



# **Sugapea laiguviiruse genoomi sekveneerimine ja annoteerimine**

Bakalaureusetöö

Üliõpilane: Martin Frei

Juhendajad: Merike Sõmera, Cecilia Sarmiento

Õppekava: Rakenduskeemia, toidu- ja geenitehnoloogia

Tallinn 2023



# **Cynosurus mottle virus genome's sequencing and annotation**

Bachelor's Thesis

Student: Martin Frei

Supervisors: Merike Sõmera, Cecilia Sarmiento

Programme: Chemistry, food and gene technology

Tallinn 2023

## Autorideklaratsioon

Kinnitan, et olen koostanud antud lõputöö iseseisvalt ning seda ei ole kellegi teise poolt varem kaitsmisele esitatud. Kõik töö koostamisel kasutatud teiste autorite tööd, olulised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on töös viidatud.

Martin Frei

Töö vastab bakalaureusetööle esitatavatele nõuetele.

Juhendaja: Merike Sõmera

Töö on lubatud kaitsmisele.

Kaitsmiskomisjoni esimees: Vello Tõugu

[allkiri ja kuupäev]

Annotatsioon.....	6
Abstract .....	7
Lühendite loetelu .....	8
Sissejuhatus.....	10
1. Kirjanduse ülevaade .....	11
1.1. Sobemoviirused.....	11
1.2. Genoomi organisatsioon ja ekspressioon .....	11
1.1.2. Virion .....	13
1.1.3. Infektsioon ja sümptomid .....	14
1.3. Sugapea laiguviirus.....	15
1.3.1 Peremehed ja vektorid.....	15
1.3.2 Sümptomid .....	15
1.4. Genoomi sekveneerimis meetodid ja analüüs.....	16
1.4.1 <i>Primer Walking</i> .....	16
1.4.2 <i>Basic Local Alignment Search Tool (BLAST)</i> .....	17
1.4.3 <i>5' RACE</i> .....	17
2. Töö eesmärgid.....	18
3. Eksperimentaalne osa .....	19
3.1. Materjalid.....	19
3.1.1 Viiruse materjal .....	19
3.1.2 Kasutatud reagensid .....	19
3.2. Meetodid.....	19
3.2.1 Pöördtranskriptsioon ehk cDNA 1-ahela süntees .....	19
3.2.2 PCRIks valmistamine ja PCR.....	20
3.2.3 Geelelektroforees.....	20
3.2.4 PCR produktide puhastamine geelist .....	20
3.2.5 Genoomi 5' otsa sekveneerimine RACE-meetodil.....	21
3.2.6 Genoomide assambleerimine ja omavaheline võrdlus.....	22
3.2.7 Vastete otsimine .....	22
4. TULEMUSED JA ARUTELU .....	24
4.1. RT-PCR .....	24
4.2. 5'RACE .....	25
4.3. Kloonide genoomide assambleerimine ja omavaheline võrdlus .....	26

4.4.	Genoomi annoteerimine .....	27
4.5.	Võrdlus teiste sobemoviiruste genoomidega .....	27
4.6.	Fülogeneetiline puu.....	28
5.	Kokkuvõte.....	30
6.	Kasutatud kirjandus.....	32
7.	Lisad.....	34

## Annotatsioon

Sugapea laiguviirus (*Cynosurus mottle virus*, CnMoV) on positiivse polaarsuse RNA genoomiga viirus, mis nakatab kõrrelisi. Selles töös sekveneeriti neli CnMoV isolaati, mis pärinesid Saksamaalt Julius Kühn'i Instituudist Dr. Annette Niehl'i laborist. Sekveneerimisele järgnev joendus näitas suurt sarnasust Leibniz'i instituudi kogus leiduvate CnMoV genoomidega, mille järjestused ilmusid GenBank'is selle projekti käigus. CnMoV genoomi annoteerimisel leiti sobemoviirustele iseloomulikud viis avatud lugemisraami. Määrati nende genoomis paiknemise kohad, suurused, transleeritavate valkude pikkused ja valkude molekulmassid. CnMoV genoomi ja sellelt transleeritud valke võrreldi teiste sobemoviiruste genoomide ja valkudega. Kõige suuremat identsust nähti keraheina laiguviiruse genoomiga ning sellelt transleeritud P2A polüproteiini, RNA sõltuva RNA polümeraasi ja kattevalguga. Sobemoviiruste RNA sõltuvate RNA polümeraaside fülogeneesipuu arvutus näitas CnMoV polümeraasi suurimat lähedust keraheina laiguviiruse polümeraasiga. Saadud tulemused viitavad sellele, et need kaks viirust, millel on sarnane peremeeste ring ning kattuvad geograafilised levialad, omavad tõenäoliselt ühist eellasliiki.

## Abstract

Cynosurus mottle virus (CnMoV) is a virus with a positive polarity RNA genome that infects grasses. In this work, four CnMoV isolates from the laboratory of Dr. Annette Niehl of the Julius Kühn Institute in Germany were sequenced. The alignment that followed sequencing showed a high similarity with the CnMoV genomes found in the collection of the Leibniz Institute that were released in GenBank while the current project was carried out. The annotation of the CnMoV genome revealed five open reading frames typical for sobemoviruses. Their locations in the genome, their sizes, the lengths of the translated proteins as well as proteins' molecular weights were determined. The CnMoV genome and its translated proteins were compared with those of other sobemoviruses. The greatest identity was observed in the case of *Cocksfoot mottle virus* genome, P2A polyprotein, RNA-dependent RNA polymerase, and coat protein. The phylogeny tree calculation based on the RNA-dependent RNA polymerases of sobemoviruses showed the closest relationship of the CnMoV polymerase to the *Cocksfoot mottle virus* polymerase. The obtained results suggest that these two viruses, with similar host ranges and overlapping geographical distributions, probably share a common ancestral species.

## Lühendite loetelu

BLAST - *Basic Local Alignment Search Tool* (ingl. k.)- kohalikul joondusel põhinev otsingutööriist

bp - *Base pair* (ingl. k.), aluspaar

CfMMV - *Cocksfoot mild mosaic virus* (ing.k rahvusvaheliselt käibiel olev ametlik liiginimetus)

CfMV - *Cocksfoot mottle virus* (ing.k. rahvusvaheliselt käibiel olev ametlik liiginimetus), keraheina laiguviiirus

CnMoV - *Cynosurus mottle virus* (ing.k. rahvusvaheliselt käibiel olev ametlik liiginimetus), sugapea laiguviiirus

CP - *Coat protein* (ingl. k.),kattevalk

dNTP - *Deoxynucleoside triphosphate* (ingl. k.), deoksüribonukleotiid

DSMZ - *Deutsche Sammlung von Mikroorganismen und Zellkulturen* (sks. k.), Saksa mikroorganismide ja rakukultuuride kolleksioon (Leibniz'i instituut)

EDTA - *Ethylenediaminetetraacetic acid* (ingl. k.), etüleen-diamiin-tetratselaat

GCFV - *Ginger chlorotic fleck virus* (ing.k rahvusvaheliselt käibiel olev ametlik liiginimetus)

ICTV - *International Committee on Taxonomy of Viruses* (ingl. k.), Rahvusvaheline Viiruste Taksonoomia Komitee

IYMV - *Imperata yellow mottle virus* (ing.k. rahvusvaheliselt käibiel olev ametlik liiginimetus)

gRNA - *Genomic RNA* (ingl. k.), genoomne RNA

mRNA - *Messenger RNA* (ingl. k.), informatsiooni RNA

ORF - *Open reading frame* (ingl. k.), avatud lugemisraam

PCR - *Polymerase chain reaction* (ingl. k.), polümeraasi ahelreaktsioon

PLYV - *Papaya lethal mottle virus* (ing.k. rahvusvaheliselt käibiel olev ametlik liiginimetus)

PRO - *Serine protease* (ingl. k.), seriinproteas

RdRp - *RNA-dependent RNA polymerase* (ingl. k.), RNA-sõltuv RNA polümeraas

RGMoV - *Ryegrass mottle virus* (ing.k rahvusvaheliselt käibiel olev ametlik liiginimetus)

RYMV - *Rice yellow mottle virus* (ing.k rahvusvaheliselt käibiel olev ametlik liiginimetus)

SCMoV - *Subterranean clover mottle virus* (ing.k. rahvusvaheliselt käibiel olev ametlik liiginimetus)



SCPMV - *Southern cowpea mosaic virus* (ing.k. rahvusvaheliselt käibiel olev ametlik liiginimetus)

SDS - *Sodium dodecyl sulfate* (ingl. k.), naatrium dodetsüülsulfaat

SeMV - *Sesbania mosaic virus* (ing.k. rahvusvaheliselt käibiel olev ametlik liiginimetus)

sgRNA - *Subgenomic RNA* (ingl. k.), subgenoomne RNA

TAE buffer - *Tris base, Acetic acid, EDTA* (ingl. k.), Tris-atsetaat-EDTA puhver

TM - *Transmembrane domain* (ingl. k.), transmembraanne domään

UTR - *Untranslated region* (ingl. k.), mittetransleeritav regioon

VPg - *Viral protein genome-linked* (ingl. k.), viraalne genoomiseoseline valk

## Sissejuhatus

*Cynosurus mottle virus* (CnMoV) ehk sugapea laiguviirus on üheahelaline +RNA viirus, mille genoomis esinevad 5 avatud lugemisraami (ORF1; ORFx; ORF2a; ORF2b; ORF3), selle kattevalgu 180 koopiast ehitatud virion on ikosaeedrilise struktuuriga ja kattevalgu ühikud on omavahel paigutatud T=3 sümmeetriaga. Veel aastal 2000 kuulus CnMoV oletatavalt (nn. *tentative species*) sobemoviruse perekonda, kuid praeguste andmete järgi ei kuulu see ühtegi perekonda, kuna selle viiruse genoom on uurimata. Esimene oletatav leid pärineb Iirimaa aastast 1972, kui John. A. Brook leidis kahest uuritud taime liigist (*Cynosurus cristatus* L. ; *Lolium perenne* L.) viiruse, mille ta nimetas *Lolium mottle virus*. Edaspidi leitud sarnane viirus sai nimeks *Cynosurus mottle virus*.

Viimase 50 aasta jooksul alates esimesest CnMoV mainimisest on sarnaste bioloogiliste ja biokeemiliste omadustega viirust leitud ja uuritud Uus-Meremaal, Saksamaal ja Inglismaal. Võib oletada, et on selle viirus levik on laiem, sest puuduvad tuvastamiseks vajalikud testid. CnMoV nakatab kõrrelisi peamiselt mehhaanilise ülekande kaudu.

Uus-meremaa ja inglise isolaate pole kogudes säilinud, kuid see-eest on säilitatud mitmed saksa isolaadid Julius Kühni'i instituudi labori kogus ning Leibniz'i instituudi saksa mikroorganismide ja koekultuuride (DSMZ) viiruste kogus Saksamaal. Meie labori pikaajase huvi tõttu sobemoviiruste perekonna viiruste suhtes on meie laboris sekveneeritud DSMZ-kogust pärinev isolaat DSMZ-PV-0728, mis oli kogutud Betzdorf/Sieg linna lähedalt W. Huth poolt 1983. a. ning inokuleeritud paljundamiseks ja säilitamiseks nisu taimedesse. Käesoleva bakalaureusetöö raames sekveneerisime Julius Kühni'i instituudi kogust pärinevad isolaadid ja võrdlesime nende genoome DSMZ isolaadiga DSMZ-PV-0728 ning isolaadiga DSMZ-PV-1185, mis paralleelselt käesoleva tööga sekveneeriti DSMZ-kogu haldaja (Leibniz instituut) poolt.

Selle lõputöö eesmärgiks oli CnMoV Julius Kühni'i instituudi labori kogust pärinevate saksa isolaatide genoomide sekveneerimine ning CnMoV genoomi ja sellelt transleeritavate valkude järjestuste võrdlemine teiste sobemoviiruste genoomide ja neilt transleeritavate valkudega.

# 1. Kirjanduse ülevaade

## 1.1. Sobemoviirused

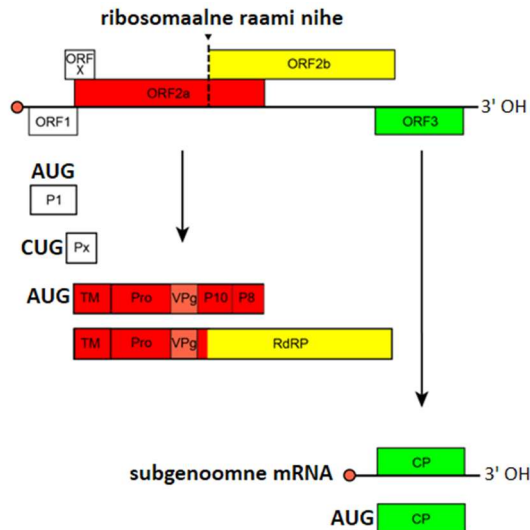
1977. aastal soovitas Roger Hull luua uus taimeviiruse grupp põhinedes sarnasustel genoomse RNA ja viiruskapsiidi omadustes, ja samuti viirusosakeste jaotuses rakus. Kapsiidi omadustest vaadeldi sedimentatsiooni koefitsienti, kapsiidi stabiilsust, kattevalgu ühiku massi (Hull, 1977). *Sobemovirus* perekond kui taksonoomiline ühik loodi ametlikult Rahvusvahelise viiruste taksonoomia komitee (*International Committee on Taxonomy of Viruses*, ICTV) poolt 1995-ndal aastal (Tamm ja Truve, 2000).

*Sobemovirus* perekond on saanud oma nime oma tüüpliigi järgi - *Southern bean mosaic virus* (SBMV). *Sobemovirus* perekond omakorda kuulub *Sobemoviridae* sugukonda, mis kuulub *Sobelivirales* seltsi.

Praeguste uuendatud ICTV andmete alusel, sisaldab *Sobemovirus* perekonna nimekiri 20 viiruse liiki (Lisa 1, tabel 1; leitav aadressil: <https://ictv.global/taxonomy>). Nimekiri sisaldab ainult neid liike, mille näidiste genoomid on sekveneeritud. Varasem süstemaatika, mis arvestas genoomi tüüpi ja viiruspartikli omadusi, lubas sekveneerimata liikide klassifitseerimist „oletatavate“ (*tentative*) liikide kategooriasse. Näiteks 2000. a. *Sobemovirus* perekonna nimekiri sisaldas 3 oletatavat liiki, mida 2022. a. nimekiri enam ei sisalda: *Cocksfoot mild mosaic virus* (CfMMV), *Cynosurus mottle virus* (CnMoV) ja *Ginger chlorotic fleck virus* (GCFV) (Tamm ja Truve, 2000).

## 1.2. Genoomi organisatsioon ja ekspressioon

Sobemoviiruse genoom on polüsitroonne, üheaheelaline positiivse polaarsusega RNA. Genoomi kuuluvad 5' ja 3' mittekodeerivad otsad ja 5 avatud lugemisraami (ORF)- ORF1; ORFx, ORF2a; ORF2b; ORF3. ORFide genoomne paigutus ja nende translatsioon on illustreeritud joonisel 1. Genoomi mittekodeeriva 5' otsa küljes asub viraalne genoomiga seotud valk (VPg), mis on vajalik translatsioonikompleksi moodustumiseks. Genoomi mittekodeerivas 3' otsas on mõnede sobemoviiruste (*Rice yellow mottle virus* (RYMV); *Cocksfoot mottle virus* (CfMV)) puhul arvutiprogrammidega modelleerides leitud potentsiaalne tRNA-sarnane struktuur (*tRNA-like structure*, TLS) (Sõmera *et al.*, 2015).



**Joonis 1. Sobemoviiruse genoomi translatsiooni ja transkriptsiooni kaart** - valk P1 ; ORFx - valk X ; ORF2a - TM (transmembraanne domään), Pro (seriinproteaaas), VPg (viiruse genoomiga seotud valk), P10 (10 kDa valk), P8 (8 kDa valk) / RdRP (RNA-sõltuv RNA polümeraas); ORF3 - CP (kattevalk). Pilt võetud ja modifitseeritud: <https://ictv.global/report/chapter/solemoviridae/solemoviridae/sobemovirus>

**ORF 1 kodeerib P1 valku.** ORF1 asub nõrgas initsiatsioonikontekstis ja seetõttu seda transleeritakse vähe. Valgu P1 ekspresseerimine on vajalik süsteemse infektsiooni korral, millest võiks selguda, et P1 valk võib olla viiruse liikumisvalk (*movement protein*, MP). Samas P1 mängib supressori rolli RNA vaigistamises (Sõmera *et al.*, 2015). ORF1 järgnevatelt lugemisraamidelt ORFx ja ORF2a/2ab valkude translatsioon toimub lekkivalt skanneerides, mis kätkeb nähtust, mille puhul ribosoom translatsiooni initsiatsiooni jooksul mõnikord jätab vahele nõrga initsiatsioonikoodoni mRNA-l (kirjeldatud aadressil: <https://viralzone.expasy.org/1976>). Selline mehhanism võimaldab mRNA-ga seondunud 40S ribosoomidel algatada translatsiooni järgnevast ORF-ist, mööda minnes eelneva ORFi alguskoodonist.

**ORF X kodeerib Px valku.** Asub ORF1 ja ORF2a vahel, osaliselt nendega kattudes. ORFx algab - mitte-AUG initsiatsioonikoodoniga (*non-AUG*), mis paikneb paar koodonit enne ORF1 lõppu ja lõpeb kattudes ORF2a 5'-poolsete koodonitega (kattuvuse pikkus sõltub liigist) (Ling *et al.*, 2013).

**ORF 2a kodeerib transmembraanset domääni (TM); seriinproteaaasi (Pro); viiruse genoomiga seotud valku (VPg); RNA-siduvat domääni P10 (10 kDa valk); ATP domääni P8 (8 kDa valk).** TM - funktsiooniks on polüproteiini ankurdamine rakumembraanidesse, et hõlbustada proteolüütilist töötlemist ja tõenäoliselt ka viiruse miinusahela sünteesi. **Pro** - katalüüsib kahe spetsiifilise aminohappejäägi vahel paiknevate spetsiifiliste peptiidsidemete hüdrolüüsi. **VPg** - vajalik translatsiooni initsiatsiooniks. Funktsiooniks on võime siduda paljusid erinevaid partnereid ja täita erinevaid funktsioone viiruse elutsükli jooksul. RYMV, SBMV, CfMV ja *Ryegrass mottle virus* (RGMoV) VPg-d sisaldavad mitmeid fosforüülitud jääke, mis on tõenäoliselt seotud korrastamata VPg valkude voltimise ja lahtivoltimise reguleerimisega interaktsiooni määramise ajal. **P8** - SeMV P8 on võimeline siduma RNA-d ja DNA-d *in vitro* sisaldades RNA-d siduvat piirkonda. Samas vastutab P8 valk P10 Mg<sup>2+</sup>-sõltuva ATPaasi aktiivsuse eest prekursorvalgus P10-P8. **P10** - SeMV P10 interakteerub tugevalt P1-ga, see interaktsioon on seotud viiruse liikumiskompleksi

moodustamisega. Samas P10 interakteerub ka RdRp-ga, see interaktsioon on seotud viiruse replikatsiooniga (Sõmera et al., 2015).

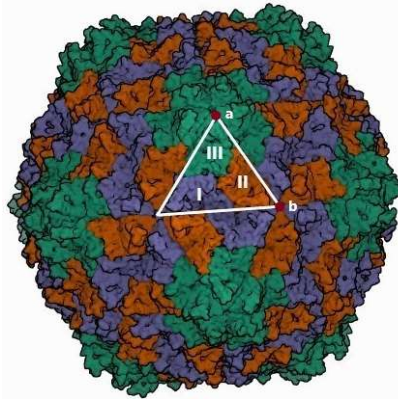
**ORF2b - kodeerib RNA sõltuvat RNA polümeraasi (RdRp)** - Transleerimiseks on vaja, et algaks -1 programmeeritud ribosomaalne raaminihe (*programmed ribosomal frameshift*, PRF), mis lõpetab ORF2a transleerimise pärast kahte kolmandikku ja jätkab transleerimist ORF2b raamist. PRF-signaali suunavad transleerivad ribosoomid ümber uude lugemisraami (+1 või -1 nukleotiidi võrra, sobemoviiruste puhul -1 positsiooni). PRF signaal võimaldab ribosoomidel mööduda 0-raamis kodeeritud *in-frame* stopp-koodonist ja jätkata teisest lugemisraamist pikendatud liitvalgu sünteesi. -1 PRF koosneb kolmest olulisest elemendist: seitsmest nukleotiidist koosnev "libe" koht, kus tegelikult toimub lugemisraami translatsiooniline nihe; lühike vahejärjestus, ja allavoolu asuv PRF stimuleeriv RNA sekundaarstruktuur (Dinman, 2012). Sobemoviiruse perekonna liikme *Sesbania mosaic virus* (SeMV) RdRp suudab sünteesida RNA-d praimerist sõltumatul viisil (Govind ja Savithri, 2010). SeMV RdRp ja P10 kunstlikes tingimustes koos-ekspressioon põhjustab märkimisväärselt suuremat polümeraasi aktiivsust, kui ainult RdRp ekspressioon. Samas, ka SeMV RdRp ja Pro vahel eksisteerivad mõõdukad interaktsioonid. Nende interaktsioonide täpsed rollid ei ole veel välja selgitatud. *In vitro* testid ei ole näidanud RdRp koostoimet VPg, CP või P8-ga (Sõmera et al., 2015).

**ORF3** - kodeerib kattevalku (*coat protein*, CP). Kattevalgu translatsioon toimub subgenoomselt RNAlt (sgRNA). Kattevalgu monomeeridest moodustub viiruse kapsiid ehk virion.

### 1.1.2. Virion

Röntgenkristallograafia (tehnikat biomolekulide 3D-struktuuri määramiseks) abiga on teada sobemoviiruste virionide struktuurid. Sobemoviiruste virionide ruumilised struktuurid on väga sarnased, isegi peaaegu identsed, kuigi CP järjestuste sarnasus on suhteliselt väike, nt. SeMV ja *Southern cowpea mosaic virus* (SCPMV) CP järjestuste vaheline identsus on umbes 63%, teiste CP-de omavaheline identsus on palju väiksem (Sõmera et al., 2021).

Sobemoviiruse virion omab ikosaedriline kuju, diameetriga ca. 30 nm (Joonis 2). Kapsiid koosneb umbes 180 CP molekulist ja on ehitatud T=3 sümmeetrilise järgi. CP monomeeridel on 3 erinevat konformatsiooni (Joonisel 2 näidatud erinevate värvidega). Kokku nad moodustavad ühe ikosaedriline subühiku (Joonisel 2: valge kolmnurk). Seega, sobemoviiruse T=3 struktuuri moodustamiseks kulub 60 ikosaedriline subühikut. Heksameerid ja pentameerid erinevad üksteist kuju poolest: heksameerid on tasapinnalised ja pentameerid on oluliselt painutatud. Nende kombinatsioon annab ikosaedrile oma vormi (Sõmera et al., 2015).



**Joonis 2. Sobemoviiruse virioni struktuuri pilt.** T=3 sümmeetria - CP monomeeride konformatsioonid: I (lilla), II (punane), III (roheline); Valge kolmnurk - ikosaedriline subühik; **a** - pentameeri keskpunkt; **b** - heksameeri keskpunkt. Pilt võetud ja modifitseeritud <https://www.rcsb.org/structure/1F2N>.

Virioni stabiilsus sõltub pH-st, katioonidest, mis moodustavad soolasilla kapsiidi valguga ja viiruse genoomse RNA vahel. Kui pH suureneb või väheneb katioonide kogus, siis toimub virioni paisumine ja stabiilsus väheneb. Lisaks välistele teguritele sõltub virioni stabiilsus valk-valk interaktsioonist ikosaedri subühikute vahel ja CP-RNA interaktsioonist (Sõmera *et al.*, 2015).

### 1.1.3. Infektsioon ja sümptomid

**Infektsioon:** Sobemoviiruste liikidel on erinevad levialad, mõned levivad peaaegu kogu maailmas, teised on piiratud kontinendiga või isegi konkreetse piirkonnaga (Sõmera *et al.*, 2015). Levikut mõjutavad erinevad tegurid, näiteks uus kohanenud taimeliik võib potentsiaalselt saada mõnede viiruste peremeheks, suurendades sellega viiruse levikut, põllumajanduse laienemisega seotud uute bioomide loomine annab viirustele uue leviku koha, jne. Peamiseks sobemoviiruse ülekandeallikaks nimetatakse mehaanilist taimede vigastamist ja viirusnakkuseta taimede kokkupuudet nakatanud taimematerjaliga. Erinevad putukate liigid, mis tekitavad toitumisel mehhaanilisi vigastusi, võivad toimida sobemoviiruse vektoritena. Vaadates sügavamale viiruste levimise võimalustesse tuleb arvestada, et erinevad viiruse liigid käituvad erinevalt. Näiteks RYMV mehaaniline inokuleerimine võib looduses toimuda imetajate (Sarra ja Peters, 2003), lindude (Peters *et al.*, 2012), pinnasevee ja tuule abil (Sarra *et al.*, 2004). *Subterranean clover mottle virus* (SCMoV) põhiline levikutee on mehaaniline edasikandmine inimeste või loomade abil, jalatsite või kapjadega. Nakatunud taime peale astudes taim vigastatakse ja viiruse osake, mis jääb kapjade või kingade peale kantakse edasi, auto ratastega samamoodi. Mõned viiruseosakesed on leitavad vees ja pinnases, mille tõttu saavad viirused "liikuda" pikkasid vahemaasid, heaks näiteks on *Papaya lethal yellowing virus* (PLYV) (Camarço *et al.*, 1998). Mehhaanilise ülekande tõestamiseks on laborikatse tingimustes inokuleerimiseks võimalik kasutada nakatunud taimest eraldatud taimemahla, segades seda saviga (tseliidiga). Inokuleerimine toimub taimemahlast, puhverlahusest ja saviosakestest valmistatud inokulumiga terve taime lehti hõõrudes.

**Sümptomid:** Sobemoviiruste sümptomid sõltuvad konkreetsetest viiruse liikidest ja nende poolt nakatunud taimede liikidest. Sümptomid, mis on iseloomulikud kogu sobemoviiruste perekonna liikmetele on kloroos, nekroos, laigulisus, juhtsoonte kahjustamine ja steriilsus. Mõned infektsioonid võivad olla asümptomaatilised.

### 1.3. Sugapea laiguviirus

Sugapea laiguviiruse (CnMoV) esimeseks võimalikuks leiuks arvatakse olevat 1967. a leiud Iirimaaalt Caherconlish, Louisberg, Eire ja Limerick lähistelt ning neid on kirjeldatud *Lolium mottle virus* nime all sugapea (*Cynosurus rictatus* L.) ja raiheina (*Lolium perenne* L.) taimedest (A'Brook, 1969). Nakatatud taimed näitasid nõrka klorootilist laigulisust. Hilisemates uurimises ei suudetud näidata enam infektsiooni raiheinas.

Lisaks Iirimaaale, on CnMoV olemas teadaolevalt veel mitmetes riikides nagu Uus-Meremaa, Saksamaa, Inglismaa. Samas võib oletada, et need riigid ei ole ainukesed kus on CnMoV levinud, kuna teistes riikides ei ole seda uuritud.

#### 1.3.1 Peremehed ja vektorid

**Peremehed:** Eksperimentaalselt on kindlaks tehtud mitmeid taimeliike, mis on CnMoV peremehed (Catherall *et al.*, 1977; Mohamed, 1978). Nakatumise sümptomid on järgnevad:

Tugev laigulisus (kloroos) - *Agrostis puchella* Kunth, *Agrostis tenuis* Sibth, *Agrostis stolonifera* L., *Anthoxanthum puellii* Lecoq et Lamotte, *Apera spica-venti* (L.) Beauv., *Avena sativa* L., *Bromus secalinus* L., *Bromus unioloides* B.H.K., *Cynosurus cristatus* L., *Dactylis glomerata* L., *Lagurus ovatus* L., *Holcus lanatus* L., *Hordeum sativa* L., *Phleum arenarium* L., *Poa annua* L., *P. paniculatum* Huds., *Triticum aestivum* L.

Keskmise intensiivsusega laigulisus - *Avena fatua* L., *Chloris virgata* Swartz, *Hordeum vulgare* L., *Phalaris minor* Retz., *Phalaris tuberosa* L., *Secale cereale* L.

**Vektorid:** Toominga-lehetäi *Rhopalosiphum padi* L. on näidatud CnMoV vektorina Uus-Meremaal: kui lehetäid olid söönud nakatatud kaera kakskümmend neli tundi ja ümber tõstetud tervetele kaheksa-päevaste kaerataimedele, kus nad viibisid viis päeva, siis kolme nädala jooksul ilmnisid CnMoV sümptomid kolmel taimel kuuest, lehetäidega mitte kokku puutunud kontrolltaimed jäid sümptomiteta (Mohamed, 1978).

Seevastu Catherall (1977) uuring inglise isolaadiga näitas, et vektoriteks on mardikad: harilik viljakukk *Oulema melanopa* L. ja sinine viljakukk *O. lichenis* Voet, aga toominga-lehetäi *Rhopalosiphum padi* L ei ole. Iga viirus inokuleeriti nisu taimedesse ja selle liigi nakatunud lehti kasutati inokulaadi allikana.

#### 1.3.2 Sümptomid

CnMoV välised sümptomid võivad erineda peremehest sõltuvalt (Joonis 3). Uurimine 1978-ndast aastast näitab, et viirusega nakatatud raiheina taimed omasid klorootiliseid laiike lehtede alumisel küljel või klorootiliseid triipe lehtede basaalosas (Joonis 3b). Nakatatud sugapea taimed omakorda näitasid mosaiigi sümptomeid (Joonis 3c) või klorootilist laigulisust (Joonis 3d). Vanematel lehtedel esines mõnikord nekrootilisi laiike ja triipe. Välitingimustest kogutud taimed olid kasvus pidurdunud ja väheste pähikutega. Nakatatud valge kasteheina taimedel esines tumeroheline mosaiik. Nakatatud hariliku kasteheina taimedel olid tugevalt laigulised lehed (Mohamed, 1978).



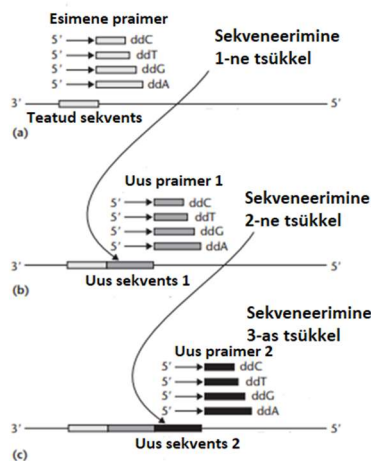
**Joonis 3. Sugapea laiguviiruse infektsiooniga taimed** a - terve raiheinas leht; b - raihein (klorootiline triibutamine/laigulisus); c - sugapea (mosaiik); d - sugapea (tugev laigulisus); Sümptomid kaeras pärast mehaanilist inokuleerimist: e - sugapeast; f - raiheinast. (Mohamed, 1978)

CnMoV viiruseosakesi on leitud parenhüümirakkude tsütoplasmas, vakuoolides ning floeemikoos. Viiruse osakesed olid kas tsütoplasmas fibrillaarsete kogumites või olid need vakuoolides. Viiruse osakesi ei ole leitud nakatunud raku organelidest ega tuumast (Mohamed, 1981).

## 1.4. Genoomi sekveneerimis meetodid ja analüüs

### 1.4.1 Primer Walking

*Primer walking* meetod leiutati 1997-ndal aastal Fraser'i ja Fleischmann'i poolt. Meetod annab võimaluse sekveneerida tundmatu järjestusega pikad DNA ahelad samm sammu haaval (Joonis 4). Sekveneerimise protsess algab järjestusest, mis on teada (Sverdlov ja Azhikina, 2005).



**Joonis 4. Primer walking** - (1) Esimene praimer, mis on komplementaarne "teatud" järjestusele (hall kast) kasutatakse alustamaks Sanger meetodil tundmatu ala sekveneerimist. (2) Uus komplementaarne praimer on moodustatud äsja sekveneeritud "ue" järjestuse põhjal. (3) Protsessi saab korrata nii kaua kui on vaja. Pilt võetud ja modifitseeritud - (Sverdlov ja Azhikina, 2005).



### **1.4.2 Basic Local Alignment Search Tool (BLAST)**

Programm, mis võrdleb nukleotiidseid (RNA või DNA) või valgujärjestusi. Teostatakse andmebaasi vastu, et tuvastada sarnasusi teiste teadaolevate järjestustega ja arvutada kattuvuse statistiline olulisus (leitav aadressil: [blast.ncbi.nlm.nih.gov/Blast.cgi](http://blast.ncbi.nlm.nih.gov/Blast.cgi)).

### **1.4.3 5' RACE**

Viirusgenoomide 5' otsa nukleotiidse järjestuse teada saamiseks kasutatakse cDNA otste kiiret amplifikatsiooni (*rapid amplification of cDNA ends, RACE*).

5' RACE - algab pöördtranskriptsiooniga mRNAst, kasutades uuritava järjestuse või geeni anti-sense ahelale vastavat oligonukleotiidi praimerit (GSP1) mis tunneb ära teadaoleva järjestuse. Primer seostub mRNA-ga ja RT ensüüm lisab aluspaarid 3' praimerile otsale genereerides spetsiifilist üheaheelalist cDNA-d. Saadud cDNA-le lisatakse 3'-terminaalne järjestus - näiteks homopolümeerne saba, mis on moodustatud terminaalse deoksünukleotidültransferaasi (TdT) ensüümi abil. cDNA esimese ahela sünteesile järgneb PCR, mis kasutab teist antisense-ahelale vastavat geenispetsiifilist praimerit (GSP2), mis seob 3' otsa GSP1-ga, ja sense- (forward) universaalset praimerit (UP), mis seob cDNA-de 3'-otstele lisatud homopolümeerse sabaga, et amplifitseerida cDNA produkti 5' otsast (leitav aadressil: <https://www.thermofisher.com>).

## 2. Töö eesmärgid

Käesoleva töö eesmärkideks oli:

- Sugapea laiguviiiruse (CnMoV) Julius Kühn'i instituudi laborist pärinevate isolaatide sekveneerimine;
- CnMoV genoomi annoteerimine ja analüüs;
- CnMoV genoomi ja valkude võrdlemine teiste sobemoviiiruste genoomide ja valkudega.

## 3. Eksperimentaalne osa

### 3.1. Materjalid

#### 3.1.1 Viiruse materjal

CnMoV isolaatid (1,2,4,5) on saadud Saksamaa Julius Kühn Instituudist Dr. Annette Niehl'i käest.

DSMZ-PV-0728 isolaat pärineb TalTech taimeviroloogia labori kolleksioonist ja on saadud DSMZ-kogust.

Töös kasutatud praimerid on näidatud Lisas 2 (Tabel 2).

#### 3.1.2 Kasutatud reagensid

Agaros; ampitsilliin; *blunting enzyme*; *Celite™*; CnMoV spetsiifilised praimerid ja praimeri paarid Cn5pF2+Cn1R, Cn2L+Cn2R, Cn3L+Cn3R, Cn4L+Cn4R, Cn5L+Cn5R, Cn6L+Cn6R, Cn7L+Cn7R, Cn8L+Cn3pR2; *DreamTaq Green PCR Master mix 2x (Thermo Scientific™)*; *E. coli DH5alpha*; etanool 70%; etiidumbromiid; eluatsiooni puhver; E1 - bakterite resuspensioonipuhver (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 100 µg/ml RNase A); E2 - bakterite lüüsilahus (200 mM NaOH, 1% SDS); E3 - aluselise lüüsi neutralisatsioonipuhver (3.2 M KOAc/HOAc, pH 5.5); FADF puhver; Gel/ *PCR purification Kit (FavorPrep™)*; isopropanool; *Maxima RT* (pöördtranskriptaas); mQ-vesi; NTI ja NT3 puhvrid; *pJET 1.2 blunt cloning vector*; *RNase inhibitor*; *RNeasy Plant mini kit (Qiagen™)*; T4 DNA ligaasi; 1xTAE puhver; 10 mM dNTP.

### 3.2. Meetodid

#### RNA eraldus

Saksamaalt saadud uuritav taimematerjal uhmerdati peeneks pulbriks lisades lämmastikku peenestamise lihtsustamiseks ja intaktse RNA eraldamiseks. Saadud proovidest puhastati RNA kasutades *RNeasy Plant mini kit (Qiagen™)*.

Pärast puhastamist mõõdeti RNA kontsentratsioonid NanoDrop1000 masinaga.

#### 3.2.1 Pöördtranskriptsioon ehk cDNA 1-ahela süntees

Iga Julius Kühn' proovi (1, 2, 4, 5) taimematerjalist eraldatud RNA-lt teostati cDNA 1-ahela süntees viiruse-spetsiifilise materjali kopeerimiseks, kasutades CnMoV 3' otsaga seonduvat praimerit Cn3pR2.

Ependorfis segati 10 mM dNTP (1 ul) , spetsiifiline as-primer (Cn3pR2) 20 pmol/ul (1 ul), eelnevalt eraldatud RNA (0.5 ug) , mQ vett (11.5 ug) nii, et kokku oleks 14 ul. Protsess teostati jääl, vältides RNA lagunemist. Saadud segu pandi 65 °C juurde 5-ks minutiks, RNA sekundaarstruktuuri lõhkumiseks. Seejärel asetati proovid tagasi jää peale, et vältida uue sekundaarstruktuuri moodustumist, seejärel lisati 6 ul ettevalmistatud *mastermix*'i (MM, koostisosad - 5x puhver (4ul), *RNase inhibitor* (1ul), *Maxima RT* (1ul)). Järgnevalt inkubeeriti reaktsioone cDNA sünteesiks 1 tund 55 °C juures, reaktsiooni inaktiveerimiseks kuumutati proove 5 minutit 85 °C juures.

### 3.2.2 PCRiks valmistamine ja PCR

**PCRiks valmistamine:** Iga 1-ahela sünteesireaktsioonist tehti 8 erinevat PCR reaktsiooni, mis katsid kogu viiruse genoomi (ca. 4 kb), kasutades järgnevaid praimerite kombinatsioone: **1.**Cn5pF2+Cn1R; **2.**Cn2L+Cn2R; **3.**Cn3L+Cn3R; **4.**Cn4L+Cn4R; **5.**Cn5L+Cn5R; **6.**Cn6L+Cn6R; **7.**Cn7L+Cn7R; **8.**Cn8L+Cn3pR2.

PCR-ks kasutati *DreamTaq Green PCR Master mix 2x* (ThermoScientific™) protokoll. Reaktsiooni mahuks oli 20 ul. Selleks segati kokku: *DreamTaq Green PCR Master Mix (2x)* - 10ul; *Forward primer* (20 pmol/ul) - 1ul; *Reverse primer* (20 pmol/ul) - 1ul; *Template DNA* (1-ahela sünteesiprodukt) - 2ul; nukleasivaba vesi - 6 ul. Korraliku segamise tagamiseks segu vorteksiti ja tsentrifuugiti.

**PCR:** Reaktsioonisegud pandi termotsüklerisse *Eppendorf Mastercycler®* programmiga, mis koosneb 5 etapist - algne denaturatsioon 3 minutit temperatuuriga 95 °C, sellele järgnes 35 tsükli kolmet etapiga: denaturatsioon 95 °C juures 30 s, praimer seondumine 50 °C juures 30 s ja DNA süntees 72 °C juures 1 minut. Viimaseks etapiks oli lisa süntees 72 °C juures 5 min.

### 3.2.3 Geelelektroforees

Geelelektroforeesi kasutati kahe eesmärgiga: esimene oli lahutada PCR produktid foreesil fragmentideks, et seejärel uurida meie otsitava viiruse järjestuse olemasolu taimeproovis; teine oli saadud DNA fragmendid geelist puhastada ja sekveneerida. Iga kloonile vastavaid PCRi produkte käsitleti teistest eraldi, et vältida erinevatelt kloonidelt pärinevate proovide ristisaastumist.

Nende eesmärkide saavutamiseks valmistati agarosgeel - agaros segatud 1xTAE puhvriga kuumutamise teel. Peale paari minutit jahtumist valati geel spetsiaalse ristkülikukujulise vormi sisse ja seejärel lisati etiidiumbromiid (2 ul 10 mg/ml kontsentratsiooni 50 ml geeli kohta), mis seondub DNAGA ja on vajalik UV valguse all DNA fragmentide vaatlemiseks. Tardunud geel pandi elektroforeesi vanni, mis oli täidetud 1xTAE puhvriga. PCR produktid ja marker pipeteeriti geeli kaevukestesse. Suurusmarkerina kasutati - *1kb DNA ladder* (5ul; ThermoScientific). Geelelektroforees viidi läbi 180 voldi juures 40 minuti vältel. Saadud geel pildistati ultravioletvalguse all kasutades ImageQuant 400 masinat.

### 3.2.4 PCR produktide puhastamine geelist

DNA geelist puhastamiseks kasutati *FavorPrep™ GEL/PCR purification Kit*. DNA fragmendid lõigati geelist (umbes 100 mg) ja asetati mikrotsentrifuugi tuubide sisse. Järgnevalt lisati igale geelitükile 500 ul FADF puhvrit. Saadud proov inkubeeriti raputades 55 °C juures geeli täieliku lahustumiseni. Proov pipeteeriti FADF kolonni, tsentrifuugiti 11 000 rpm 30 sekundit ja eemaldati supernatant. Samasse kolonni lisati 750 ul pesupuhvrit (etanooliga), tsentrifuugiti 11 000 rpm 30 sekundit ja eemaldati supernatant. Kolonn tsentrifuugiti maksimaalse kiirusega (13 200 rpm) 3 minutit, et täielikult kuivatada kolonni maatriks. FADF kolonn tõsteti ümber uude mikrotsentrifuugi tuubi. FADF kolonnile tilgutati 40 ul mQ-vett (mQ filtriga puhastatud) membraani keskele ja tsentrifuugiti, et koguda elueerimispuhvril lahustunud DNA. Pärast puhastamist mõõdeti kontsentratsioonid NanoDrop abil.

Proovide näidised saadeti Tartu ülikooli genoomikakeskusesse sekveneerimiseks. Iga PCRi reaktsioon sekveneeriti nii *sense* kui *antisense* suunas, kasutades samasid praimereid, millega PCR teostati. Juhul kui proovides esines ootuspärasest erinevaid lisaprojekte, siis sekveneeriti need ka.

### 3.2.5 Genoomi 5' otsa sekveneerimine RACE-meetodil

Viiruse kloonide genoomi 5' otsa järjestuse uurimiseks olid läbi viidud 5' RACE reaktsioonid, mille produktide puhastamine, kloneerimine, bakteritesse transformeerimine, koloonia-PCR, miniprep-DNA eraldamine ja sekveneerimine viidi läbi bakalaureusetöö autori poolt.

**Geelelektroforees:** Ühe geeli valmistamiseks kasutati 50 ml 1xTAE + 0,5 g agaros + 2ul EtBr (10 mg/ml). Marker: **Generuler low range DNA Ladder** (Thermo Scientific™). Geelelektroforeesile järgnes ca. 200 bp bändide geelist lõikamine ja puhastamine.

**Geelist puhastamine:** Puhastamiseks kasutati järgnevat protokoll: **NucleoSpin Gel and PCR Clean-up**. 100 mg geelitükile lisati 400 ul NT1 puhvrit ja inkubeeritiseda 50 °C juures 5-10 min nõrgalt vorteksides, kuni geel sulas täielikult. Järgmiseks, pipeteeriti lahus teisse kolonni ja fuugiti 11 000 g juures 30 s, pärast mida supernatant eemaldati. Edasi lisati kolonni 700 ul NT3 puhvrit ja fuugiti 11,000 g juures 30 s, eemaldati supernatant ja korrati seda etappi veelkord. Sellele järgnes kolonni kuiv fuugimine 11.000 g 1 min. Viimaseks etapiks oli kolonni ümbertõstmise uue ependorfi sisse, kolonni lisati 20 ul ME eluatsioonipuhvrit, 1 min pärast fuugiti, et DNA kolonnist kätte saada.

Kõikidel proovidel mõõdeti kontsentratsioon, arvatati kui mitu ul oleks vaja võtta 0.15 pmol vektorplasmidi otste kohta kloneerimise ligatsioonireaktsioonideks.

**Ligeerimine:** Ligeerimist toimetati jääs, kasutati geelist puhastatud materjali (kloonid 1, 4, 5), 2x BF, *Blunting enzyme*, mQ H<sub>2</sub>O. Segamisele järgnes inkubatsioon 70 °C juures 5 min, mille lõpus proovid pandi tagasi jää peale, kus neile lisati juurde 1 ul *pJET 1.2 blunt cloning vector* ja 1 ul T4 DNA ligaasi. Proovid inkubeeriti toatemperatuuril 30 min. *E. coli* DH5alpha tüve kompetentseted bakterid sulatati jää peal. Lisati tehtud ligatsioonisegu, siis hoiti seda 42 °C juures 90 sek ja jahutati jääs. Seejärel külvati ligatsiooniseguga transformeeritud bakterid LB söötmega petri tassidele, kuhu oli lisatud 100 ng/ml ampitsilliini. Bakteritassid pandi inkubaatorikappi 37 °C juurde üleöö hoidmiseks.

**Kolooniate PCR:** Spetsiaalse pulgaga võeti igalt tassilt 5 erinevat kolooniat ja kanti need paljundamiseks uuele LB-söötme ja ampitsilliiniga petri tassile. Nendest kolooniatest pärinevale materjalile tehti nn. kolooniate PCR. Selleks viidi pisut bakterikolooniast pärinevat materjali eelnevalt valmistatud PCR *mastermixi* sisse (pcr tuubis). MASTER MIX sisaldas 20 ul kohta: 10 ul 2x DreamTaq Green PCR Master Mix, 1 ul *pJET Forward primer* (10 pmol/ul); *pJET Reverse primer* (10 pmol/ul); 8 ul mQ vett.

Materjali ümber tõstmise meetod: 1. Võeti spetsiaalne külvinõel; 2. leiti üksik koloonia petri tassi peal; 3. võeti külvinõelaga koloonia; 4. uue petri tassi numereeritud lahtrisse tehti külvinõelal oleva bakterimaterjaliga õrn joon; 5. külvinõel loputati PCR tuubis, kuhu oli eelnevalt pipeteeritud 20 ul *master mix*'i; 6. korratati nii palju kui oli kolooniaid, mida oli vaja testida; 7. ümber külvatud kolooniatega petri tass pandi inkubeerimiskappi 37 °C juures üleöö bakterite kasvatamiseks.

Kolooniate PCR segud inkubeeriti järgneva PCRi programmiga, mis koosnes 5 etapist - algne denaturatsioon 3 minutit temperatuuriga 95 °C, sellele järgnes 25 tsükli kolmes etapis: denaturatsioon 94 °C juures 30 s, praimerid seondumine 60 °C juures 30 s ja DNA süntees 72 °C juures 30 s. Viimaseks etapiks oli lisasüntees 72 °C juures 5 min.

**Geelelektroforees:** Geelelektroforeesi teostati teada saamiseks, millistesse bakterikolooniatesse õnnestus transformeerida plasmid, mis sisaldas ligeeritud viirusjärjestust. Geelelektroforeesi tulemuste alusel valiti bakterikoloonias edasiseks kasvatamiseks ja miniprep-DNA eraldusteks.

Bakterikultuuride kasvatamiseks kolooniatest inokuleeriti igast individuaalsest kolooniast pisut materjali 2 ml LB söötmesse, mis sisaldas ampitsilliini 100 ng/ml. Bakterite vedelkultuure inkubeeriti 37 °C juures üleöö raputades 180 rpm.

**Miniprep-DNA eraldamine aluselise lüüsi meetodil:** 1. Valati 1,5 ml bakterikultuuri epsi, fuugiti täispööratel lauafuugis 2 min, eemaldati supernatant vaakumiga. 2. Resuspendeeriti bakterisade 300 ul E1 lahuses, segiti ja hoiti toatemperatuuril 5 min. 3. Lisati E2 lahust 300 ul, segati ettevaatlikult tuubi paar korda tagurpidi pöörates, pandi jää peale 1 min. 4. Fuugiti 4 °C juures 15 min. 5. 800 ul supernatanti tõsteti ümber uude eppendorfi, lisati DNA sadestamiseks isopropanooli 560 ul (0.7 vol), segati, inkubeeriti toatemperatuuril 5 min. 6. Fuugiti 10 min ja eemaldati supernatant. 7. Lisati 1 ml külma 70 % EtOH, fuugiti 5 min, eemaldati supernatant. 8. DNA sade kuivatati lambi all 30 min. 9. sade resuspendeeriti 50 ul mQ H<sub>2</sub>O ja mõõdeti kontsentratsioonid. Eraldatud miniprep-DNA saadeti sekveneerima Tartusse pJET-vektori UNI-praimeriga.

### 3.2.6 Genoomide assambleerimine ja omavaheline võrdlus

RT-PCRi produktide järjestused, mis olid sekveneeritud Tartus, võeti serverist programmi *WinSCP* abiga.

Saadud sekvensid sisestati programmi *Geneious Prime* (programmi versioon 2020.1.2) koos meie poolt järjetatud DSMZ kloni järjetusega (nn. D-klon) ning DSMZ poolt GenBank'i lisatud genoomide järjestustega.

Miniprep-DNA järjestustest otsiti esmalt VecScreen programmiga pJET-vektorile vastav järjestus ja see eemaldati sekveneerimisel saadud järjestusest. Igale kloonile vastavad PCRi produktide sekvensid kaardistati (funktsioon: *map*) CnMoV sarnaseima DSMZ-kloni genoomile ja selle põhjal assambleeriti igale kloonile vastav genoomi järjestus.

Pärast andmete töötlemist saadud konsensusjärjestusi kasutati genoomi annoteerimiseks (funktsioon: *annotate*). Sobemoviirustele iseloomulikud ORFid mis võiksid valke kodeerida, valiti välja järjestushomoloogia alusel võrreldes teiste sobemoviiruste genoomidega, valitud ORFid transleeriti (funktsioon: *translate*). Erinevate kloonide samade valkude omavahelist sarnasust võrreldi joonduse alusel kasutades *MUSCLE* algoritmi.

### 3.2.7 Vastete otsimine

Edasi teostati CnMoV genoomi ja sellelt transleeritavate valkude võrdlus teiste sobemoviiruste genoomide ja nendelt transleeritud valkudele *Geneious Prime* programmi abil. Selleks teostati

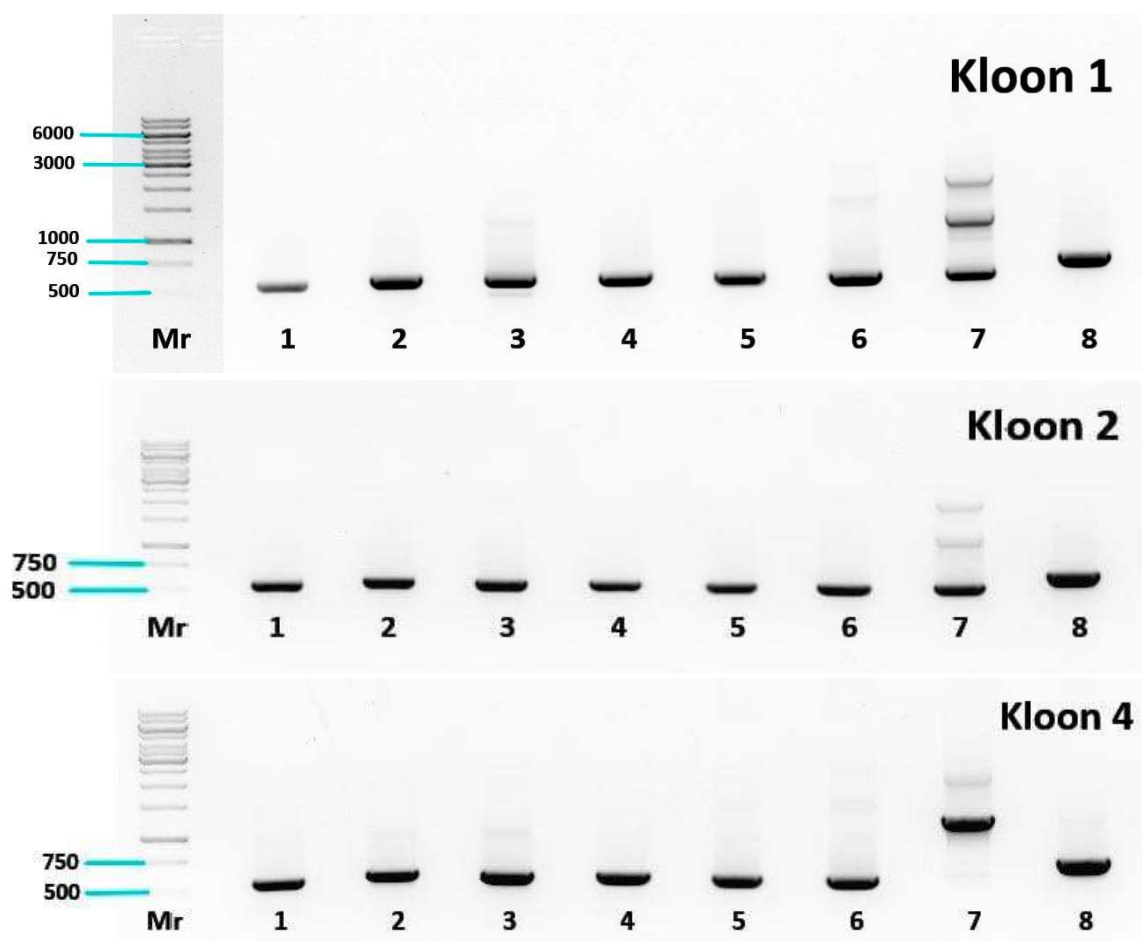
genoomide joondus ja genoomidelt transleeritavate valkude joondused, kasutades *MUSCLE* algoritmi.

## 4. TULEMUSED JA ARUTELU

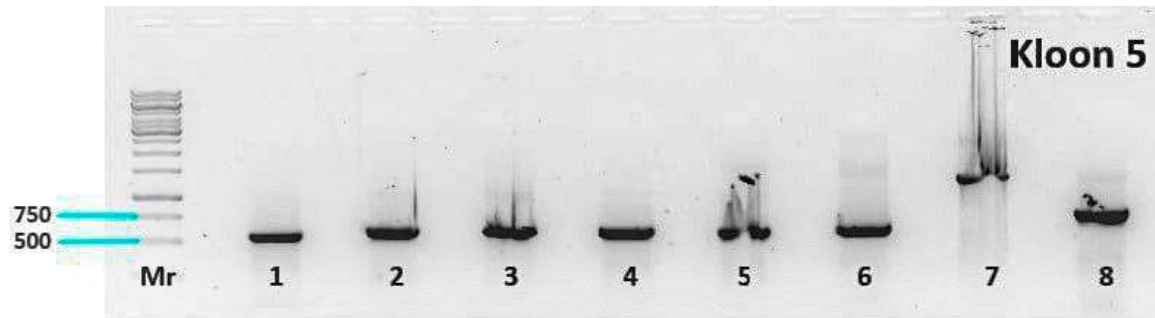
### 4.1. RT-PCR

**RNA eraldamine:** kloonide (1; 2; 4; 5) RNA kontsentratsiooni mõõtmise tulemused olid piisava kvaliteediga edasiseks töötlemiseks. Eraldatud RNAsid kasutati viiruse järjestuste amplifitseerimiseks.

**Elektroforees RT-PCR produktidele:** Saadud geeli vaatlemine (joonis 5) näitas, et kõik soovitud viirus isolaatide („kloonide“) genoomifragmendid õnnestus paljundada. Markeriga võrreldes selgus, et suur osa fragmentidest on oodatud suurusega, ca. 600 bp (Lisa 2). Võib rõhutada, et ainult praimeri paar 7 andis lisaks ebaspetsiifilisi fragmente, teiste praimeri paaridega neid praktiliselt ei tekkinud. Kloonide 4 ja 5 puhul eraldati geelist ca 1.2 kb fragment, kuna 600 bp fragment puudus. Bändid olid selgelt nähtavad ja kättesaadavad, mis võimaldas geelist puhastamist alustada.





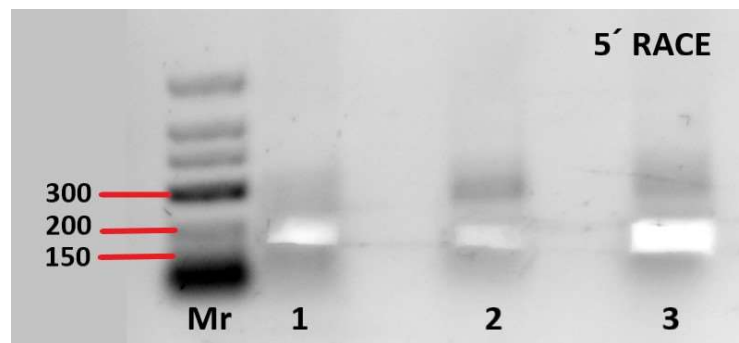


**Joonis 5. Kloonide RT-PCR-id.** CnMoV kloonidelt 1, 2, 4 ja 5 RT-PCR-il paljundatud fragmendid. Iga geeli vasakpoolsele rajale on kantud suurusmarker (Mr: 1kb DNA ladder, ThermoScientific), numbrid vasakul näitavad markeri fragmentide suurust aluspaarides (bp). Esimese ahela süntees viidi läbi CnMoV 3'otsa spetsiifilise praimeriga Cn3pR2. Sünteesitud esimeselt ahelalt viidi läbi PCR reaktsioonid. 1.-8. radadel on näha PCRi produktid, kui kasutati järgmisi praimeride paare: **1.** Cn5pF2 ja Cn1R; **2.** Cn2L ja Cn2R; **3.** Cn3L ja Cn3R; **4.** Cn4L ja Cn4R; **5.** Cn5L ja Cn5R; **6.** Cn6L ja Cn6R; **7.** Cn7L ja Cn7R; **8.** Cn8L ja Cn3pR2.

Kontsentratsioonide mõõtmise tulemusena selgus, et RT-PCR produktide eraldamine geelist õnnestus ning kontsentratsioonid on sobisid eraldatud DNA sekveneerimisele saatmiseks.

## 4.2. 5'RACE

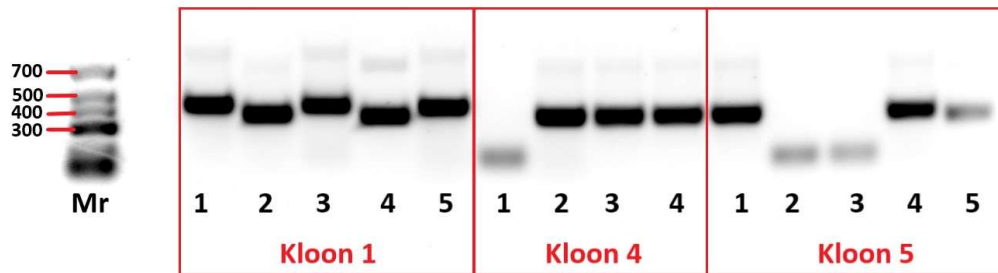
**Produktide puhastamine geelist:** saadud fragmendid (joonis 6) lõigati geelist edasiseks analüüsiks. Markeriga võrreldes need oodatava suurusega. Geeli pilti vaadates võib oletada et 5'RACE reaktsioonid õnnestusid. Kloon 2 5'RACE oli tehtud varasemalt eraldi ja seda ei ole käesolevas töös näidatud.



**Joonis 6. 5'RACE geelelektroforees.** Geeli vasakpoolsele rajale on kantud suurusmarker (Mr: LR DNA ladder, Thermo Scientific), numbrid vasakul näitavad markeri fragmentide suurust aluspaarides (bp). Pilt oli tehtud pärast mažoorsete RT-PCR fragmentide geelist välja lõikamist, et UV ekspositsiooni mõju sekveneerimise tulemusele minimeerida. Numbrid näitavad kloonide: 1 - kloon 1; 2 - kloon 4; 3 - kloon 5.

**Koloonia-PCR:** tuvastati sobiva suurusega inserti olemasolu plasmiidides, mis pärinesid bakterikolooniatest 1.1; 1.2; 1.3; 1.4; 1.5; 4.2; 4.3; 4.4; 5.1; 5.4; 5.5 (Joonis 7). Koloonia-PCR alusel otsustati paljundada materjal kolooniatest 1.2; 1.3; 1.4; 1.5; 4.2; 4.3; 4.4; 5.1; 5.4; 5.5. Plasmiidse DNA eralduseks sooritati miniprep-DNA eraldused. Kontsentratsioonid olid üldjoontes head. Kuna

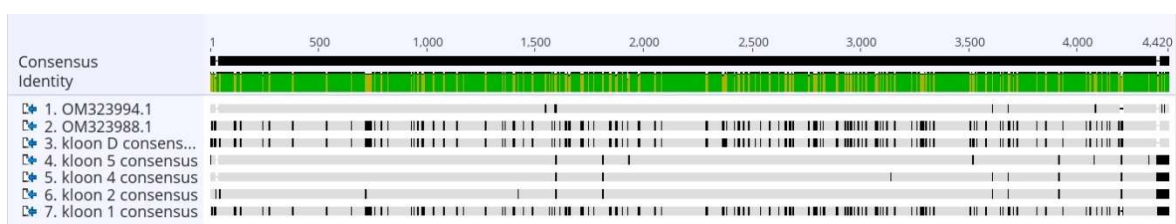
fragmentide suurused varieerusid pisut, saadeti kõik paljundatud plasmiidse DNA-d edasiseks sekveneerimiseks. Proove lahjendati kontsentratsioonini 300 ng/ul ja saadeti sekveneerima.



**Joonis 7. Koloonia-PCR-i geelelektroforees.** Geeli vasakpoolsele rajale on kantud suurusmarker (Mr: *LR DNA ladder*, ThermoScientific), numbrid vasakul näitavad markeri fragmentide suurust aluspaarides (*bp*). Joonis on jagatud kolmeks osaks: Kloon 1, viis kolooni; kloon 4, neli kolooni; ja kloon 5, viis kolooni.

### 4.3. Kloonide genoomide assambleerimine ja omavaheline võrdlus

Sekveneerimise tulemuseks olid sekventsides ferogrammid. Iga klooni kaardistamine näitas, et PCR produktid kuulusid CnMoV-le ja olid usaldusväärse kvaliteediga (Lisa 3). Saadud fragmendid katavad kogu genoomi (Lisa 3). Kaardistamisel kasutati kõige sarnasemat CnMoV DSMZ-järjestust: kas OM323994.1 või OM323988.1 lähtuvalt kõigi kloonide omavahelise joonduse tulemusest (Joonis 8). Iga sekveneeritud isolaadi ehk klooni kaardistamise abil saadi sellele vastav konsensusjärjestus. Konsensusjärjestuste ja referentstgenoomide omavaheline võrdlus *multiple alignment* funktsiooni abil näitas, et kloon 1 (ja kloon D) oli sarnasem referentsjärjestusele DSMZ-PV-0728 (OM323988.1, Lisa 4, joonis 2 A.) ja kloonid 2, 4 ja 5 olid sarnasemad referentsjärjestusele DSMZ-PV-1185 (OM323994.1, Lisa 4, joonis 2 B). Lisaks võrreldi nende genoomide identsusprotsente, millest selgus et üleval mainitud 2 grupi sisemine identsus on päris suur, keskmiselt 99.8%. Kahe grupi omavaheline identsus oli keskmiselt 96.5% (Lisa 5).



**Joonis 8. Genoomide omavaheline joondamine.** Ülemisel real on näidatud konsensusjärjestus, roheline joon näitab kloonide omavahelist identsust. Järjestuste 3' otstel asetsevad mustad alad näitavad sekveneerimata piirkonda, mis vajab 3' RACE abil määramist.

**SNP:** kloonidel vaadeldi genoomi muutumiskohtasid (Lisa 6). Kuna RT-PCR produkti ei kloonitud, vaid sekveneeriti otse peale genoomilt paljundamist, siis need erinevused on tõenäoliselt päriselt olemas. Kloonitud 5' RACE produkte sekveneeriti iga klooni kohta mitu, et kindel olla saadud tulemuses.

#### 4.4. Genoomi annoteerimine

Sekveneeritud ala sisaldas kõik lugemisraame, seega oli võimalik arvutada leitud ORFide pikkus, asukoht, transleeritava valgu pikkus ja molekulmass. Kuna kloonide omavaheline sarnasus oli väga suur, on need andmed näidatud DSMZ-PV-0728 (OM323988) isolaadi andmete põhjal (tabel 1). Nagu on tabelist näha, siis CnMoV ORFide asukohad ja pikkus sarnanevad teiste sobemoviiruste ORFide pikkustega ja asukohtadega, millest võib oletada et CnMoV võiks kuuluda sobemoviiruste perekonda. Tähtis on mainida, et siiani pole CnMoV taksonoomiliselt klassifitseeritud, kuigi oli oletatud et ta on sobemoviirus.

Tabel 1. CnMoV ORFide analüüsi tulemused

ORF	Pikkus (bp)	Asukoht	Valgu aminohapped	Molekulmass (kDa)
ORF1	330	124 - 453	109	12,17
ORFx*	384	712 - 329	128	14,306
ORF2a	1983	477 - 2459	660	72,771
ORF2b**	1584	2015 - 3598	527	59,304
ORF3	777	3447 - 4223	258	28,125

ORFide asukoht on näidatud vastavalt DSMZ-PV-0728 referentsgenoomile.

\* ORFx algab alternatiivse translatsiooni initsiatsiooni koodoniga

\*\* ORF2b algus on määratud raaminihke signaaliga (UUUAAAC)

#### 4.5. Võrdlus teiste sobemoviiruste genoomidega

GenBank andmebaasist võeti 20 sobemoviiruse genoomi, mis edasi sisestati *Genious Prime* programmi: AF208001 (*Subterranean clover mottle virus*); AJ608206 (*Rice yellow mottle virus*); AM940437 (*Sowbane mosaic virus*); AM990928 (*Imperata yellow mottle virus*); AY004291 (*Sesbania mosaic virus*); DQ680848 (*Cocksfoot mottle virus*); DQ875594 (*Southern bean mosaic virus*); EF091714 (*Ryegrass mottle virus*); HM754263 (*Velvet tobacco mottle virus*); JF495127 (*Soybean yellow common mosaic virus*); JN620802 (*Artemisia virus A*); JQ782213 (*Lucerne transient streak virus*); JX123318 (*Papaya lethal yellowing virus*); KC577469 (*Rottboellia yellow mottle virus*); KC577470 (*Solanum nodiflorum mottle virus*); KC778720 (*Turnip rosette virus*); KR996515 (*Cymbidium chlorotic spot virus*); LC019764 (*Blueberry shoestring virus*); M23021 (*Southern cowpea mosaic virus*); MT989351 (*Snake melon asteroid mosaic virus*). CnMoV näitena kasutati OM323988.

Kõikidest nende viiruste genoomidest kopeeriti ORFid (ORFx otsiti käsitsi, sest mitte kõikidel genoomidel ei ole see annoteeritud) ning transleeriti kõigist kopeeritud ORFidest valgud. ORF2b alguspunktiks valiti libisev järjestus UUUAAAC.

**Analüüs:** Kõikide sobemoviiruste genoomide joondati üksteisega (*MUSCLE*). Suurim identsus (48.4 %) tuvastati CnMoV ja CfMV genoomide vahel (Lisa 7). Samuti, kõikidelt genoomidelt transleeritud samad valgud joondati üksteisega (*MUSCLE*).

Nende joonduste alusel otsiti üles CnMoV genomist transleeritud valkudest üles sobemoviiruste omased konserveerunud motiivid:

P1 - tsinksõrme motiiv  $_{60}\text{CxxC}_{63}\dots_{91}\text{CxxC}_{94}$ ;

P2a võimalikud proteaasi lõikamiskohad polüproteiini domäänide vahel: (transmembraanne ankur, TM - E/M (pos. 138-139), seriinproteaasi domään, Pro - E/S (pos. 322-323), VPg ja C-terminaalne domään - proteolüüsi kohta ei saa määrata selle tõttu, et järjestus ei ole konserveerunud (joonis 9: konsensusjärestuse piirkond 500...600);

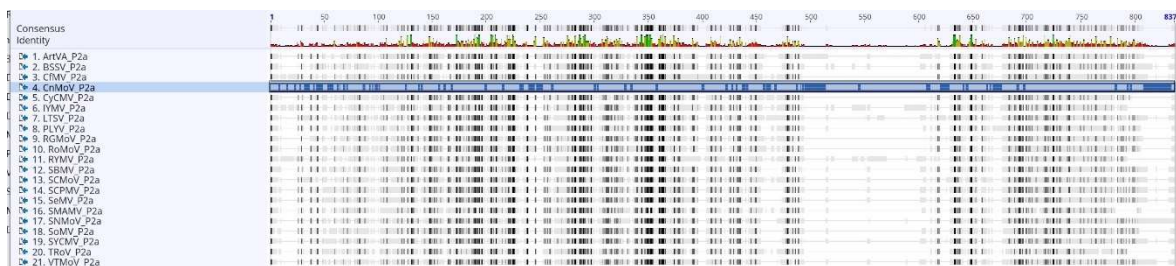
Transmembraanne ankur (kaks heeliksiit) pos. 12-32 ja 38-56;

Seriinproteaasi katalüütiline triaad (HDS) H - pos. 178; D - pos. 211; S - pos. 279;

VPg motiiv -  $_{387}\text{WADxxDxDDE}_{396}$ ;

P2b - RdRp motiiv GDD asukoht P2ab suhtes pos. 890 -892;

CP N-terminus on rikas aluseliste aminohapete pooltest:  $_{3}\text{KKxxxxxxKxRRxxxxxxKxRxRRRxR}_{30}$ .



Joonis 9. Sobemoviiruste P2a polüproteiinide MUSCLE joondus. Sinise taustaga on nädatud uuritav CnMoV.

Joonduste põhjal arvatati programmi poolt sobemoviiruste valkude omavahelised identsusprotsendid (Lisa 7). Selgus, et CnMoV valkudele kõige sarnasemad valgud on: *Lucerne transient streak virus* (LTSV) P1 15,2%; *Imperata yellow mottle virus* (IYMV) Px 26%; RYMV P2a 29,4% ja CfMV P2a 29,1%; CfMV P2ab 44,3% ja RYMV P2ab 43%; CfMV P2b (RdRp) 60,1% ja SCPMV 58,6%; CfMV CP 37,2%. On näha, et CnMoV valgud on üldiselt kõige sarnasemad CfMV valkudega.

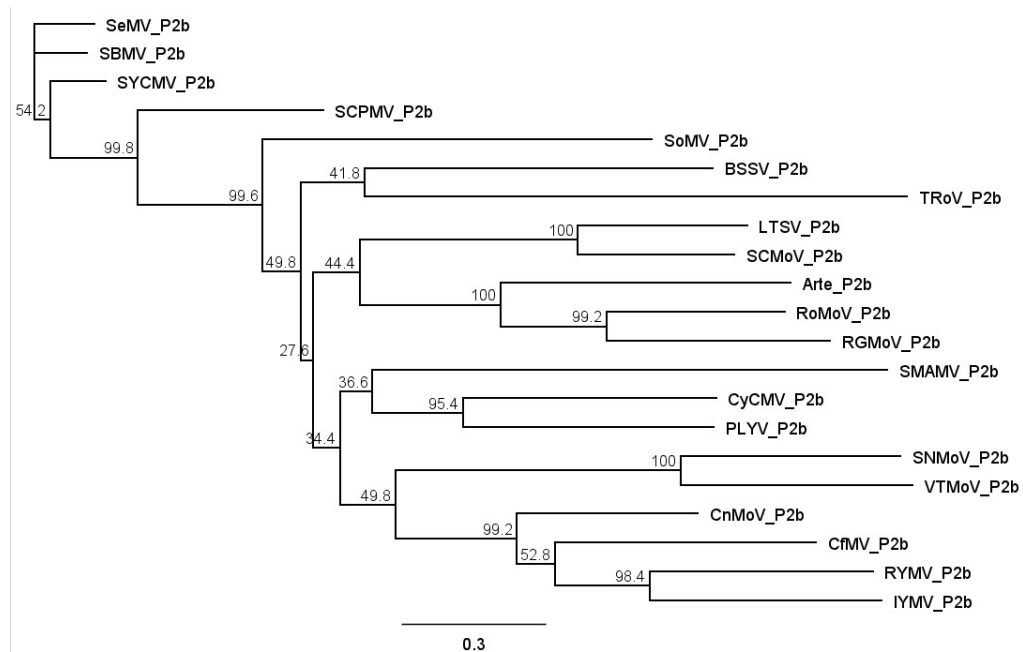
#### 4.6. Fülogeneetiline puu

Fülogeneetilise puu arvutamiseks valiti valk P2b ehk viiruse RNA sõltuv RNA polümeraas (RdRp), kuna see on kõige konserveerunud viirusvalk.

Fülogeneetiline analüüs näitas (joonis 10), et CnMoV kõige lähedasem viirus on keraheina laiguviirus (CfMV), millega CnMoV-il kattuvad geograafilised levialad ja mis on samuti kõrrelisi nakatav viirus. Analüüsi järgi võib CnMoV omada CfMV-ga ühist eellasliiki. Lisa 7 esitatud valkude joondusel põhinev identsuse tabel juba näitas nende kahe viiruse RdRp-de suurimat omavahelist sarnasust, seega oli nende fülogeneetiline lähedus oodatav. Suurim omavaheline sarnasus (37,2%) on ka CnMoV ja CfMV CP valkude vahel. Teiste sobemoviiruste valkude võrdlused (Lisa 7) näitasid suuremaid identsusprotsente LTSV P1 (15,2%), RYMV P2a (29,4%) ja IYMV Px (26,0%) suhtes kui

CfMV vastavate valkude osas (8,0%, 29,1% ja 19,7%). Fülogeneetilise analüüs näitab, et CnMoV ja CfMV moodustasid grupi RYMV-ga ja IYMV-ga, mis on ka kõrrelisi nakatavad viirused.

Kuna selgus, et CnMoV ja CfMV on fülogeneetiliselt lähedased liigid, siis oleks huvitav edasi uurida kattuvusi näiteks nende vektorite ja levalade osas, kuna viimased CnMoV uuringut viidi läbi 40 aastat tagasi.



**Joonis 10. Fülogeneetiline puu.** Puu konstrueeriti suurima tõepära (*maximum-likelihood*) meetodil 500 kordusega, kasutades GeneiousPrime'i integreeritud PhyML 3.0 programmi (Guindon jt 2010). Aminohapete asendusmaatriksiks valiti LG (Guindon ja Gascuel 2003). Numbrid harude juures näitavad mitu korda 100-st oli toetus fülogeneetilise grupeerumise kohta sama.

## 5. Kokkuvõte

Sugapea laiguviirus (CnMoV) on positiivse polaarsuse RNA genoomiga viirus, mis nakatab kõrrelisi ja mis pole siiani taksonoomiliselt klassifitseeritud, kuigi on oletatud, et tegemist on sobemoviirusega. Meie laboris oli mass-sekveneerimise meetodil järjestatud Leibniz'i instituudi kogust pärinev DSMZ-PV-0728 isolaat (D-kloon), kuid laborikogus leidis veel neli sekveneerimata isolaati Julius Kühn'i instituudi kogust. Käesoleva töö esimeseks eesmärgiks oli Julius Kühn'i instituudi kogust pärinavate CnMoV isolaatide sekveneerimine. Neli Saksamaalt saadud isolaati sekveneeriti alustades 5' otsast (5' RACE strateegiaga). Sekventsid olid assambleeritud ja jäi puudu vaid 3' otsa regiooni. Selle töö käigus ilmusid GenBank'is teiste CnMoV isolaatide järjestused DSMZ-PV-728 ja DSMZ-PV-1185. Seega oli võimalik neid võrrelda meie poolt saadud sekventsidega. Joondus näitas, et Kloon 1 ja DSMZ-PV-0728 omasid omavahel 99,8% identsust, kloonid 2, 4, 5 ja DSMZ-PV-1185 omasid keskmiselt 99,8% identsust. Kahe grupi omavaheline identsus oli keskmiselt 96,5%. Teiseks ülesandeks oli CnMoV genoomi annoteerimine ja analüüs. Analüüsi tulemuseks näidati, et ORFide paiknemisalad sarnanevad sobemoviiruste ORFide pikkustega ja asukohtadega. Kolmandaks eesmärgiks oli CnMoV genoomi ja valkude võrdlemine juba kirjeldatud (kokku kakskümmend) sobemoviiruste järjestustega. Sobemoviiruste valkude omavaheline identsuse protsendi arvutamine näitas, et suurimat sarnasust CnMoV valkude suhtes näitasid järgmised valgud: *Lucerne transient streak virus* (LTSV) P1 15,2%; *Imperata yellow mottle virus* (IYMV) Px 26%; *Rice yellow mottle virus* (RYMV) P2a 29,4% ja *Cocksfoot mottle virus* (CfMV) P2a 29,1%; CfMV P2b 60,1%, ning CfMV CP: 37,2%. Viimaseks arvutati fülogeneesipuu kõige konserveerunuma valgu põhjal, milleks on P2b ehk RNA sõltuv RNA polümeraas. Arvutus näitas, et CnMoV paikneb CfMV lähedal ja nii RYMV kui ka IYMV on nendega lähedalt seotud. Kuigi CnMoV 3' otsa järjestus pole 3'RACE-ga kinnitatud, näitab selle genoomi mitmete isolaatide järjestamine, et CnMoV kuulub sobemoviiruse perekonda ja võib CfMV-ga omada ühist eellasliiki. Tulevikuks jääb CnMoV genoomi 3' otsa sekveneerimine. Samuti oleks vaja uurida, millised on CnMoV vektorid, kuna uuringud selle kohta on tehtud rohkem kui 40 aastat tagasi ja nende uuringute tulemused on vastuolulised.

## **6. Tänuavaldused**

Tahaksin tänada selle töö juhendajaid Merike Sõmera't ja Cecilia Sarmiento't, nende suur töö eest. Ilma nendeta see töö ei oleks kunagi valminud. Samuti tahan tänada kõiki õpetajaid, kes õpetasid mind need aastad. Lisaks olen väga tänulik õppespetsialist Marju Hõimla'le Loodusteaduskonna dekanadist, kes alati vastas minu sõnumitele ja aitas ülikooliõpinguid puudutava informatsiooniga.

## 7. Kasutatud kirjandus

Basic Local Alignment Search Tool. (accessed on 2021). Available online: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Brook, J.A. 1972. "Lolium mottle virus." *Plant Pathology*, 21, 118-120. <https://doi.org/10.1111/j.1365-3059.1972.tb01740.x>

Catherall, P.L.; Andrews, P.A. and Chamberlain, J.A. 1977. "Host-ranges of cocksfoot mottle and cynosurus mottle viruses." *Annals of Applied Biology*, 87, 233-235. <https://doi.org/10.1111/j.1744-7348.1977.tb01880.x>

Camarço, R.F.E.A.; Lima, J.A.A. and Pio-Ribeiro, G. 1998. "Transmissão e presença em solo do "papaya lethal yellowing virus" *Fitopatologia Brasileira*, 23, 453-458.

Descriptions of Plant Viruses. (accessed on 2021). Available online: <http://www.dpvweb.net>

Dinman, J.D. 2012. "Control of gene expression by translational recoding." *Advances in Protein Chemistry and Structural Biology*, 86, 129-149. <https://doi.org/10.1016/B978-0-12-386497-0.00004-9>

Govind, K. and Savithri, H.S. 2010. "Primer-independent initiation of RNA synthesis by SeMV recombinant RNA-dependent RNA polymerase." *Virology*, 401, 280-292. <https://doi.org/10.1016/j.virol.2010.02.025>

Hacker, D.L. and Sivakumaran, K. 1997. "Mapping and expression of southern bean mosaic virus genomic and subgenomic RNAs." *Virology*, 234, 317-327. <https://doi.org/10.1006/viro.1997.8667>

ICTV taxonomy browser (accessed on 2021). Available online: <https://ictv.global/taxonomy>

ICTV, Family Solemoviridae, Genus Sobemovirus. (accessed on 2021) Available online: <https://ictv.global/report/chapter/solemoviridae/solemoviridae/sobemovirus>

Ling, R.; Pate, A.E.; Carr, J.P. and Firth, A.E. 2013. "An essential fifth coding ORF in the sobemoviruses." *Virology*, 446, 397-408. <https://doi.org/10.1016/j.virol.2013.05.033>

Mohamed, N.A. 1978. "Physical and chemical properties of Cynosurus mottle virus." *Journal of General Virology*, 40, Issue 2. <https://doi.org/10.1099/0022-1317-40-2-379>

Mohamed, N.A. 1981. "An ultrastructural study of three grasses naturally infected with Cynosurus mottle virus." *Journal of Phytopathology*, 100, 121-129. <https://doi.org/10.1111/j.1439-0434.1981.tb04629.x>

Mohamed, N.A. and Mossop, D.W. 1981. "Cynosurus and cocksfoot mottle viruses: a comparison." *Journal of General Virology*, 55, Issue 1. <https://doi.org/10.1099/0022-1317-55-1-63>

Mohamed, N.A. 2012. "Cynosurus mottle virus, a virus affecting grasses in New Zealand." *New Zealand Journal of Agricultural Research*, 21, 709-714. <https://doi.org/10.1080/00288233.1978.10427472>



Paul, H.L.; Querfurth, G. and Huth, W. 1980. "Serological studies on the relationships of some isometric viruses of Gramineae." *Journal of General Virology*, 47, 67-77. <https://doi.org/10.1099/0022-1317-47-1-67>

Peters, D.; Engels, C. and Sarra, S. 2012. "Natural spread of plant viruses by birds." *Journal of Phytopathology*, 160, 591-594. <https://doi.org/10.1111/j.1439-0434.2012.01937.x>

Sarra, S. and Peters, D. 2003. "Rice yellow mottle virus is transmitted by cows, donkeys, and grass rats in irrigated rice crops." *Plant Disease*, 87, 804-808. <https://doi.org/10.1094/PDIS.2003.87.7.804>

Sarra, S.; Oevering, P.; Guindo, S. and Peters, D. 2004. "Wind-mediated spread of Rice yellow mottle virus (RYMV) in irrigated rice crops." *Plant Pathology*, 53, 148-153. <https://doi.org/10.1111/j.0032-0862.2004.00981.x>

Sverdlov, E. and Azhikina, T. 2005. "Primer walking". *Wiley Online Library*. (accessed on 2021). Available online: <https://doi.org/10.1038/npg.els.0005382>

Sömera, M.; Sarmiento, C. and Truve, E. 2015. "Overview on sobemoviruses and a proposal for the creation of the family Sobemoviridae." *Viruses*, 7, 3076-3115. <https://doi.org/10.3390/v7062761>

Sömera, M.; Fargette, D.; Hébrard, E.; Sarmiento, C. and the ICTV Consortium. 2021. "ICTV Virus Taxonomy Profile: Solemoviridae 2021." *Journal of General Virology*, 102, Issue 12. <https://doi.org/10.1099/jgv.0.001707>

Ziegler, A.; Cowan, G. and Torrance, L. 2009. "Comparative sequence analysis and serological and infectivity studies indicate that cocksfoot mild mosaic virus is a member of the genus Panicovirus." *Archives of Virology*, 154, 1545-1549. <https://doi.org/10.1007/s00705-009-0468-8>

Tamm, T. and Truve, E. 2000. "Sobemoviruses." *Journal of Virology*, 74, 6231-6241. <https://doi.org/10.1128/jvi.74.14.6231-6241.2000>

5' RACE System for Rapid Amplification of cDNA Ends. (accessed on 2021). Available online: <https://www.thermofisher.com/ee/en/home/references/protocols/nucleic-acid-amplification-and-expression-profiling/cdna-protocol/5-race-system-for-rapid-amplification-of-cdna-ends.html>

## 8. Lisad

### Lisa 1. Sobemoviiruse liigid

Tabel 1. Sobemoviiruste taksonoomiline nimekiri (2021 väljalase\*)

Liik	Lühend	GenBank kirje
<i>Artemisia virus A</i>	ArtVA	JN620802
<i>Blueberry shoestring virus</i>	BSSV	LC081344
<i>Cocksfoot mottle virus</i>	CfMV	Z48630
<i>Cymbidium chlorotic mosaic virus</i>	CyCMV	LC019764
<i>Imperata yellow mottle virus</i>	IYMV	AM990928
<i>Lucerne transient streak virus</i>	LTSV	JQ782213
<i>Papaya lethal yellowing virus</i>	PLYV	JX123318
<i>Physalis rugose mosaic virus</i>	PhyRMV	MK681145
<i>Rice yellow mottle virus</i>	RYMV	AJ608206
<i>Rottboellia yellow mottle virus</i>	RoMoV	KC577469
<i>Ryegrass mottle virus</i>	RGMoV	EF091714
<i>Sesbania mosaic virus</i>	SeMV	AY004291
<i>Sobemovirus smamv</i>	SMAMV	MT989351
<i>Solanum nodiflorum mottle virus</i>	SNMoV	KC577470
<i>Southern bean mosaic virus</i>	SBMV	DQ875594
<i>Southern cowpea mosaic virus</i>	SCPMV	M23021
<i>Sowbane mosaic virus</i>	SoMV	AM940437
<i>Soybean yellow common mosaic virus</i>	SYCMV	JF495127
<i>Subterranean clover mottle virus</i>	SCMoV	AF208001
<i>Turnip rosette virus</i>	TRoV	KC778720

\*<https://ictv.global/taxonomy>

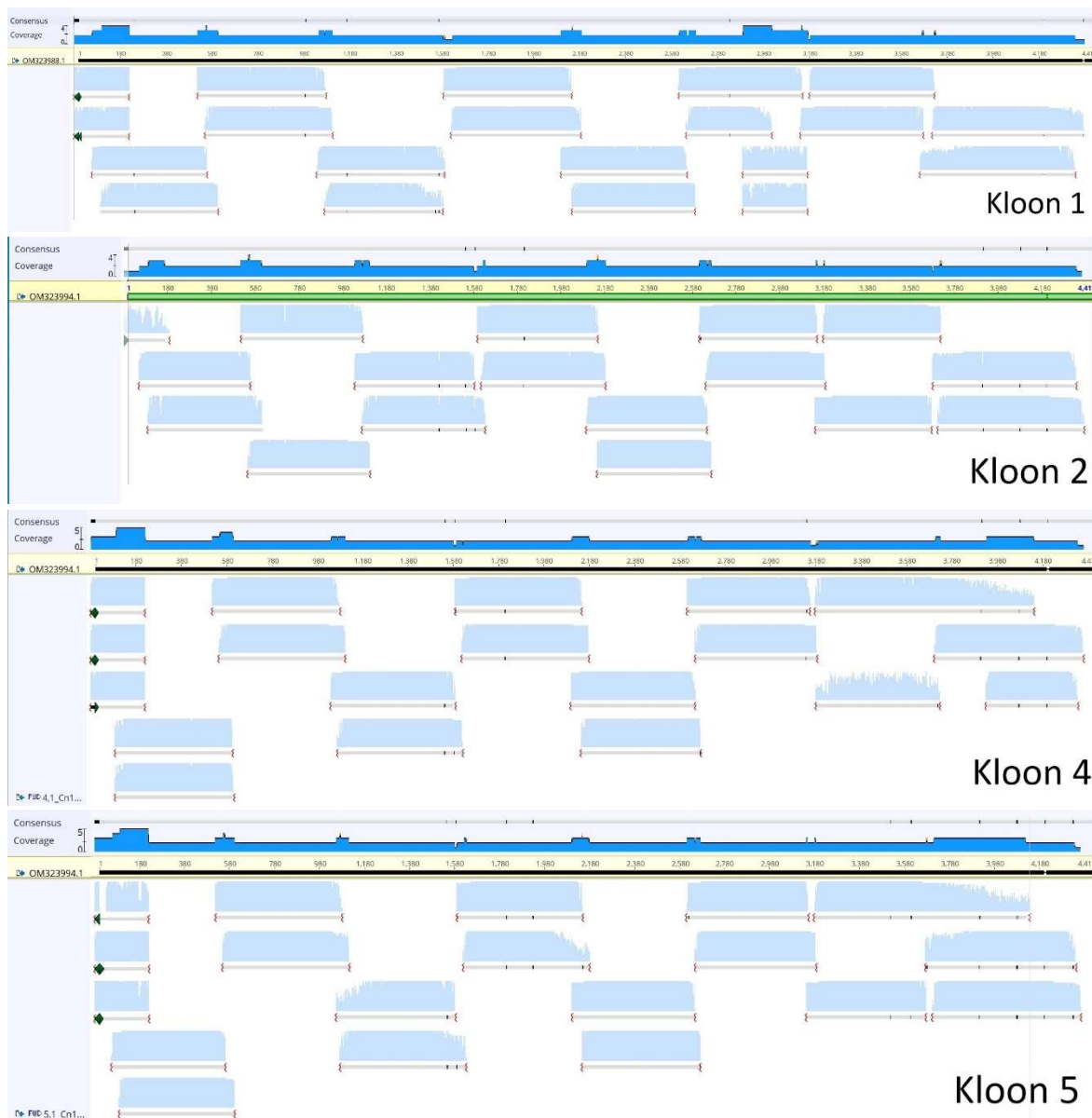
## Lisa 2. Kasutatud praimerid

Tabel 2. Praimeri paarid mida analüüsidel kasutati

Praimer	järjestus	koht	Bp
ADA nested	AGTTTAGAATCCTCAGAGGTC		245
Cn153R	AAAATAGGGCTCGTCAGTCAGG	224-202	
Cn5pF2	GTGAACTGAAATCGAGATAATAATCTTTCG	50-76	565
Cn1R	GGGATCGCCAACCATAATC	614-596	
Cn2L	AGGAGCGTGTTGAGGAACAG	508-527	611
Cn2R	ATGGCGAAGTCCAACATCTC	1118-1099	
Cn3L	ATGGCCAAACGAGGTAAGTG	1036-1055	601
Cn3R	AACTTCGGTATGAGGGGTCA	1636-1617	
Cn4L	CCACAGGTTAGCCCACTTTG	1578-1597	615
Cn4R	CTCAGGGAAGGTCTCTGCTG	2192-2173	
Cn5L	ATGCGATTCTTCAAGTTGG	2091-2110	595
Cn5R	CAGAAGAGATAAGGCGAAAACG	2685-2664	
Cn6L	TGACCCAGTACGCTTGTTTG	2603-2622	598
Cn6R	CAATGCGGAGTATGCCTCTT	3200-3181	
Cn7L	CATAGCCATGGGTGATGATTC	3134-3154	599
Cn7R	GAACAGAGCAGGCGTCACTT	3732-3713	
Cn8L	CTCGGTGAGTTCGTTTCAGC	3664-3682	713
Cn3pR2	AGCGCGTTTTGTGAAACCGAC	4376-4356	

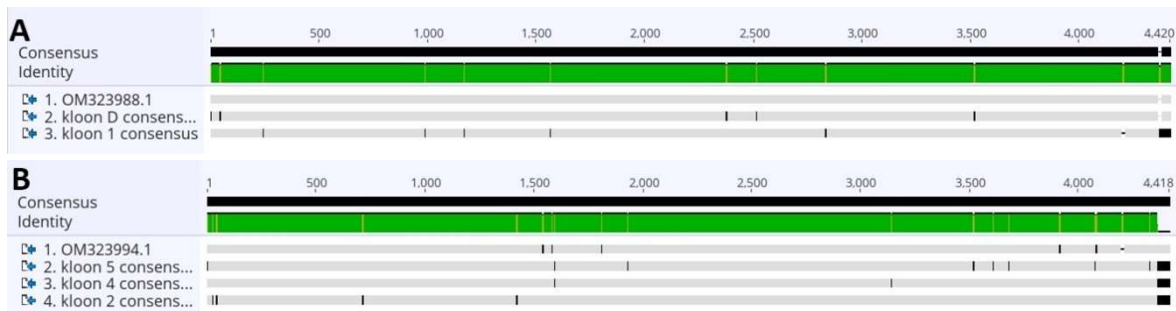
Praimerite positsioonid on näidatud DSMZ-PV-0728 isolaadi – OM323988.1 (GenBank järjestus) suhtes.

### Lisa 3. Genoomi kaardistamine



**Joonis 1. Sekvenside assembleerimine.** Ülevalt alla on näidatud isolaadid 1;2;4;5. Rohelise noolekesega on näidatud 5’RACE produktile kinnitunud adapter-praimer järjestus, mis eemaldati enne konsensusjärjestuse moodustamist. Helesinised ristkülikud näitavad sekveneerimise signaali kvaliteetsust (mida homogensem, seda parem on materjal). Mustad täpid konsensusjärjestustes (hall pidevjoov) näitavad nukleotidi erinevust referentsgenoomiga (OM323988 isolaat 1 puhul; OM323994 isolaatide 2, 4 ja 5 puhul).

## Lisa 4. Konsensusjärjestuste võrdlus



**Joonis 2. Konsesusjärjestuste omavaheline MUSCLE joondus kahes gruppis.** Grupp A all on kloon1, meie laboris sekveneritud DSMZ-PV-0728 isolaat (kloon D) ja Saksamaal sekveneritud DSMZ-PV-0728 isolaat (OM323988). Grupp B all on kloonid 2, 4, 5 ja DSMZ-PV-1185 isolaat (OM323994). Musta kriipsuga on näidatud genoomi muutumiskohad.

## Lisa 5. CnMoV isolaatide (kloonide) genoomide identsus protsentides

	OM323994.1	OM323988.1	kloon D co...	kloon 5 con...	kloon 4 con...	kloon 2 con...	kloon 1 con...
OM323994.1		96.6%	96.5%	99.7%	99.8%	99.8%	96.5%
OM323988.1	96.6%		99.9%	96.5%	96.6%	96.6%	99.8%
kloon D consensus	96.5%	99.9%		96.5%	96.5%	96.5%	99.7%
kloon 5 consensus	99.7%	96.5%	96.5%		99.8%	99.8%	96.4%
kloon 4 consensus	99.8%	96.6%	96.5%	99.8%		99.9%	96.5%
kloon 2 consensus	99.8%	96.6%	96.5%	99.8%	99.9%		96.5%
kloon 1 consensus	96.5%	99.8%	99.7%	96.4%	96.5%	96.5%	

**Joonis 3. Meie laboris sekveneritud kloonide ja DSMZ kogu haldaja poolt sekveneritud CnMoV genoomide omavaheline identsus (arvutatud MUSCLE joonduse põhjal).**

## Lisa 6. Leitud SNP-id

Tabel 3. Julius Kuhn'i instituudi kogust pärinevate kloonide genoomides identifitseeritud nukleotiidsete muudatuste (SNP) kohad

Kloon*	tüüp	muudatus	koht
Kloon 1	Deletsion	G	4201
	SNP	C - T	2834
	SNP	T - C	1567
	SNP	G - A	1169
	SNP	G - A	988
Kloon 2	Insertion	G	4202-4201
	SNP	C - T	4080
	SNP	A - G	3911
	SNP	T - C	1811
	SNP	T - C	1544
Kloon 4	Insertion	G	4202-4201
	SNP	C - T	4080
	SNP	A - G	3911
	SNP	C - T	3139
	SNP	T - C	1811
	SNP	G - A	1586
	SNP	T - C	1544
Kloon 5	SNP	A - G	4328
	Insertion	G	4202-4201
	SNP	C - T	4080
	SNP	T - C	4076
	SNP	A - G	3911
	SNP	G - A	3608
	SNP	T - C	3516
	SNP	G - A	1930
	SNP	T - C	1811
	SNP	G - A	1586
	SNP	T - C	1544

\*Kloon 1 võrreldi DSMZ-PV-0728 (OM323988) suhtes ja kloonid 2, 4 ning 5 võrreldi DSMZ-PV-1185 (OM323994) suhtes.

## Lisa 7. CnMoV genoomide ja valkude võrdlus teiste sobemoviiruste valkudega.

Tabel 4. CnMoV genoomide ja valkude identsus (protsentides) teiste sobemoviiruste valkudega

Viirused	Genoomid	P1	Px	P2a	P2ab	P2b	CP
ArtVa	32,8	8,0	7,6	21,0	33,4	46,4	13,9
BSSV	33,3	8,1	4,9	19,7	32,4	46,1	19,9
CfMV	48,4	8,0	19,7	29,1	44,3	60,1	37,2
CyCMV	33,7	11,1	8,9	21,6	34,5	50,2	23,7
IYMV	39,8	6,4	26,0	25,1	40,8	55,9	31,6
LTSV	32,8	15,2	7,5	24,3	36,1	51,2	20,2
PLYV	34,4	6,7	6,5	21,3	32,9	45,0	21,1
RGMoV	40,5	9,8	7,8	22,9	33,8	47,2	14,8
RoMoV	33,8	6,1	10,4	21,5	33,6	47,7	15
RYMV	34,4	9,5	19,5	29,4	43,0	56,0	29,9
SBMV	32,9	7,8	5,8	24,0	36,1	49,7	20,7
SCMoV	34,7	8,4	7,4	22,8	35,0	51,6	19,9
SCPMV	33,2	6,6	8,3	22,4	34,7	58,6	21,9
SeMV	34,8	8,9	4,5	24,9	36,3	49,0	21,3
SMAMV	35,3	5,6	9,7	20,5	32,2	44,8	16,4
SNMoV	34,5	8,4	6,9	22,5	32,8	45,1	22
SoMV	35,1	8,5	6,0	23,3	35,3	49,6	17,1
SYCMV	34,4	6,5	7,8	24,2	35,8	49,3	21,4
TRoV	33,3	9,3	5,6	19,4	31,7	45,9	22,3
VTMOV	35	11,9	10,1	24,0	33,0	43,0	23,1

Kollasega on märgitud kõige suuremad identsusprotsendid.