackie A virus

CVB – coxsackie B virus

CTLA4 – cytotoxic T-lymfocyte-associated protein 4

DNA – deoxynukleová kyselina

dNTP – deoxyribonukleotid trifosfát

HEV – Human Enterovirus

HLA – Human Leukocyte Antigens, hlavní histokompatibilní komplex člověka

IAA - Insulin Antibody, autoprotilátka proti insulinu

ICA – Islet Cell Antibody, antigen ostrůvkových buněk

IFIH1 - interferon induced with helicase C domain 1

INS – insulin gene

IRES – Internal Ribosome Entry Site, část genomu enteovirů

NGS – Next-generation sekvenování

NJ - Neighbour Joining metoda

PCR – Polymerase Chain Reaction

PTPN22 - protein tyrosine phosphatase nonreceptor type 22

RNA – ribonukleová kyselina

RT–PCR – Reverse-Transcriptase Polymerase Chain Reaction

T1D – type 1 diabetes, diabetes 1. typu

Obsah

1. Úvod

Diabetes 1. typu je polygenní, multifaktoriální onemocnění, vyskytující se především u dětí a adolescentů, ale i u dospělých. Jeho podstatou je autoimunitní destrukce beta buněk pankreatických ostrůvků, která začíná měsíce až roky před propuknutím klinických příznaků onemocnění.

Genetická složka je dnes poměrně dobře charakterizována a podílí se na rozvoji diabetu 1. typu zhruba 50%. Genetickou predispozici udávají HLA geny, konkrétně *HLA-DQB1, -DQA1, -DRB1* a dále jsou to polymorfismy v genech *INS*, *CTLA4, PTPN22, IFIH1* a další, slaběji asociované geny.

Jako u každého multifaktoriálního onemocnění mají roli vlivy environmentální. Za důležité jsou považovány především virové infekce a složení stravy. Z virových infekcí jsou hlavními podezřelými enteroviry. Ačkoli bylo provedeno mnoho studií zabývajících se potenciální spojitostí enterovirových infekcí s rozvojem autoimunity a následným propuknutím diabetu, etiologický podíl enterovirů v patogenezi diabetu 1. typu není stále objasněn: z longitudinálních studií vyplývá, že relevantní pro objasnění vztahu enterovirů a diabetu je detekce enterovirů v krvi a jejích částech a vzorcích tkáně, naopak ve vzorcích stolice nebyla nalezena zvýšená frekvence enterovirů u pacientů oproti kontrolním subjektům. Situace je enormně komplikována počtem různých virových typů, dlouhou dobou mezi infekcí a propuknutím nemoci a heterogenitou charakteru patologické imunitní odpovědi, která nakonec způsobí autoimunitní diabetes.

Virologický výzkum je v současnosti podstatně měněn nástupem metod sekvenování nové generace, které posouvá možnosti charakterizace komplexních vzorků na novou úroveň. Studium enterovirů v patogenezi diabetu není výjimkou a předkládaná práce je součástí tohoto úsilí.

1.1. Cíle diplomové práce

Praktickým cílem této diplomové práce je vytvořit postupy a nastavit analytické parametry, které pomohou charakterizovat kmeny enterovirů ve vzorcích stolice dětí s genetickým rizikem diabetu 1. typu, a to pomocí kombinace NGS (next-generation sekvenováním) se Sangerovým sekvenováním.

2. Literární přehled

2.1. Diabetes 1. typu

Diabetes 1. typu, také nazývaný insulin-dependentní diabetes mellitus či juvenilní diabetes, je celoživotní autoimunitní metabolická porucha. Je způsoben insulinovou deficiencí v důsledku destrukce beta buněk pankreatických ostrůvků. Skrytý patogenetický proces, který začíná roky před propuknutím klinických symptomů, vede k nedostatečné syntéze insulinu a nakonec končí neschopností kontroly krevní glukosy (Coppieters, Boettler, & von Herrath, 2012).

Incidence diabetu 1. typu celosvětově narůstá. Tato choroba představuje celkem 5-10% všech případů diabetu a vyskytuje se převážně u dětí a adolescentů, kteří následně potřebují celoživotní léčbu injekčním insulinem (Wu, Ding, Gao, Tanaka, & Zhang, 2013).

2.1.1. Patogeneze diabetu 1. typu

Diabetes 1. typu je autoimunitní choroba. Již poměrně dlouho před jeho manifestací dochází k selektivnímu masivnímu ničení beta buněk pankreatických ostrůvků vlastním imunitním systémem (Daneman, 2006). Důsledkem této destrukce je absolutní nedostatek insulinu. Již před propuknutím klinických projevů diabetu cirkulují v krvi jedince protilátky, jejichž cílem se stávají autoantigeny beta buněk: protilátky jsou užitečnými markery prediabetického procesu (Hober & Sane, 2010), nejsou však destruktivní. To, že nejsou destruktivní, je dobře vidět u dětí diabetických matek. Děti sice mají přenesené mateřské protilátky, ale nemají diabetes.

2.1.2. Faktory ovlivňující vznik diabetu 1. typu

2.1.2.1. Genetické faktory

HLA je nejsilnějším genetickým faktorem pro vznik diabetu 1. typu a je také znám déle než všechny ostatní s diabetem asociované geny. Mimo HLA bylo ale do dnešního dne odhaleno nejméně 40 dalších oblastí v lidském genomu, které jsou asociovány s diabetem 1. typu. O mnoha z nich je známo, že jsou významné při antivirové odpovědi (Tauriainen, Oikarinen, Oikarinen, & Hyöty, 2011).

2.1.2.1.1. HLA

HLA lokus je lokalizován na chromozomu 6. HLA 2. třídy. HLA DR a DQ vykazují nejsilnější genetickou asociaci s diabetem 1. typu (Nokoff & Rewers, 2013).

2.1.2.1.2. PTPN22, INS, CTLA-4, IFIH1

Nejvýznamnější z hlediska rozvoje autoimunitních procesů a rozvoje diabetu 1. typu jsou hned po HLA geny *PTPN22* (protein tyrosine phosphatase nonreceptor type 22), *INS* (insulin gene), *CTLA-4* a *IFIH1* (interferon induced with helicase C domain 1) (Steck et al., 2012).

Přítomnost alely 1858T genu *PTPN22* je relativně silně asociována s rozvojem diabetu 1. typu. *PTPN22* kóduje lymfoidně specifickou fosfatázu, která je vytvářena v lymfocytech a je inhibitorem T buněčné aktivace (Bottini et al., 2004). Substituce jedné aminokyseliny u *PTPN22* pravděpodobně mění vazbu k intracelulární kináze Csk, což vede k poklesu inhibice T buněčné aktivace a podpoře rozvoje multiorgánové autoimunity. U jedinců s vysoce rizikových genotypem pro vznik diabetu 1. typu je *PTPN22* alela 1858T nezávisle asociována s rozvojem persistentní ostrůvkové autoimunity (Steck et al., 2009).

INS kóduje preproinsulin, který je přeměněn na proinsulin a po odstranění C-peptidu vzniká insulin. Polymorfismy -23HphI a +1140A/C jsou asociovány s diabetem 1. typu (Barratt et al., 2004). Někteří autoři (Lempainen et al., 2012) se domnívají, že rizikový *INS* genotyp se pravděpodobně účastní indukce a rané fáze beta buněčné autoimunity a rizikový *PTPN22* v pozdějších fázích.

IFIH1 je cytoplazmatická helikáza, která hraje roli při detekci intracelulární virové dvouřetězcové RNA pikornavirů (Kato et al., 2006). Navázání dvouřetězcové RNA na tento receptor spustí uvolnění prozánětlivých cytokinů, jako jsou např. interferony. Ty se projevují silnou antivirovou aktivitou chránící neinfikované buňky a indukující apoptózu u infikovaných buněk (Bouças, de Oliveira, Canani, & Crispim, 2013). V případě polymorfismu v tomto genu, dojde k narušení této protivirové aktivity.

2.1.2.2. Environmentální faktory

Genetická predispozice je důležitá pro propuknutí diabetu 1. typu, ve stabilních populacích se však její míra v čase nemění. Genetické predispozice proto nevysvětlují rapidní nárůst výskytu diabetu 1. typu, kterého jsme v posledních dekádách svědky.

Existuje množství studií a důkazů o roli environmentálních faktorů v souvislosti s rozvojem tohoto onemocnění (Roivainen & Klingel, 2010). Tyto faktory, především virové infekce a složení stravy, hrají významnou roli v patogenezi diabetu 1. typu, jelikož mohou ovlivňovat aktivaci T-buněk (Achenbach et al., 2005). Za nejpravděpodobnější kandidáty, kteří mohou spustit autoimunitní proces a následný rozvoj diabetu 1. typu, dnes považujeme enterovirové infekce (Knip, 2011).

Jak bude uvedeno podrobněji dále, několik sérotypů coxsackie virů a echovirů je asociováno s rozvojem diabetu, jak se ukázalo na základě mnoha průřezových a prospektivních studií (Hober & Sauter, 2010; Roivainen, 2006; Richer & Horwitz, 2009; Tauriainen, Oikarinen, Oikarinen, & Hyöty, 2011). V lidském pankreatu některé enteroviry vykazují silnou afinitu k ostrůvkům, což naznačuje, že přímá interakce viru s ostrůvky může mít zvláštní význam v patogenezi diabetu 1. typu. (Merja Roivainen & Klingel, 2010)

2.1.3. Diagnostika diabetu 1. typu

Diabetes 1. typu je diagnostikován na základě měření glykémie a přítomnosti symptomů. Diabetes je u dětí obvykle doprovázen charakteristickými symptomy jako je polyurie (nadměrné močení), polydipsie (stavy nadměrné žízně), rozmazané vidění, ztráta hmotnosti. Dále je přítomna glykosurie (přítomnost glukosy v moči) a ketonurie (přítomnost ketolátek v moči). V případě, kdy jsou ketony přítomny v krvi či moči, je léčba bezodkladná (Craig, Hattersley, & Donaghue, 2009).

Diabetes 1. typu je asociován s výskytem řady protilátek, které jsou zároveň biomarkery destrukce beta buněk pankreatických ostrůvků. K diagnostice jsou nejčastěji využívané protilátky proti třem antigenům: 65 kD isoformě dekarboxylázy kyseliny glutamové (GADA – glutamic acid decarboxylase antibodies), tyrosin-fosfatáze IA2 (Islet antigen 2) a proti insulinu (IAA – Insulin autoantibody). Dále se ještě vyšetřují protilátky proti ostrůvkovým buňkám pankreatu (ICA – Islet-cell antibodies) a ZnT8, což je zinkový transportér, který je exprimovaný pouze v pankreatu (Seissler & Scherbaum, 2006).

Sérologické markery autoimunitního patologického procesu GAD, IA-2 nebo insulinové protilátky, jsou přítomné u 85-90% nově diagnostikovaných pacientů (Sabbah et al., 2000). Proporce těchto markerů je závislá na věku pacienta, počtu a kvalitě používaných testů a etnické příslušnosti (Barker et al., 2004).

Přítomnost jedné či více z těchto protilátek může předcházet klinické manifestaci diabetu 1. typu o měsíce až léta, přičemž počet a persistence protilátek vypovídá o pravděpodobnosti rozvoje klinického onemocnění (Daneman, 2006). Protilátky tak umožňují detekci prediabetické autoimunity tehdy, kdy je environmentálními vlivy nastartována. Vyšetřování autoprotilátek je tak zásadním nástrojem ve studiích zjišťujících vztah environmentálních vlivů – včetně virů – k rozvoji prediabetické autoimunity a diabetu.

2.2. Enteroviry

2.2.1. Zařazení

Rod enterovirů patří do rodiny *Picornaviridae*, řádu *Picornavirales*. Do rodu enterovirů řadíme následující druhy: lidské enteroviry druhů A-D, lidské rhinoviry druhů A-C, prasečí enterovirus B a opičí enterovirus A. Dnes již známe přes 100 sérotypů lidských enterovirů a toto číslo stále roste od doby, kdy se používá k určení metoda založená na sekvencích nukleových kyselin (Roivainen and Klingel 2010). Lidské enteroviry byly dříve děleny na polioviry, Coxsackie A a B viry, echoviry a další nepojmenované, očíslované sérotypy. Toto rozdělení bylo založeno zejména na základě růstu virů na buněčných kulturách (Hyoty and Taylor 2002).

Většina znalostí o enterovirové biologii pochází z poznatků o polioviru a několika málo dalších prototypických kmenů enterovirů (Stene and Rewers 2012).

Tab. 1: *Přehled Enterovirů vycházející z (Stellrecht K.A. 2011)*

druhy lidských enterovirů	sérotypy
HEV A	Lidské Coxsackie viry: A2-8, A10, A12, A14,A16
	Enteroviry: 71, 76, 89-92
HEV B	Lidské Coxsackie viry: A9, B1-6
	Lidské echoviry
	Enteroviry: 71, 76, 89-92
HEV C	Lidské Coxsackie viry: A1, A11, A13, A17, A19-22, A24
	Lidské polioviry: 1-3
	Enteroviry: 71, 76, 89-92
HEV D	Enteroviry: 68, 70, 94
Lidský rhinovirus A	75 druhů sérotypů
Lidský rhinovirus B	25 druhů sérotypů
Lidský rhinovirus C	

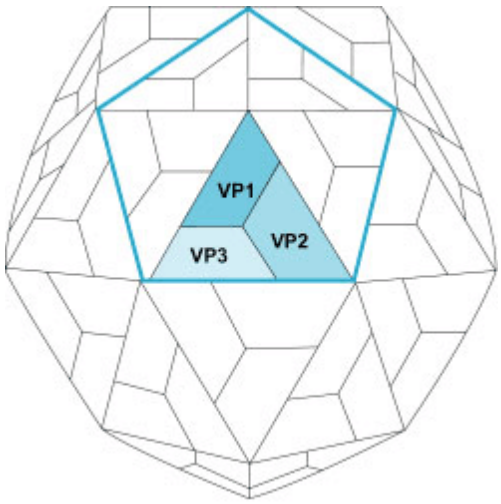
2.2.2. Stavba

Enteroviry jsou jednořetězcové neobalené RNA viry s dvacetistěnnou symetrií. Jejich jednořetězcová RNA molekula je přibližně 7500 nukleotidů dlouhá, má jeden otevřený čtecí rámec, který má na 5´ a 3´ koncích nekódující oblasti. Pikornavirová genomická RNA je unikátní tím, že na 5´ konec je kovalentně navázán VPg protein (Flanegan, Petterson, Ambros, Hewlett, & Baltimore, 1977). VPg různých pikornavirů se liší délkou, od 22 do 24 aminokyselin a je kódován jediným virovým genem. V 5´ oblasti se nachází IRES (Internal Ribosome Entry Site), neboli vnitřní místo pro vstup ribozomu. Tato sekvence umožňuje nasednutí ribozomu a směruje mRNA k místu na ribozomu, kde poté dochází k translaci (Forss & Schaller, 1982).

Na 3´ konci je netranslatovaná oblast čítající 70-100 nukleotidů a poly-A ocásek (Stellrecht K.A. 2011). Ten zodpovídá za větší stabilitu virové RNA. Čtecí rámec je translatován do polypeptidových prekurzorů, které jsou následně naštěpeny virovou proteázou. Polypeptid je rozdělen do tří funkčních oblastí, označovaných jako P1 až P3 (Fields, 2007).

Oblast P1 kóduje virové kapsidové proteiny VP1 až VP4 (Fields, 2007). VP proteiny tvoří vrcholy dvacetistěnu (viz obrázek 1). V molekulární diagnostice enterovirů se využívají

zejména dvě oblasti: identifikace na základě VP1 proteinu, který koreluje s charakterem jednotlivých sérotypů a 5´ netranslatované oblasti (Thoelen et al., 2004).

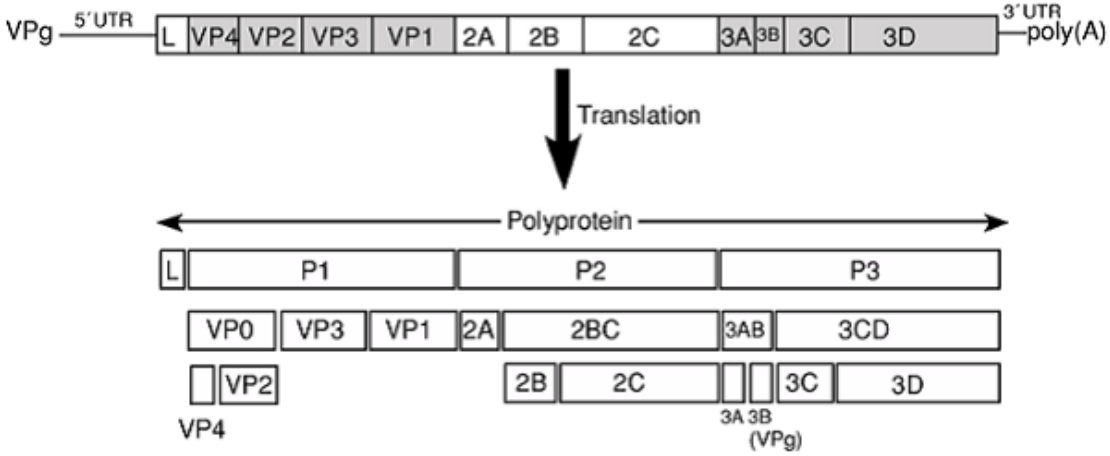


Obr. 1 *Enterovirová kapsida, kde P proteiny tvoří vrcholy dvacetistěnnu. (Alex I. Donaldson, "Foot-and-mouth disease," in AccessScience, ©McGraw-Hill Companies, 2008. cit. 27.4.2012 z http://www.accessscience.com)*

Nestrukturální proteiny jsou nezbytné pro enterovirový životní cyklus. Oblasti P2 a P3 kódují oblasti potřebné k produkci proteinů (2A proteáza, 3C proteáza a 3CD proteáza) a replikaci (2B, 2C, 3AB, 3B, Vpg, 3CD proteázu a 3D polymerázu) (Fields, 2007)

Protein 2A je chymotrypsinu podobná proteáza, která katalyzuje štěpení mezi VP1/2A a tím uvolňuje P1 strukturální proteinový prekurzor od zbytku polypeptidového řetězce (Toyoda et al., 1986). Protein 2B hraje neznámou roli v syntéze RNA, indukuje permeabilitu buněčných membrán a částečně je zodpovědný za proliferaci membránových veziklů (Johnson & Sarnow, 1991). Protein 2C je vysoce konzervovaný mezi enteroviry a jeho funkce je při replikaci RNA a nejspíše i při stabilizaci struktury virionu (Ryan & Flint, 1997).

Protein 3AB je prekurzorem 3B, malého polypeptidu kovalentně připojeného k VPg proteinu na 5´ netranslatované oblasti pikornavirové RNA molekuly. Protein 3C je chymotrypsinu podobná serinová proteáza, občas ve formě svého prekurzoru 3CD, je zodpovědná za primární štěpení mezi P2 a P3 (mezi 2C a 3A) a sekundárně štěpí P1 a P2 polyprotein. Protein 3D je RNA dependentní RNA polymeráza (Fields, 2007). Procesy štěpení polyproteinu jsou znázorněny na obrázku 2.



Obr. 2 *Enterovirový genom. Je zde zobrazen enterovirový genom a jeho proteinové produkty, které u enterovirů vznikají postupným štěpením původního polyproteinu (Fields, 2007).*

2.2.3. Životní cyklus

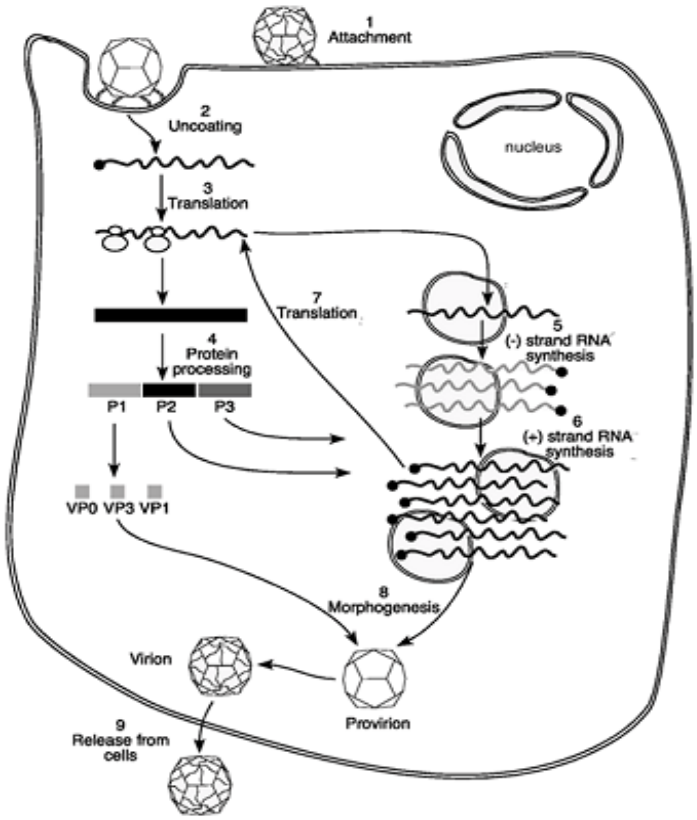
Replikace pikornavirů probíhá v cytoplazmě. Prvním krokem je přichycení viru na buněčný receptor. Jednořetězcová RNA poté vstupuje do buňky mechanismem endocytózy, je translatována a tím vzniknou virové proteiny nezbytné pro replikaci genomu a produkci nových virových partikulí. Virové proteiny jsou syntetizovány z polyproteinového prekurzoru, který je posléze štěpen. Štěpením nejprve vzniknou dvě proteázy 2A a 3C nebo 3CD. 2A katalyzuje štěpení mezi VP1/2A, zatímco 3C, někdy i ve formě svého prekurzoru 3CD, je zodpovědný za ostatní štěpení. Aktivita, zodpovědná za finální maturační štěpení mezi VP4/VP2, probíhá po složení virové partikule (Fields, 2007).

Mezi syntetizovanými proteiny je i virová RNA dependentní RNA polymeráza a příslušné proteiny nutné pro replikaci genomu. Replikace genomu je vykonávána RNA dependentní RNA polymerázou (3D) s pomocí virových a hostitelských faktorů. Nejprve je syntetizována negativně orientovaná kopie, která je následovně použita jako templát pro nový genomický RNA řetězec. Malý protein (VPg) je asociován s 5´ koncem genomu a podílí se na replikačním a skládacím procesu, kde je genomická RNA balena dovnitř kapsidy.

Výsledkem napadení buňky tímto virem je její lýza.

Replikace virové RNA začíná během 1 hodiny po infekci buňky a trvá okolo 4 hodin. Po této době buňka lyzuje a je z ní uvolněn téměř milion virových partikulí.

U buněk napadených enterovirem neprobíhá účinně translace buněčné mRNA, protože virové proteázy inaktivují buněčný komplex připojující čepičku na 5´ konec, který je potřebný pro navázání buněčné mRNA k ribozomům.



Obr. 3: *Obecný přehled replikačního cyklu pikornavirů. (1) Virus se váže k buněčnému receptoru a (2) genom je obnažen. VPg je odstraněn od virové RNA, která je poté translatována (3). Vznikající polyprotein je štěpen na individuální virové proteiny (4). RNA syntéza probíhá v membránovém veziklu, což není na obrázku zaznamenáno. Virový pozitivně orientovaný řetězec RNA je kopírován virovou RNA polymerázou do mínus řetězců RNA (5), které jsou poté kopírovány a produkují další pozitivní řetězce (6). Při brzké infekci je nově syntetizovaná RNA translatována a dochází k produkci virových proteinů (7). V pozdějších fázích infekce dochází k morfogenetickému utváření partikulí. Dovnitř těchto partikulí jsou baleny pozitivně orientované řetězce (8). Nově syntetizované virové partikule opouštějí buňky při jejím lyzování (9) (Fields, 2007).*

2.2.3.1. Persistentní infekce

Persistentní nebo pomalá virová infekce může být také důležitá při rozvoji autoimunity (Wu et al., 2013). Persistentní infekce je obecně spojována s imunitně zprostředkovanou destrukcí cílových orgánů.

Frekvence detekce enterovirů v pankreatu a intestinální mukóze (Oikarinen et al., 2012) napovídá, že se může jednat o persistentní infekce. Dříve byla tato persistence dokumentována v případě srdeční tkáně, kde vede k chronické myokarditidě a kardiomyopatii. U myších modelů persistentní enterovirová infekce způsobuje nejen zánětlivou myopatii a srdeční poškození, ale také poškození centrálního nervového systému (Chapman & Kim, 2008).

Persistentní infekce je charakterizována pomalou replikací virového kmene, který může mít v genomu specifické delece (Chapman & Kim, 2008). Tento případ persistujícího viru je nejspíše případem i enteroviry indukované kardiomyopatie a dle tohoto příkladu se domníváme, že stejně tak to probíhá i v případě enterovirové infekce a diabetu. Jelikož místem persistujících enterovirů by mohl být pankreas nebo, jak naznačují nedávné studie, střevní sliznice (Maarit Oikarinen et al., 2012; Nurminen, Oikarinen, & Hyöty, 2012).

2.2.4. Přenos enterovirové infekce

Jediným přirozeným hostitelem lidských enterovirů jsou lidé (D Richman 2002). Enterovirové infekce se přenášejí fekálně-orální či orálně-orální cestou, ale také přímým kontaktem se sekrety z očních či kožních lezí.

Voda, jídlo a půda kontaminovaná infekčními výkaly jsou zdroji nákazy, které vytvářejí mnoho příležitostí pro přenos infekce a tím rychlé epidemické rozšíření za krátkou časovou periodu. Enteroviry byly izolovány ze všech typů vod – zemních i odpadních vod, z mořské vody i z pitné vody.

Enteroviry jsou velmi odolné organismy schopné vzdorovat i vysokým koncentracím NaCl a vysokým teplotním změnám. Přežívají i v prostředí gastrointestinálního traktu, kde jsou stabilní při pH 3-5 po dobu 1-3 hodin (Rajtar, Majek, Polański, & Polz-Dacewicz, 2008).

V oblastech mírného pásu se enterovirové infekce vyskytují nejčastěji v létě a na podzim (v 70-80% případů). V oblastech tropického pásu se enterovirové infekce vyskytují v průběhu celého roku, jejich nárůst však může být zaznamenán v období dešťů (Stellrecht K.A. 2011).

2.2.5. Patogeneze

Enterovirové infekce patří mezi jedny z nejběžnějších virových infekcí. Ve většině případů jsou příznaky subklinické nebo mírné se symptomy nachlazení a průjmů, oční infekce, kožní onemocnění apod. V některých výjimečných případech však může dojít k závažnějším onemocněním myokarditis či neurologickým obtížím.

K primární infekci a replikaci virů dochází ve stěně tenkého střeva a faryngu. Rozšiřuje se do poměrně velké části gastrointestinálního epitelu. Tato infekce má krátkou inkubační dobu, 1-3 dny. Ačkoli je virová replikace omezena povrchem intestinálního traktu, účinky mohou být více generalizované. Generalizované infekce probíhají z počátku tak, že virion pronikne epitelovým povrchem. Zde dochází k replikaci viru. Potom migrují místními lymfatickými uzlinami. Některé viry jsou zde pohlceny makrofágy a inaktivovány, jiné však vstoupí do krevního oběhu, což má za následek primární virémii. Z krve mohou prostoupit do jiných orgánů, jako jsou játra, slezina, kostní dřeň či cévní endotelium, kde se znovu množí. Velké množství virů, které je pomnoženo v těchto orgánech, se může zpátky vrátit do krevního oběhu a dochází k sekundární virémii. (Leslie Collier 2000).

Epidemiologická data ukazují, že viry (enteroviry či cytomegalovirus) mohou přispívat k patogenezi diabetu. Na základě seroepidemiologických studií se zdá, že zejména enterovirus může vyvolávat diabetes 1. typu. Několik faktorů reguluje hostitelovu vnímavost k enterovirovým infekcím. Obecné rizikové faktory vážných enterovirových onemocnění zahrnují nízký věk, mužské pohlaví, humorální imunodeficienci, nedostatek mateřských protilátek a krátkou dobu kojení (Enterovirus suirveillance-United States. 2006, Khetsurani N; Karita Sadeharju et al., 2007; Nurminen et al., 2012).

2.3. Enteroviry a diabetes 1. typu

2.3.1. Spojitost virové infekce a diabetu 1. typu

Zájem o viry pochází z pozorování, že některé způsobují diabetu podobnou chorobu u zvířat. Domněnka, že enviromentální faktory přispívají k rozvoji diabetu i u lidí, se objevila už koncem 19. století, kdy epidemie příušnic v malé norské vesnici byla spojena s propuknutím dětského diabetu. Od této doby se pátrá po infekčním agens, které by spouštělo ostrůvkovou autoimunitu a klinický počátek onemocnění (J. W. Yoon, 1990). Dalším krokem byly pokusy na různých zvířecích modelech, u nichž se zkoumal mechanismus virově indukovaného

diabetu (Jaïdane et al., 2009) a ukázalo se, že několik virů má opravdu schopnost způsobit diabetes u zvířat, a to vzájemně odlišnými mechanismy (J.-W. Yoon & Jun, 2006; Nurminen et al., 2012).

Jestliže by enteroviry způsobovaly diabetes 1. typu, je logické předpokládat, že by měly stejné patogenní rysy, jež jsou známé od jiných enterovirových chorob. Příkladem jsou například onemocnění způsobená polioviry, které patří mezi nejlépe charakterizované enteroviry a jejich patogeze je široce studována (Mueller, Wimmer, & Cello, 2005).

2.3.2. Metody detekce enterovirů

Sérotypy enterovirů se navzájem liší v ochotě proliferovat na buněčných kulturách, ale obecně platí, že pokusy diagnostikovat infekci přímou izolací viru vesměs selhávají. Nepřímá diagnostika pomocí protilátek byla hlavním způsobem zjišťování infekce po několik dekád, ale ani ta není zcela snadná. Dnes se využívají metody založeny na PCR a sekvenování. Všechny metody přinesly důkazy pro i proti podílu enterovirů na etiopatogenezi diabetu 1. typu.

2.3.2.1. Sérologické důkazy

Možný vztah mezi enterovirovými infekcemi a diabetem 1. typu je zkoumán více než 40 let. První výzkum zabývající se frekvencí protilátek proti coxsackievirům B u nově diagnostikovaných diabetických pacientů oproti kontrolním subjektům byl proveden a publikován už roku 1969 (Gamble, Kinsley, FitzGerald, Bolton, & Taylor, 1969). Od té doby bylo provedeno mnoho podobně navržených studií v mnoha zemích potvrzující toto pozorování, avšak ne všechny sérologické studie hledající asociaci enterovirové infekce a autoimunity našly pozitivní výsledky.

Rozpor mezi jednotlivými studiemi může být způsoben především samotnými metodami stanovení protilátek proti enteroviru. Sérotypová specifita mnoha sérologických testů je nízká a identifikace asociace mezi specifickými virovými sérotypy a diabetem 1. typu je komplikována dalšími faktory. Mezi ně patří zkřížená reaktivita mezi enterovirovými sérotypy a mezi příbuznými pikornaviry; heterotypické protilátky odpovídají na sekundární či pozdější infekce; individuální variace v protilátkové odpovědi; existence odlišných kmenů v rámci sérotypu.

2.3.2.2. Detekce enterovirové RNA

Ačkoli jsou mezi sérologickými důkazy rozpory, souvislost mezi enterovirovými infekcemi a rozvojem diabetu podporují i studie, které využívají citlivé metody přímé diagnostiky jako je detekce enterovirového genomu pomocí RT-PCR (Reverese Transcriptase-Polymerase Chain Reaction). (Lauwers, Bissay, & Rombaut, 1998)

Detekce enterovirového genomu je obvykle prováděna pomocí amplifikace 5´netranslatované oblasti, která je konzervativní napříč enterovirovými skupinami. Taková RT-PCR detekuje RNA všech známých enterovirů. Sekvenování této části virového genomu však nikdy nedává jasné informace o sérotypu. (Andréoletti et al., 1997; Clements, Galbraith, & Taylor, 1995). Molekulární klasifikace enterovirů se místo toho zaměřuje na oblast kódující protein VP1, tedy hlavní antigenní místo – jeho míry polymorfismu je však taková, že v tomto genu nelze ukotvit žádné primery, které by potenciálně mohly sloužit ke generické detekci enteroviru.

V současnosti je k dispozici relativně dosti údajů o asociaci enteroviru s prediabetem; enterovirus byl ve studiích detekován v různých druzích vzorků, tedy v různých fázích infekce.

2.3.2.2.1. Detekce enterovirové RNA ve stolici

Enterovirová RNA ve stolici, jako odraz primární replikace viru ve střevě, se neliší frekvencí ani kvantitou mezi případy s prediabetem a kontrolami, jak bylo ukázáno v několika studiích.

Například v norské populaci byla frekvence enterovirové RNA ve stolici u pacientů před sérokonverzí (43 z 339 vzorků) 12,7%, zatímco u kontrol (94 z 692 vzorků) to bylo 13,6%; rozdíl nebyl signifikantní (Tapia et al., 2011).

2.3.2.2.2. Detekce v střevní stěně

Existují dvě finské studie zabývající se přítomností enterovirové RNA ve stěně střevní, kde tento virus je patrně schopen persistovat, což může být významným patogenetickým faktorem při rozvoji diabetu 1. typu. V první z nich byly shromážděny vzorky biopsie střevní sliznice od 39 diabetických pacientů, 41 nediabetických pacientů a 40 pacientů s celiakií. Tyto vzorky autoři testovali třemi metodami na přítomnost enterovirové RNA: in situ hybridizací, imunohistochemickým značením a RT-PCR. Výsledky ukázaly, že 74% diabetických pacientů bylo pozitivní na přítomnost enterovirů oproti 29% pozitivitě u nediabetických kontrol a 45% pozitivitě u pacientů s celiakií. Tyto poznatky jsou klinicky signifikantní.

Přítomnost enterovirové RNA byla asociována se zvýšenou zánětlivou aktivitou (buněčnou i protilátkovou) ve stěně střevní sliznice. (Maarit Oikarinen et al., 2012; Sarmiento et al., 2013)

Zdá se, že problematika je kontroverzní, protože obdobná studie konkurenční finské skupiny používající italské vzorky podobnou asociaci nenašla.

2.3.2.2.3. Detekce v pankreatu

Zkoumání vzorků pankreatu je velmi obtížné, neboť získat tyto vzorky od žijících pacientů je pro ně velmi riskantní (Krogvold et al., 2014). Existuje však studie, která pracovala se vzorky pankreatu, získaných autopsií, tedy že se vzorky pankreatické tkáně odebraly zemřelým pacientům. Tímto způsobem bylo například postupně odebráno 72 vzorků od diabetických a 163 vzorků od nediabetických pacientů. Vzorky byly imunofluorescenčně značeny proti insulinu, glukagonu, VP1 proteinu a dvouřetězcové RNA. VP1 pozitivní buňky byly detekovány ve 44 případech (61%) diabetických pacientů oproti 3 pozitivitám z 50 u nediabetických dětí (Richardson, Willcox, Bone, Foulis, & Morgan, 2009).

2.3.2.2.4. Detekce enterovirové RNA v krvi

Několik studií detekovalo enterovirový genom v krvi diabetických pacientů. Není však známo, zda-li se jednalo o persistentní nebo akutní infekci. (Tauriainen et al., 2011) Enterovirová RNA byla detekována v séru dlouho před diagnostikováním diabetu. Jednalo se o pozitivitu u 2,7% vzorků (36 z 1326 vzorků pacientů) oproti 1,9% (19 z 993 vzorků) u kontrol. Frekvence pozitivity na enteroviry byla analyzována během rozdílných stádií preklinického procesu onemocnění. Tato frekvence vrcholila 6 měsíců před tím, než se objevily první protilátky, kdy byla pozitivita u pacientů nalezena v 15,2% případů, zatímco u kontrol to bylo 3,3% (S. Oikarinen et al., 2011).

Ze srovnání dat ze studií frekvence viru ve stolici, střevní stěně a krvi je zřejmé, že potenciálně diabetogenní virus patrně proniká několika imunitními a mechanickými bariérami, aby se následně šířil krevním řečištěm do cílového orgánu, pankreatu. Alternativním vysvětlením je vyplavování viru nebo jeho RNA z míst sekundární replikace, jakými by mohly být například lymfatické uzliny gastrointestinálního systému s potenciálním propojením k pankreatu. Možných vysvětlení patogeneze diabetu ve spojistosti s enterovirem je několik.

2.3.3. Možné role enterovirů v patogenezi diabetu 1. typu

Z publikovaných nálezů je zřejmé, že enterovius proniká do Langerhansových ostrůvků krátce po manifestaci diabetu, při ní, dokonce v souvislosti s iniciací autoimunitního procesu. Pro vysvětlení role právě při iniciaci autoimunity je několik teorií.

2.3.3.1. Přímá lýze beta buněk způsobená enterovirovou infekcí

Enteroviry jsou známé svou cytolytickou aktivitou, takže poté, co se dostanou do pankreatických ostrůvků, mohou destruovat beta buňky produkující insulin mechanismem virové cytolýzy. Navíc důsledkem infekce je zánět vyvolaný aktivací vrozeného a adaptivního imunitního systému (Merja Roivainen & Klingel, 2010).

2.3.3.2. Virem navozená imunitní odpověď proti infikovaným beta buňkám

Dalším možný mechanismus patogeneze enterovirů může představovat akutní či persistentní infekce v pankreatu vyvolávající autoimunitní proces. Tento proces způsobují autoreaktivní T buňky či autoprotilátky. Aktivátory autoreaktivních T buněk mohou být molekulární mimikry či procesy bystander aktivace.

2.3.3.2.1. Molekulární mimikry

Koncept molekulárních mimiker popisuje situaci, kdy podobnost sekvenční nebo strukturní mezi virem a autoantigeny, vede k funkční T a B buněčné zkřížené reaktivitě. Výsledkem je tkáňové poškození a přetrvávání autoimunitní odpovědi.

Imunogenní epitopy enterovirového kapsidového proteinu VP1 a prokapsidového proteinu VP0 mají sekvenční podobnost s diabetes asociovaným epitopem v tyrosinové fosfatáze IA-2 a heat shock proteinu 60 (Härkönen et al., 2003).

Dalším důkazem existence molekulárních mimiker může být sekvenční homologie mezi karboxylázou 65 (GAD65) a 2C proteinem coxsackievirů B, která následně vede ke zkřížené reaktivitě (Nurminen et al., 2012).

2.3.3.2.2. Bystander aktivace

Tento mechanismus je založen na nespecifické aktivaci autoreaktivních T-buněk, které unikly selekci v thymu, v prostředí chronického zánětu vyvolaného viry.

Mechanismy bystander aktivace mohou přispívat k agresivní destrukci ostrůvků pankreatu. Virová infekce vede k aktivaci antigen-prezentujících buněk jako jsou dendritické buňky, následnék aktivaci autoreaktivních T buněk a k autoimunitní odpovědi (Fujinami, von Herrath, Christen, & Whitton, 2006). Spustit bystander aktivaci mohou i virově specifické T buňky, které migrují do pankreatu, kde se setkávají s virem infikovanými. CD8+ T buňky rozpoznají infikované buňky a uvolní cytotoxická granula. Výsledkem je zabití infikovaných buněk. Vše poté vede k zánětu, který může vést k zabíjení i neinfikovaných okolních buněk.

Dalším mechanismem je hyperexprese CXCL10 v samých ostrůvkových buňkách při infekci – CXCL 10 může být chemoatraktantem a aktivátorem autoreaktivních T buněk a marofágů. Takové buňky zvyšují koncentraci cytokinů, INF gama a NO, což může vést k bystander zabíjení neinfikovaných okolních ostrůvkových buněk (Tanaka, Aida, Nishida, & Kobayashi, 2013).

2.3.4. Diabetogenní enteroviry

Není zřejmé, zdali děti geneticky predisponované k diabetu jsou častěji infikovány běžně kolujícími kmeny enterovirů nebo zdali jsou náchylnější specificky k infekci diabetogeními variantami viru. Není známo ani, co dělá enteroviry diabetogeními. Současné znalosti naznačují, že diabetogenní vlastnosti enterovirů nejsou definovány sérotypem, ale spíše jinými charakteristikami virových kmenů. Existuje několik studií, kde byly molekulární determinanty pankreatotropních virových kmenů zkoumány s využitím analýzy kompletních genomových sekvencí (enterovirus infection in human pancreatic islet cells, islet tropism in vivo and receptor involvement in cultured islet beta cells, Ylipaasto 2004) (analysis of pancreas tissue in a child positive for islet cell antibodies, oikarinen m, 2008). Zatím však nebyl odhalen žádný faktor vyvolávající diabetes (Nurminen, Oikarinen, & Hyöty, 2012).

2.3.5. Longitudinální studie virů v patogenezi diabetu 1. typu

Rozsáhlé prospektivní studie jsou nezbytné ke stanovení kauzálního vztahu mezi enterovirovými infekcemi a rozvojem prediabetické autoimunity a/nebo vývojem od autoimunity ke klinickému diabetu. Několik prospektivních studií se zabývá přirozenou progresí destrukce beta buněk a jejími prediktory či akcelerátory.

2.3.5.1. DiMe

První prospektivní studií zabývající se rolí enterovirové infekce v souvislosti indukce či akcelerace diabetu byla studie DiMe (Childhood Diabetes in Finland). Probíhala v letech 1987 až 1993 a odstartovala novou éru výzkumu patogeneze diabetu 1. typu. Sérologickými a molekulárními metodami bylo zjištěno, že u pacientů, u kterých se projevila subklinická destrukce beta buněk nebo klinický diabetes, byl vyšší výskyt enterovirových infekcí než u jejich nediabetických vrstevníků (Hyöty et al., 1995). Vyšší výskyt infekcí byl pozorována jak před klinickým diabetem, tak již o několik let předtím. Navíc se ukázalo, že výskyt enterovirových infekcí koinciduje se vzrůstem autoimunitních markerů pro rozvoj diabetu, především se jedná o protilátky proti ostrůvkovým buňkám (ICA) a dekarboxyláze kyseliny glutamové (GADA) (Hiltunen et al., 1997; Lönnrot, Salminen, et al., 2000; M Roivainen et al., 1998).

2.3.5.2. DIPP

Tyto poznatky byly potvrzeny i následující prospektivní studií DIPP (Finnish Diabetes Prediction and Prevention trial), která byla započata roku 1994. DIPP je studie sledující kohorty finských dětí s genetickou predispocizí pro vznik diabetu. U těchto dětí se sledoval výskyt s diabetem asociovaných protilátek v intervalech 3 až 12 měsíců od narození do 15. roku dítěte (Kupila et al., 2001). Byl pozorován větší výskyt enterovirových infekcí v období, kdy byly detekovány první protilátky. Frekvence enterovirových infekcí během následujícího období byla vyšší u dětí, které vykazovaly znaky beta buněčné autoimunity, než u kontrolních subjektů odpovídající časem narození, pohlavím a HLA-DQB alelami asociovanými s diabetem 1. typu (Lönnrot, Korpela, et al., 2000). Pacienti i kontrolní subjekty této studie měli podobné frekvence adenovirových infekcí, což naznačuje, že děti, u kterých se vyvinuly protilátky, nejsou více obecně náchylnější k jakýmkoli virovým infekcím, ale že efekt enteroviru je specifický.

Sérologicky bylo ověřeno, že frekvence enterovirových infekcí byla vyšší u pacientů v porovnání s nediabetickými dětmi průběhu šesti měsíční periody, která předcházela prvnímu výskytu protilátek asociovaných s diabetem (K Sadeharju et al., 2001). Tato časová asociace mezi enterovirovými infekcemi a počátkem beta buněčné autoimunity podporuje hypotézu, že enteroviry mohou být jedním z iniciačních faktorů vzniku autoimunity (Salminen et al., 2003).

2.3.5.3. DAISY

DAISY (Diabetes and Autoimmunity Study in the Young) studie probíhá od roku 1993 v Coloradu (USA). Do této studie bylo zařazeno 2365 dětí geneticky predisponovaných k diabetu 1. typu. Žilní krev a rektální výtěry byly sbírány každé 3 až 6 měsíců po sérokonverzi ostrůvkových autoprotilátek (GAD, insulin,IA-2) do doby diagnostikování diabetu.

U 140 dětí proběhla sérokonverze a opakovaně vykazovaly pozitivitu na ostrůvkové autoprotilátky. Z těchto dětí došlo u 50 k manifestaci diabetu. Riziko vývoje klinických projevů diabetu 1. typu u pacientů, kterým byly odebrány vzorky krve, ve kterých byla detekována enterovirová RNA, vzrostlo ve srovnání se vzorky, které byly negativní na enterovirovou RNA. Přítomnost enterovirové RNA v rektálních výtěrech nepředpovídá vznik diabetu 1. typu (Stene et al., 2010).

2.3.5.4. MIDIA

MIDIA studie pochází z Norska (Tapia et al., 2011). Na rozdíl od ostatních nepotvrdila, že enterovirová RNA je ukazatelem budoucího vývoje autoimunity proti ostrůvkům. Jako materiál byly použity vzorky stolic, které byly měsíčně sbírány od norských dětí s vysokou genetickou predispozicí k diabetu 1. typu. Četnost výskytu enterovirové RNA ve vzorcích stolice u pacientů před sérokonverzí byla u 43 dětí z 339 (12,7 %) a u nediabetických dětí to bylo v 92 případech z 692 (13,6 %). Rozdíl od ostatních studií však spočívá v tom, že enterovirová RNA byla detekována pouze ze vzorků stolic, což, jak se ukázalo, není tím správným místem pro hledání korelace mezi enterovirovou infekcí a následným vývojem protilátek. Vhodnější materiálem pro hledání důkazů je krev, popř. krevní sérum.

Tab. 2: *Přehled vybraných studií zabývajících se detekcí enterovirů v různých typech vzorků. Zároveň je zde vidět i počet subjektů zapojených do studie a počet pacientů po sérokonverzi a počet pacientů, u kterých propukl klinický diabetes.*

studie	subjekty studie	intervaly mezi odběry vzorků	typ vzorků a použitých testů	počet testovaných subjektů / subjekty, u kterých proběhla sérokonverze / počet subjektů, u kterých propukl T1D
DAISY	Děti geneticky predisponovaní k rozvoji autoimunity a vzniku T1D z Colorada	3 - 6 měsíců po sérokonverzi protilátek	vzorky séra a rektálních výtěrů, RT-PCR	2 365 / 140 / 50
DIPP	Děti, které vykazovaly opakovaně pozitivitu na enterovirovou RNA, z Finska	intervaly po 6 měsících	sérum, vzorky stolice, sérologie a RT-PCR	237 / 41 / -
DiMe	Sourozenci nově diagnostikovaných dětí s T1D ve Finsku	intervaly po 6 měsících	sérum,sérologie a RT-PCR	765 / 23 / 3
MIDIA	Děti genově predisponovaní k rozvoji autoimunity a vzniku T1D z Norska	intervaly po 3 měsících do 1 roku věku dětí	vzorky stolice, RT-PCR	339 / 43 / -

3. Materiál a metody

3.1. Vzorky pacientů s genetickou predispozicí k diabetu 1 . typu

Celkem jsme testovali 22 vzorků, které všechny vykazovaly pozitivitu na enterovirus; cílem bylo získat kompletní či téměř kompletní sekvenci virové RNA.

Z 22 virů jich 10 bylo pěstováno na buněčných kulturách Hela buněk, zbylých 12 vzorků pochází z reálných vzorků stolic. Všechny vzorky pocházejí z Finska, kde byly získány od dětí s genetickou predispozicí pro diabetes 1. typu, které jsou součástí kohorty DIPP.

Deset vzorků izolovaných na buněčných kulturách obsahovaly viry Coxsackie B3. Tento sérotyp se ukázal jako velmi důležitý v recentní sérologické studii (Laitinen et al., 2014), kde byly negativně asociovány s rizikem diabetu - patrně skrze virovou interferenci nebo zkříženou reaktivitu protilátek. Při užití virus neutralizačního testu je vhodné užít viry, které skutečně v populaci cirkulují a testovat imunitu proti nim - z této studie pochází i námi testovaných deset kmenů.

Zbývajících 12 vzorků jsou vzorky stolice jako materiálu pro přímou detekci bez předchozí izolace na buněčné kultuře. Tyto sloužily pro demonstraci efektivity zvoleného postupu analýzy.

3.2. Next-generation sekvenování

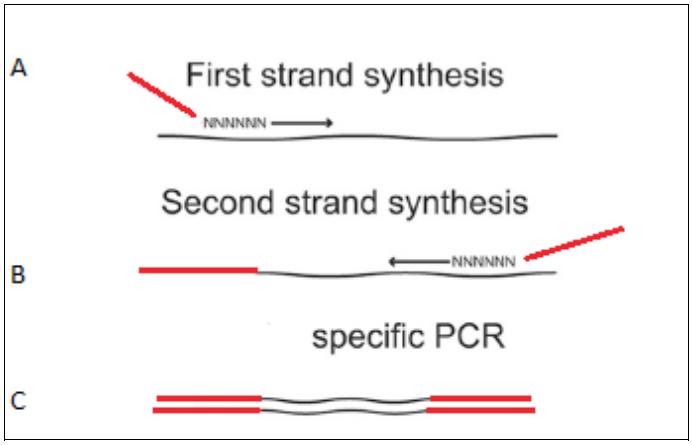
Nová technika sekvenování DNA, která je známá jako next-generation sekvenování (NGS) nebo také sekvenování nové generace, poskytuje vysokou rychlost a výkonost při produkci enormního množství sekvencí. Tato metoda skýtá mnoho možných využití jak na výzkumném tak na diagnostickém poli.

Sekvenování celých virových genomů je obtížným úkolem, jelikož se musíme potýkat s přítomností kontaminujících nukleových kyselin hostitelských buněk a jiných agens ve virových vzorcích (Barzon, Lavezzo, Militello, Toppo, & Palù, 2011).

3.2.1. Postup

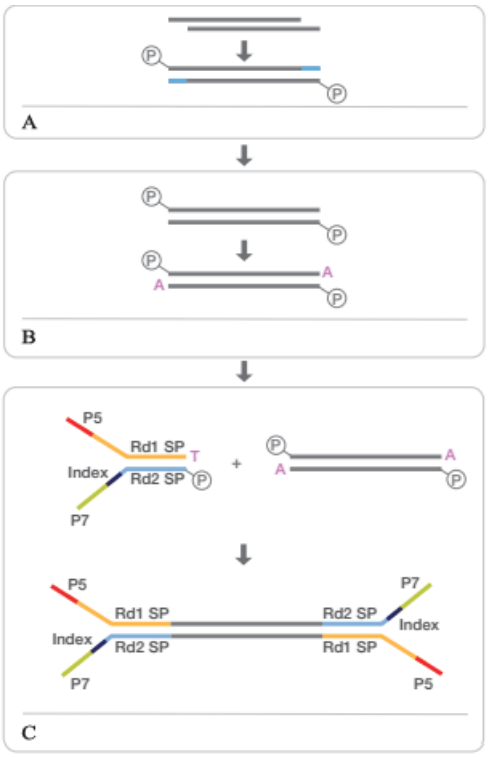
Prvním krokem krokem po odebrání vzorků je obohacení virové frakce a izolace celkové RNA. Vzorky stolic jsou převedeny do 10% suspenze v roztoku HBBS (Hanks ballanced salt sollution). Ze stolic byly odseparovány ostatní komponenty (zbytky potravy) od virové frakce

provedením 3x mírné centrifugace. Bakterií ze vzorku jsme se zbavili následnou filtrací přes bakteriální filtr (0,45 mikrometrů). Poté byla provedena ultracentrifugace po dobu 3 hodin, 10°C při 80000 g. Pro samotnou izolaci virů jsme využili QIAamp Viral RNA mini kit. Po získání virové RNA ze vzorků jsme je přepsali pomocí reverzní traskriptázy do cDNA za využití náhodných primerů s tagem. Tím jsme získali jednořetězcovou DNA. Druhý řetězec DNA jsme nasyntetizovali pomocí Klenowova fragmentu s použitím stejných náhodných primerů s tagem jako v prvním kroku. Výsledkem je tedy dvouřetězcová DNA s tagy na koncích, kterou podrobíme 30 cyklům PCR. Tato amplifikace následující po reverzní transkripci náhodnými primery slouží ke zvýšení celkového množství nukleové kysliny vstupující do sekvenace. Po PCR se produkty vizualizují na elektroforéze, kde ze z gelu vyříznou pouze produkty dlouhé 400 – 600 bp.



Obr. 4 *Příprava vzorků před NGS. A) Reverzní transkripce virové RNA do cDNA pomocí náhodných primerů s tagy. B) Dosyntetizování druhého řetězce Klenowovým fragmentem se stejnými náhodnými primery. C) Amplifikování DNA během 30 cyklů PCR.*

V několika následných krocích popsaných na obrázku 5 je na fragmenty DNA připojen adaptor.



Obr. 5 *Úprava konců fragmentů kitem TrueSeq. A) Úprava konců enzymatickou reakcí, při níž jsou připojeny fosfáty na konce fragmentu. B) Navázaní adeninu na fragmenty C) Ligace adaptoru, který využívá adenin k tomu, aby se připojil. Převzato z http://www.illumina.com.*

Samotná sekvenace všech amplifikovaných částí DNA probíhá najednou v přístroji MiSeq platforma Illumina za využití kitu TruSeq PCR-free. Princip metody je založen na sekvenaci syntézou (Ansorge, 2009). Fragmenty DNA jsou sekvenovány báze po bázi pomocí čtyř odlišných fluorescenčních barviv a vytváří shluky. Postupně dochází k navázání jednotlivých fluorescenčně značených bází na templát. Po každém kole syntézy jsou shluky detekovány laserem přístroje, který podle určitého fluorescenčního barviva pozná, o kterou bázi se jedná (http://www.giga.ulg.ac.be 11).

Výhodou této metody je, že nám umožňuje identifikovat viry ve vzorcích, aniž bychom věděli o jaké viry se jedná. To je rozdíl oproti doposud využívaným metodám specifického PCR, kdy jsme museli vědět, jaký virus hledáme a poté jsme ho specificky detekovali.

3.3. Skládání kontigů de novo

Zatímco virové genomy mají tisíce až stovky tisíc bází, úseky přečtené sekvenováním nové generace mají délku mezi desítkami a několika málo sty bází. Protože jsou virové genomy

velmi polymorfní, je obvyklým postupem jejich rekonstrukce de novo sestavení (de novo assembly) z jednotlivých čtení z NGS (Paszkiewicz & Studholme, 2010).

De novo sestavení genomu je proces, pomocí něhož spojíme jednotlivá čtení do jedné dlouhé souvislé sekvence, kterou nazýváme kontig. Ten sdílí stejné nukleotidové sekvence jako originální templát DNA, z něhož byla čtení sekvence odvozena.

Ačkoli větší proporce genomu může být obsažena ve větších kontizích, při sestavování jsou přítomny mezery mezi kontigy a velký počet krátkých čtení. Mezery se nám podařilo překlenout pomocí Sangerova sekvenování, jak je popsáno podrobně dále v této práci.

Před samotným sestavením virového genomu, jsme museli výstupní data z NGS upravit a filtrovat, teprve poté jsme mohli přistoupit k samotnému sestavení. Pro tuto úpravu jsme vybrali program Galaxy, který je popsán v následující kapitole. Pro zpracování a sestavení čtení do kontigů jsme využili program Velvet, jehož algoritmus je založen na de Bruijnových grafech.

3.3.1. Úprava dat v Galaxy

Galaxy je platforma pro interaktivní analýzu genomických dat. Galaxy slučuje práci již existujících databází pro anotaci genomů s jednoduchým webovým portálem, kde uživatelé mohou využívat vzdálených datových zdrojů společně s nezávislými zdroji, čili s vlastními daty. Srdcem Galaxy je flexibilní systém historie, kde jsou ukládány jednotlivé kroky daného uživatele. Lze mít kontrolu nad tím, co se se sekvencemi děje v jednotlivých krocích a tuto změnu zobrazit *(Giardine, 2005)*.

Data získaná pomocí NGS jsou ve formátu FastQ. Tento textový formát je založen na uchování biologických sekvencí (nukleotidové sekvence) společně s jejich odpovídajícími hodnotami kvality. Každý FastQ soubor podává informaci o sekvenci, sestává ze čtyř řádek, kde každý má svou vypovídací hodnotu. První řádek začínajíc @ je následován identifikátorem daného čtení. Druhý řádek je sama vlastní nukleotidová sekvence. Třetí řádek je pouhé +, které může být následováno nějakou přidanou informací o sekvenci a čtvrtý řádek kóduje hodnoty kvality pro sekvenci a musí obsahovat stejný počet symbolů jako je bází v sekvenci (Cock, Fields, Goto, Heuer, & Rice, 2010).

Prvním krokem, který následuje po získání dat, je příprava a kontrola kvality dat po sekvenování. Tyto kroky mají svojí typickou posloupnost a to: rozbor výsledku sekvenování, výpočet a vizualizace přehledu statistických údajů o hodnotách a kvality a nukleotidové distribuci sekvence, dále odstranění konců sekvencí, pokud je to nezbytné a filtrování čtení dle hodnot kvality (Blankenberg et al., 2010).

Kritéria pro třídění a úpravu sekvencí jsou nastavena tak, abychom zachovali, co největší počet sekvencí, ale zároveň byla zachována kvalita dat. Postup práce v programu Galaxy je dále popsán v kapitole Výsledky.

3.3.2. Velvet

Velvet (Zerbino & Birney, 2008) je nástroj sloužící k de-novo sestavení genomických fragmentů, které jsou výstupem metody sekvenování nové generace.

Velvet zpracovává čtení sekvencí do kontigů a odstraňuje chyby, jak je popsáno níže. Kontig je sada překrývajících se segmentů DNA, které dohromady vytváří konsensuální oblast DNA.

Algoritmus, podle kterého Velvet funguje, je založen na principu de Bruijnových grafů, které pracují s krátkými k-mery. Zadaná hodnota k-meru představuje délku fragmentů sekvencí, ze kterých je poté sestaven unikátní kontig, bez ohledu na to, jak často se vyskytuje.
Nejnáročnější část konstrukce de Bruijnových grafů sestává z „nařezání“ všech čtení (hashing) dle toho, jaká je zadaná hodnota k-meru (v našem případě se jedná o hodnotu 57). Tento proces je však poměrně časově náročný v porovnání s obecným párovým porovnávání všech sekvencí (pairwise alignment), zvláště v případě, kdy máme velké požadavky na vysokou míru pokrytí.

Tento program je složen ze dvou na sobě závislých programů, a to velveth a velvetg.

3.3.2.1. Velveth

Velveth pomáhá připravit data pro následující program velvetg. V tomto programu musíme specifikovat údaje o formátu čtení, tedy v našem případě se jedná o fastq formát. Dále specifikujeme délku *k*-meru. My jsme zvolili hodnotu 57, což je nejvyšší možná nastavitelná hodnota. A dále vybíráme údaj o typu čtení, v našem případě se jedná o krátká čtení (short reads).

Příkazový řádek pak vypadá následovně:

velveth ./output_directory 57 -fastq -short (náš soubor)

3.3.2.2. Velvetg

Velvetg je stěžejní částí Velvetu. Velveth připravuje sekvence na samotnou práci ve Velvetg. Velvetg pracuje podle algoritmu de Bruijnových grafů, odstraňuje chyby v sekvencích, jak je popsáno výše.

Pro velvetg je nutné nastavit parametry tak, aby se nám sekvence sestavily správně pod sebe a nedocházelo k jejím velkým ztrátám. Po testování různých hodnot, jsme došli k závěrečným parametrům.

Příkazový řádek vypadal následovně:

velvetg ./ -exp_cov 100 -ins_length 300 -cov_cutoff 5 -min_contig_lgth 150 -unused_reads yes -amos_file yes

- *exp_cov* = expected coverage – 100, tato hodnota vyjadřuje očekávané pokrytí unikátních oblastí

- *ins_length* = insert length – 300, je předpokládaná délka fragmentu – zjištěno na základě PCR a elektroforézy nasyntetizovaných úseků cDNA z našich vzorků.

- *cov_cutoff* = minimální počet čtení, která mohou vytvořit kontig - touto hodnotou požadujeme, aby kontig, který vznikne, byl pokryt minimálně pěti překrývajícími se sekvencemi.

- *min_contig_lgth* = minimum contig length - 100 bází – tento parametr určuje minimální délku kontigu, který je exportován jako soubor contigs.fa.

- *unused reads* – tento příkaz zadáváme proto, aby se nám vyexportovaly i sekvence, které nesplňují kritéria pro zařazení do kontigu. Tímto souborem se zabýváme,

abychom zjistili, zdali čtení, která byla vyřazena na základě našich kritérií, neobsahují některé enterovirové sekvence. V případě, že je neobsahují, nás zajímá, o jaké sekvence se jedná, zdali jsou virového původu a jakého, či bakteriálního.

- *amos file* – tento příkaz je nutný k tomu, aby se nám výsledky exportovaly jako soubor velvet_asm.afg, jelikož pouze v tomto případě je můžeme zobrazit pomocí nástroje Tablet.

3.3.2.3. Odstraňování chyb ve Velvetu

Algoritmus pro odstraňování chyb ve Velvetu se jmenuje Tour Bus. Je zaměřený na odstraňování chyb bez toho, aby byly porušeny spoje v de Bruijnových grafech.

Chyby jsou ve Velvetu odstraňovány na základě topologických znaků. Chybná data vytváří tři typy struktur: „tips“, které jsou vytvořeny chybami na koncích jednotlivých čtení, dále „bulges“ nebo „bubbles“ (vyboulení), které jsou způsobeny chybami uvnitř čtení. Třetí strukturou jsou chybná spojení zapříčiněné klonovacími chybami či dalekým/chybným slučování konců (tips). Tyto tři znaky/struktury jsou odstraňovány postupně a ne najednou.

Odstraňování tips

- „Tip“ je řetězec nodů, který je přerušený na jednom konci. Tip vzniká, když je nízká kvalita konců čtení, které se s ničím nepřekrývají. Může být odstraněn pouze v případě, když je kratší než 2*k.* Když je delší než 2k, představuje pravou sekvenci nebo také akumulaci chyb, které jsou jen velmi těžko rozlišitelné od nové sekvence (Zerbino & Birney, 2008).

Odstraňování bublin

- Tyto struktury vznikají uvnitř dlouhé sekvence, či v případě, když se konce dvou čtení překryjí náhodně a nesprávně. Nejsou však opravovány odstraněním nechtěných dat, ale promítnutím jedné větve do druhé a dojde k „přemapování“ na novou větev.

Odstraňování chybných spojení

- Chybná spojení jsou odstraňována po Tour Bus. Tato nechtěná spojení nevytvářejí žádnou rozpoznatelnou smyčku či strukturu, tudíž nemohou být identifikována z jejich topologie, jako je tomu u tips či bublin. Proto je Velvet odstraňuje na základě

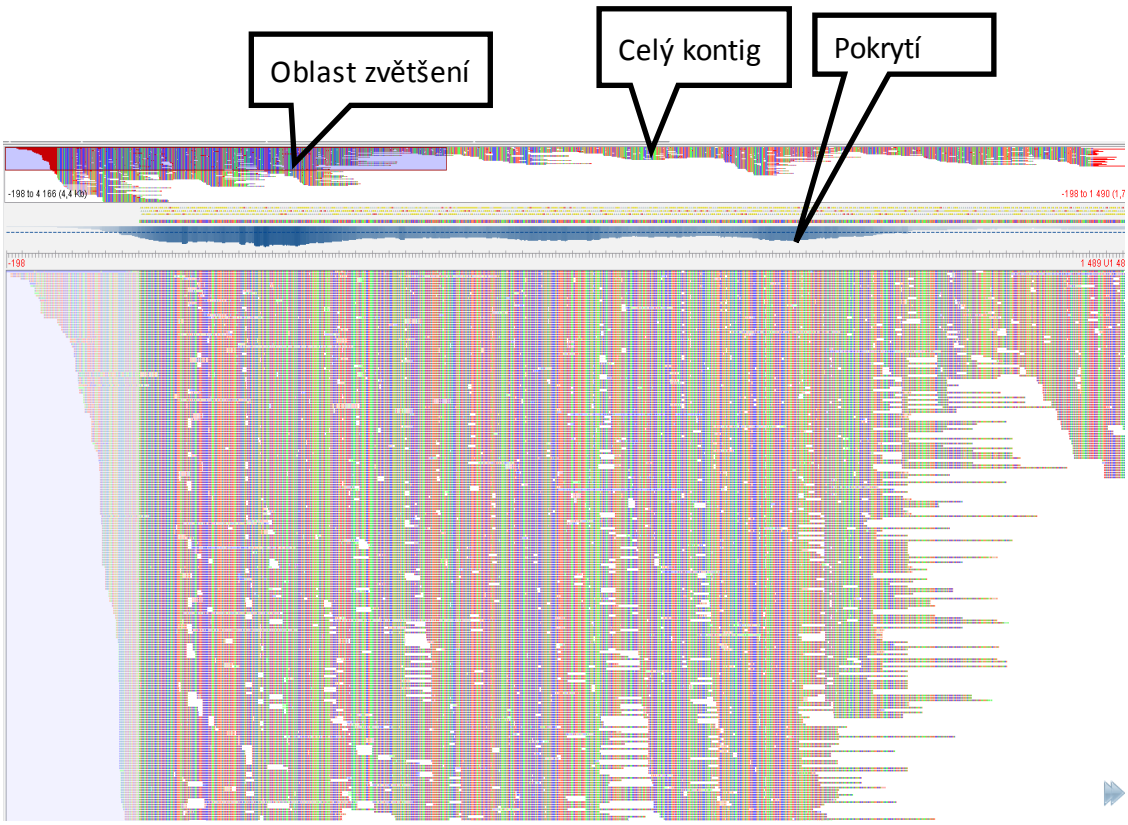
nastavení tzv. coverage cutoff, což je hodnota, která nám udává minimální počet překrývajících se čtení, která mohou vytvořit kontig.

Po skončení oprav jsou eliminovány kontigy s nízkým pokrytím, tedy ty které se neintegrovaly do větších kontigů.

3.4. Tablet

Tablet (= Next Generation Sequence Assembly Visualization) je vysoce výkonný grafický prohlížeč kontigů z NGS. Zde můžeme vidět délku jednotlivých čtení, z kolika sekvencí se nám daný kontig složil a také je zde grafické i procentuální zobrazení tzv. mismatchů (nukleotidových neshod), tedy zastoupení bází, které s ostatními v daném kontigu nekorespondují. V tomto programu si můžeme zobrazit všechna čtení v daném kontigu.

Také zde je možnost měnit kontrast mezi variantními a nevariantními nukleotidy. Tento kontrast nám projasní báze, které se odlišují od konsenzuální sekvence, a tedy mohou představovat potenciální jednonukleotidové polymorfismy či sekvenční chyby (Milne et al., 2010). Zde již můžeme také vidět konsenzuální sekvenci vytvořenou z daného kontigu.

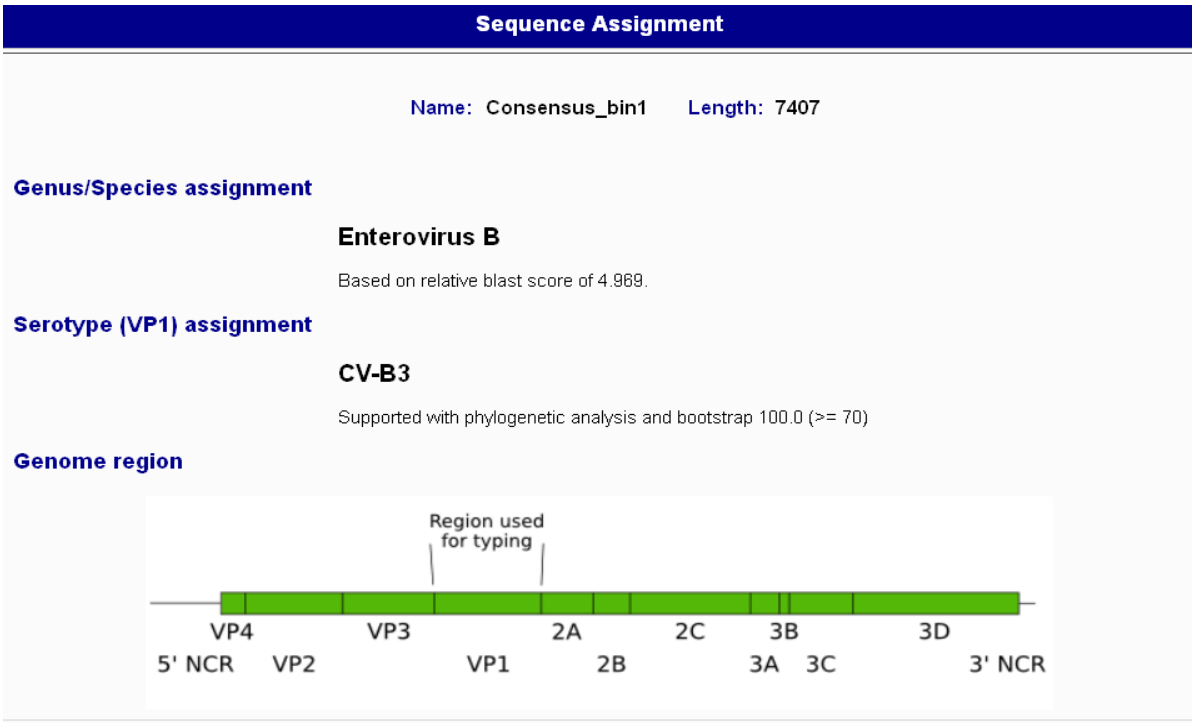


Obr. 6 *Výstup prohlížeče Tablet. Zde jsou vidět jednotlivá čtení pod sebou, která vytváří kontig.*

3.5. Enterovirus genotyping tool

Tento webový nástroj využívá fylogenetické metody k tomu, aby identifikoval v nukleotidové sekvenci enterovirový genotyp. Je to volně přístupný nástroj pro vyhledává společných sekvencí s enteroviry. Tento nástroj je založený na algoritmu BLAST v databázi GenBank, kde je vždy žádaná sekvence proti řadě referenčních sekvencí virů z rodiny Picornaviridae (Kroneman et al., 2011).

Pomocí tohoto nástroji jsme zjistili, zdali námi sestavené sekvence odpovídají některým enterovirovým a jak dalece se tyto sekvence překrývají. Možným výsledkem je jak grafické zobrazení sekvencí shodných s referenčními sekvencemi, jak se zobrazeno na obrázku 7, tak také excelová tabulka, která nám dává informace o tom, s jakými enteroviry je náš genom (či části genomu) nejpodobnější.



Obr. 7 *Záznam vyhledávání vzorku 1 v nástroji genotyping enterovirus tool. Zde je vidět, že námi celoosekvenovaný genom je téměř totožný s coxsackievirem B3.*

3.6. Sestavení virové sekvence

Novou enterovirou sekvenci (ve většině případů části enterovirové sekvence) jsme porovnali s vhodnou referenční sekvencí. Jednotlivé kontigy byly sestaveny pod referenční sekvenci pomocí softwaru Sequencher (Sequencher® version 5.2 sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA http://www.genecodes.com). Soubory s namapovanými kontigy jsme si importovali do Bioedit, kde jsme následně identifikovali mezery a jejich počet mezi kontigy. Ty jsme poté pomocí Sangerova sekvenování doplnili. Celý proces doplňování mezer je popsán v následující kapitole.

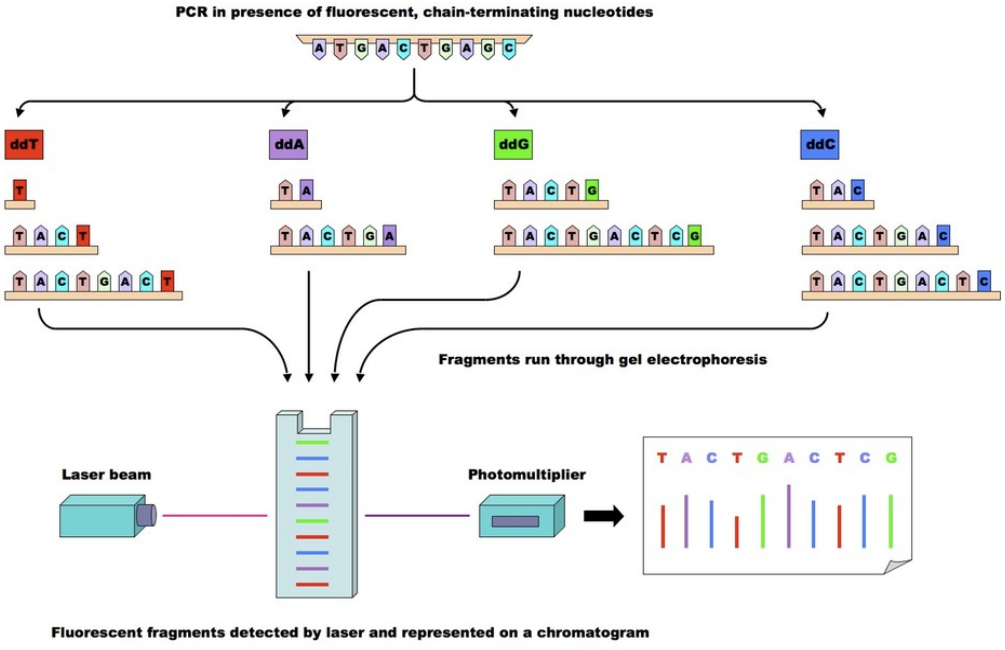
3.7. Sangerovo sekvenování

Metoda Sangerova sekvenování je jednou z nejvlivnějších inovací v biologickém výzkumu od doby, kdy byla poprvé představena v roce 1977 (Sanger, Nicklen, & Coulson, 1977). Tato metoda se také označuje jako dideoxy metoda sekvenování.

Principem Sangerovy metody je terminace syntézy nových DNA řetězců prostřednictvím náhodné inkorporace modifikovaných nukleotidů, tzv. dideoxynukleotidů(ddNTPs), separace těchto řetězců a jejich vizualizace. Nejprve je zkoumaný úsek DNA namnožen pomocí

polymerázové řetězové reakce na vysoký počet kopií (řádově miliardy z jedné molekuly). Poté replikační enzym DNA polymeráza syntetizuje komplementární vlákno k sekvenované DNA. Syntéza probíhá obdobně jako při replikaci ve směru 5'→3'. K syntéze využívá DNA polymeráza primer zhruba 20 bází dlouhý, komplementární ke konkrétní části templátu, a také jednotlivé deoxynukleotidy (dNTPs), které zařazuje na základě principu komplementarity bází do nově vznikajícího řetězce. V původním Sangerově provedení s radioaktivně značeným dATP (značení prostřednictvím 32P) je výše zmíněná reakční směs rozdělena do čtyř zkumavek označených G, A, T,C a do každé z nich je přidán i příslušný ddNTP a to v koncentraci řádově nižší (specificky dle konkrétní použité DNA polymerázy), než je koncentrace ostatních dNTP. Do zkumavky označené G je tedy přidán ddGTP, do zkumavky A je přidán ddATP atd (Martínek, Stehlík, Grossmann, Ka, & Vaněček, 2013).

V dnešní podobě tohoto sekvenování se využívají fluorescenčně značené ddNTP, které jsou smíchány s neznačenými, neterminujícími nukleotidy v jednokolové sekvenační reakci . Kapilární elektroforéza potom sekvence na základě jejich délky analyzuje a provede následující vyhodnocení terminačních bází (Pettersson, Lundeberg, & Ahmadian, 2009).



Obr. 8 *Princip Sangerova sekvenování a následné analýzy na automatickém sekvenátoru. Převzato z http://www.vce.bioninja.com.au/aos-3-heredity/molecular-biology-technique/sequencing.html.*

3.7.1. Laboratorní přístroje a pomůcky

Přístroje	
Vortex ZX3	P-Lab; Česká republika
Centrifuga B4i/BR4i JOUAN	Trigon-plus; Francie
Mikrocentrifuga Mini spin plus	Eppendorf; Německo
Cycler	Eppendorf; Německo
Labcycler	
17G centrifuga, Universal 32	Hettick
Biomek 3000	Beckman Coulter
Pipety	Finnpipette; Finsko
	Eppendorf; Německo

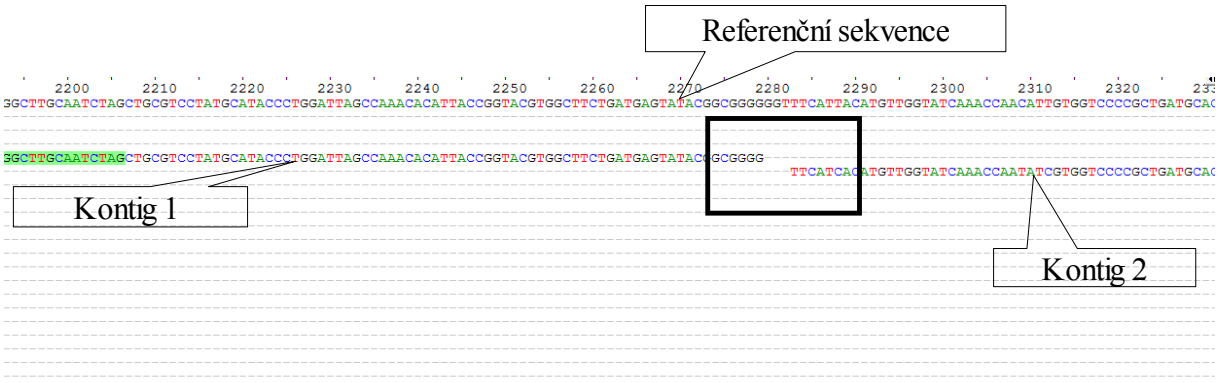
Spotřební materiál	
96-jamkové destičky pro PCR	Applied Biosystem;USA
Krycí fólie pro PCR	Thermo Scientific
Pipetovací špičky 10µl, 100µl, 200µl, 1000µl	Eppendorf; Německo

3.7.2. Návrh primerů

Primery navrhujeme na úseky, kde potřebujeme dosekvenovat mezery mezi jednotlivými kontigy, které na sebe nenavazují. Primery jsou dlouhé cca 18-24 bází, jejich teplota tání se pohybuje mezi 50-55°C a obsah GC párů je přibližně 50%. Potřebujeme zjistit, zdali se nám primery nespecificky nenasedají na jiné než enterovirové sekvence. Toho jsme docílili otestováním sekvencí primerů v databázi GenBank pomocí nástroje NCBI Blastu. Pokud byl výsledek takový, že sekvence primerů jsou unikátní pro enteroviry a splňují požadavky popsané výše, mohli jsme je použít.

Pro výpočet teploty tání jsme použili metodu zahrnující do svého výpočtu koncentraci soli přítomné v reakci [$MgCl_2$], obsah G-C párů a délku primeru (N). Výpočet platí pro primery délky 14 – 70 nukleotidů. Vzorec je následující (Stephenson, 2010):

$$Tm\ (°C) = 81{,}5 + 16{,}6\ (\log\ [MgCl_2^+]) + 0{,}41\ (\%G+C) – (500/N)$$



Obr. 9 *Mezera, kterou jsme pomocí primerů překlenuli.*

3.7.3. PCR reakce

Polymerázová řetězová reakce (PCR – z anglického Polymerase Chain Reaction) byla vynalezena roku 1983 (Mullis & Faloona, 1987) Kary Mullisem. Princip PCR metody je založen na specifickém pomnožení konkrétního hledaného úseku DNA, který je ve výchozím materiál.

Chemikálie

PCR voda	Braun; Německo
pufr 10x konc.	Applied Biosystem; USA
MgCl2 (25mM)	Applied Biosystem; USA
dNTP	Sigma; USA
AmpliTaq Gold polymeráza 5U/µl	Applied Biosystem; USA

Postup

Pro přípravu reakční směsi pro PCR reakci jsme byly smíchány následující chemikálie:

	µl na vzorek
PCR voda	1,55
pufr 10x konc.	1
MgCl2 (25mM)	1
dNTP (0,5 mM konc.)	0,4
primer forward (2,5 µM)	2,5
primer reverse (2,5 µM)	2,5
AmpliTaq Gold polymeráza 5U/µl	0,05
DNA	1
Celkem objem	10

Jako templát jsme použili naředěnou cDNA vzniklou náhodnou reverzní transkripcí RNA získané z příslušného vzorku.

Připravili jsme si reakční směs pro PCR pro všechny vzorky společně do 7 ml zkumavky, počítali jsme i s pipetovací chybou, tudíž jsme vždy připočítali 15% objemu každé reagencie.

Amplifikovali jsme 27 úseků, překlenující vnitřní mezery, každý úsek v duplikátu (tedy 46 reakcí). Stejný postup jsme opakovali i při amplifikování úseků v oblastech 5´ a 3´ konců.

Dále jsme reakční směs s cDNA zvortexovali, krátce centrifugovali a vložili do cycleru, který běžel s následujícím teplotním programem:

95°C	10 min (počáteční denaturace)	
96°C	15 s (denaturace)	10x
61°C	30 s (anelace primeru)	
72°C	1 min (syntéza)	
96°C	15 s (denaturace)	35x
56°C	30 s (anelace primeru)	
72°C	1min (syntéza)	
72°C	10 min (elongace)	
4°C	stále	

3.7.4. Elektroforéza

2,5% agaróza	Lonza; USA
0,5x TBE (Tris báze, k. boritá, EDTA 05, M, pH 8, dest. H2O)	
Velikostní marker 2-Log DNA Ladder	Biolabs; USA
GelRed	Biotium; USA
Nanášecí barvivo (bromfenolová modř + xylencyanol + 40% sacharoza)	

Produkty PCR reakce byly vizualizovány na gelové elektroforéze. Principem metody je pohyb záporně nabitých molekul DNA v elektrickém poli, kdy putují směrem k anodě. Rychlost pohybu je nepřímo úměrná velkosti fragmentů DNA.

Při gelové elektroforéze jsme používali 2,5% agarózu, TBE, nanášecí barvivo (bromfenolová modř + xylencyanol + 40% sacharoza) a jako měřítko velikosti produktu nám sloužil velikostní marker (hmotnostní standard, DNA ladder).

Postupovali jsem tím způsobem, že jsme produkt PCR smíchali 1:1 s nanášecím barvivem. Do jamky gelu jsme pak pipetovali 9 mikrolitrů směsi s barvivem. Zbytek PCR produktu jsme si uschovali a pokud se zde nacházel produkt, poté jsme vzorek podrobili Sangerovu sekvenování.

Elektroforéza probíhala 20 minut při 8 V/cm. Gel byl zdokumentován fotografováním na UV transiluminátoru. Obrázky elektroforézy jsou zobrazeny v kapitole Výsledky na straně 61 a 63.

3.7.5. Přečistění před sekvenační reakcí

Přečištění produktů před sekvenační reakcí je nezbytné k tomu, aby se ze vzorku odstranily všechny chemikálie, které zde byly jak po PCR, tak po elektroforéze, především zbylé primery a neinkorporované nukleotidy.

Chemikálie

destilovaná voda	
voda	
Ampure	Agentcourt; Beckman Coulter; USA
Absolutní etanol	Penta; Česká republika
Injekční voda	Braun; Německo

Postup

	příprava 96 jamkové desky
Ampure	2,8 ml
75% etanol	40 ml
injekční voda	8-10 ml

Čištění probíhá na přístroji Biomek 3000, který zvládne veškerou mechanickou činnost a posloupnost pipetování ve správném pořadí. Pro přečištění jsou nezbytné tři chemikálie: Ampure, etanol (absolutní etanol ředěný injekční vodou) a injekční voda. Aby se tyto

chemikálie nemísily mezi sebou, je zapotřebí omývání špiček. To je zajištěno promýváním špiček ve dvou velkých 96 jamkových deskách s vodou (do jedné patří voda destilovaná a do druhé voda odpadní). Dále k vybavení přístroje Biomek patří magnetická destička. Princip čištění spočívá v tom, že Ampure obsahuje magnetické kuličky, které na sebe navážou DNA. Když se pak deska se vzorky dá do magnetické destičky, kuličky s navázanou DNA se přichytí na stěnu jamky a dojde k odmytí všech nepotřebných chemikálií.

3.7.6. Sekvenační reakce

Chemikálie

injekční voda	Braun; Německo
BigDye Terminator sekvenační pufr	Applied Biosystem; USA
sekvenační směs Big Dye Terminator v3.1	Applied Biosystem; USA

Postup

V jedné sekvenační reakci byly smíchány následující chemikálie:

	µl na 1 vzorek, 12,5% konc.
injekční voda	3,38
BigDye Terminator sekvenační pufr	1,88
sekvenační směs BigDye Terminator	0,25
sekvenační primer konc. 2,5 µM	2,5
PCR produkt	2
celkem	10

Sekvenační reakce běžela v cycleru s následujícím teplotním programem:

96°C	1 min	
96°C	10 s	25x
55°C	5 s	
60°C	4 min	
72°C	7 min	
15°C	stále	

3.7.7. Přečištění po sekvenační reakci

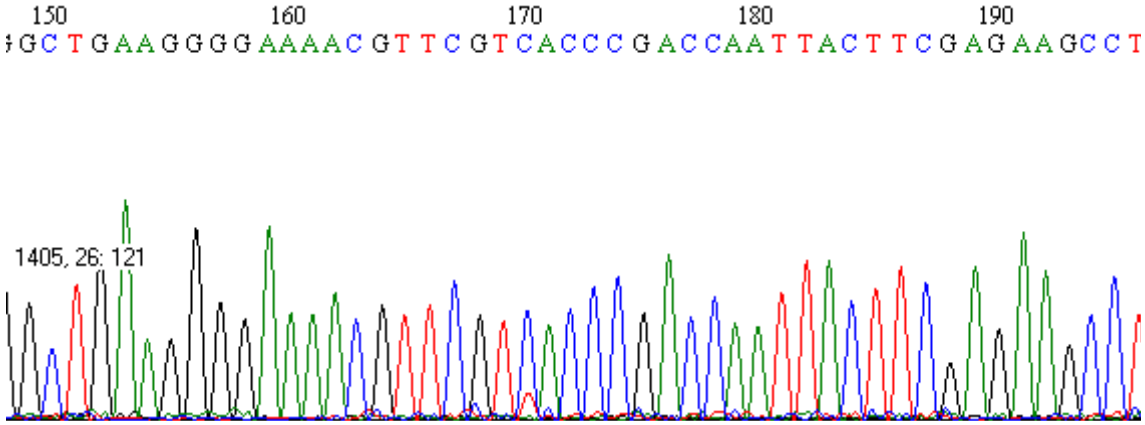
Po sekvenační reakci je stejně jako po PCR nutné produkty přečistit od zbylých chemikálií.

Cleanseq	Agentcourt, Beckman Coulter; USA
Absolutní etanol	Penta; Česká republika
injekční voda	Braun; Německo
EDTA	Promega; USA

Přečišťuje se opět na stroji Biomek 3000, za použití chemikálie Cleanseq , dále je potřeba 80% etanol a nakonec je potřeba 0,05M roztoku EDTA.

3.7.8. Analýza dat na automatickém sekvenátoru

Produkty cyklického sekvenování jsou v kapilárním sekvenátoru elektroforeticky rozděleny podle délky a detekovány pomocí laserového detektoru, který je napojen na počítač. Elektroforéza probíhá v tenké kapiláře naplněné gelem. Kapilárou procházejí různě dlouhé PCR produkty v závislosti na jejich velikosti. Laser zachycuje fluorescenci markerů navázaných na ddNTP navázaných na jednotlivé PCR produkty a následně dochází ke stanovení pořadí nukleotidů. Když jsou fluorescenční markery excitovány laserem, jsou detekovány 4 různé produkty (tedy 4 druhy nukleotidů) a intenzita fluorescence je přeložena do tzv. datového píku.



Obr. 10 *Záznam elektroforeogramu z automatického sekvenátoru.*

Data z automatického sekvenátoru s příponou ABI můžeme zobrazit v programu Bioedit. Vidíme, jakou kvalitu mají jednotlivé nukleotidy a to dle rozložení a výšky píků na záznamu ze sekvenátoru. Podle toho pak můžeme upravit sekvence ze Sangerova sekvenování tak,

abychom si byli jisti, že se skutečně jedná o to dané a správné pořadí nukleotidů, tzn., vymažeme části se špatným či překrývajícím se signálem.

3.7.9. Sekvenování 3´ a 5´ konců

Podobně jako jsme sekvenovali vnitřní mezery mezi kontigy, jsme postupovali i v případě sekvenování 3´a 5´konců. Jediný rozdíl byl v tom, že jsme si v případě 3´ konců museli forward primer navrhnout podle referenční/referenčních sekvencí, reverse primer jsme pak navrhovali dle námi osekvenavané sekvence. Stejně tak v případě 5´ konců jsme si reverse primer navrhli též podle referenční sekvence a forward primer podle naší sekvence. Nebylo to však zapotřebí u všech vzorků: 3´ konec jsme dosekvenovávali u 7 vzorků z 9 a 5´ konec u 6 vzorků z 9.

Provedli jsme výběr referenčních sekvencí, podle kterých jsme navrhovali primery. V tabulce 3 jsou sekvence, podle kterých jsme navrhovali 5´ konce a v tabulce 4 jsou sekvence, podle kterých jsme navrhovali 3´ konce.

Tab. 3 *Přehled sekvencí, podle kterých jsme navrhovali reverse primery na dosekvenování 5´konce.*

druh		referenční číslo
Enterovirus A	Human coxsackievirus A2	AY421760.1
	Human coxsackievirus A3	AY421761.1
	Human coxsackievirus A8	AY421766.1
	Human coxsackievirus A10	AY421767.1
Enterovirus B	Human coxsackievirus B1	M16560
	Human echovirus 26	AY302557
	Human coxsackievirus B4	AY302550
Enterovirus C	Human coxsackievirus A1	AF499635
	Human coxsackievirus A11	AF499636
	Poliovirus 1	KF537633
	Poliovirus 2	JX275380
	Human coxsackievirus A20	DQ358078.1
Enterovirus D	Human enterovirus D	NC001430
	Enterovirus 68	AY426531
Enterovirus E	Bovine enterovirus	NC001859

Tab. 4 *Přehled sekvencí, podle kterých jsme navrhovali forward primery při dosekvenovávání 3´ konce.*

druh		referenční číslo
Enterovirus B	Human coxsackievirus B1	M16560
Enterovirus C	Poliovirus 1	KF537633

3.8. Fylogenetická analýza

Pokrok v sekvenačních technologiích a záplava sekvenčních dat poskytuje mnoho příležitostí ke studiu evoluce genů a proteinových rodin, společně s fylogenetickými vztahy mezi druhy. Každá rodina homologních sekvencí poskytuje mnoho znaků, které jsou potenciálními zdroji cenných informací pro vytvoření fylogenetického stromu.
Fylogenetické analýze předchází řada kroků: upravení výsledné konzenzuální sekvence genomů jednotlivých vzorků v programu Bioedit, vybrání vhodných refrenčních sekvencí k tvorbě multiple alignment, následný import sekvencí do programu Mega5, který slouží ke konstrukci fylogenetických stromů. V tomto programu jsme ověřili vhodnost sekvencí k vytvoření důvěryhodného fylogenetického stromu pomocí eliminování duplikátních sekvencí a výpočtem distančních vzdáleností mezi sekvencemi.

3.8.1. Bioedit

Bioedit je biologický počítačový nástroj určený pro práci se sekvencemi. Poskytuje základní funkce pro úpravy proteinových a nukleotidových sekvencí, uspořádání sekvencí (tzv. alignment), umožňuje manipulace a analýzy sekvencí. Bioedit není nejvýkonnější sekvenční analyzační program, ale nabízí nám rychlé a snadné funkce pro editování a anotování sekvencí. Bioedit je zároveň propojen s některými externími sekvenčními analyzačními programy. V našem případě bylo nejdůležitější propojení s databází GenBank.

Do tohoto programu jsme si importovali sekvence sestavené do kontigů, které prošly procesem úpravy, jak je popsáno výše v této práci. Tyto sekvence sestavené pod jednou referenční sekvencí, která sloužila jako lešení, netvořily jeden nepřerušovaný virový genom. Proto jsme mezery překlenuli pomocí Sangerova sekvenování.

Data ze Sangerova sekvenování jsme si také importovali do Bioeditu, abychom celý genom složili. Sestavování fragmentů v Bioeditu lze dělat pouze manuálně. Musíme přesně znát místa, podle kterých jsme si navrhovali primery, a tam poté dosazujeme fragmenty, které byly osekvenovány.

Po sestavení celého virového genomu bylo zapotřebí sekvence translatovat, abychom se přesvědčili, zda se zde vyskytuje jeden smysluplný čtecí rámec (jelikož pikornaviry mají jeden čtecí rámec), a zda jsou zde nějaké delece či inzerce. Při nalézání správného čtecího rámce jsme se opět opírali o sekvenci referenční, jelikož v té je jeden nepřerušovaný čtecího rámec.

Po nalezení správného čtecího rámce jsme genom anotovali, jelikož bez toho nelze provést fylogenetickou analýzu jednotlivých proteinů. Anotaci jsme provedli v Bioeditu, který je propojen a databází GenBank. V této databázi jsme našli nejbližší referenční sekvenci k našim sestaveným genomům. Tu jsme anotovali jako první. Jednotlivé proteinové sekvence jsme si označili jinou barvou pro přehlednost. Tato funkce je vhodná pro rychlou orientaci mezi jednotlivými sekvencemi proteinů. Poté jsme si podle referenční sekvence anotovali naši sekvenci, jelikož jsou téměř shodné.

3.8.2. Vybrané referenční sekvence pro multiple alignment

Důležitým krokem při plánování konstrukce fylogenetických stromů je zcela jistě výběr referenčních sekvencí. My jsme vybrali 40 referenčních sekvencí, tak aby u každé enterovirové skupiny bylo několik zástupců. Vybírali jsme opět z databáze GenBank, kritériem výběru byly celoosekvenované genomy. Vybrané sekvence společně s jejich referenčními čísly jsou v tabulce 12.

Všechny tyto sekvence jsme si nahráli do programu Bioedit. Anotovali jsme je všechny stejným způsobem jako jsme to dělali v případě referenční sekvence, která sloužila jako lešení pro sestavení námi osekvenovaného genomu. Stejně jsme si i barevně rozlišili všechny proteiny v sekvencích.
Fylogenetické analýze jsme podrobili pouze vybrané proteiny, a ne celé genomy. Proto jsme z hromadného alignmentu referenčních sekvencí vybrali úseky, které v dané sekvenci

odpovídají vybranému proteinu. Pro naši fylogenetickou analýzu jsme vybrali proteiny: VP1, 2A a 2C.

3.8.3. Mega5

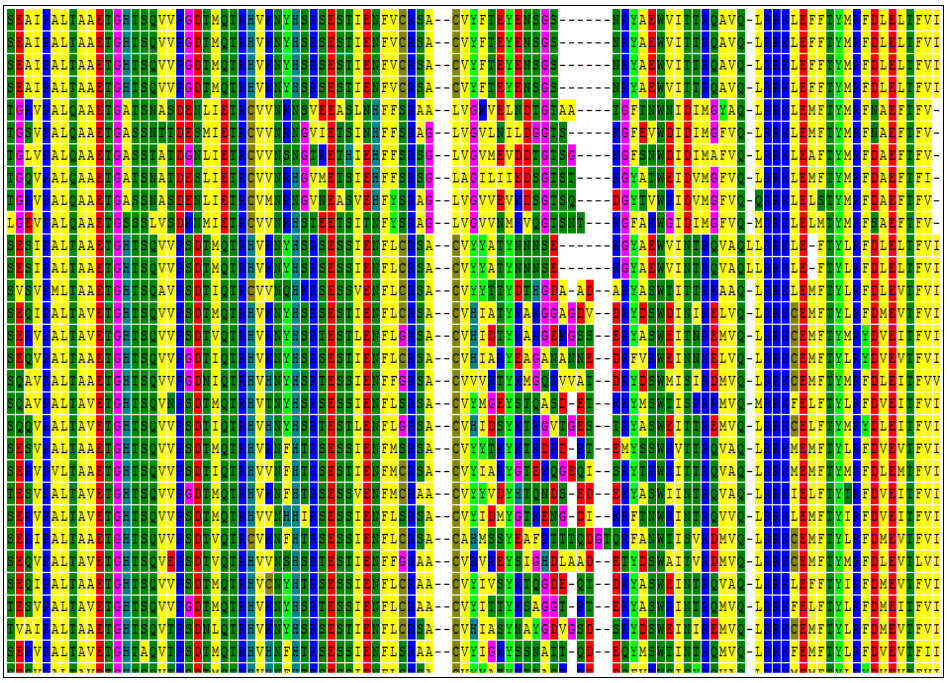
Mega 5 (Molecular Evolutionary Genetics Analysis version 5), je software určený ke srovnávací analýze molekulárních sekvenčních dat (Tamura et al., 2011). Při konstrukci fylogenetických stromů máme na výběr z celé řady rekonstrukčních metod. Jednotlivé metody jsou ve své podstatě matematickými algoritmy, které sestrojí fylogenetický strom. Podrobný popis těchto matematických modelů/algoritmů není cílem této práce. Níže je stručný popis metod, který je nezbytný k pochopení biologických výsledků.

Mega 5 nám dovoluje nejen vytvářet fylogenetické stromy, ale jsme v něm schopni provádět i hromadné uspořádání sekvencí (multiple sequence alignment) jak pomocí ClustalW tak pomocí MUSCLE. Neznámé sekvence jsou nejprve spolu s referenčními enterovirovými kmeny upraveny do „multiple sequence alignment“ a poté fylogeneticky analyzovány a přiřazeny do skupiny k nejpříbuznějšímu referenčnímu kmeni, který nám tímto určuje druh nalezeného enteroviru. Správné uspořádání sekvencí je jedním z nejdůležitějších kroků ve fylogenetické analýze.

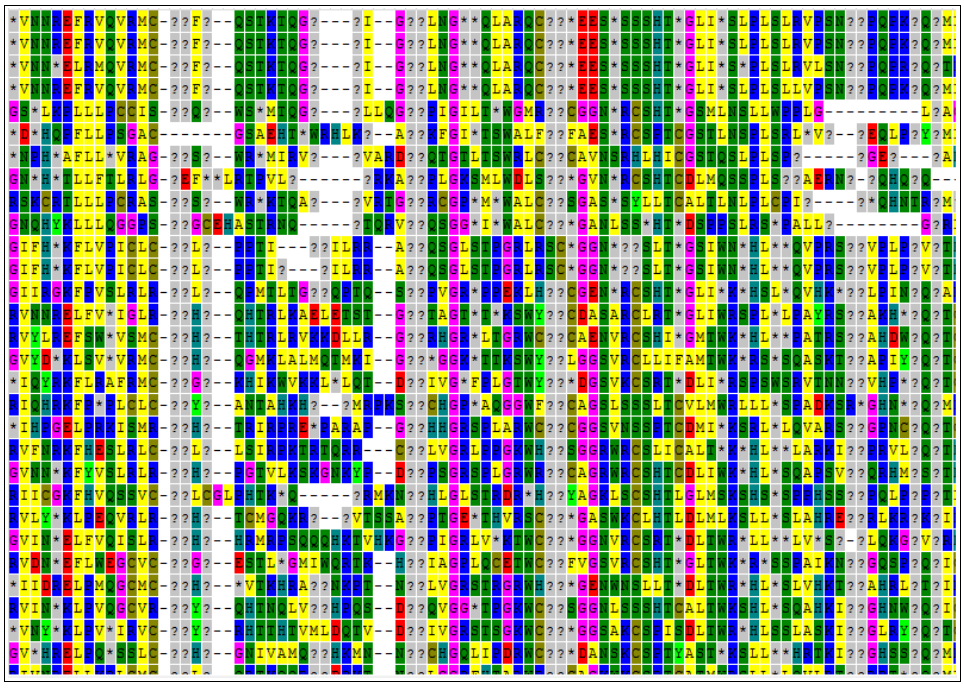
3.8.3.1. Alignment sekvencí

MEGA nabízí 2 metody pro řazení nukleotidových či aminokyselinových sekvencí pod sebe, tzn. vytváření alignmentu: ClustalW (Thompson, Higgins, & Gibson, 1994) a MUSCLE (Multiple Sequence Comparison by Log-Expectation) (Edgar, 2004). Ačkoli je ClustalW obecně používanější, MUSCLE je o něco přesnější (Nuin, Wang, & Tillier, 2006) a je 2-5 krát rychlejší v případě přiměřené velikosti dat. Hlavní výhodou MUSCLE je zvládnutí zpracování velkého množství dat. Například při zpracování 5000 sekvencí o průměrné délce 350 nukleotidů byl MUSCLE 80 000krát rychlejší než ClustalW (Edgar, 2004).

Sekvence, které kódující proteiny je dobré srovnávat na základě sekvence aminokyselin, abychom dosáhli správného složení a zachování čtecího rámce (Hall, 2005). Výstup po hromadném sestavení sekvencí založeném na srovnání kodónů je vidět na obrázku 11. Pokud by byly použity sekvence nukleotidové, algoritmus by zanášel do alignmentu velké množství arteficiálních substitucí, včetně stop kodónů. To je zobrazeno na obrázku 12.



Obr. 11 *Sestavení „multiple sequence alignment“ na základě kodónů. Na tomto obrázku je zřetelné, že se nám v alignmentu již nevyskytují žádné stop kodóny. Proto jsme s takto uspořádanými sekvencemi prováděli další analýzu.*



Obr. 12 *Sestavení „multiple sequence alignment“ na základě nukleotidů. Na obrázku je vidět, že když využije uspořádání sekvencí na základě nukleotidů, tak po translaci těchto sekvencí se v nich vyskytuje řada stop kodónů.*

3.8.3.2. Eliminování duplikátních sekvencí

Ačkoli se snažíme vyhnout zahrnutí dvou stejných sekvencí, může se stát, že máme mezi vybranými sekvencemi dvě stejné. Abychom to byli schopni odhalit, použili jsme funkci pairwise distances, neboli srovnání všech dvojic sekvencí, kdy se srovnávají všechny se všemi.

3.8.3.3. Výpočet distancí

Evoluční vzdálenosti dovoluje odhadnout míru evoluční rozdílnosti sekvencí, která je reprezentována fylogramem. Tato definice evoluční vzdálenosti podněcuje vývoj evolučních modelů poskytujících způsoby odhadu evolučních vztahů a vytváření teorií o vývoji molekulárních sekvencí, spojujíc fylogenetiku s evoluční biologií *(Nei & Kumar, 2000)*. Tyto metody vycházejí z matice distancí, která udává vzájemné vzdálenosti mezi všemi dvojicemi taxonomických jednotek, pro které konstruujeme fylogenetický strom.

3.8.3.3.1. Distanční metody

Distance, nebo také vzdálenost, je v těchto metodách vyjádřena jako frakce míst, které se liší mezi dvěma sekvencemi v hromadném alignmentu. Je zřejmé, že když se dvojice sekvencí liší v 10% jejich míst, jsou si více příbuzné, než kdyby se lišily ve 30%. Také dává smysl, že čím dále v čase jsou sekvence od svého evolučního předka, tím se od sebe více odlišují. Toto však ve fylogenetice neplatí zcela stoprocentně. Může nastat případ, kdy se jedna linie vyvíjí rychleji než druhá nebo se dvě linie vyvíjejí stejně rychle, avšak liší se nikoli časovou vzdálenosti, ale mnohočetnými substitučními událostmi. Každá nukleotidová substituce zvyšuje početní rozdílnost mezi těmito liniemi a tedy je oddaluje od společného předka.

Existují dvě nejpopulárnější distanční metody UPGMA (Unweighted Pair-Group Method with Arithmetic Mean) a Neighbor Joining, jedná se o algoritmické metody (tzn. využívají řadu výpočtů k odhadu fylogenetického stromu). Výpočty zahrnují manipulace se vzdálenostními maticemi, které jsou odvozené od multiple alignmentu. Z tohoto alignmentu oba programy vypočítají pro každý pár sekvencí vzdálenost taxonu, nebo míru rozdílnosti a zaznamenají ji do vzdálenostní matice (Hall, 2011).

3.8.3.3.2. P - distance

Délky vnitřních větví mají indikovat míru genetické změny mezi evolučním předkem a jeho evolučním potomkem. Z toho bychom usuzovali, že způsob výpočtu délky vnitřních větví jsou založeny na počtu rozdílů v sekvencích. Obvykle však vyjadřujeme délku vnitřních větví jako proporci, raději než počet míst, u kterých došlo ke změně. S tímto poměrem rozdílnosti mezi sekvencemi pracuje model p-distance (Hall, 2011).

Tato vzdálenost je poměrem (p) nukleotidových míst, v kterých jsou dvě srovnávané sekvence rozdílné. Je toho dosaženo podělením počtu nukleotidových rozdílů celkovým počtem nukleotidů sekvencí. Nedochází zde k žádným opravám v místech s hromadnými substitucemi, substituční mírou chyb (například rozdíly v tranzicích a transverzích) nebo rozdíly v evoluční rychlosti mezi místy .

3.8.3.3.3. Jukes Cantor model

V Jukes Cantorově metodě (Jukes & Cantor, 1969) je frekvence nukleotidové substituce stejná pro všechny čtyři nukleotidy. Tento model produkuje maximální pravděpodobnostní odhad pro počet nukleotidových substitucí mezi dvěma sekvencemi. Předpokládá rovnost substituční rychlosti mezi místy, stejnou nukleotidovou frekvenci a neopravuje vyšší míru mutací typu tranzicí (mutace, kdy dochází k záměně jednoho nukleotidu za jiný. V případě tranzice jde o záměnu purinové báze za purinovou nebo pyrimidinové za pyrimidinovou) porovnání se substitucemi typu transverze (záměna purinové báze za pyrimidinovou nebo pyrimidinové za purinovou).

3.8.4. Výpočet fylogenetického stromu (topologie, délka větví)

Výpočet fylogenetického stromu je velmi obtížnou úlohou, z hlediska relevantnosti výsledků a jejich interpretace. Při vytváření fylogenetických stromů neexistuje jen jedno úplně správné řešení, snažíme se spíše vyjádřit tu nejpravděpodobnější verzi. My jsme pro výpočet fylogenetického stromu využili metodu Neighbor Joining.

3.8.4.1. Neighbor Joining metoda

NJ (Neighbor-Joining) metoda (Saitou & Nei, 1987) pracuje s maticemi vzdáleností, tedy na základě distančních dat. Z řady těchto matic sestaví fylogenetický strom. Přímo počítá vzdálenosti od interních uzlů stromu.

Na začátku se vytvoří jeden hvězdicový strom, kde je jeden vnitřní vrchol, a všechny řešené taxonomické jednotky jsou reprezentovány pomocí listů. Tento strom se postupně rozkládá shlukováním nejbližších taxonomických jednotek tak, aby se v každém kroku co možná nejvíce zmenšila celková délka stromu.

NJ první počítá pro každý taxon jeho vlastní síť rozdílnosti od všech ostatních taxonů. A poté přepočítá souhrn jednotlivých vzdáleností od tohoto taxonu. To je potom využito k novému výpočtu správné vzdálenostní matice. NJ potom najde dvojici taxonů s nejnižší vzdáleností, jejíž spojení nejvíce zmenší délku stromu, tj. součet délek všech jeho větví, a vytvoří vnitřní uzel, od něhož se oba taxonu oddělují. Vzdálenost dvou taxonů od vnitřního uzlu nemusí být identická. NJ nepředpokládá, že jsou všechny taxony stejně vzdáleny od kořene stromu.

3.8.5. Stanovení spolehlivosti topologie větví (bootstrap)

Nejčastěji používanou metodou při stanovování spolehlivosti topologie větví je bootstrapping. Jeho princip poprvé popsal Bradly Efron, profesor Stanfordské univerzity, v roce 1979. Jednalo se tehdy o jednu z prvních metod, která ve statistice nahrazovala tradiční algebraické výpočty počítačovými simulacemi na pozorovaných datech. Bootstrap přinesl možnost odhadnout přesnost libovolného odhadu libovolného parametru. Přitom spočívá v prosté myšlence mnohonásobného opakování jednoduchého algoritmu.
Při stanovování spolehlivosti topologie větví při fylogenetické analýze je ke každému uzlu původního kladogramu uvedeno příslušné procento, tzv. bootstrapping hodnota. Jestliže se pro určitý uzel tato hodnota blíží 100, má existence tohoto uzlu velmi silnou podporu ve výchozích datech, jestliže je naopak nízká, je nízká i podpora existence daného uzlu (Soltis & Soltis, 2003).

Bootstrapping hodnotu nelze v žádném případě interpretovat jako pravděpodobnost existence daného větvení vývojové linie. Je to vždy pouze vyjádření stupně podpory pro existenci daného větvení v našich výchozích datech. Jestliže je tato hodnota nízká, je to třeba chápat jako signál, že pro spolehlivou rekonstrukci kladogeneze je třeba získat více vstupních dat. Nízká hodnota bývá většinou způsobena malým množstvím výchozích dat nebo absencí fylogenetického signálu ve vstupních datech. Druhý případ nastává tehdy, jestliže v daném úseku proběhlo příliš velké množství evolučních změn takže dané sekvence jsou již tzv. substitučně nasycené. Jinou možnou příčinou nízkých bootstrapping hodnot může být

existence konfliktu mezi dvěma protichůdnými signály, například jestliže daný úsek genu vznikl fúzí dvou různých genů s různou evoluční historií. Bootstrapping hodnoty nezávisí pouze na síle fylogenetického signálu, ale také na množství druhů zahrnutých do analýzy. Z tohoto důvodu není možné jednoduše porovnávat na základě bootstrapping hodnot spolehlivost dvou stromů obsahujících jiné počty druhů (Efron, Halloran, & Holmes, 1996).

4. Výsledky

Celkem jsme testovali 22 vzorků, jež byly pozitivní na enterovirovou RNA. Prvních deset vzorků obsahovalo enteroviry, které byly pěstovány na buněčných kulturách. Ostatní vzorky, tedy 11 až 22, pocházejí ze stolic. Všechny vzorky byly získány od dětí geneticky predisponovaných k diabetu 1. typu.

Hlavní náplní této diplomové práce je bioinformatická analýza těchto dat, následné sestavení virového genomu de novo a u vzorků 1 až 9 doplnění mezer v genomech, jež nebyly pokryty těmito daty, Sangerovým sekvenováním. Celé virové genomy jsme získali u vzorků 1 až 9, u vzorků 10 až 22 byla zpracována data z NGS, ale nedoplňovali jsme je Sangerovým sekvenováním.

4.1. Bioinformatická analýza dat z NGS

Po osekvenování fragmentů virů metodou NGS jsme získali data, u kterých jsme zkontrolovali, zdali jsou vhodná k použití pro sestavení do kontigů. Data jsme kontrolovali v programu Galaxy, který nám přehledně umožňuje získat informace o počtu vstupních dat, jejich kvalitě, rozložení jednotlivých bází v sekvencích apod. Pomocí tohoto nástroje jsme byli schopni identifikovat oblasti sekvencí, které bylo potřeba odstranit kvůli jejich špatné kvalitě, která by se následně projevila při špatném sestavení do kontigů a pozdějším namapování na enterovirový genom. Jelikož se jedná o vzorky, které nepocházejí ze stejných zdrojů (vzorky 1-10 byly získány z buněčných kultur, zatímco virová RNA vzorků 11-22 byla izolována přímo ze vzorků stolic), bylo potřeba s daty pracovat odlišně, minimálně v první fázi procesu, který se týká filtrace a úpravy dat.

4.1.1. Filtrace a úprava sekvencí

Úpravu a filtraci jsme prováděli na vzorcích 11 – 22, data prvních deseti vzorků byla dobré kvality.

Úpravu jsme prováděli na základě informací o vstupních sekvencích. Tyto informace jsme získali na základě grafů. Ty jsou zobrazeny na obrázcích 13 a 14. Získali jsme informace o délce sekvencí, která byla maximálně 250. Rozložení nukleotidů v sekvencích, které je znázorněno na obrázku 13, nebylo rovnoměrné v celé délce sekvencí. Prvních 30 bází má značně vychýlené hodnoty. Jedná se o tagy, neboli identifikátory sloužící k vzájemnému

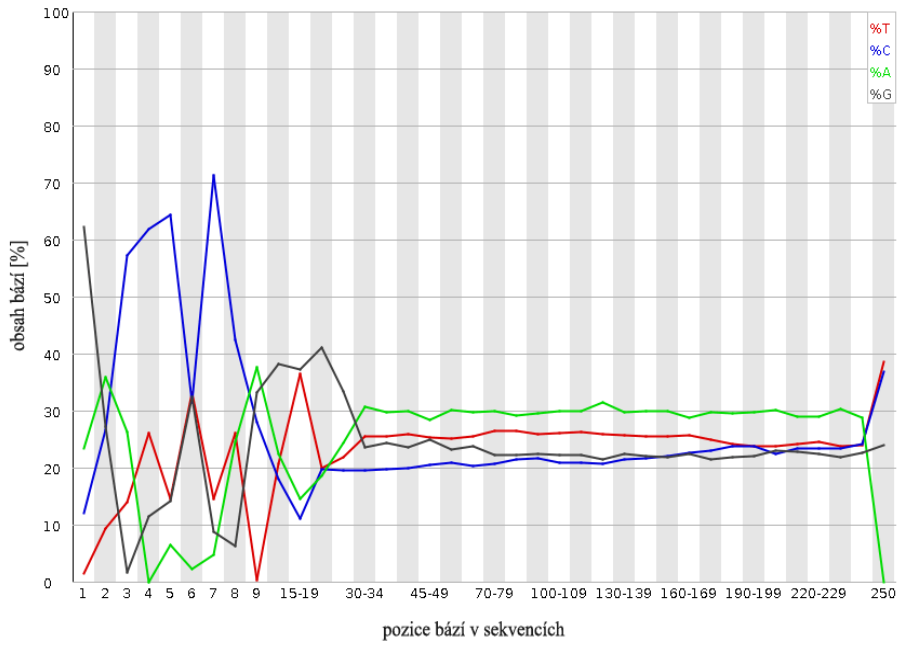
rozlišení současně sekvenovaných vzorků. Pro vyrovnání hodnot obsahu bází na 5´konci jsme jich 30 odstranili. Taktéž byl nestejnoměrný obsah bází na 3´ konci, proto jsme odstranili 25 bází z tohoto konce. Po úpravách konců se nám změnila délka sekvencí, jak je vidět na obrázku 13B, kdy z maximální délky 250 bází klesla na 194 bází.

Nejprve bylo aplikováno zkrácení čtení podle kvality. Klouzavé okno, které čítá vždy deset nukleotidů, postupuje po sekvenci a pokud není průměrná hodnota kvality v tomto okně vyšší než 20 odstraní poslední 2 nukleotidy. Klouzavé okno postupuje z 3´ i 5´ konce sekvence. Tím jsou eliminovány především báze konců s nízkou kvalitou.

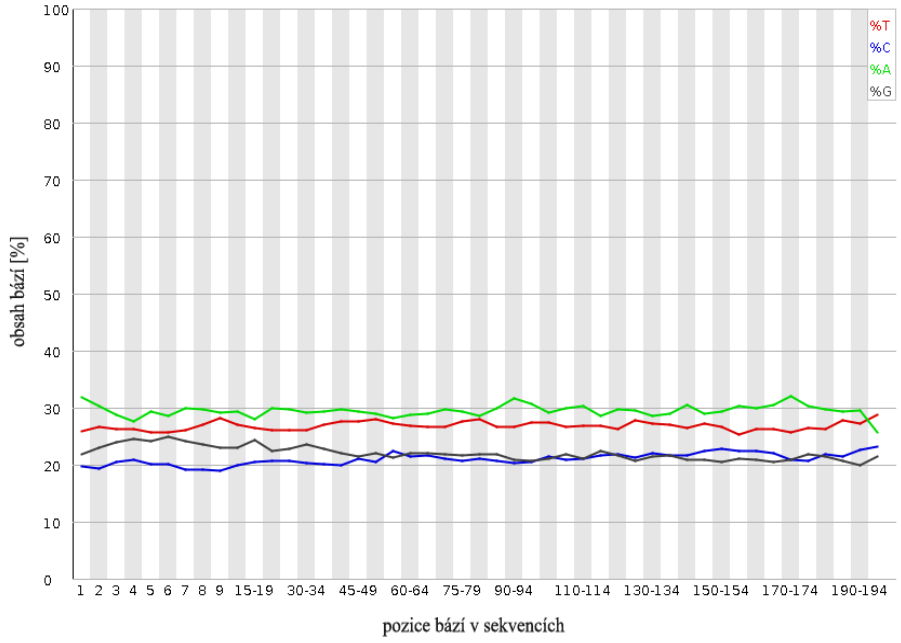
Následně jsme zbylá čtení filtrovali podle délky a průměrné kvality. Minimální délka sekvence byla stanovena na 50 bází. Sekvence rovny či kratší než 50 bází nejsou vhodné k sestavování do kontigu. Skóre kvality nebo také průměrná kvalita, byla nastavena na hodnotu 14. Tím se odfiltrovala čtení, která měla nízkou kvalitu bází uvnitř. Zbylý počet čtení, který prošel filtrací u vzorků 11 - 22 je zaznamenán v tabulce 5.

Výsledné grafické zobrazení filtrování sekvencí je zobrazeno na obrázku 14 kde obrázek A zobrazuje skóre kvality bází v sekvencích před úpravou konců a filtrací. Na obrázku B je skóre kvality sekvencí po úpravě a filtraci.

Obrázek 14 je krabicovým grafem. Percentil krabicového grafu je v rozmezí 25 - 75, medián je vyznačen čárkou uvnitř krabice. S úpravou sekvencí zde pozorujeme zkrácení sekvencí.

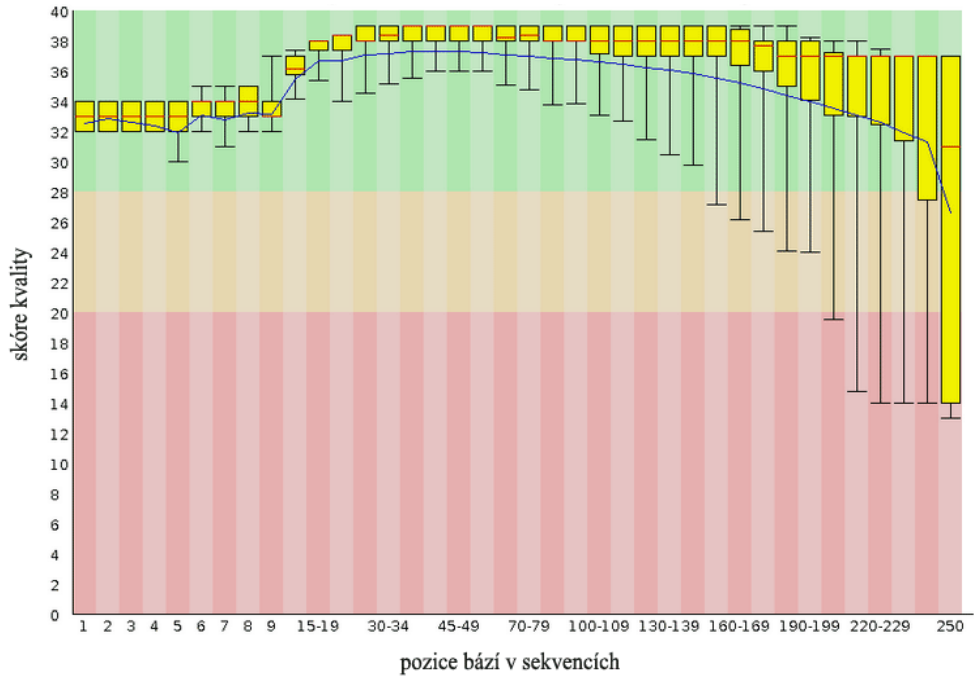


A

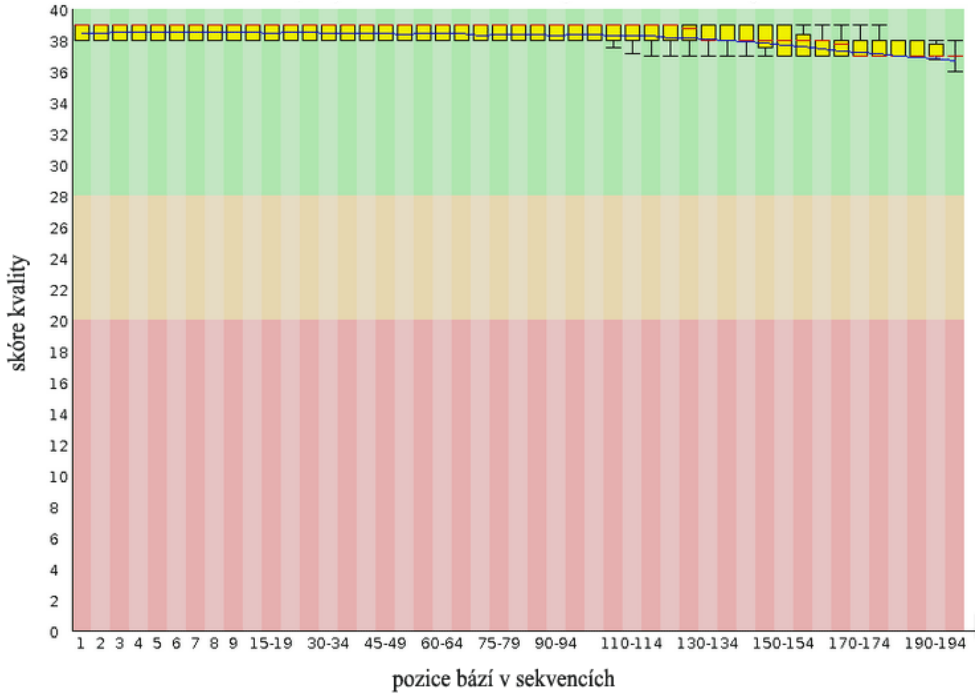


B

Obr. 13 *Obsah nukleotidů v sekvencích před úpravou a po úpravě v programu Galaxy. Na obrázků A je na pozicích 1-20 vidět nerovnoměrné rozložení bází, je to způsobenou tím, že se jedná o identifikátor, použitý při NGS. Na 3´ konci je také nerovnováha nukleotidů. Na obrázku B jsou je rozložení nukleotidů v sekvencích po odstranění 30 bází z 5´ konce a 25 bází z 3´ konce.*



A



B

Obr. 14 *Skóre kvality v sekvencích před a po úpravě v programu Galaxy. Krabicový graf značí kvalitu bází v jednotlivých úsecích sekvencí. Obrázek A zaznamenává kvalitu bází před úpravou v programu Galaxy, obrázek B po úpravě. Při srovnání obrázků je vidět, že byly odstraněny báze 3´ a 5´ konců a výsledné skóre kvality u čtení je o řád vyšší.*

Výsledné počty vstupních sekvencí a sekvencí, které prošly úpravou a filtrací jsou zaznamenány v tabulce 5. Navzdory nízké nastavené hodnotě kvality bylo odfiltrováno cca 50% sekvencí. Tento úbytek je v případě analýzy dat z NGS přípustným. V případě nižšího nastavení bychom byli nuceni pracovat s nedůvěryhodnými daty. Původní počet čtení u posledních dvanácti vzorků byl u poloviny z nich vyšší než 30 000 čtení. Po úpravě a filtraci je počet čtení u všech vzorků nižší než 30 000.

Tab. 5 *Přehled počtu čtení před a po upravě v programu Galaxy. zaznamenává vstupní počty čtení zpracovaných v programu Galaxy; počet čtení, který splňoval nároky filtrace a procentuální vyjádření úbytku čtení po filtraci.*

	původní počet čtení	počet čtení po filtraci v Galaxy	zbylých čtení po filtraci v Galaxy [%]
vzorek 11	60827	28054	46
vzorek 12	35028	18018	51
vzorek 13	23123	9938	43
vzorek 14	30407	15516	51
vzorek 15	30209	11528	38
vzorek 16	24078	13464	56
vzorek 17	12633	5608	44
vzorek 18	38621	28396	74
vzorek 19	20625	10218	50
vzorek 20	35800	19918	56
vzorek 21	6999	4342	62
vzorek 22	9893	5080	51

Takto upravené a vyselektované sekvence jsme dále zpracovávali pomocí programu Velvet.

4.1.2. Sestavení sekvencí do kontigů

Dalším krokem po filtraci a úpravě bylo sestavení sekvencí do kontigů pomocí programu Velvet. Tento program na základě nastavených parametrů dokáže sestavit řadu překrývajících se čtení, které vznikají při NGS, do jedné sekvence – kontigu. Parametry jsme museli optimalizovat:

Parametry byly nastaveny na hodnoty:

velveth ./output_directory 57 -fastq -short (náš soubor)

velvetg ./ -exp_cov 100 -ins_length 300 -cov_cutoff 5 -min_contig_lgth 150 -unused_reads yes -amos_file yes

Tyto parametry jsou popsány v kapitole Velvet na str. 34 - 37.

Coverage cutoff, neboli minimální počet překrývajících se čtení, která mohou vytvořit kontig, byl nastaven na hodnotu 5. Když tuto hodnotu zvýšíme, připravujeme se o značné množství sekvencí, nebo spíše poskládaných kontigů. Když naopak hodnotu snížíme, riskujeme tím to, že se nám za sebe nesmyslně poskládají navzájem nesouvisející kontigy spojené ostrovy extrémně nízkého pokrytí.

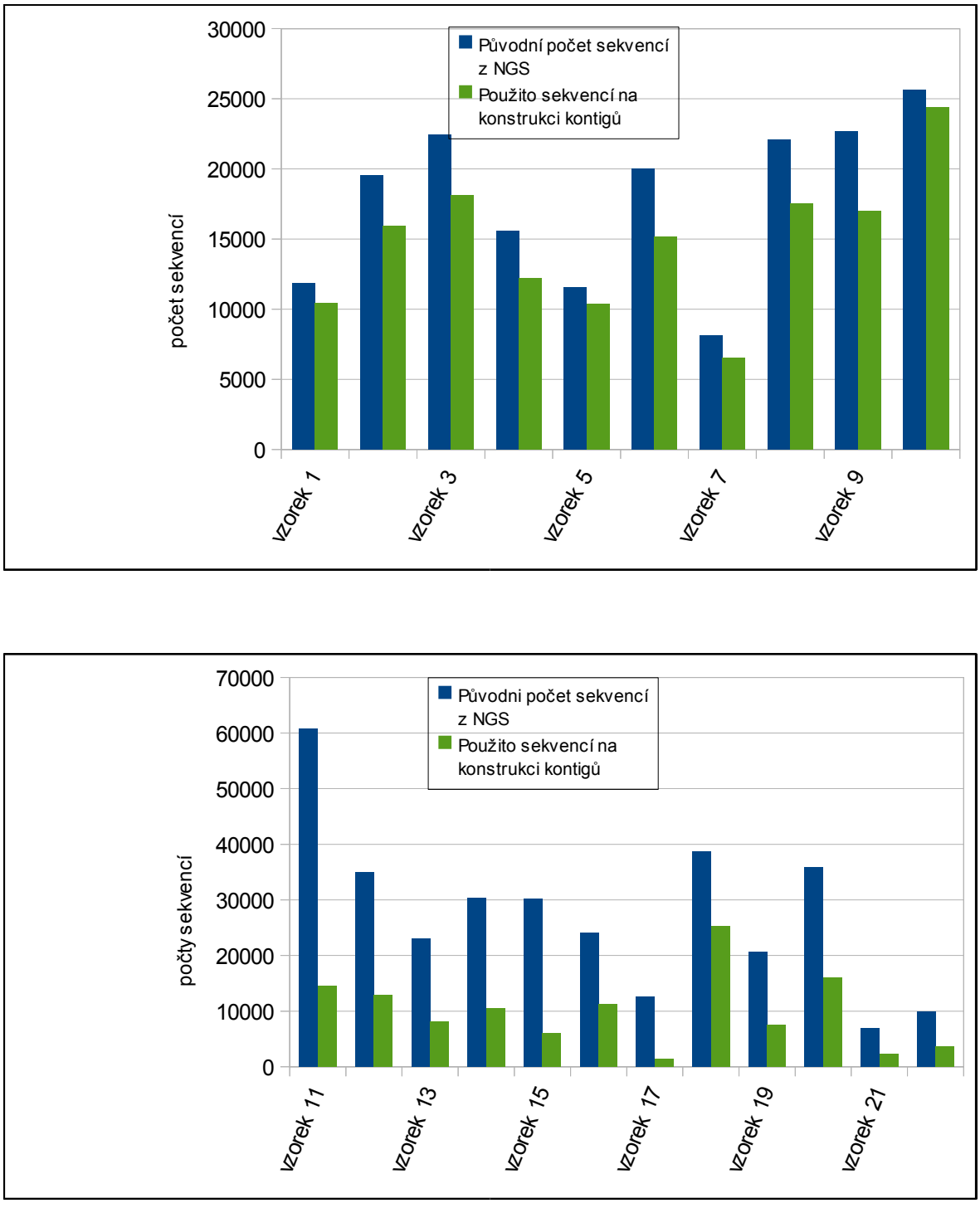
Minimální délku kontigů (min_contig_length) jsme museli v některých případech upravovat, a to na základě výsledků, které jsme dostali. Některá čtení byla kratší než námi nastavená a vzniklo nám velké množství krátkých kontigů, které se překrývaly v jednom úseku, tudíž bylo jasné, že se jedná o tutéž oblast a měl by vzniknout pouze 1 kontig. To jsme upravili tím, že jsem minimální délku kontigu snížili na hodnotu 80. V tomto případě se nám množství krátkých čtení snížilo, jelikož byla sestavena do kontigů.

U vzorků jedna až deset bylo použito původní množství sekvencí vyprodukovaných NGS. Množství sekvencí vzorků 11-22 odpovídá výstupnímu množství sekvencí z programu Galaxy. Vzorky 11-22 se liší od prvních deseti vzorků v počtu sekvencí zpracovaných Velvetem do kontigů oproti původnímu počtu sekvencí jednotlivých vzorků. Je to způsobeno vlastním původem vzorků. Jak již bylo zmíněno dříve, prvních deset vzorků pochází z buněčných kultur, zatímco zbytek byl izolován ze vzorků stolic, což vytváří velkou heterogenitu v kvalitě a tedy v následném vyselektování sekvencí při filtraci.

Vstupní počet sekvencí a počet sekvencí zpracovaných do kontigů u jednotlivých vzorků je zaznamenán v tabulce 6.

Tab. 6 *Počty sekvencí před a po zpracování v programu Velvet. Je zde zaznamenává průběh zpracování sekvencí pomocí programu Velvet, tzn. kolik bylo vstupních sekvencí, které program zpracovával a kolik jich bylo využito ke konstrukci kontigů, dále kolik ze vstupního počtu čtení vytvořil kontigů a jaká byla jejich průměrná délka.*

	počet sekvencí	počet kontigů	průměrná délka kontigu	použito sekvencí na konstrukci kontigů	čtení sestavených do kontigů [%]
vzorek 1	11830	42	717	10410	88
vzorek 2	19539	25	2222	15965	82
vzorek 3	22443	27	1843	18111	81
vzorek 4	15597	48	617	12189	78
vzorek 5	11571	20	1411	10351	89
vzorek 6	20027	23	1404	15147	76
vzorek 7	8119	11	7007	6532	80
vzorek 8	22115	65	734	17554	79
vzorek 9	22697	56	754	17026	75
vzorek 10	25640	55	1122	24392	95
vzorek 11	28054	191	221	14532	52
vzorek 12	18018	81	740	12968	72
vzorek 13	9938	54	1178	8056	81
vzorek 14	15516	56	1779	10567	68
vzorek 15	11528	27	905	6017	52
vzorek 16	13464	33	2637	11310	84
vzorek 17	5608	27	509	1427	25
vzorek 18	28396	83	542	25261	89
vzorek 19	10218	94	633	7585	74
vzorek 20	19918	59	1335	16087	81
vzorek 21	4342	13	1711	2294	53
vzorek 22	5080	55	1122	3573	70



Obr. 15 *Počet použitých sekvencí na konstrukci kontigů. Zobrazení kolik sekvencí z původního počtu vyprodukovaných NGS (modré sloupce) bylo použito na konstrukci kontigů (zelené sloupce). Je patrný rozdíl počtu sekvencí sestavených do kontigů u vzorků 1-10 a 11-22. Je to způsobeno úpravou a filtrací sekvencí vzorků 11-22.*

Rozdíl mezi čteními, která byla zařazena do kontigů u vzorků jedna až deset a 11 až 22 je na první pohled zřejmý jak dokumentuje obrázek 15. Musíme si uvědomit, že v případě vzorků 11 až 22 sekvence prošly filtrací a úpravou v programu Galaxy. Původní počet čtení u těchto vzorků dosahoval v některých případech až 60 000 čtení. Po filtraci byla čtení všech vzorků nižší než 30 000 čtení. Počet čtení, který byl použit ke konstrukci kontigů, se pak u všech vzorků pohybuje od 50 do 90%.

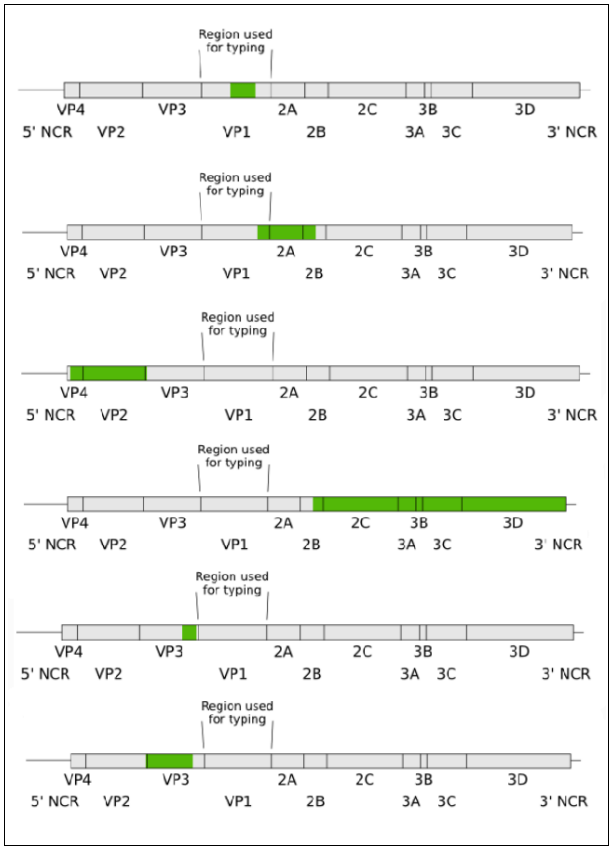
Výstupem programu Velvet jsou složky, které obsahují sekvence kontigů a také složky obsahující nevyužitá čtení. Mezi nevyužitými sekvencemi se mohou vyskytovat takové, které mohou být enterovirové. Proto jsme i tuto složku analyzovali v databázi Enterovirus genotyping tool. Některé z těchto sekvencí vykazovaly podobnost s enterovirovými. Ty jsme poté zahrnuli do sestavení genomu de novo.

V prvním kroku po sestavení kontigů ve Velvetu nás zajímalo, jaká je délka jednotlivých kontigů a zdali jsou identické s nějakou již známou enterovirovou sekvencí. Ne všechny vytvořené kontigy se však namapují na enterovirový genom.

Celý soubor s kontigy jsme analyzovali prostřednictvím databáze Enterovirus genotyping tool, která nám vyhledala nejbližší enterovirové sekvence, se kterými jsou kontigy nejvíce podobné. Výstup z tohoto nástroje je zobrazen na obrázku 16.

4.1.3. Identifikace kontigů jednotlivých vzorků

Kontigy vzorku 1 pokrývají téměř celý genom coxsackieviru B3, jak je patrné z obrázku 17. Proto se tato sekvence stala nejvhodnější referenční sekvencí pro sestavení genomu viru.



Obr. 16 *Namapování kontigů vzorku 1 na referenční sekvenci. Výstup nástroje Enterovirus Genotyping tool. Lze pozorovat, že kontigy nám pokrývají téměř celý genom referenční sekvence, kterou je coxsackievirus B3*

Pomocí tohoto nástroje získáme informace o délce kontigů a kam se mapují na referenční sekvenci. Sekvence pokrývající oblast proteinu VP1 jsou nezbytné k určení sérotypu viru. Aby byl určen sérotyp musí sekvence oblast VP1 překrývat minimálně 100 bázemi. Tyto sekvence byly sérotypicky stanoveny jako coxsackievirus B3. U sekvencí nepřekrývajících tuto oblast je určen rod či kmen sekvence, na kterou se namapuje hledaná sekvence nikoliv však sérotyp. Stejný postup analýzy v Enterovirus Genotyping tool byl použit pro všechny vzorky (1 – 22).

Tab. 7 *Souhrn informací o kontizích jednotlivých vzorků získaných na základě analýzy v programu Enterovirus genotyping tool.*

	jméno	délka	sérotyp	pokrytí (čtení)	míra shody	referenční sekvence	začátek	konec
Vzorek 1	NODE_28	350	CV-B3	79	7,33	NC_001472	2430	2783
	NODE_1	323	CV-B3	414	4,67	NC_001472	2821	3144
	NODE_10	613		160	4,52	NC_001472	1683	2296
	NODE_7	976		346	4,41	NC_001472	744	1720
	NODE_9	179		437	3,82	AB426610	2247	2426
	NODE_6	767	CV-B3	256	3,35	NC_001472	3173	3931
	NODE_8	3369		196	2,62	NC_001472	3930	7296
	NODE_12	773		101	1,36	NC_001472	-3	769
Vzorek 2	NODE_4	313	CV-B3	375	6,7	NC_001472	2301	2614
	NODE_1	2272		325	3,65	NC_001472	58	2331
	NODE_2	4274	CV-B3	357	3,26	NC_001472	2610	6875
	NODE_15	431		200	2,18	NC_001472	6903	7333
Vzorek 3	NODE_1	2430	CV-B3	406	8,98	NC_001472	627	3057
	NODE_3	450	CV-B3	349	4,79	NC_001472	3007	3448
	NODE_2	1475		293	2,1	NC_001472	3399	4874
	NODE_9	2465		278	1,73	NC_001472	4866	7332
	NODE_4	459		717	1,33	NC_001472	121	577
Vzorek 4	NODE_14	608	CV-B3	307	7,67	NC_001472	2608	3216
	NODE_6	314	CV-B3	351	6,99	NC_001472	2301	2615
	NODE_5	925		400	5,63	NC_001472	1424	2349
	NODE_12	667		204	3,1	NC_001472	3273	3940
	NODE_3	246		176	2,95	NC_001472	1127	1373
	NODE_17	229	Unassigned	22	2,89	AF114383	2470	2699
	NODE_9	2018		180	2,81	NC_001472	3876	5894
	NODE_7	1500		150	2,29	NC_001472	5850	7349
	NODE_15	1155		93	1,9	NC_001472	136	1291
Vzorek 5	NODE_8	385	CV-B3	324	7,22	NC_001472	2230	2615
	NODE_5	533	CV-B3	252	7,12	NC_001472	2609	3142
	NODE_4	610		266	6,03	NC_001472	1567	2177
	NODE_1	4201	CV-B3	195	2,79	NC_001472	3138	7330
	NODE_11	156		227	2,71	NC_001472	2119	2275
	NODE_3	1461		201	2,4	NC_001472	128	1589
Vzorek 6	NODE_1	1454	CV-B3	445	7,42	NC_001472	1489	2943
	NODE_4	697		451	2,9	NC_001472	785	1482
	NODE_6	1943	CV-B3	335	2,75	NC_001472	2942	4875
	NODE_3	265		246	1,89	NC_001472	570	835
	NODE_5	2496		262	1,76	NC_001472	4865	7362
	NODE_7	600		258	1,37	NC_001472	-12	585
Vzorek 7	NODE_1	7057	CV-B3	123	3,62	NC_001472	181	7229
Vzorek 8	NODE_2	1304		592	5,51	NC_001472	771	2075
	NODE_26	406	CV-B3	343	5,06	NC_001472	2784	3190
	NODE_1	509	CV-B3	277	4,37	NC_001472	2076	2585
	NODE_4	288	CV-B3	349	3,63	NC_001472	3140	3419
	NODE_3	3945		292	1,77	NC_001472	3368	7312
	NODE_8	350		341	1,33	NC_001472	417	767
	NODE_20	201		495	1,18	NC_001472	120	320
Vzorek 9	NODE_3	354	CV-B3	258	7,13	NC_001472	2610	2964
	NODE_20	313	CV-B3	427	6,04	NC_001472	2301	2614
	NODE_27	460		389	5,12	NC_001472	1856	2316
	NODE_33	309		377	3,34	NC_001472	1553	1862
	NODE_1	443		682	3,21	NC_001472	4582	5025
	NODE_5	1523	CV-B3	311	2,95	NC_001472	3131	4645
	NODE_4	2165		296	2,42	NC_001472	5065	7231
	NODE_28	255		176	2,27	AB426610	5011	5266
	NODE_6	1230		278	2,11	NC_001472	167	1397
	NODE_7	804		53	2,11	NC_001472	725	1529
	NODE_24	186		208	1,8	NC_001472	32	217

Tabulka 7 je souhrnem informací o sestavených kontizích jednotlivých vzorků, které se týkají množství a délky kontigů, nejbližší nalezené sekvence a popř. jejího sérotypu, dále číslo pod kterým je referenční sekvence uložena v databázi GenBank. Sloupce začátek a konec dávají informace o přesném úseku na referenční sekvenci, na nějž se namapoval daný kontig.

Všechny kontigy vzorků 1-9 spadají do rodu Enteroviru B. U kontigů, jež se namapují na referenční sekvenci, může být určen rod či druh této sekvence, zatímco při určení sérotypu musí kontig přesahovat oblast proteinu VP1 referenční sekvence minimálně 100 bázemi. U kontigů, které překrývají oblast VP1, lze určit sérotyp daného genomu. V tabulce je vidět, že sérotyp je charakterizován u všech vzorků shodně a je jím sérotyp coxsackieviru B3.

Každý řádek v tabulce odpovídá jednomu kontigu, který se namapoval na enterovirový genom.

Sloupec referenční sekvence je označení sekvence v databázi GenBank, pod kterým si referenční enterovirový genom můžeme najít a použít sekvenci pro další analýzu, konkrétně sestavení kontigů jednotlivých vzorků pod tuto referenci.

Míra shody je číslo, jež vyjadřuje do jaké míry je úsek, který se namapoval na určitou část enterovirového genomu, shodný s touto sekvencí. Číslo je bráno jako relevantní, když je vyšší než 1.

Ve sloupci začátek jsou u kontigu 12 vzorku 1 a kontigu 7 vzorku 6 hodnoty záporné, což vypovídá o tom, že se nám podařilo osekvenovat začátek genomu viru dále, než je u původní referenční sekvence coxsackieviru B3.

Na základě informací z tabulky 7 jsme jako referenční sekvenci použili sekvenci coxsackieviru B3. Ta nám sloužila jako lešení pro sestavení genomu z jednotlivých kontigů vzorků.

Tab. 8 *Souhrn informací o kontizích jednotlivých vzorků získaných na základě analýzy v programu Enterovirus genotyping tool*

	jméno	délka	Rod/druh	sérotyp	Pokrytí (čtení)	míra shody	referenční sekvence	začátek	konec
Vzorek 10	NODE_8	1763	Enterovirus B	CV-B3	402	8,92	NC_001472	846	2608
	NODE_2	176	Enterovirus B	CV-B3	529	5,15	AB426611	2587	2763
	NODE_21	278	Enterovirus B	CV-B3	694	4,13	NC_001472	2823	3101
	NODE_6	926	Enterovirus B	CV-B3	526	3,6	NC_001472	3078	3994
	NODE_9	1486	Enterovirus B		336	2,79	NC_001472	4752	6238
	NODE_11	200	Enterovirus B		360	2,76	NC_001472	648	848
	NODE_10	720	Enterovirus B		534	2,33	NC_001472	3983	4704
	NODE_12	1172	Enterovirus B		344	2,14	NC_001472	6188	7358
	NODE_35	195	Enterovirus B		124	1,27	NC_001472	52	247
	NODE_22	268	Enterovirus B		450	1,11	NC_001472	379	647
Vzorek 11	NODE_2	4766	Enterovirus A	CV-A16	291	7,26	NC_001612	92	4864
	NODE_3	1656	Enterovirus A		272	3,94	NC_001612	4984	6642
	NODE_7	220	Enterovirus A		342	3,56	NC_001612	4813	5033
	NODE_4	679	Enterovirus A		92	2,23	NC_001612	6592	7271
	NODE_433	302	Enterovirus A		5	1,52	NC_001612	6592	6894
Vzorek 13	NODE_30	1037	Enterovirus B	E-30	39	6,81	NC_001472	1898	2923
	NODE_81	421	Enterovirus B		7	2,54	NC_001472	3647	4068
	NODE_8	1015	Enterovirus B		14	1,97	NC_001472	369	1384
	NODE_37	625	Enterovirus B		7	1,6	NC_001472	6522	7147
Vzorek 14	NODE_108	318	Enterovirus A	CV-A2	8	4,44	NC_001612	2353	2668
	NODE_37	760	Enterovirus A		5	1,82	NC_001612	1432	2198
Vzorek 15	NODE_4	3776	Enterovirus A	CV-A2	154	3,08	NC_001612	89	3875
	NODE_2	3217	Enterovirus A		114	2,93	NC_001612	4098	7315
	NODE_11	234	Enterovirus A		319	2,57	NC_001612	3869	4103
Vzorek 16	NODE_34	1515	Enterovirus A		11	4,02	NC_001612	4937	6452
	NODE_62	188	Enterovirus B	CV-B1	6	2,54	NC_001472	2518	2706
	NODE_45	307	Enterovirus A	CV-A16	6	2,32	NC_001612	2936	3243
	NODE_19	986	Enterovirus A		11	1,79	NC_001612	179	1170
Vzorek 18	NODE_99	245	Enterovirus A	CV-A10	6	3,29	NC_001612	3124	3369
	NODE_46	724	Enterovirus A	CV-A10	15	2,18	NC_001612	2496	3223
	NODE_133	998	Enterovirus A		5	1,54	NC_001612	1245	2246
	NODE_164	233	Enterovirus A		10	1,06	NC_001612	426	659

Vstupní data prvních devíti vzorků se liší od zbylých. Tomu odpovídají výsledky získané z databáze Enterovirus genotyping tool. Rozdíly nacházíme mezi rody a sérotypy jednotlivých vzorků. U posledních dvanácti vzorků se vyskytují rody Enterovirus A a B, zatímco u prvních devíti vzorků to byl shodně rod Enterovirus B. U 6 vzorků (12, 17, 19, 20, 21 a 22) nebyly nalezeny kontigy, které by se namapovaly na známou enterovirovou sekvenci, zahrnutou v této databázi. U těchto vzorků byla původní kvantita enterovirů nižší o více jak tři řády než u vzorků ostatních.

Vytvořené kontigy nepokryly všechny oblasti enterovirového genomu. Po namapování všech kontigů jednotlivých vzorků k referenční sekvenci jsme názorně viděli, které oblasti chybí. Namapování jsme provedli v programu Sequencher. Výsledný soubor ze Sequencher jsme pak zobrazili v Bioeditu, kde jsme pracovali se sekvencemi dále. Mezery mezi namapovanými kontigy jsme se u vzorků 1-9 rozhodli přemostit pomocí Sangerova sekvenování. U vzorků 10-22 jsme mezery nesekvenovali z důvodu, že mezer se zde vyskytuje nepoměrně více a tudíž by byla analýza velmi náročná. Dalším zpracováním těchto vzorků se předkládaná diplomová práce nezabývá.

4.1.4. Přemostění mezer mezi kontigy

Soubor namapovaných kontigů na referenční sekvenci jsme zobrazili v programu Bioedit. Zde jsme si spočítali množství mezer mezi jednotlivými kontigy ve vzorcích. Počet chybějících úseků jednotlivých vzorků 1-9 je zaznamenán v tabulce 9. Vyhledávali jsme vhodné sekvence pro navržení primerů. Na každou mezeru jsme navrhli vždy jeden pár primerů Každá dvojice primerů je navržena vždy ze sekvence kontigů mezi nimiž je mezera. Tabulka 9 obsahuje sekvence primerů, které byly navrženy na tyto úseky. Postup návrhu primerů je popsán v kapitole Návrh primerů na str. 31, 32. U vzorku 2 se nevyskytovaly žádné mezery, vytvořil se nám jeden kontig dlouhý 7007 bází. U tohoto vzorku jsme sekvenovali 3´ a 5´ konce.

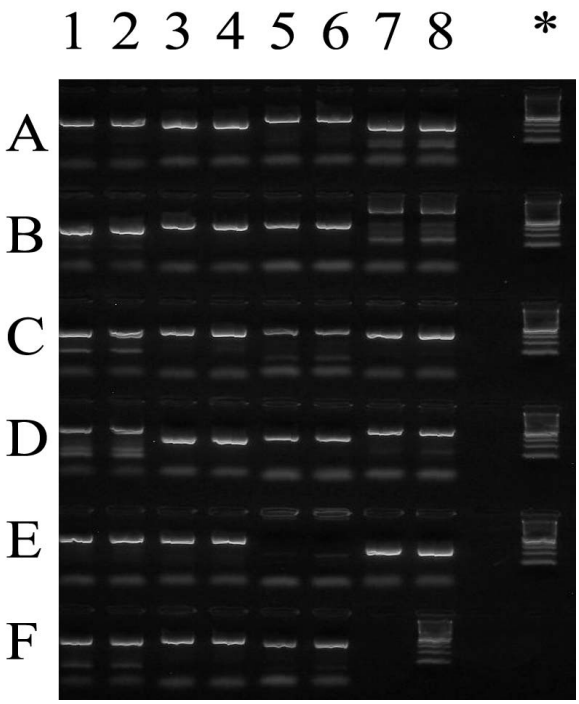
Mezi chybějící úseky jsme v této fázi nepočítali oblasti 3´ a 5´ konců. Těmi jsme se zabývali až v následujícím kroku.

Tab. 9 *Přehled počtu mezer a sekvencí primerů využitých k přemostění těchto mezer.*

	Počet úseků mezi kontigy, které je třeba přemostit Sangerovým sekvenováním	číslo mezery	Sekvence forward primeru	Sekvence reverse primeru
Vzorek 1	4	1	ACTGGGCGCTAGCACTCT	GGCTCTTCACACCATGTCAGTA
		2	GGGATGTAGGCTTGCAATCTAG	GGGCCGGTATAGCTTCAGAATT
		3	GGGCACACATCTCAAGTTGTTCCAG	CTCCGACCAGCCATCGTAAAAG
		4	CGTTCACCCAGTGATGCCAATGAGA	CCACACCGTTGTCTAGTTCGGTAA
Vzorek 2	0	genom neobsahoval mezery, byl plně pokryt čteními z NGS		
Vzorek 3	3	1	GCACAACCCAAGTGTAGATCA	CAATTGTCACCATAAGCGGCCAG
		2	GGCTGCGCTGGAAGAGAAA	GGCCCTTGAAACAGGGCTT
		3	GGGGGAGTGCTTCAATGAAGT	CGCACCGAATGCGGAGAA
Vzorek 4	2	1	GCGCTAGCACTCTGGTATCA	AGTGCATCAGGTAGTTTCCACC
		2	GGCTGGAAGATGATGCCATGGAA	CTTGCAGGCGTTGGTCATCT
Vzorek 5	2	1	GCGCTAGCACTCTGGTATCA	CCGGAGGACTACCAATTAGCTC
		2	GGGTATGGTCTGATCATGACACCA	TACCACACCGTTGTCTAGTTCGG
Vzorek 6	3	1	GCACAACCCAAGTGTAGATC	CACCGGATGGCCAATCCAATA
		2	GCCGGACTGGGTATACCATACA	CCCGTTGTACTCAGCGCACAT
		3	GGGCACTTGCTAGGAGATTCCACT	CCCACACTATGTCTGTGGTTGT
Vzorek 7	2	1	CAGCCTGTGGGTTGTACCCA	ACTGGGGTTGTGCGGAGCGAAA
		2	GGGGATGATGTGATTGCGTCCTA	CCGCACCGAATGCGGAGAATTTA
Vzorek 8	3	1	GGCGCTAGCACACTGGTAT	CTGCCCACTGGCATGTGGGTAT
		2	GTAGGGACTGGACCAACAAATTC	CCTTCAGTCCAGAATACACTGGGATT
		3	GGGTTGATCATGACACCAGCT	CCACACCGTTGTCTAGTTCGGTAA
Vzorek 9	4	1	GCGCTAGCACTCTGGTATCA	CTGCGCAACTTCCATGGTGGTA
		2	GGGTGCCTATTGGTAGTGTGTGTA	CCGTTGTACTCAGCGCACAT
		3	GGCAGTGCAATTGAGAAGAAAGC	TCCACGCCTTGACGTGCTTT
		4	GGGAGTGCTTTAACGAGGTGACAT	CCACACCGTTGTCTAGTTCGGTAA

Úseky mezi kontigy jsme amplifikovali pomocí metody PCR za využití navržených primerů. Metoda PCR společně s amplifikačním programem, který jsme používali, je vysvětlena v kapitole Metody na stránce 32 a 33.

Výsledky amplifikace jsme zjistili pomocí elektroforézy a vizualizovali je na transiluminátoru. Výsledný elektroforetický záznam je znázorněn na obrázku 17.



Obr. 17 *Elektroforetický záznam. Je zde zobrazeno 46 reakcí, kdy jsme testovali v duplikátech 23 oblastí genomu viru, které nebyly pokryty daty z NGS. Duplikáty jsou vždy v řadě vedle sebe. Velikost produktů jsme zjistili dle velikostního DNA markeru, který je na konci každého řádku. Tento marker ukazuje velikosti od 100 do 1000 bp, nejsilnější proužek pak značí velikost 500 bp. Tyto produkty byly dále využity pro sekvenování pomocí Sangerova sekvenování.*

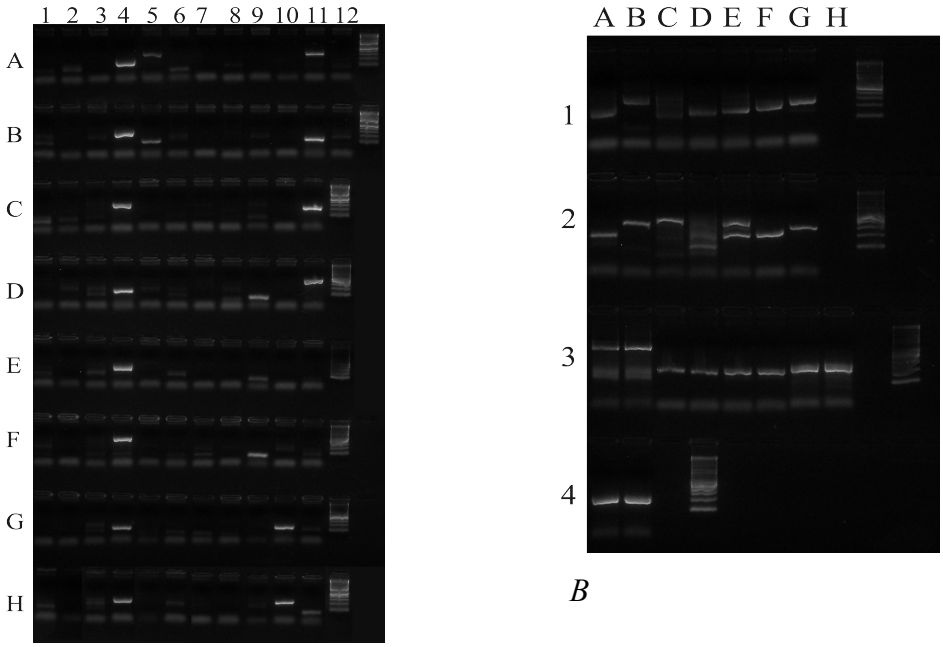
Výsledný elektroforetický záznam ukázal, že primery byly navrženy správně a komplementárně nasedaly na úseky mezi kontigy. Vzorky na pozicích E 5, 6 v tomto pokusu nebyly namnoženy, proto jsme je museli amplifikovat s novým párem primerů.

Amplifikované úseky jsme podrobili přečištění před sekvenační reakcí, sekvenační reakci, přečištění po sekvenační reakci a nakonec analýze na automatickém sekvenátoru.

Při sekvenování 3´ a 5´ konců jsme stáli před problémem využít velmi komplikované metody inverzního PCR pro sekvenování konců genomů, nebo se pokusit konce genomů sekvenovat podobně jako mezery mezi kontigy. Zvolili jsme si druhou možnost. Navrhli jsme si primery pro sekvenování konců genomů. V tomto případě jsme využili při návrhu jednoho z dvojice primerů referenční genom. Podrobnější popis návrhu primerů pro sekvenování 3´ a 5´ konce je v kapitole Metody.
Sestavili jsme si tabulku vybraných referenčních sekvencí pro návrh primerů, viz tabulka 3 a 4. Zkoušeli jsme všechny naše vzorky se všemi vybranými primery pro 5´ konec a všechny vzorky s vybranými primery pro 3´ konec. Konkrétní sekvence navržených primerů jsou v tabulce 10. Výsledné elektroforetické záznamy jsou na obrázku 18.

Tab. 10 *Sekvence primerů navržených na sekvenování konců genomů. Tyto primery jsou navrženy z referenčních sekvencí, druhý primer je vždy navržen z kontigu, nejbližšího k danému konci*

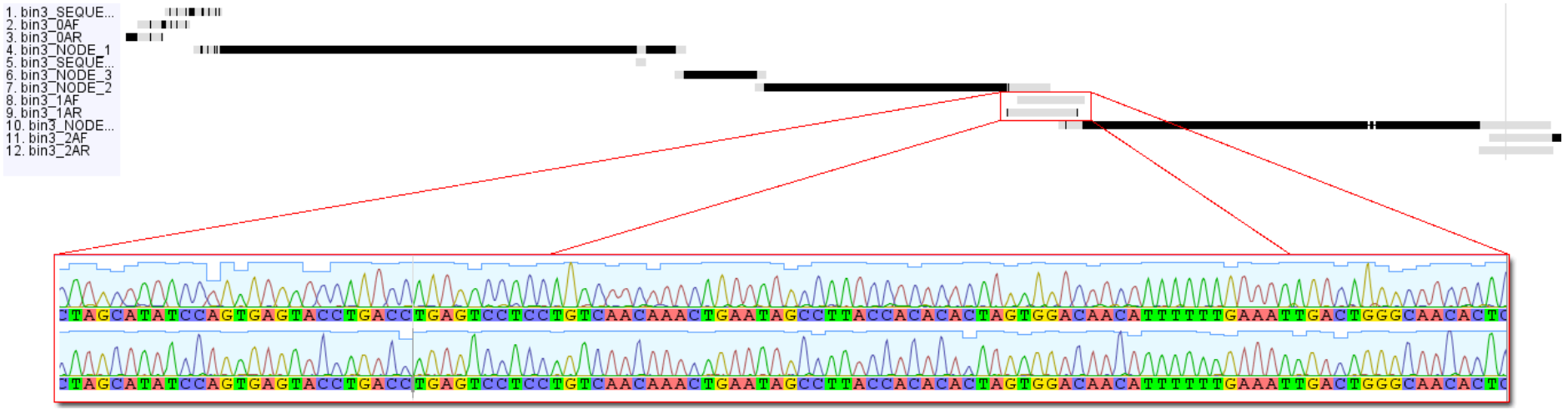
koncové primery_5´	
bovine_ent	atccggtgggtgtattagggcc
poliovirus2	taccctacaacagtatgaccc
poliovirus1	cccctacaacagtgtaacccaa
enterovirusD	cctctacaaaatctaatgccccagg
echovirus26	ccgcaccgaatgcggagaattta
coxsackie B1	acccccactgcaccgttatctagtt
coxsackie A2	cccaccagtaattcacacgaccaga
coxsackie A11	cccctacatcagtatcacccaa
coxsackie A10	caccagtcatttgcacgaccaggtt
coxsackie A8	accagatttctggtggggttcagt
coxsackie A3	ccaccagtcatttacacgaccaga
coxsackie A1	cctacaacaccacaacccaagcca
enterovirus_d68	ggtccccaagtgaccaaaatttacc
coxsackie_B4	accgaacgcggagaatttacccccta
coxsackie_A20	cccctacaacagtataacccaatcc
koncové primery_3´	
coxsackie B1	ttaaaacagcctgtgggttgyw
poliovirus1	ttaaaacagctctggggttg



Obr. 18 *Elektroforetický záznam amplifikace úseků na 5´ a 3´ koncích genomů. Obrázek A je elektroforetickým záznamem aplifikování 5´ konců, obrázek B je záznamem amplifikace 3´ konců.*

Sekvenovali jsme 5´ konce u vzorků 2, 4, 5, 6 a 8. U všech těchto vzorků se nám amplifikovaly úseky, jejichž primery byly navržené podle referenční sekvence echoviru 26, jež odpovídá sloupci 4 na obrázku 18A. U vzorků 1, 2, 4 a 5 byla též vhodnou referenční sekvencí i sekvence coxsackieviru B1 (odpovídá sloupci 9 pozice D až F) a u vzorků 2, 4 a 5 se amplifikovaly úseky s primery navrženými podle coxsackieviru A1 (sloupec 10, pozice G, H a sloupec 11, pozice A až D).

U vzorků 1, 2, 3, 4, 5, 6, 8 a 9 jsme sekvenovali 3´ konec. Na obrázku 18B je vidět, že se nám naamplifikovaly všechny zkoušené úseky. Po analýze na automatickém sekvenátoru, jsme však zjistili, že se nám lépe osekvenovaly úseky, které byly amplifikovány s použitím primerů navržených podle sekvence polioviru 1.

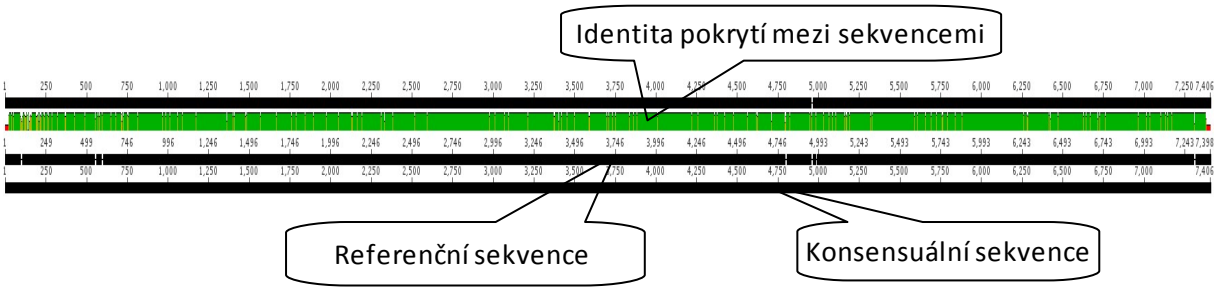


Obr. 19 ***Sestavený virový genom.*** *Sestavení virového genomu, kde jako lešení sloužila referenční sekvence. Zvětšený úsek zaznamenává mezeru mezi kontigy, která byla překlenuta pomocí Sangerova sekvenování.*

Obrázek 19 dokumentuje sestavený virový genom. Referenční sekvence nám sloužila jako lešení pro sestavení genomu. Data z NGS však nepokryla celou délku genomu, tudíž jsme museli mezery mezi kontigy překlenout pomocí Sangerova sekvenování.

Sekvence pokrývající oblasti mezi dvěma kontigy, získané obousměrnou sekvenací, jsme upravili do jedné výsledné konsenzuální sekvence, a tu pak uspořádali do příslušných úseků v genomu viru. Z takto poskládaného genomu jsme potřebovali získat jednu výslednou konsenzuální sekvencí, která má po translaci jeden otevřený čtecí rámec bez stop kodonů.

Na obrázku 20 vidíme konsensuální sekvenci osekvenovaného genomu společně s referenční sekvencí. Je zde znázorněna míra identity mezi oběma sekvencemi.



Obr. 20 *Srovnání virového genomu s referenční sekvencí.*

Srovnání virového genomu s referenční sekvencí jsme vyjádřili v procentech. Tyto hodnoty srovnání jsou zaznamenány v tabulce 11. Zde lze porovnat hodnoty shody jednotlivých genomů s referenční sekvencí, kterou je coxsackievirus B3. Míra shody [%], která se dá považovat za relevantní, je vyšší než 70. U vzorků 1-9 je míra shody vysoká.

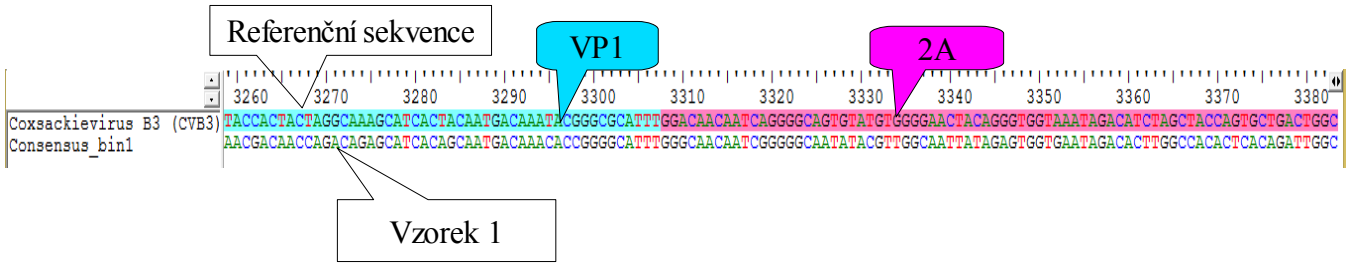
Tab. 11 *Míra identity konsenzuální sekvence daného vzorku s referenční sekvencí*

	vzorek 1	vzorek 2	vzorek 3	vzorek 4	vzorek 5	vzorek 6	vzorek 7	vzorek 8	vzorek 9
identita s referenční sekvencí [%]	97,3	97,2	86,9	79	97,5	86,3	80	86,6	97,8

4.2. Fylogenetická analýza

Po získání všech konsensuálních sekvencí virových genomů, jsme je porovnali s ostatními sekvencemi enterovirových skupin, abychom mohli provést fylogenetickou analýzu. Proto jsme z databáze GenBank vybrali vhodné zástupce, a to takové, aby jejich genom byl plně osekvenován a byly zde rozlišitelné jednotlivé proteinové sekvence. Vybrané sekvence společně s jejich identifikačními čísly jsou zaznamenané v tabulce 12. Tyto vybrané sekvence jsme poté podrobili multiple alignmentu pomocí ClustalW, jenž je součástí programu Bioedit, a vybrali jsme si jen proteiny, s nimiž jsme pracovali při fylogenetické analýze. Jednalo se o proteiny VP1, 2A a 2C.

Stejné proteinové sekvence jsme vybrali i v případě našich genomů. Byli jsme toho schopni na základě porovnání s referenční sekvencí coxsackieviru B3, jež byl v databázi GenBank anotován. Obrázek 21 zobrazuje jeden z našich genomů společně s jeho referenční sekvencí, která je anotována a proteinové produkty jsou zde charakterizovány. Proteinové sekvence jsme si rozlišili barevně. Tyto proteiny námi osekvenovaného genomu jsou shodné s proteiny referenční sekvence.

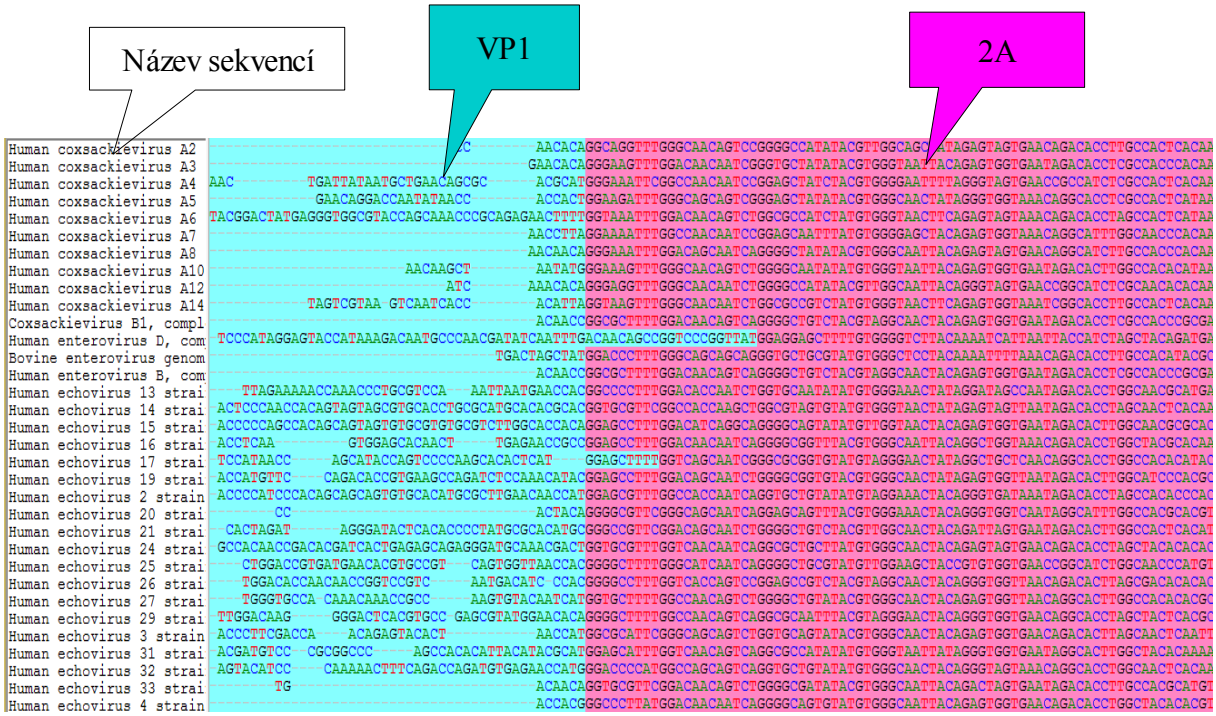


Obr. 21 *Srovnání referenční sekvence s genomem viru. Obrázek znázorňuje alignment osekvenovaného genomu s referenční sekvencí a rozhraní proteinů VP1 a 2A na sekvenci, které je totožné s rozhraním proteinů genomu enteroviru.*

Tab.12 *Všechny referenční sekvence, použité k porovnání s našimi sekvencemi.*

druh	referenční číslo	název viru/kmene viru
Enterovirus A	Popset_40068428	Human coxsackievirus A2 kmen, Fleetwood
		Human coxsackievirus A3 kmen, Olson
		Human coxsackievirus A4 kmen, High Point
		Human coxsackievirus A5 kmen, Swartz
		Human coxsackievirus A6 kmen, Gdula
		Human coxsackievirus A7 kmen, Parker
		Human coxsackievirus A8 kmen, Donovan
		Human coxsackievirus A10 kmen, Kowalik
		Human coxsackievirus A12 kmen, Texas-12
		Human coxsackievirus A14 kmen, G-14
Enterovirus B	M16560	Coxsackievirus B1
	NC001472	Human enterovirus B
	M16572	Coxsackievirus B3
	Popset_34485417	Human echovirus 13 kmen, Del Carmen
		Human echovirus 14 kmen, Tow
		Human echovirus 15 kmen, CH 96-51
		Human echovirus 16 kmen, Harrington
		Human echovirus 17 kmen, CHHE-29
		Human echovirus 19 kmen, Burke
		Human echovirus 2 kmen, Cornelis
		Human echovirus 20 kmen, JV-1
		Human echovirus 21 kmen, Farina
		Human echovirus 24 kmen, DeCamp
		Human echovirus 25 kmen, JV-4
		Human echovirus 26 kmen, Coronel
		Human echovirus 27 kmen, Bacon
		Human echovirus 29 kmen, JV-10
		Human echovirus 3 kmen, Morrisey
		Human echovirus 31 kmen, Caldwell
		Human echovirus 32 kmen, PR-10
		Human echovirus 33 kmen, Toluca-3
		Human echovirus 4 kmen, Pesacek
		Human echovirus 6 kmen, D´Amor
		Human echovirus 7 kmen, Wallace
	Popset_375004670	Human coxsackievirus B5 kmen, CVB/CC10/16
		Human coxsackievirus B5 kmen, CVB/CC10/17
Enterovirus C	JX174176	Human coxsackievirus A1 isolate HT-THLH02F/XJ/CHN/2011
	AY184219	Human poliovirus 1 strain Sabin 1
	D00625	Human poliovirus 2 genomic RNA
Enterovirus D	NC001430	Human enterovirus D
Enterovirus E	D00214	Bovine enterovirus genomic RNA

Mezi vybranými sekvencemi se vyskytuje i bovinní enterovirus, který je lidským enterovirům příbuzně vzdálen. Zahrnuli jsme ho do fylogenetické analýzy proto, abychom při konstrukci fylogenetických stromů je tzv. zakořenili. Při zakořenění přidáme druh, který nepatří do studované skupiny a odvětvil se od společného předka dříve, než se vzájemně odvětvily ostatní druhy.



Obr. 22 *Multiple alignment vybraných referenčních sekvencí. Na obrázku je vidět, jak jsme si barevně jednotlivé proteiny od sebe odlišili.*

Všechny sekvence použité v multiple alignmentu byly již anotovány v databázi GenBank. Proto jsme si mohli jednotlivé proteinové sekvence od sebe odlišit, což je vidět na obrázku 22. Zde je viditelné i rozhraní proteinů VP1 a 2A, což jsou proteiny, které jsme následně použili k fylogenetické analýze, společně ještě s proteinem 2C.

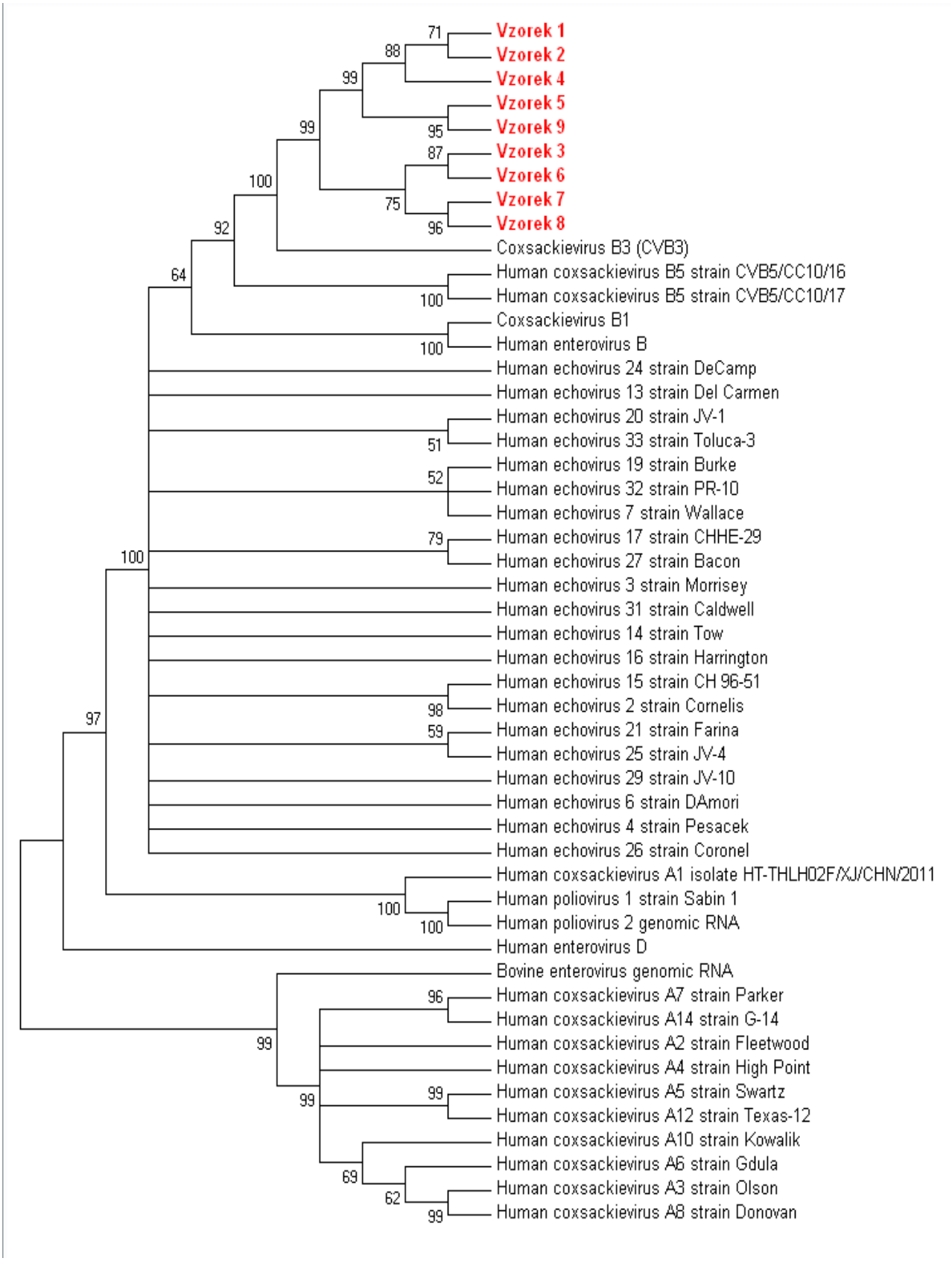
Na základě parametrů, jež jsou podrobně popsány v kapitole Mega5, pro vytvoření fylogenetické analýzy jsme vytvořili tři fylogenetické stromy.

Do fylogenetické analýzy jsme zahrnuli tři virové proteiny: VP1, 2A a 2C. VP1 patří mezi strukturální proteiny a určuje enterovirový sérotyp. Proteiny 2A a 2C patří mezi proteiny nestrukturální. Výsledné fylogenetické stromy jsou zobrazeny na obrázcích 23 až 25.

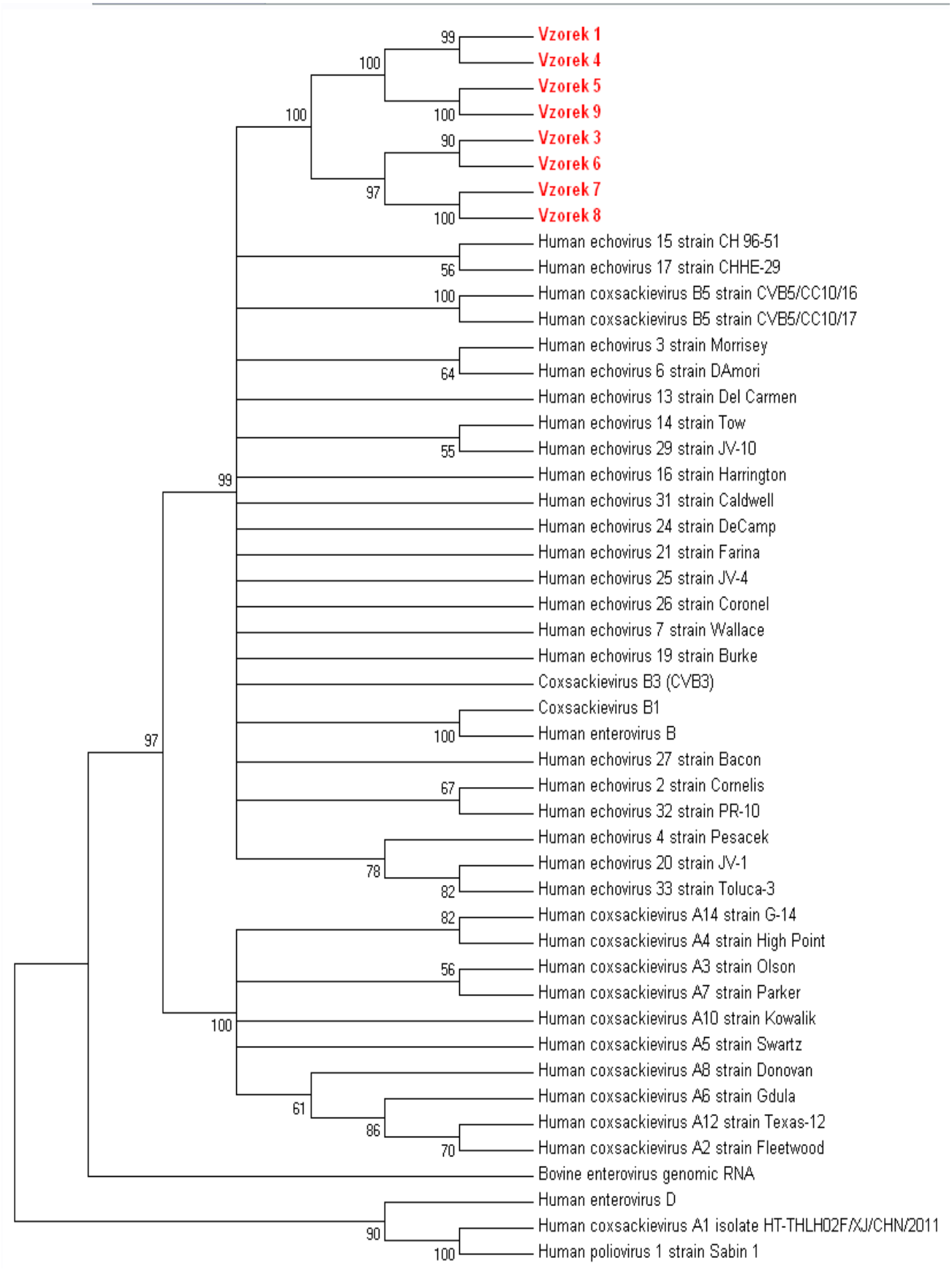
Na základě fylogenetické analýzy proteinu VP1 jsme zjistili, že naše izoláty patří do jedné monofyletické skupiny nejpříbuznější s coxsackievirem B3, všechny vzorky vytváří jednu monofyletickou skupinu. Tento výsledek odpovídá analýzám provedeným už v databázi Enterovirus genotyping tool, kde všechny naše vzorky byly sérotypicky nejpodobnější coxsackieviru B3.

V případě analýzy proteinu 2A už nebyl coxsackievirus nejpříbuznější. Všechny vzorky opět vytvářejí monofyletickou skupinu.

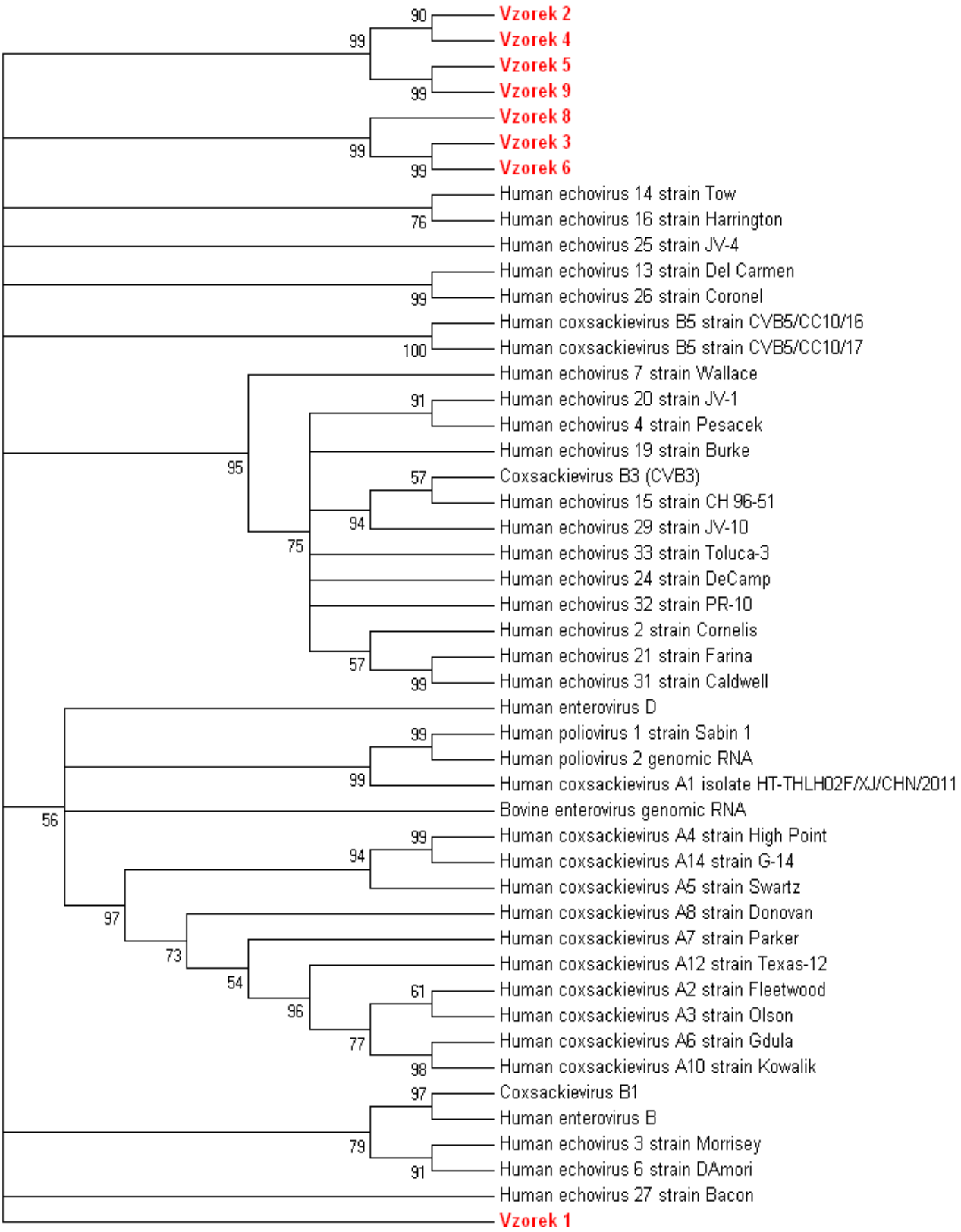
U proteinu 2C se nám vzorek 1 oddělil od ostatních vzorků, které vytvářejí monofyletickou skupinu. To značí, že u tohoto vzorku mohlo dojít k rekombinaci. To je u enterovirů velmi běžný jev. Vzorek 7 jsme z analýzy sekvence proteinu 2C vyřadili protože, se v něm vyskytovalo hodně stop kodónů, což může být následkem chybné sekvenace či rekombinace viru, která naruší čtecí rámec.



Obr. 23 *Srovnání proteinu VP1*



Obr. 24 *Srovnání proteinu 2A*



Obr. 25 *Srovnání proteinu 2C*

5. Diskuze

Hledání příčinné souvislosti mezi enterovirovými infekcemi a rozvojem diabetu 1. typu je již po mnoho let záměrem mnoha výzkumných skupin. Zatím se však role enterovirů v etiologii tohoto onemocnění nepodařila řádně vysvětlit, navzdory sérologickým důkazům (Stene & Rewers, 2012) i přímé detekci enterovirové RNA v různých typech tkání (Yeung, Rawlinson, & Craig, 2010). Sérologické důkazy vykazují ve svých závěrech značnou heterogenitu. Přímá detekce RNA je ve svých závěrech konkrétnější: enterovirová RNA ve vzorcích stolice není častější u pacientů oproti kontrolám, zatímco detekce enterovirů ve tkáni pankreatu či ve vzorcích krve vyšší frekvenci u pacientů vykazuje.

S vývojem nových přístupů a metod molekulární biologie získáváme nové možnosti, jak zkoumat souvislost mezi enteroviry a diabetem 1. typu. Především se jedná o průlomovou metodu NGS.

5.1. Role NGS ve výzkumu viromu

Ještě před několika lety ve výzkumu virových sekvencí nebyla jiná možnost, než virus izolovat na tkáňové kultuře či se pokusit suklonování jeho sekvence náhodně fragmentované nukleové kyseliny viru. Tehdy to znamenalo použít klasické buněčné klonování; sekvenování několika set subklonů bylo prací na týdny a šance, že že ani jeden neobsahuje virovou sekvenci, nebyla zanedbatelná (Victoria, Kapoor, Dupuis, Schnurr, & Delwart, 2008).

Sekvenování nové generace přineslo revoluční změnu i do výzkumu viromu. Fragmentace, částečná amplifikace a následná sekvenace na moderních NGS platformách nyní ze vzorku poskytuje desetitisíce až milióny čtení pocházejících ze všech nukleových kyselin, které jsme ve vzorku přítomny.

Na rozdíl od analýzy transkriptomu nebo genomu lidí či laboratorních organismů není analýza viromu závislá na předchozí znalosti sekvence viru. Zatímco lidský genom je polymorfní v každé cca tisící bázi (bereme-li pouze signifikantní frekvenční polymorfismy), viry se mohou lišit o desítky procent své sekvence, i když jsou stejného druhu a sérotypu. Virus žijící v neustálém zápasu s imunitou neustále mutuje; ty varianty, které jsou selekčně výhodné, pak predominují.

V analýze viromu neexistuje nic, co by bylo stálé, neměnné a spolehlivé, nic podobného stabilní lidské referenční sekvenci nebo invariantním úsekům genu 16S bakteriální

ribozomální RNA. Proto je analýza viromu technicky i bioinformaticky zajímavým problémem.

Hodnocení, sestavení a doplnění dat ze sekvenace viromu se věnovala i má práce.

5.2. Hodnocení dat z NGS

Samotná optimalizace procesů během NGS je náročná, stejně tak práce s výstupními daty této metody. Právě analýzou těchto dat a jejich optimalizací jsme se zabývali v této práci. Je to problematika velice komplikovaná, jelikož neexistuje pouze jedno správné řešení a doposud chybí publikované práce pro analýzu viromu tímto způsobem.

Usnadněním při rozpoznávání nového viru je sestavení čtení, která se překrývají, do jednoho dlouhého kontigu. Výstupy z programu Galaxy a Velvet jsou výsledné kontigy. Čím je nižší komplexita nukleových kyselin ve zkoumaném vzorku, tím je počet čtení, která jsou sestavena do kontigu, vyšší (Li & Delwart, 2011).

Tím míním, že pokud je ve vzorku přítomen nižší počet organismů, ale ve vysoké kvantitě, bude reprezentace (pokrytí) jejich sekvencí vysoká. Naopak vzorek s vysokou diverzitou organismů, kde každý je přítomen v nižší kvantitě, bude spíše podléhat stochastickým jevům; některé části genomů pak pokryty nebudou, jiné řídce.

To odpovídá i našim výsledkům, kdy se sekvence enterovirů pěstovaných na buněčných kulturách sestavily do kontigů daleko delších než sekvence pocházející z reálných vzorků stolic. Také první zmiňované pokryly téměř celý genom referenční sekvence, zatímco mezi kontigy zbylých vzorků byly četné mezery.

Při filtraci a úpravě dat reálných vzorků stolic v programu Galaxy, byl úbytek sekvencí cca 50%. To je v kontextu množství dat ze vzorku s velkou diverzitou, jako je stolice, přípustný úbytek. Téměř stejný úbytek sekvencí (52,5%) filtrováním byl uveden i v obdobné studii (Donaldson et al., 2010).

Při sestavování sekvencí v jasně definovaných kontizích skončilo 50 – 80% dat. Sekvence, které nebyly sestaveny do kontigů, jsme též podrobili namapování na enterovirový genom. Mezi nimi se skutečně vyskytovaly i některé enterovirové. Ty jsme následně využili při sestavování genomu de-novo.

To, že některé sekvence nenalezly své protějšky a do kontigů se nesestavily, může mít několik důvodů. Prvním je nízká kvantita příslušného organismu. Druhým je specifický

metagenomový problém při infekci několika příbuznými viry najednou mohou být do jednoho kontigu sestaveny homologní oblasti obou dvou virů, kdežto heterologní oblasti se vydělí jako nesestavená čtení. Konečně třetí důvod je čistě technický – programy pro de-novo assembly se signifikantně liší úspěšností svých algoritmů (Miller, Koren, & Sutton, 2010), ačkoli právě Velvet si v nezávislém hodnocení vede poměrně dobře.

V případě některých vzorků byla tato hodnota nižší. Následně jsme zjistili, že vzniklé kontigy se vůbec nenamapovaly na enterovirovou sekvenci. V porovnání s ostatními vzorky byla i původní kvantita enterovirové RNA v těchto vzorcích o tři řády nižší.

Proč v těchto vzorcích nenalezlo NGS žádný enterovirus, ačkoli v real-time PCR detekován byl, jakkoli jej bylo málo? Je třeba mít na paměti, že hodnocení viromu v naší, ale i většině ostatních studií, je založeno na poměrně omezeném počtu čtení – mezi 100 000 a miliónem. Je-li ve vzorku přítomen např. bakteriofág ve vysoké kvantitě několika miliónů na mikrolitr, není velká šance, aby se do náhodného procesu přípravy vzorků vnutil enterovirus o koncentraci deset kopií na mikrolitr – signál bude prostě tvořen bakteriofágy. To je v přímém kontrastu ke specifickému PCR, kde si primer nakonec svůj cíl většinou najde, je-li přítomen, a úspěšně jej detekují.

Identifikace sérotypu jsme prováděli v programu Enterovirus genotyping tool, kde se vzniklé kontigy namapovaly na referenční enterovirovou sekvenci. Zde byl u prvních deseti vzorků určen sérotyp coxsackieviru B3. U zbylých vzorků se druhy a sérotypy referenčních enterovirů lišily. To lze odůvodnit tím, že se jednalo o enteroviry pocházející z reálných vzorků stolic zdravých jedinců, kde se vyskytují různé druhy enterovirů.

V naší laboratoři se toto nastavení nyní používá univerzálně na všechny vzorky, které byly sekvenovány NGS.

5.3. Charakterizace virového genomu

Bez znalosti celých či téměř celých enterovirových genomů bychom viry nebyli schopni charakterizovat na úrovni sekvence aminokyselin. Pouze v tom případě můžeme pozorovat změny mezi jednotlivými kmeny virů, které mohou hrát roli v jejich rozdílné virulenci a v našem případě i v jejich potenciální schopnosti způsobit diabetes.

K získání celého virového genomu, jak jsme se sami přesvědčili, nestačí vzorky sekvenovat pouze NGS metodou. Ve většině případů data nepokryjí celou délku genomu. Proto jsme mezery přemostili Sangerovým sekvenováním. Touto metodou je v některých případech

nutno resekvenovat i oblasti pokryté daty z NGS s velkou variabilitou mezi sekvencemi či nízkým pokrytím.

Fylogenetická analýza byla provedena u vybraných proteinů (VP1, 2A a 2C). Na základě této analýzy jsme zjistili, že naše vzorky pocházejí ve fylogenetickém stromě ze společného uzlu a utvářejí monofyletický vztah. Obecně představují evoluční skupinu se společnou historií a jsou díky původu všechny příbuzné.

Přítomnost stejný neklasifikovatelných sekvencí u různých pacientů může vést k detekci nové virové rodiny, avšak je potřeba získat přímý důkaz virové replikace, jako je amplifikace in vitro či in vivo a sérokonverze u exponovaných jedinců (Li & Delwart, 2011).

Detekce viru v krvi či tkáni může být považována za důkaz virové replikace. Je obtížné říci, jaké množství virových částic se může pasivně dostat do vnitřních tkání a orgánů. Virová replikace na lidských buněčných kulturách také může odrážet tropismus virů k dané tkáni, ale protože překážky, které se v hostitelích vyskytují, mohou být v buněčných kulturách obejity a mnoho virů může růst i v kulturách, které nepocházejících od daného druhu, in vitro. Proto replikace nemůže být považována za rozhodující důkaz.

Autoimunitní choroby, jako diabetes 1. typu, mohou být indukovány viry, které nejsou zatím identifikovány. Metagenomika virů nabízí jednoduchý nástroj k identifikování kandidátních patogenů pro tato onemocnění. Otázkou však zůstává, zdali tyto "nově" objevené viry, které byly nalezeny v klinických vzorcích pacientů, nejsou pouze náhodně spojovány s onemocněním. Mohou pouze odrážet neškodnou a běžnou infekci. Prokázání jisté patogenicity viru lze až detekcí virově specifických protilátek, které prokáží replikaci viru u daného jedince.

5.4. Studium viromu

Ve srovnání s bakteriálním mikrobiomem je studium viromu, tedy souboru všech virů ve vzorku, nesnadným úkolem. U virů neexistuje konzervovaná oblast společná všem virům, která by dovolovala širokou amplifikaci všech virových genomů. Studium lidského viromu zahrnuje popis virové komunity v lidském těle, včetně bakterifágů, a jejich vztah ke zdraví a nemoci.

Lidský mikrobiom je úplná populace mikrobů (bakterií, parazitů, hub a virů), která kolonizuje lidské tělo. V případě metagenomických studií bakterií je analýza ulehčena využitím univerzálního a konzervovaného cíle, kterým je hlavně 16S rRNA gen. Tento gen má

konzervované oblasti, které mohou být využity pro PCR primery a mezi těmito oblastmi leží variabilní sekvence, které dovolují identifikaci rodu a druhu (Petrosino, Highlander, Luna, Gibbs, & Versalovic, 2009; Hamady & Knight, 2009).

V případě metagenomických studií virů došlo ke zlepšení až s příchodem NGS. Tím se obešla potřeba předchozí virové amplifikace in vitro nebo in vivo. Využitím NGS se metagenomika virů zaměřuje především na objev virů, které mohou být novými patogeny, sekvenování virových variant známých virů, což může vést k lepšímu pochopení jejich evoluce, nebo také zkoumání virové ekologie. Také lze dosáhnout objektivního průzkumu virové komunity bez snížení diverzity během amplifikace v buněčných kulturách.

V diagnostice může být metagenomika viromu důležitá při systematických analýzách vzorků sbíraných od pacientů s neobjasněnou chorobou, zvláště v kontextu šíření onemocnění a epidemie (Svraka et al., 2010).

V souvislosti s propuknutím onemocnění je nezbytné, aby bylo charakterizováno infekční agens, nejen k lepšímu porozumění choroby, ale především k vývoji specifických diagnostických testů a kontrolních měření.

Využití virové metagenomiky pro diagnostické účely se zdá být slibné, avšak je potřeba vyřešit současné překážky, jako je vysoká složitost a relativně nízká rychlost a obtížná interpretace dat. Protokol náhodné amplifikace a konstrukce DNA knihovny je nyní složitý, ale pravděpodobně bude zjednodušen a urychlen v blízké budoucnosti. Samotný čas sekvenace by měl být také redukován. Postup, který byl popsán v této práci, také přispívá k jistému posunu dopředu v aplikaci NGS a vyhodnocování jeho dat v oblasti identifikace virů.

5.5. Aplikace využití vytvořeného protokolu

Stejný postup, který jsme optimalizovali na vzorcích popsaných v této diplomové práci, jsme použili i při analýze vzorků stolic, jež pocházely od zdravých malawských dětí ve věku od 6 až 12 měsíců. Analýza těchto vzorků byla provedena, kvůli závažnému podezření na více virů ve vzorku a možnou simultánní infekci.

Vzorků bylo celkem 16, vždy jeden vzorek od jednoho dítěte. Všechny tyto vzorky byly pozitivní na enterovirus, a byly sekvenovány pomocí metody NGS. Následná analýza byla provedena tak, jak je popsáno v této diplomové práci. Byla provedena filtrace sekvencí v Galaxy na základě kvality sekvencí, odstraněny báze na koncích sekvencí, které měly nerovnoměrný obsah bází, a dále byly sestaveny do kontigů pomocí programu Velvet.

Rozdíl, oproti vzorkům, které jsme zkoumali a jež jsou popsány v této práci, byl u vzorků z Malawi ve výstupních datech. Získávali jsme velký počet sekvencí, které nevytvořily kontig, ačkoli se namapovaly na referenční genom a byly téměř identické. V tomto případě jsme mezery mezi kontigy nepřeklenovali. Zajímalo nás, jaké všechny viry jsme schopni ze vzorku identifikovat. Pro identifikaci všech virů v těchto vzorcích jsme museli kromě identifikace kontigů v Enterovirus genotyping tool využít lokální databáze NCBI Blast. Tímto způsobem se nám podařilo identifikovat řadu virů a sestavit jejich genomy de novo, viz. tabulka 13.

Tím jsme dokázali, že i u vzorku s vysokou diverzitou, jako je stolice dětí, jejichž hygienické návyky jsou stížené místními podmínkami oproti dětem evropským, jsme schopni identifikovat velké množství virů, bez podezření na jejich přítomnost ve vzorku.

Tab. 13 *Předpokládané sérotypy podle analýzy sekvencí. Přehled virů identifikovaných ve vzorcích stolic získaných od malawských dětí. Vytvořený protokol dokáže oddělit i vícečetné směsné virové populace v jednom vzorku.*

Vzorek	Enterovirus	Parechovirus	Cosavirus	Další
1	Echovirus 24	Parechovirus (1,2 a 3)	Cosavirus A19	
2	Coxsackie A22			Bocavirus
3	Enterovirus B			
4	Echovirus 11, Coxsackie A9		Cosavirus A	
5	Enterovirus B, Echovirus 19	Parechovirus (1,4 a 5)		
6	Coxsackie A24,Poliovirus 2			
7	Enterovirus 71			Bocavirus, Norovirus
8	Coxackie A5		Cosavirus A	
9	Enterovirus 88, Echovirus 3		Cosavirus A	Bocavirus
10		Parechovirus (1,3 a 4)		Bocavirus
11	Enterovirus71,CoxsackieB4			
12	Enterovirus B,Echovirus 25			
13	Enterovirus79,Coxsackie A4			
14	Coxsackie B5		Cosavirus A	
15	Enterovirus79,CoxsackieA4			
16	Enterovirus71,CoxsackieB4			

5.6. Další možné využití NGS

NGS technologie se stávají dostupnější v posledních několika letech a stále dochází k jejímu vývoji a zlepšování. Je široce využívaná v mnoha projektech: celogenomové sekvenování, metagenomice, objevu malých RNA a RNA sekvenování. Společným znakem je extrémní výkonost generování dat.

Potenciální využití NGS a následné sestavování genomů de novo z biologických vzorků se v budoucnosti může uplatnit v mnoha oblastech dalšího výzkumu virů. Konkrétním přínosem může být studium interakcí všech virů ve vzorku s ohledem na klinický stav pacienta. To znamená, že pokud budeme znát celý virom od pacienta s neznámou chorobou, můžeme vyvozovat možné příčiny vzniku jeho onemocnění v souvislosti s virovou infekcí. Další otázkou zůstává, zdali přítomnost koinfekce může zhoršit symptomy. Jestliže se rozvoj onemocnění zvyšuje v kontextu jiných infekcí, potom pouze celkový počet infekcí a konkrétní kombinace virů může být asociována s těmito symptomy. K rozluštění těchto komplexnícxh interakcí může právě přispět studium viromu ve vzorcích pacientů více než testy zaměřující se na jeden virus.

Virologické aplikace NGS jsou slibné a výsledků je dosahováno v objevu a charakterizaci nových virů, detekci neznámých virových patogenů v klinických vzorcích, vyšetřování virové diverzity, evoluce, rozšíření a hodnocení lidského viromu (Barzon et al., 2013).

6. Závěr

Teoretická část diplomové práce shrnuje dosud známé poznatky o možném vlivu enterovirů v etiologii onemocnění diabetu 1. typu, jehož incidence celosvětově narůstá. Jako jedním ze správných řešení sledování vlivu virů na onemocnění, jsou longitudinální studie, kdy jsou vzorky odebírány predisponovaným jedincům, stejně jako kontrolním subjektům v určitém časovém rozmezí.

V praktické části jsme se zabývali vytvořením strategie zpracování výstupních dat z NGS. Zpracovávali jsme 22 vzorků. Data prošla procesem filtrace, úpravy konců sekvencí, sestavení sekvencí do kontigů a namapování kontigů na referenční genom Tato data však nepokrývají celý virový genom. Proto jsme mezery mezi jednotlivými kontigy překlenuli pomocí Sangerova sekvenování. U vzniklých genomů jsme identifikovali jednotlivé proteinové sekvence. Poté jsme provedli fylogenetickou analýzu proteinů VP1, 2A a 2C.

Výsledkem praktické části je protokol nyní používaný v analýze dalších vzorků viromu ve studiích vztahu viromu k prediabetické autoimunitě.

7. Reference

Achenbach, P., Bonifacio, E., Koczwara, K., & Ziegler, A. (2005). Natural History of Type 1 Diabetes THE NATURAL HISTORY OF TYPE 1 DIABETES, (12), 25–31.

Badorff, C., Lee, G. H., Lamphear, B. J., Martone, M. E., Campbell, K. P., Rhoads, R. E., & Knowlton, K. U. (1999). Enteroviral protease 2A cleaves dystrophin: evidence of cytoskeletal disruption in an acquired cardiomyopathy. *Nature Medicine*, *5*(3), 320–6. doi:10.1038/6543

Barratt, B. J., Payne, F., Lowe, C. E., Hermann, R., Healy, B. C., Harold, D., … Todd, J. A. (2004). Remapping the insulin gene/IDDM2 locus in type 1 diabetes. *Diabetes*, *53*(7), 1884–9.

Barboni, E., Manocchio, I., & Asdrubali, G. (n.d.). [Observations on diabetes in cattle due to experimental epizootic aphthae (Preliminary note)]. *Nuovi Annali D'igiene E Microbiologia*, *17*(3), 223–6.

Barker, J. M., Barriga, K. J., Yu, L., Miao, D., Erlich, H. A., Norris, J. M., … Rewers, M. (2004). Prediction of autoantibody positivity and progression to type 1 diabetes: Diabetes Autoimmunity Study in the Young (DAISY). *The Journal of Clinical Endocrinology and Metabolism*, *89*(8), 3896–902. doi:10.1210/jc.2003-031887

Barzon, L., Lavezzo, E., Militello, V., Toppo, S., & Palù, G. (2011). Applications of next-generation sequencing technologies to diagnostic virology. *International Journal of Molecular Sciences*, *12*(11), 7861–84. doi:10.3390/ijms12117861

Barzon, L., Lavezzo, E., Costanzi, G., Franchin, E., Toppo, S., & Palù, G. (2013). Next-generation sequencing technologies in diagnostic virology. *Journal of Clinical Virology : The Official Publication of the Pan American Society for Clinical Virology*, *58*(2), 346–50. doi:10.1016/j.jcv.2013.03.003

Bergholdt, R., Brorsson, C., Palleja, A., Berchtold, L. A., Fløyel, T., Bang-Berthelsen, C. H., … Pociot, F. (2012). Identification of novel type 1 diabetes candidate genes by integrating genome-wide association data, protein-protein interactions, and human pancreatic islet gene expression. *Diabetes*, *61*(4), 954–62. doi:10.2337/db11-1263

Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., & Nekrutenko, A. (2010). Manipulation of FASTQ data with Galaxy. *Bioinformatics (Oxford, England)*, *26*(14), 1783–5. doi:10.1093/bioinformatics/btq281

Bottini, N., Musumeci, L., Alonso, A., Rahmouni, S., Nika, K., Rostamkhani, M., … Mustelin, T. (2004). A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nature Genetics*, *36*(4), 337–8. doi:10.1038/ng1323

Bouças, A. P., de Oliveira, F. dos S., Canani, L. H., & Crispim, D. (2013). The role of interferon induced with helicase C domain 1 (IFIH1) in the development of type 1 diabetes mellitus. *Arquivos Brasileiros de Endocrinologia E Metabologia*, *57*(9), 667–76.

Cai, Q., Yameen, M., Liu, W., Gao, Z., Li, Y., Peng, X., … Lin, T. (2013). Conformational plasticity of the 2A proteinase from enterovirus 71. *Journal of Virology*, *87*(13), 7348–56. doi:10.1128/JVI.03541-12

Cock, P. J. a, Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Research*, *38*(6), 1767–71. doi:10.1093/nar/gkp1137

Coppieters, K. T., Boettler, T., & von Herrath, M. (2012). Virus infections in type 1 diabetes. *Cold Spring Harbor Perspectives in Medicine*, *2*(1), a007682. doi:10.1101/cshperspect.a007682

Daneman, D. (2006). Type 1 diabetes. *Lancet*, *367*(9513), 847–58. doi:10.1016/S0140-6736(06)68341-4

Donaldson, E. F., Haskew, A. N., Gates, J. E., Huynh, J., Moore, C. J., & Frieman, M. B. (2010). Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. *Journal of Virology*, *84*(24), 13004–18. doi:10.1128/JVI.01255-10

Efron, B., Halloran, E., & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(23), 13429–34.

Fields, B. (2007). *Fields virology*. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Filippi, C. M., & von Herrath, M. G. (2010). 99th Dahlem conference on infection, inflammation and chronic inflammatory disorders: viruses, autoimmunity and immunoregulation. *Clinical and Experimental Immunology*, *160*(1), 113–9. doi:10.1111/j.1365-2249.2010.04128.x

Flanegan, J. B., Petterson, R. F., Ambros, V., Hewlett, N. J., & Baltimore, D. (1977). Covalent linkage of a protein to a defined nucleotide sequence at the 5'-terminus of virion and replicative intermediate RNAs of poliovirus. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(3), 961–5.

Forss, S., & Schaller, H. (1982). A tandem repeat gene in a picornavirus. *Nucleic Acids Research*, *10*(20), 6441–50.

Fujinami, R. S., von Herrath, M. G., Christen, U., & Whitton, J. L. (2006). Molecular mimicry, bystander activation, or viral persistence: infections and autoimmune disease. *Clinical Microbiology Reviews*, *19*(1), 80–94. doi:10.1128/CMR.19.1.80-94.2006

Gamble, D. R., Kinsley, M. L., FitzGerald, M. G., Bolton, R., & Taylor, K. W. (1969). Viral antibodies in diabetes mellitus. *British Medical Journal*, *3*(5671), 627–30.

Giardine, B. (2005). Galaxy: A platform for interactive large-scale genome analysis. *Genome Research*, *15*(10), 1451–1455. doi:10.1101/gr.4086505

Hall, B. G. (2005). Comparison of the accuracies of several phylogenetic methods using protein and DNA sequences. *Molecular Biology and Evolution*, *22*(3), 792–802. doi:10.1093/molbev/msi066

Hall, B. G. (2011). Phylogenetic Trees Made Easy: A How-To Manual.

Hamady, M., & Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research*, *19*(7), 1141–52. doi:10.1101/gr.085464.108

Hober, D., & Sauter, P. (2010). Pathogenesis of type 1 diabetes mellitus: interplay between enterovirus and host. *Nature Reviews. Endocrinology*, *6*(5), 279–89. doi:10.1038/nrendo.2010.27

Härkönen, T., Paananen, A., Lankinen, H., Hovi, T., Vaarala, O., & Roivainen, M. (2003). Enterovirus infection may induce humoral immune response reacting with islet cell autoantigens in humans. *Journal of Medical Virology*, *69*(3), 426–40. doi:10.1002/jmv.10306

Hiltunen, M., Hyoty, H., Knip, M., Ilonen, J., Reijonen, H., Viihiisalo, P., … Akerblom, H. K. (1997). Islet Cell Antibody Seroconversion in Children Is Temporally Associated with Enterovirus Infections, 554–560.

Hober, D., & Sane, F. (2010). Enteroviral Pathogenesis of Type 1 Diabetes. *Discovery Medicine*, *10*(51), 151–160.

Hober, D., & Sauter, P. (2010). Pathogenesis of type 1 diabetes mellitus: interplay between enterovirus and host. *Nature Reviews. Endocrinology*, *6*(5), 279–89. doi:10.1038/nrendo.2010.27

Hyöty, H., Hiltunen, M., Knip, M., Laakkonen, M., Vähäsalo, P., Karjalainen, J., … Hovi, T. (1995). A prospective study of the role of coxsackie B and other enterovirus infections in the pathogenesis of IDDM. Childhood Diabetes in Finland (DiMe) Study Group. *Diabetes*, *44*(6), 652–7.

Chapman, N. M., & Kim, K. S. (2008). Persistent coxsackievirus infection: enterovirus persistence in chronic myocarditis and dilated cardiomyopathy. *Current Topics in Microbiology and Immunology*, *323*, 275–92.

Jaïdane, H., Sané, F., Gharbi, J., Aouni, M., Romond, M. B., & Hober, D. (2009). Coxsackievirus B4 and type 1 diabetes pathogenesis: contribution of animal models. *Diabetes/metabolism Research and Reviews*, *25*(7), 591–603. doi:10.1002/dmrr.995

Johnson, K. L., & Sarnow, P. (1991). Three poliovirus 2B mutants exhibit noncomplementable defects in viral RNA amplification and display dosage-dependent dominance over wild-type poliovirus. *Journal of Virology*, *65*(8), 4341–9.

Jukes, T., & Cantor, C. (1969). Evolution of Protein Molecules.

Kato, H., Takeuchi, O., Sato, S., Yoneyama, M., Yamamoto, M., Matsui, K., … Akira, S. (2006). Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature*, *441*(7089), 101–5. doi:10.1038/nature04734

Knip, M. (2011). Pathogenesis of type 1 diabetes: implications for incidence trends. *Hormone Research in Pædiatrics*, *76 Suppl 1*, 57–64. doi:10.1159/000329169

Krogvold, L., Edwin, B., Buanes, T., Ludvigsson, J., Korsgren, O., Hyöty, H., … Dahl-Jørgensen, K. (2014). Pancreatic biopsy by minimal tail resection in live adult patients at the onset of type 1 diabetes: experiences from the DiViD study. *Diabetologia*, *57*(4), 841–3. doi:10.1007/s00125-013-3155-y

Kroneman, A., Vennema, H., Deforche, K., v d Avoort, H., Peñaranda, S., Oberste, M. S., … Koopmans, M. (2011). An automated genotyping tool for enteroviruses and noroviruses.

Journal of Clinical Virology : The Official Publication of the Pan American Society for Clinical Virology, *51*(2), 121–5. doi:10.1016/j.jcv.2011.03.006

Kupila, A., Muona, P., Simell, T., Arvilommi, P., Savolainen, H., Hämäläinen, A. M., … Simell, O. (2001). Feasibility of genetic and immunological prediction of type I diabetes in a population-based birth cohort. *Diabetologia*, *44*(3), 290–7.

Laitinen, O. H., Honkanen, H., Pakkanen, O., Oikarinen, S., Hankaniemi, M. M., Huhtala, H., … Hyöty, H. (2014). Coxsackievirus B1 is associated with induction of β-cell autoimmunity that portends type 1 diabetes. *Diabetes*, *63*(2), 446–55. doi:10.2337/db13-0619

Lauwers, S., Bissay, V., & Rombaut, B. (1998). Development of an enterovirus specific PCR method for the quantification of enterovirus genomes in blood of diabetes patients. *Clinical and Diagnostic Virology*, *9*(2-3), 135–9.

Li, L., & Delwart, E. (2011). From orphan virus to pathogen: the path to the clinical lab. *Current Opinion in Virology*, *1*(4), 282–8. doi:10.1016/j.coviro.2011.07.006

Lloyd, R. E., Grubman, M. J., & Ehrenfeld, E. (1988). Relationship of p220 cleavage during picornavirus infection to 2A proteinase sequencing. *Journal of Virology*, *62*(11), 4216–23.

Lönnrot, M., Korpela, K., Knip, M., Ilonen, J., Simell, O., Korhonen, S., … Hyöty, H. (2000). Enterovirus infection as a risk factor for beta-cell autoimmunity in a prospectively observed birth cohort: the Finnish Diabetes Prediction and Prevention Study. *Diabetes*, *49*(8), 1314–8.

Lönnrot, M., Salminen, K., Knip, M., Savola, K., Kulmala, P., Leinikki, P., … Hyöty, H. (2000). Enterovirus RNA in serum is a risk factor for beta-cell autoimmunity and clinical type 1 diabetes: a prospective study. Childhood Diabetes in Finland (DiMe) Study Group. *Journal of Medical Virology*, *61*(2), 214–20.

Martínek, P., Stehlík, J., Grossmann, P., Ka, J., & Vaněček, T. (2013). Sekvenování – klasická metodika, 2013.

Miller, J. R., Koren, S., & Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics*, *95*(6), 315–27. doi:10.1016/j.ygeno.2010.03.001

Milne, I., Bayer, M., Cardle, L., Shaw, P., Stephen, G., Wright, F., & Marshall, D. (2010). Tablet--next generation sequence assembly visualization. *Bioinformatics (Oxford, England)*, *26*(3), 401–2. doi:10.1093/bioinformatics/btp666

Mueller, S., Wimmer, E., & Cello, J. (2005). Poliovirus and poliomyelitis: a tale of guts, brains, and an accidental event. *Virus Research*, *111*(2), 175–93. doi:10.1016/j.virusres.2005.04.008

Mullis, K. B., & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, *155*, 335–50.

Nei, M., & Kumar, S. (2000). *Molecular Evolution and Phylogenetics*.

Nokoff, N., & Rewers, M. (2013). Pathogenesis of type 1 diabetes: lessons from natural history studies of high-risk individuals. *Annals of the New York Academy of Sciences*, *1281*, 1–15. doi:10.1111/nyas.12021

Nurminen, N., Oikarinen, S., & Hyöty, H. (2012). Virus infections as potential targets of preventive treatments for type 1 diabetes. *The Review of Diabetic Studies : RDS*, *9*(4), 260–71. doi:10.1900/RDS.2012.9.260

Oberste, M. S., Maher, K., Kilpatrick, D. R., & Pallansch, M. a. (1999). Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. *Journal of Virology*, *73*(3), 1941–8.

Oikarinen, M., Tauriainen, S., Honkanen, T., Oikarinen, S., Vuori, K., Kaukinen, K., … Hyöty, H. (2008). Detection of enteroviruses in the intestine of type 1 diabetic patients. *Clinical and Experimental Immunology*, *151*(1), 71–5. doi:10.1111/j.1365-2249.2007.03529.x

Oikarinen, M., Tauriainen, S., Oikarinen, S., Honkanen, T., Collin, P., Rantala, I., … Hyöty, H. (2012). Type 1 diabetes is associated with enterovirus infection in gut mucosa. *Diabetes*, *61*(3), 687–91. doi:10.2337/db11-1157

Oikarinen, S., Martiskainen, M., Tauriainen, S., Huhtala, H., Ilonen, J., Veijola, R., … Hyöty, H. (2011). Enterovirus RNA in blood is linked to the development of type 1 diabetes. *Diabetes*, *60*(1), 276–9. doi:10.2337/db10-0186

Paszkiewicz, K., & Studholme, D. J. (2010). De novo assembly of short sequence reads. *Briefings in Bioinformatics*, *11*(5), 457–72. doi:10.1093/bib/bbq020

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A., & Versalovic, J. (2009). Metagenomic pyrosequencing and microbial identification. *Clinical Chemistry*, *55*(5), 856–66. doi:10.1373/clinchem.2008.107565

Pettersson, E., Lundeberg, J., & Ahmadian, A. (2009). Generations of sequencing technologies. *Genomics*, *93*(2), 105–11. doi:10.1016/j.ygeno.2008.10.003

Rajtar, B., Majek, M., Polański, Ł., & Polz-Dacewicz, M. (2008). Enteroviruses in water environment--a potential threat to public health. *Annals of Agricultural and Environmental Medicine : AAEM*, *15*(2), 199–203.

Richardson, S. J., Willcox, A., Bone, A. J., Foulis, A. K., & Morgan, N. G. (2009). The prevalence of enteroviral capsid protein vp1 immunostaining in pancreatic islets in human type 1 diabetes. *Diabetologia*, *52*(6), 1143–51. doi:10.1007/s00125-009-1276-0

Richer, M. J., & Horwitz, M. S. (2009). Coxsackievirus infection as an environmental factor in the etiology of type 1 diabetes. *Autoimmunity Reviews*, *8*(7), 611–5. doi:10.1016/j.autrev.2009.02.006

Roivainen, M., & Klingel, K. (2010). Virus infections and type 1 diabetes risk. *Current Diabetes Reports*, *10*(5), 350–6. doi:10.1007/s11892-010-0139-x

Roivainen, M., Knip, M., Hyöty, H., Kulmala, P., Hiltunen, M., Vähäsalo, P., … Akerblom, H. K. (1998). Several different enterovirus serotypes can be associated with prediabetic autoimmune episodes and onset of overt IDDM. Childhood Diabetes in Finland (DiMe) Study Group. *Journal of Medical Virology*, *56*(1), 74–8.

Roivainen, M. (2006). Enteroviruses: new findings on the role of enteroviruses in type 1 diabetes. *The International Journal of Biochemistry & Cell Biology*, *38*(5-6), 721–5. doi:10.1016/j.biocel.2005.08.019

Ryan, M. D., & Flint, M. (1997). Virus-encoded proteinases of the picornavirus super-group. *The Journal of General Virology*, *78 (Pt 4)*, 699–723.

Sadeharju, K., Knip, M., Virtanen, S. M., Savilahti, E., Tauriainen, S., Koskela, P., … Hyöty, H. (2007). Maternal antibodies in breast milk protect the child from enterovirus infections. *Pediatrics*, *119*(5), 941–6. doi:10.1542/peds.2006-0780

Sadeharju, K., Lönnrot, M., Kimpimäki, T., Savola, K., Erkkilä, S., Kalliokoski, T., … Hyöty, H. (2001). Enterovirus antibody levels during the first two years of life in prediabetic

autoantibody-positive children. *Diabetologia*, *44*(7), 818–23. doi:10.1007/s001250100560

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406–25.

Salminen, K., Sadeharju, K., Lönnrot, M., Vähäsalo, P., Kupila, A., Korhonen, S., … Hyöty, H. (2003). Enterovirus infections are associated with the induction of beta-cell autoimmunity in a prospective birth cohort study. *Journal of Medical Virology*, *69*(1), 91–8. doi:10.1002/jmv.10260

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–7.

Sarmiento, L., Frisk, G., Anagandula, M., Cabrera-Rode, E., Roivainen, M., & Cilio, C. M. (2013). Expression of innate immunity genes and damage of primary human pancreatic islets by epidemic strains of Echovirus: implication for post-virus islet autoimmunity. *PloS One*, *8*(11), e77850. doi:10.1371/journal.pone.0077850

Seissler, J., & Scherbaum, W. A. (2006). Autoimmune diagnostics in diabetes mellitus. *Clinical Chemistry and Laboratory Medicine□: CCLM / FESCC*, *44*(2), 133–7. doi:10.1515/CCLM.2006.025

Soltis, P. S., & Soltis, D. E. (2003). Applying the Bootstrap in Phylogeny Reconstruction. *Statistical Science*, *18*(2), 256–267.

Steck, A. K., Zhang, W., Bugawan, T. L., Barriga, K. J., Blair, A., Erlich, H. A., … Rewers, M. J. (2009). Do non-HLA genes influence development of persistent islet autoimmunity and type 1 diabetes in children with high-risk HLA-DR,DQ genotypes? *Diabetes*, *58*(4), 1028–33. doi:10.2337/db08-1179

Steck, A. K., Wong, R., Wagner, B., Johnson, K., Liu, E., Romanos, J., … Rewers, M. J. (2012). Effects of non-HLA gene polymorphisms on development of islet autoimmunity and type 1 diabetes in a population with high-risk HLA-DR,DQ genotypes. *Diabetes*, *61*(3), 753–8. doi:10.2337/db11-1228

Stene, L. C., Oikarinen, S., Hyöty, H., Barriga, K. J., Norris, J. M., Klingensmith, G., … Rewers, M. (2010). Enterovirus infection and progression from islet autoimmunity to type 1 diabetes: the Diabetes and Autoimmunity Study in the Young (DAISY). *Diabetes*, *59*(12), 3174–80. doi:10.2337/db10-0866

Stene, L. C., & Rewers, M. (2012). Immunology in the clinic review series; focus on type 1 diabetes and viruses: the enterovirus link to type 1 diabetes: critical review of human studies. *Clinical and Experimental Immunology*, *168*(1), 12–23. doi:10.1111/j.1365-2249.2011.04555.x

Stephenson, F. H. (2010). *Calculations for Molecular Biology and Biotechnology: A Guide to Mathematics in the Laboratory 2e*.

Svraka, S., Rosario, K., Duizer, E., van der Avoort, H., Breitbart, M., & Koopmans, M. (2010). Metagenomic sequencing for virus identification in a public-health setting. *The Journal of General Virology*, *91*(Pt 11), 2846–56. doi:10.1099/vir.0.024612-0

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731–9. doi:10.1093/molbev/msr121

Tanaka, S., Aida, K., Nishida, Y., & Kobayashi, T. (2013). Pathophysiological mechanisms involving aggressive islet cell destruction in fulminant type 1 diabetes. *Endocrine Journal*, *60*(7), 837–45.

Tapia, G., Cinek, O., Rasmussen, T., Witsø, E., Grinde, B., Stene, L. C., & Rønningen, K. S. (2011). Human enterovirus RNA in monthly fecal samples and islet autoimmunity in Norwegian children with high genetic risk for type 1 diabetes: the MIDIA study. *Diabetes Care*, *34*(1), 151–5. doi:10.2337/dc10-1413

Tauriainen, S., Oikarinen, S., Oikarinen, M., & Hyöty, H. (2011). Enteroviruses in the pathogenesis of type 1 diabetes. *Seminars in Immunopathology*, *33*(1), 45–55. doi:10.1007/s00281-010-0207-y

Thoelen, I., Moe, E., Lemey, P., Mostmans, S., Wollants, E., Lindberg, A. M., … Ranst, M. Van. (2004). Analysis of the Serotype and Genotype Correlation of VP1 and the 5 J Noncoding Region in an Epidemiological Survey of the Human Enterovirus B Species, *42*(3), 963–971. doi:10.1128/JCM.42.3.963

Toyoda, H., Nicklin, M. J., Murray, M. G., Anderson, C. W., Dunn, J. J., Studier, F. W., & Wimmer, E. (1986). A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. *Cell*, *45*(5), 761–70.

Victoria, J. G., Kapoor, A., Dupuis, K., Schnurr, D. P., & Delwart, E. L. (2008). Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathogens*, *4*(9), e1000163. doi:10.1371/journal.ppat.1000163

Wu, Y.-L., Ding, Y.-P., Gao, J., Tanaka, Y., & Zhang, W. (2013). Risk factors and primary prevention trials for type 1 diabetes. *International Journal of Biological Sciences*, *9*(7), 666–79. doi:10.7150/ijbs.6610

Yeung, W. G., Rawlinson, W. D., & Craig, M. E. (2010). Enterovirus infection and type 1 diabetes mellitus: systematic review and meta-analysis of observational molecular studies, (May), 1–9. doi:10.1136/bmj.d35

Yoon, J. W. (1990). The role of viruses and environmental factors in the induction of diabetes. *Current Topics in Microbiology and Immunology*, *164*, 95–123.

Yoon, J.-W., & Jun, H.-S. (2006). Viruses cause type 1 diabetes in animals. *Annals of the New York Academy of Sciences*, *1079*, 138–46. doi:10.1196/annals.1375.021

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821–829. doi:10.1101/gr.074492.107