

LuK-tutkielma

**SARS-CoV-2-virus Jyväskylässä jätevesiseurannan
perusteella**

Virpi Autio, Siiri Parviainen & Max Sundström



Jyväskylän yliopisto

Bio- ja ympäristötieteiden laitos

2.12.2022

JYVÄSKYLÄN YLIOPISTO, Matemaattis-luonnontieteellinen tiedekunta
Bio- ja ympäristötieteiden laitos
Biologian kandidaattiohjelma

Virpi Autio, Siiri Parviainen ja Max Sundström: SARS-CoV-2-virus Jyväskylässä jätevesiseurannan perusteella
Kandidaatin tutkielma: 38 s., 7 liitettä (28 s.)
Tutkielman ohjaajat: Professori Marja Tirola, FT Elina Ruuhilehto
Marraskuu 2022

Hakusanat: Jätevesi, lineaarinen sekamalli, sekvensointi, virologia, virusvariantti

Maaliskuusta 2020 lähtien SARS-CoV-2-virus on ollut osana jokaisen ihmisen arkipäivää. Pandemian aikana virus on muuntunut uusiksi virusvarianteiksi, jotka ovat vuorollaan olleet aikansa valtavariantteja. Pandemian etenemistä on seurattu monilla menetelmillä. Yhdeksi lupaavaksi seurantamenetelmäksi on osoittautunut jätevesiseuranta. Tutkimusten mukaan SARS-CoV-2-infektion saaneet henkilöt erittävät virus-RNA:ta ulosteeseensa ja tämä päättyy jätevesiin. Tällöin tutkimalla jätevesien virus-RNA:n määrää voidaan seurata infektioiden määrää väestössä. Tässä kandidaatin tutkielmassa tutkitaan jätevesien sisältämää SARS-CoV-2-viruksen määrää Jyväskylän jätevedessä ja lisäksi positiivisia SARS-CoV-2-nenänielunäytteitä. Jätevesinäytteistä tutkittiin SARS-CoV-2-positiivisuus ja näytteistä sekvensoitiin yksi Ion Torrent -menetelmällä. Potilasnäytteitä oli eri aikapisteistä ja niiden perusteella selvitettiin, mitä virusvariantteja Jyväskylässä on esiintynyt pandemian aikana. Potilasnäytteet sekvensoitiin Sanger-sekvensoinnilla. Jätevesinäytteiden SARS-CoV-2-osoituksen myötä saatiin selville C_q-arvo, joka voidaan muuntaa viruskopioiden määräksi jätevedessä muunnoskaavaa hyödyntäen. Tähän tietoon yhdistettiin Terveystieteiden ja hyvinvoinnin laitoksen (THL) COVID-19-pandemia dataa ja jätevesipuhdistamon virtaamatietoja. Lineaarinen sekamalli rakennettiin ennustamaan viruskopioiden määrää jätevedessä ja tämän mallin ennustajiksi valikoitiin mitattu fluoresenssi qPCR-osoituksessa, tehohoidossa olleet potilaat COVID-19 kanssa ja takia ja SARS-CoV-2-tapausmäärät. Lisäksi tämän tutkimuksen tuottamalle viruskopioiden määrälle jätevedessä tehtiin virtaamakorjaus. Tätä tietoa verrattiin THL:n tuottamaan virtaamakorjatun RNA:n määrään jätevedessä näytteenottopäivinä Wilcoxonin testin avulla, jonka mukaan THL:n tämän tutkimuksen tulosten välillä on tilastollisesti merkittävä ero. Tämän tutkimuksen saamat virtaamakorjatun RNA-arvot ovat keskimäärän 10-kertaa suurempia kuin THL:n tulokset. Tutkielman tulosten perusteella tulevaisuudessa jätevedessä olevan virus-RNA:n määrää voitaisiin ennustaa THL:n tuottaman datan avulla käyttäen tietoa tilastoiduista tapauksista. Tällöin malli kertoo, paljon väestössä on tilastoituja ja tilastoimattomia tapauksia perustuen virusmäärään jätevedessä. Regressiomallien R²-arvot ovat kuitenkin alhaisia mallinnettaessa COVID-19-dataa, mikä pitää ottaa huomioon tehdessä tulosten tulkintoja.

UNIVERSITY OF JYVÄSKYLÄ, Faculty of Mathematics and Science
Department of Biological and Environmental Science
Bachelor's Degree Programme in Biology

Virpi Autio, Siiri Parviainen ja Max Sundström: The study of SARS-CoV-2-virus in Jyväskylä using data collected from wastewater
Bachelor of Science Thesis: 38 p., seven supplements (28 p.)
Supervisors: Professor Marja Tirola, PhD Elina Ruuhilehto
November 2022

Keywords: Linear mixed model, sequencing, sewage, virology, virus variant

Since the start of the COVID-19 pandemic many methods have been used to monitor the progress of the pandemic and the current variants circulating the population. Wastewater has proven to be one of the most promising monitoring methods. Studies have shown that people infected with SARS-CoV-2 excrete viral RNA into sewage water. In this case, the level of viral RNA in wastewater can be used to monitor the level of infection in the population. This bachelor thesis investigates the amount of SARS-CoV-2 virus in wastewater in the treatment plant of Nenäinniemi, Jyväskylä and positive SARS-CoV-2 nasopharyngeal samples from patients. Positive sewage samples were sequenced unsuccessfully. Patient samples were taken at different time points to determine which virus variants were present in Jyväskylä during the pandemic and the samples were sequenced. The original intention was to compare the results of the nasopharyngeal sample sequences with the sequences obtained from wastewater samples. However, this could not be achieved. The detection of SARS-CoV-2 in wastewater samples provided a C_q-value that can be converted to the number of virus copies in wastewater using a transformation formula. This data was then combined with COVID-19 data from the National Institute for Health and Welfare (THL) and flow rates of the wastewater treatment plant. A linear mixed model was fitted to predict the number of viral copies in the effluent and the predictors of this model were selected based on information criteria. Predictors selected were measured fluorescence in qPCR assay, ICU patients with COVID-19 and SARS-CoV-2 case counts. The viral copy number in the wastewater produced by this study was flow corrected and this data was compared with the flow corrected RNA in the wastewater on the sampling days produced by THL using the Wilcoxon's test. The result was that there is a statistically significant difference between the results from THL and the results of this study. When comparing the flow corrected values of this study with THL's, the values were averagely ten times higher in this study. This study found that the amount of viral RNA in wastewater could be predicted in the future using data generated by THL and by using information from confirmed cases. This will help in monitoring the pandemic. However, R² values of the regression models are lower when modeling COVID-19 data, which must be considered when interpreting the results.

SISÄLLYSLUETTELO

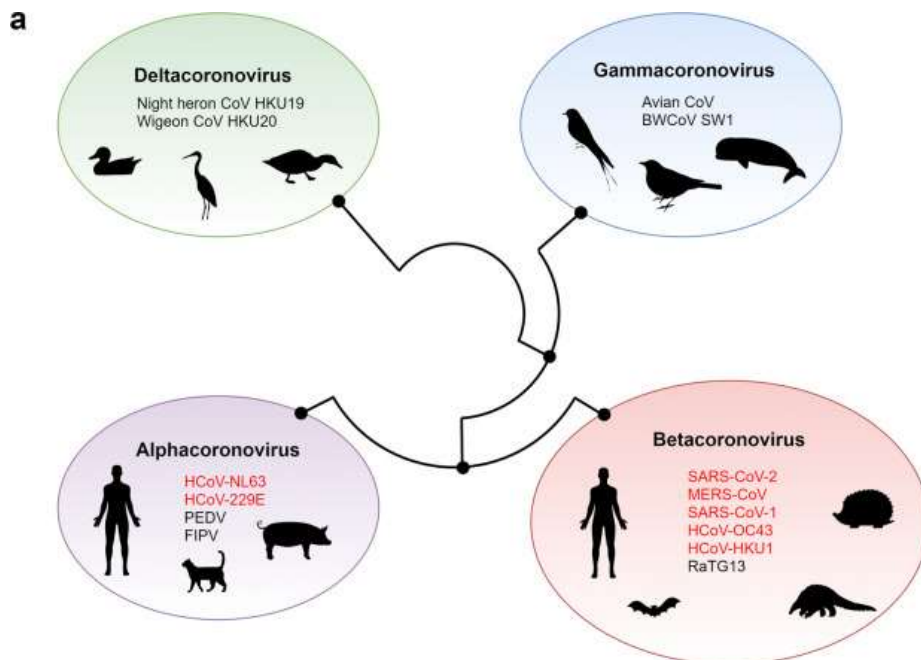
1 JOHDANTO	1
2 AINEISTO JA MENETELMÄT.....	6
2.1 Positiivisten SARS-CoV-2-nenänielunäytteiden Sanger-sekvensointi...	7
2.1.1 RT-reaktio	8
2.1.2 qPCR-reaktio	8
2.1.3 EXO/SAP-puhdistus.....	9
2.1.4 Toinen PCR.....	9
2.1.5 PCR-tuotteiden puhdistus.....	9
2.1.6 ABI-sekvensointianalyysi.....	10
2.1.7 Sanger-sekvenssien käsittely.....	10
2.2 Jätevesinäytteiden sekvensointi Ion Torrent -menetelmällä.....	12
2.2.1 Näytteiden konsentroitinto.....	12
2.2.2 RT-reaktio	13
2.2.3 Ensimmäinen PCR.....	13
2.2.4 Toinen PCR.....	14
2.2.5 PCR-tuotteiden puhdistus.....	14
2.2.6 Qubit-mittaus	15
2.2.7 Poolaus ja Tapestation.....	15
2.2.8 Ion Torrent.....	15
2.3 Tilastollinen analyysi.....	16
2.4 Systemaattinen katsaus kirjallisuudesta	18
3 TULOKSET	18
3.1 Sanger-sekvenssien tulokset.....	18
3.2 Jätevesinäytteet.....	20
3.2.1 SARS-CoV-2-osoitus qPCR-reaktiolla.....	20
3.2.2 Ensimmäinen PCR.....	20
3.2.3 Toinen PCR.....	21
3.2.4 Qubit-mittaus poolia varten ja sekvensoinnin tulokset	22
3.2.5 Tilastolliset mallit ja tunnusluvut.....	23
3.3 Systemaattisen kirjallisuuskatsauksen tulokset.....	30
TULOSTEN TARKASTELU.....	32
4.1 Sanger-sekvensoinnin tuloksista.....	32
4.2 Jätevesinäytteiden SARS-CoV-2-osoitus ja PCR-tulokset	33
4.3 Tilastolliset mallit ja tunnusluvut	34
KIITOKSET	39
KIRJALLISUUSLUETTELO	39

LIITE 1	49
LIITE 2	61
LIITE 3	66
LIITE 4	70
LIITE 5	71
LIITE 6	72
LIITE 7	75

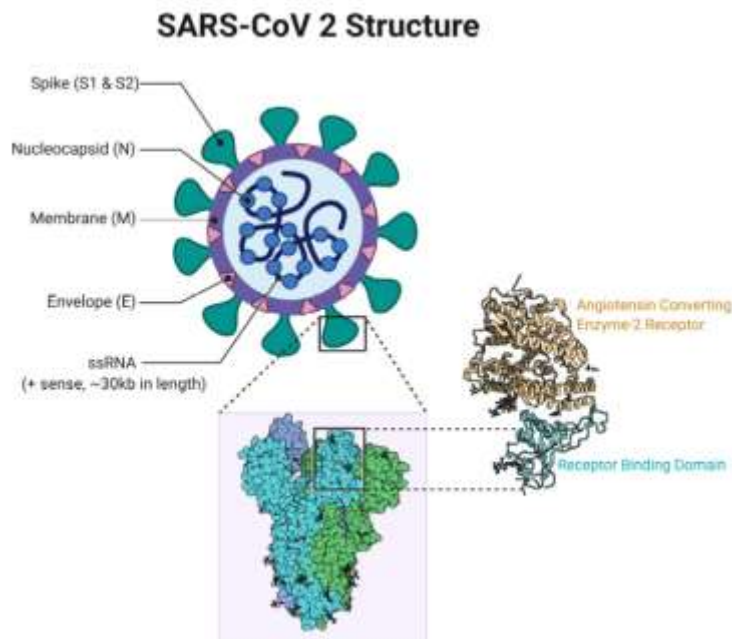
1 JOHDANTO

SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) on beetakoronavirus, joka on aiheuttanut pandemiaa vuoden 2019 joulukuusta alkaen. Ensimmäinen tapaus havaittiin Wuhanissa, Kiinassa 31. joulukuuta 2019. World Health Organization (WHO) julistikin pandemian alkaneeksi jo 11. maaliskuuta 2020 (WHO 2022). COVID-19 on aiheuttanut huolta maailmanlaajuisesti muokaten samalla yhteiskuntaa.

Koronavirukset ovat yksijuosteisia RNA-viruksia, jotka tunkeutuvat isäntäsoluun solukalvon kautta piikkiproteiinien avulla (Cavanagh 1995). Koronaviruksiin kuuluu alfa-, beeta-, gamma- ja deltakoronaviruksia (kuva 1), joista alfa- ja beetavirukset infektoivat ihmisiä (Woo ym. 2012). Beetakoronaviruksiin kuuluvat SARS-CoV-1 ja MERS-CoV aiheuttivat tautiaaltoja ennen COVID-19-pandemiaa. SARS-CoV-1-viruksen aiheuttama epidemia levisi Kiinasta neljään valtioon vuonna 2003 (WHO 2022). Epidemia kesti 6 kuukautta (CDC 2013). MERS-CoV puolestaan on lähtöisin Saudi-Arabiasta vuonna 2012, josta se levisi myöhemmin pandemiaksi (CDC 2019). SARS-CoV-2-virus aiheuttaa COVID-19-taudin, joka aiheuttaa Huangin ym. (2020) mukaan kuumetta, yskää, hengenhädistystä, väsymystä ja lihaskipuja. Lisäksi harvinaisempiin oireisiin kuuluvat muun muassa päänsärky, sekä liman kertyminen keuhkoihin (Huang ym. 2020).



KUVA 1 Koronavirusten fylogeneettinen puu (Singh ja Yi 2021)



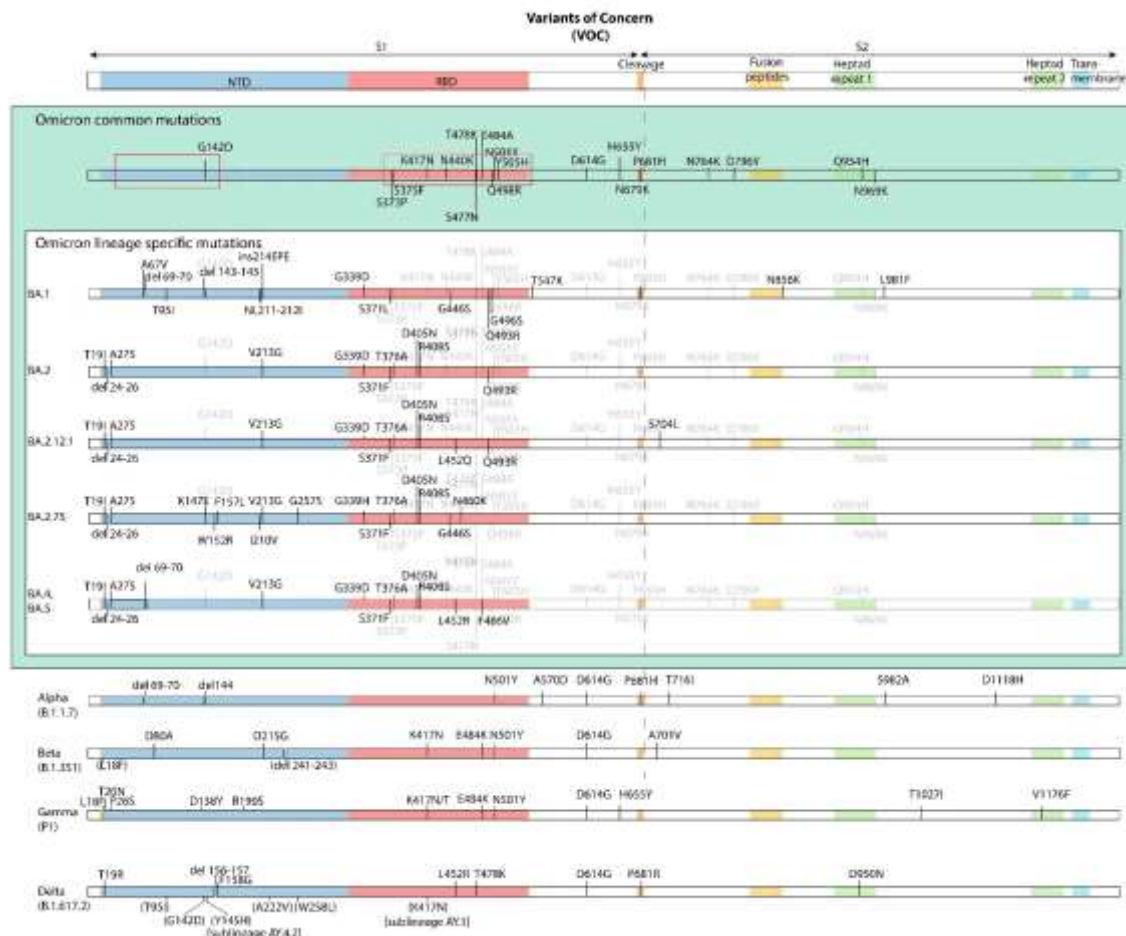
KUVA 2 SARS-CoV-2-viruksen piikkiproteiini ja ihmisen ACE-reseptori, jonka avulla se pääsee isäntäsoluun (Cascella ym. 2022).

Pandemian alusta lähtien SARS-CoV-2-virus (kuva 2) on muuntunut. SARS-CoV-2-viruksen genomi sisältää nukleokapsidia, kalvoa, vaippaa, piikkiproteiinia ja replikaasi ORF1ab koodaavat geenit (Alkhatib ym. 2021). Piikkiproteiini koostuu S1- ja S2-alaosista. S1-osassa on reseptoriin sitoutuva domeeni (RBD), N-terminaalinen domeeni (NTD) ja kaksi C-terminaalista domeenia (CTD1 ja CTD2). Suurin osa viruksen kelpoisuutta kasvattavista mutaatioista on S1-osassa. S2-osassa on kaksi toistojaksoa HR1 ja HR2, sentraalinen heliksi (*central helix*), fuusiopeptidi ja yhdistävä domeeni (*connector domain*), jonka avulla virus yhdistää itsensä piikkiproteiiniin. Piikkiproteiinsa avulla SARS-CoV-2 tarttuu isäntäsoluun. Isäntäsolun pinnalla se käyttää ACE2-reseptoria (Mistry ym. 2021). ACE2-reseptoriin tarttuminen saa aikaan muutoksia piikkiproteiinin konformaatioissa. Konformaation muutos auttaa virusta pääsemään soluun sisälle. Piikkiproteiinin S2-osan fuusiopeptidi liittyy itsensä kohdesolun solukalvoon, jonka jälkeen N- ja C-terminaalien heptad toistojaksot muodostavat helix-rakenteen ja S2-osa luhistuu. Tämä mahdollistaa solukalvojen fuusion ja viruksen pääsyn kohdesolun sisään (Simmons ym. 2013).

SARS-CoV-2-variantit voidaan jakaa alalinjoihin, joita ovat esimerkiksi alfa ja delta (Alkhatib ym. 2021). Jokaisella näistä alalinjoista on tunnusomaisia mutaatioita piikkiproteiinissa, jotka on esitetty kuvassa 3. Yksi ensimmäisenä havaituista varianteista oli alfa, joka löydettiin ensimmäisen kerran syyskuussa 2020 Iso-Britanniasta (O'Toole ym. 2021). Se sisältää piikkiproteiinissaan mutaatioita muun muassa deleetion $\Delta H69V70$, joka kasvattaa viruksen infektiivisyyttä (Meng ym. 2021). Beetavariantti havaittiin ensimmäisen kerran Etelä-Afrikassa joulukuussa 2020, mutta todisteita sen olemassaolosta on jo

vuoden 2020 ajalta (O'Toole ym. 2021). Beeta (B.1.351) on hyvin helposti tarttuva ja se kiertää vasta-ainepuolustusta muun muassa deleetion L242–L244 avulla (kuvassa 3 merkittynä L241–L243) (McCarthy ym. 2021). Gammavariantti (P.1) on myös yhdistetty vasta-ainepuolustuksen kiertämiseen, mutta sillä on vähemmän mutaatioita NTD-alueella kuin beetalla. Tämä johtaa hieman suppeampaan kykyyn kierrättää vasta-aineita kuin beeta (Alkhatib ym. 2021, Wang ym. 2021). Deltavariantilla (B.1.617.2) on korkeampi infektiivisyys sitä edeltäviin variantteihin verrattuna, koska se sisältää muun muassa mutaation L452R RBD-alueella (Liu ym. 2021). Deltavariantti sisältää myös mutaatioita, esimerkiksi T19R, G142D ja L452R, jotka lisäävät sen kykyä kierrättää monoklonaalisia vasta-aineita (Alkhatib ym. 2021, Planas ym. 2021). Tämän myötä delta nousi nopeasti valtavariantiksi Intiassa, jossa se havaittiin ensimmäisen kerran vuoden 2020 loppupuolella (Cherian ym. 2021).

Vuoden 2022 keväästä lähtien omikron ja sen alalinjat ovat olleet valtavarianttina Suomessa (THL 2022b). Omikron havaittiin ensimmäisen kerran marraskuussa 2021 Etelä-Afrikassa ja Botswanassa. Omikronin piikkiproteiinissa on monia mutaatioita, jotka ovat esiintyneet aiemmissä varianteissa kuten esimerkiksi alfan deleetio Δ H69V70 ja K417N, joka esiintyi beetassa ja gammassa (Alkhatib ym. 2021, Viana ym. 2022). Kuten kuvasta 3 näkyy, omikronista on kehittynyt monia alalinjoja. Näillä alalinjoilla on kasautunut kyky päästä isännän immuunijärjestelmä ohi monien mutaatioiden vuoksi. Erityisesti alalinjojen mutaatiot L452 ja F486 ovat tärkeitä tässä. Lisäksi on huomattu, että nämä alalinjat kykenevät ohittamaan aikaisemman omikron tartunnan synnyttämän vasta-ainepuolustuksen (Cao ym. 2022).



KUVA 3 SARS-CoV-2-piikkiproteiinin mutaatiot eri varianteissa (Swiss Institute of Bioinformatics, 2022). Tässä tutkimuksessa sekvensoidut alueet merkitty laatiikoilla kuvan yläosassa. NTD-alueen sekvensoinnin lyhenne N ja RBD-alueen sekvensoinnin lyhenne R.

SARS-CoV-2-viruksen leviämistä on seurattu jätevesiseurannalla. SARS-CoV-2-virus-RNA:n on todettu siirtyvän ulosteeseen ja eritteisiin virukseen sairastuneilla potilailla (Xiao ym. 2020). Tämän myötä SARS-CoV-2-infektioiden määrää populaatioissa voidaan seurata jätevesiä tutkimalla. Tutkimalla jätevesiä voidaan selvittää, millä alueilla infektoita on eniten ja missä pandemia on taantumassa (Daughton 2020). Suomessa jätevesiseurantaa toteuttaa Terveyden ja hyvinvoinnin laitos (THL). Jätevesinäytteitä otetaan Suomessa yhdeksältä paikkakunnalta, joista yksi on Jyväskylä. Jyväskylässä näytteet otetaan Nenäinniemen jätevedenpuhdistamolta (THL 2022a).

COVID-19-pandemian aikana monet instituutit ovat keränneet dataa muun muassa tautitapauksista, testausmääristä sekä tehohoidossa olevista potilaista. Näitä tietoja käyttäen voidaan rakentaa tilastollisia malleja, joilla voidaan arvioida esimerkiksi jätevedessä olevan virus-RNA:n määrää tai jätevesinäytteen SARS-CoV-2-positiivisuutta (Mantilla-Calderon ym. 2022). Mallinnusta voidaan tehdä lineaarisilla sekamalleilla. Lineaarilla sekamalleilla

on mallinnettu muun muassa SARS-CoV-2-viruksen leviämistä väestössä ja tutkittu miten rokotukset ja rajoitustoimenpiteet vaikuttavat leviämiseen (Moffat ym. 2022). Lineaarinen sekamalli voidaan määrittellä seuraavasti Moffat ym. (2022) mukaan

$$y_t = X_i b + U_i \times y_i + \varepsilon_i \quad i = 1, 2, \dots, j \quad (1)$$

missä X_i on design-matriisi kovarianteille, β on $p \times 1$ ulotteinen kiinteiden vaikutusten vektori, U_i on design-matriisi satunnaisvaikutuksille, y_i on q -ulotteinen satunnaisvaikutusten vektori ja ε_i on n -ulotteinen satunnaisvirheiden vektori, joka noudattaa $N \sim (0, \sigma^2 I)$. Mallin estimoinnissa voidaan käyttää joko REML- (restricted maximum likelihood) tai ML- (maximum likelihood) metodia (Moffat ym. 2022). Mallia rakentaessa pitää huomioida mahdolliset satunnaisvaikutukset sekä kiinteät kovariantit. Ottamalla satunnaisvaikutukset mukaan malliin vastemuuttujan heterogeenisyys voidaan selittää satunnaisuudella. Satunnaisvaikutusten oletetaan noudattavan normaalijakaumaa $N \sim (0, G)$, jossa G on positiivisesti määrytynyt $q \times q$ matriisi (Peng ja Lu 2012). Lineaarinen malli voidaan valita perustuen informaatiokriteereihin AIC (Akaike information criterion) ja BIC (Bayesian information criterion). Nämä informaatiokriteerit tarjoavat objektiivisen tavan valita dataan sopivimman mallin. AIC ja BIC eivät aina tarjoa sovelluksen kannalta parasta mallia, joten mallia valitessa kannattaa ottaa huomioon sovelluksen tarpeet esimerkiksi merkityksellisten selittäjien suhde vastemuuttajaan (Lindsey ja Jones 1998).

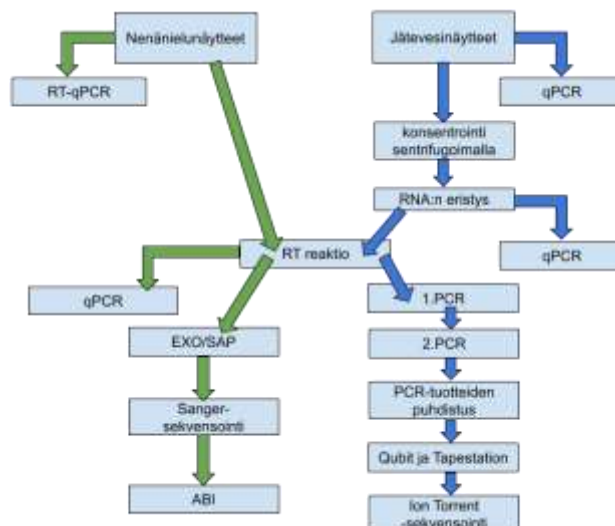
Tässä LuK-työssä tavoitteena oli tutkia COVID-19-PCR-testauksen nenänielunäytteistä ja jätevedestä löytyviä SARS-CoV-2-virusvariantteja sekä selvittää, mitkä tekijät vaikuttavat tilastollisesti jätevedessä olevaan SARS-CoV-2-viruksen määrään. Nenänielunäytteet saatiin BiopSense Oy:ltä ja ne sekvensoitiin Sanger-sekvensoinnilla. Näytteet olivat eri aikapisteistä vuoden 2021 helmikuusta lähtien vuoden 2022 huhtikuuhun asti. Jätevesinäytteet kerättiin kevään 2022 aikana Jyväskylässä, Nenäinniemen jätevedenpuhdistamolta. Nämä näytteet tutkittiin SARS-CoV-2-positiivisuuden varalta, ja näytteistä viimeinen toukokuun lopusta sekvensoitiin Ion Torrent -sekvensoinnilla. Terveystieteiden tutkimuskeskuksen (THL) saatiin COVID-19-pandemian ajalta kerättyä dataa muun muassa tehohoidossa olleista COVID-19-potilaista ja SARS-CoV-2-tapausmääristä. Nämä tiedot yhdistettiin tietoon jätevesinäytteiden qPCR (kvantitatiivinen PCR) C_q-arvoista ja jätevedenpuhdistamon virtaamatietoihin. Näiden tietojen avulla rakennettiin lineaarinen sekamalli, jonka vastemuuttuja oli jätevedessä olevan virus-RNA:n määrä. Lisäksi THL:n COVID-19-pandemia datan avulla rakennettiin itsenäinen lineaarinen sekamalli, jossa dataa oli pidemmältä aikaväliltä ja vastemuuttujana oli THL:n tekemät virus-RNA:n määrät jätevedessä.

Tutkimuksen hypoteesina oli, että nenänielunäytteet luovat ajallisen kuvan eri SARS-CoV-2-variantteista alkaen alfavariantista. Jätevesinäytteissä hypoteesina oli, että kaikki näytteet ovat SARS-CoV-2-positiivisia ja niiden valtavariantti olisi omikron. Tätä hypoteesia tukee THL:n tekemät jätevesien SARS-CoV-2-viruksen sekvensoinnit (THL 2022b).

Tilastollisessa mallinnuksessa hypoteesina oli, että SARS-CoV-2-viruksen määrään jätevedessä vaikuttaa tautitapausten määrä väestössä. Kun SARS-CoV-2-tapauksia on väestössä paljon, virus-RNA:ta eritetään enemmän jätevetteen ulosteiden mukana. THL:n tuottama COVID-19-data ei sisällä kaikkia tautitapauksia, koska Suomen koronastrategian mukaan kaikkia tapauksia ei testata virallisesti ja näin niitä ei tilastoida. Tässä työssä rakennettiin malli jäteveden sisältämälle virus-RNA:n määrälle. Jos tiedetään paljonko yksi sairastunut henkilö erittää keskimäärin virus-RNA:ta jätevetteen, niin virus-RNA:n määrää mallintavan mallin avulla voidaan laskea tapausten kokonaismäärä. Jos tapausten kokonaismäärä voitaisiin laskea, saataisiin selville myös tilastoimattomat tapaukset, joita THL:n tuottama data ei kerro. Tietämällä tapausten kokonaismäärän pandemian suuntaa on helpompi seurata ja tämä auttaa mahdollisten toimenpiteiden suunnittelemisessa.

2 AINEISTO JA MENETELMÄT

Tämän tutkielman tavoitteena oli tuottaa ajallinen kuva Jyväskylässä vallinneista SARS-CoV-2-virusvarianteista. Tässä käytettiin apuna Sanger-sekvensointia. Jätevesinäytteitä käyttäen haluttiin tutkia, paljonko jätevedessä on virus-RNA:ta ja mitä virusvariantteja jätevedestä löytyy. Jätevesinäytteitä tutkittiin hyödyntäen THL:n tilastoimaa dataa sekä omalla Ion Torrent -sekvensoinnilla. Lisäksi suoritettiin systemaattinen kirjallisuuskatsaus, jolla haluttiin arvioida käytettyjen menetelmien toimivuutta ja yleisyyttä. Kuva 4 sisältää yhteenvedon tämän tutkimuksen menetelmistä ja kuva 5 tuloksista, joihin tutkimuksessa pyrittiin.



KUVA 4 Tutkimuksessa käytetyt menetelmät



KUVA 5 Tutkimuksessa saadut tulokset.

Alkuperäinen tarkoitus oli, että nenänielu- ja jätevesinäytteiden sekvenssejä olisi vertailtu keskenään (kuva 5). Tätä ei kuitenkaan tehty, sillä sekvensoidut alueet olivat eri kohdista ja aikapisteistä. Jätevesinäytteiden sekvenssejä verrattiin sen sijaan Sanger-sekvenssien tutkimisessa käytettyihin mallisekvensseihin ja niiden avulla rakennettiin fylogeneettinen puu.

2.1 Positiivisten SARS-CoV-2-nenänielunäytteiden Sanger-sekvensointi

Ihmisnäytteitä sekvensoitiin Sanger-menetelmällä, tavoitteena saada näytteistä vertailukelpoista dataa eri SARS-CoV-2-varianteista. Tätä dataa on tarkoitus hyödyntää tutkimuksessa jätevedestä kerättyjen näytteiden vertailukohteena, jotta jätevesistä kerättyjä näytteitä voidaan yhdistää eri virusvariantteihin. Näistä ihmisnäytteistä osoitettiin SARS-CoV-2-nukleinihapot RT-PCR-reaktiolla. Tämä osoitus tehtiin cheMagic360-laitteella PerkinElmerin RT-qPCR-menetelmällä, jossa sekä käänteinen transkriptio että kvantitatiivinen PCR tapahtuivat samassa reaktiossa. Näytteiden RNA eristettiin käyttämällä PerkinElmerin kittiä Viral DNA/RNA 300 Kit H96 (cat # CMG-1033-S). SARS-CoV-2-Plus (N/ORF1ab/E/S 69_70del) RT-qPCR Reagent (PerkinElmer, USA) kittiä käytettiin viruksen RNA:n osoituksessa. Osoitus tehtiin kitin ohjeiden mukaisesti. Kitin toiminta perustuu TaqMan RT-PCR-tekniikkaan eli siinä tuotetaan neljä aluetta SARS-CoV-2:n genomista: nukleokapsidin N-geeniä, aluetta ORF1ab, E-geeniä sekä lisäksi alfavariantille tyypillistä deletiota S-geenissä. Kontrollina käytetään ihmisen RPP30-geeniä (fluoresenssi Cy5). Fluoresoivat TaqMan-koettimet (yksinaihaista DNA:ta) sitoutuvat vastinsekvensseihin fluoresoivilla leimoilla RNA:han. Kun PCR-reaktio purkaa DNA/RNA kaksoisnauhan, TaqMan leima vapautuu vaimentajastaan (Black Hole Quencher) ja fluoresenssi voidaan havaita tietyllä aallonpituudella.

Nenänielunäytteiden piikkiproteiinin Sanger-sekvensointia varten konsultoitiin THL:n Soile Blomqvistia. Hän suositteli käyttämään Artic-tietokannan sekvensointialukkeita (<https://community.artic.network/t/sars-cov-2-v4-1-update-for-omicron-variant/342>). Alukkeet tilattiin Metabion-yrityksestä (<https://www.metabion.com>) 0,020 µM skaalassa.

2.1.1 RT-reaktio

Protokollassa tehtiin ensin RT-reaktio ja qPCR neljällä alukeparilla, jonka jälkeen kukin syntynyt tuote sekvensoitiin neljällä alukkeella. RT-reaktio aloitettiin lisäämällä vettä alukkeisiin, sillä ne saapuvat kuivana yritykseltä. Vettä lisättiin valmistajan ohjeiden mukaan siten, että jokaisen alukkeen konsentraatio oli 100 µM. Tämän jälkeen alukkeet laimennettiin 10 µM ja niistä tehtiin kolme mixiä. Reagentit olivat Maxima Minus First Strand cDNA synthesis Kit with dsDNAase kitin reagentteja (Thermo Scientific, USA). RT-reaktiota varten valmistettiin kolme mixiä. Ensimmäisessä mixissä oli reagentit genomisen DNA:n poistamiseen (10x puskuria 4 µl, dsDNAase 4µl ja vettä 12 µl). Toisessa mixissä oli dNTP mix 4 µl, vettä 12 µl, 2 µl 78R aluketta ja 2 µl 84R aluketta (ensimmäinen ja viimeinen aluke sekvenssissä, sekä nukleotidit RT-reaktiota varten). Kolmannessa mixissä oli reaktiopuskuri 5x 8 µl, Maxima RT-entsyymi 9,5 µl ja vettä 8 µl (entsyymejä ja puskuria RT-reaktiota varten).

Mixien valmistuksen jälkeen laitettiin eristettyä RNA:ta 2,5 µl 96-kuoppalevyn kaivoihin, joissa on 2,5 µl Mix 1. Eristetyn RNA:n lisäämisen jälkeen näytteitä inkuboitiin +37 °C 30 minuuttia. Inkuboinnin jälkeen lisättiin 2,5 µl Mix 2 ja 2,5 µl Mix 3. Seuraavaksi näytteet sentrifugoitiin ja vortexoitiin. Sekoittamisen jälkeen inkuboitiin +50 °C lämpölevyllä 30 minuutin ajan. Tämän jälkeen inkuboitiin +85 °C 5 minuuttia.

2.1.2 qPCR-reaktio

RT-reaktion jälkeen jokaiselle kahdeksalle näytteelle suoritettiin qPCR-reaktio. Ensin valmistettiin qPCR mix. Siinä oli 2x Maxima MasterMix 100 µl, Forward-aluketta (10 µM) 4µl, Reverse-aluketta (10 µM) 4 ja vettä 76 µl. Kahdeksalla näytteellä ajettiin neljä forward aluketta ja neljä reverse aluketta, eli alukepareja oli neljä kappaletta. Jokaiseen näytteeseen laitettiin 0,5 µl forward ja reverse aluketta. PCR-ajo tehtiin seuraavan protokollan mukaan:

1. 95 °C 5 minuuttia
2. 95 °C 30 sekuntia,
3. 52 °C 1 minuutti
4. 72 °C 1 minuuttia
5. Levyn luku (qPCR)
6. 72 °C 5 minuuttia
7. 12 °C 5 minuuttia.

Vaiheet 2–5 toistettiin 35 kertaa.

2.1.3 EXO/SAP-puhdistus

PCR-tuotteet puhdistettiin EXO/SAP-puhdistuksella. Tässä menetelmässä ensin valmistettiin Master mix SAP/Exonuclease I: 4,2 µl Exonuclease (ref EN0581, Thermo Scientific, USA), 12,5 µl Shrimp alkaline phosphatase SAP (ref EF0651, Thermo Scientific, USA) ja 192 µl vettä. Tästä 24 PCR-putkeen jaettiin 8,33 µl ja lisättiin 25 µl PCR-tuotetta jokaiseen. Seuraavaksi suoritettiin nopea sentrifugointi. Nopean sentrifugoinnin jälkeen suoritettiin ajo PCR-laitteella seuraavan protokollan mukaisesti:

1. 37 °C 60 minuuttia
2. 85 °C 15 minuuttia,
3. 4 °C pito.

2.1.4 Toinen PCR

EXO/SAP-puhdistuksen jälkeen suoritettiin Sanger-sekvensointiin liittyvä PCR-ajo, jossa kutakin syntynyttä tuotetta sekvensoitiin neljällä sille alukeparille osuvalla alukkeella.

1. Alukepari 1: 71 L+76 R alt1, pituus 1854 emäsparia (myöhemmin bp)
Sekvensoinnissa käytetyt alukkeet: 71 L, 76R alt1, 72 L ja 73 R
2. Alukepari 2: 74 L+78 R, pituus 1584 bp
Sekvensoinnissa käytetyt alukkeet: 74 L, 78 R, 75 L ja 77 R
3. Alukepari 3: 77 L+81 R, pituus 1666 bp
Sekvensoinnissa käytetyt alukkeet: 77 L, 81 R, 79 L ja 79 R
4. Alukepari 4: 80 L+84 R, pituus 1594 bp
Sekvensoinnissa käytetyt alukkeet: 80 L, 84 R, 82 L ja 82 R

Sekvensointia varten valmistettiin mix, jossa yhteen sekvensointireaktioon tarvittiin 0,25 µl Big Dye terminator mix (BigDye Terminator v3.1 Applied Biosystems, ThermoFisher Scientific, USA), 1,9 µl 5x sekvensointipuskuria (Applied Biosystems BigDye terminator v3.1. cycle sequencing Kit, ThermoFisher Scientific, USA) ja 5,9 µl steriiliä vettä. Tämän jälkeen kaivoihin lisättiin 1,5 µl templaattia per kaivo, sekä 0,5 µl haluttua aluketta. Mix pipetoitiin 96-kuoppalevylle ja PCR-ajo suoritettiin seuraavan protokollan mukaan:

1. 96 °C 1 minuuttia
2. 96 °C 10 sekuntia
3. 52 °C 5 sekuntia
4. 60 °C 4 minuuttia

Vaiheet 2–4 toistettiin 24 kertaa.

2.1.5 PCR-tuotteiden puhdistus

Sekvensointituote puhdistettiin käyttämällä EDTA/Etanoli/Natrium-asetaatistaostusta. Tämän avulla PCR-tuotteesta puhdistetaan alukkeet,

nukleotidit ja entsyymit. Puhdistukseen tarvittiin 0,125 M EDTA, joten tehtiin 1:4 laimennos 0,5 M säilytysliuoksesta. 0,125 M EDTA:ta laitettiin mixiin 2000 µl ja 3M Na-asetaattia 2000 µl. Tätä liuosta laitettiin 3,33 µl ja 100 % etanolia 42 µl jokaiseen 96 näytteeseen. Näytteitä inkuboitiin 15 minuutin ajan huoneenlämmössä, jonka jälkeen sentrifugoitiin +6 °C 3000 rpm 30 minuutin ajan (Beckman Allegra Centrifuge, USA). Tämän jälkeen supernatantti poistettiin näytteistä laittamalla käsipaperia näytteiden alle ja kääntämällä 96-kuoppalevy ylösalaisin papereiden päälle sekä lisäksi sentrifugoimalla 900 rpm viiden sekunnin ajan. Seuraavaksi kuoppalevyn näytteisiin lisättiin 50 µl pakastekylmää 70 % etanolia ja sentrifugoitiin 3000 rpm 15 minuutin ajan. Supernatantti poistettiin jälleen sentrifugoimalla kuoppalevy ylösalaisin 900rpm viiden sekunnin ajan. Seuraavaksi näytteitä inkuboitiin 30 minuutin ajan huoneenlämmössä, jonka jälkeen näytteet resuspensoitiin 7 µl Hi-Di-formamidilla. Näytteet säilöttiin jääkaapissa ABI- sekvensointianalyysejä varten.

2.1.6 ABI-sekvensointianalyysi

Sekvensointianalyyseissä käytettävä ABI-sekvensointilaite on laserpohjainen ja se käyttää fluoresoivia leimoja analysoidakseen sekvensointireaktion tuotteita niiden kulkiessaan geelin läpi. Sekvensointiajon tietojen keräämisen jälkeen analyysiohjelma tunnistaa ja seuraa geelin näytekaistat ja normalisoi sekä integroi raakatiedot lopullisen sekvenssin kromatogrammiin. Geelitiedosto koostuu tietokoneella luodusta kuvasta sekvensointigeelin mukaan, jossa esiintyy fluoresoivia DNA-juovia. Tällä kuvalla voidaan tarkastella ja muokata seurantarivejä, jotka analyysi on luonut ja käyttänyt datapisteiden keräämiseen jokaisesta näytteestä. Yksittäiset näytetiedostot tallennettiin jokaisesta analysoidusta näytteestä ja niihin kuuluivat kromatogrammi, raakatiedot, sekä huomautukset, tiedot näytteestä ja sekvenssiajosta. Yleensä sekvensointireaktion tuotteet erottuvat helposti ja analyysiohjelmisto tulkitsee oikean nukleotidisekvenssin. Epäselviä emäksiä esiintyy yleensä lähellä sekvenssin loppua ja alkua. Näitä muokattiin tai poistettiin tarpeen mukaan ennen kuin tiedot vietiin lisävertailuihin tai linjauksiin.

2.1.7 Sanger-sekvenssien käsittely

ABI-raakadataa muokattiin BioEdit-ohjelmalla (BioEdit 7.2.5. 12/11/2013) poistamalla huonoa sekvenssiä näytteiden sekvenssien alkupäästä ja lopusta näytteen onnistumisen mukaan. Näytteistä moni oli epäonnistunut ja osa poistettiin sekvenssin huonon luotettavuuden vuoksi. Eri alukkeilla saadut sekvenssinpätkät asetettiin järjestykseen ja koottiin yhtenäisemmäksi sekvenssiksi. Sekvenssit löytyvät liitteestä 2.

Fylogeneettinen puu (kuva 6) luotiin Molecular Evolutionary Genetics Analysis -ohjelmalla (MEGA 11.0.11.). Sekvenssit käsiteltiin ensin BioEditin Clustal W Multiple Alignment -toiminnolla, jonka jälkeen se tuotiin MEGA-ohjelmaan. Käsittely ennen MEGA-ohjelmaan tuontia tapahtui linjaamalla näytesekvenssit ja varianttien mallisekvenssit (taulukko 1) keskenään.

Mallisekvenssit valittiin ViralZone-sivuston (<https://viralzone.expasy.org/9556>) perusteella ja ne ladattiin GenBank-sivulta alkaen 21500 bp. Linjauksen jälkeen mallisekvensseistä poistettiin ne alueet, joita ei löytynyt näytesekvensseistä. Lisäksi kaikki ne alueet, joita ei löytynyt kaikista sekvensseistä poistettiin. Etenkin näytteiden 4, 5 ja 6 takia poistettiin noin 500 bp pituinen alue, koska ne olivat erilaisia näytteisiin 1, 2 ja 3 sekä mallisekvensseihin verrattuna. Lisäksi näytteen 6 lyhyen pituuden takia, muutkin linjatut sekvenssit lyhenivät jopa noin tuhat emäsparia. Myös villityypin mallisekvenssin vuoksi oli otettava parikymmentä emäsparia pois, sillä se ei täsmännyt muihin mallisekvensseihin eikä näytesekvensseihin tällä kohtaa (kohdalla 23600 bp). Lopulta kaikkien tyhjien alueiden poistamisen jälkeen sekvenssit linjattiin uudelleen. Linjatut sekvenssit löytyvät liitteestä 3. Taulukon sekvensseistä kaikki muut, paitsi villityypin sekvenssi on valittu ViralZone-sivustolta. SARS-CoV-2-piikkiproteiini alkaa kohdasta 21 563 bp ja loppuu kohtaan 25 384 bp (Snap Gene, 2022). Tältä alueelta ladattiin sekvenssit GenBank-sivustolta.

TAULUKKO 1 Linjauksessa käytetyt mallisekvenssit. Villityyppejä ei löytynyt sivulta, haettu GenBankista: "Severe Acute Respiratory Syndrome Coronavirus Wild Type".

Variantti	Sekvenssi
Villityyppi	GenBank: MZ646043.1
Alfa	GenBank: MZ344997.1
Beeta	GenBank: MW598419.1
Delta	GenBank: MZ009823.1
Gamma	GenBank: MW642250.1
Omikron BA1	GenBank: OL672836.1
Omikron BA2	GenBank: OM371884.1
Omikron BA4	GenBank: ON373214.1
Omikron BA5	GenBank: ON249995.1

Nenänielunäytteet olivat seuraavista aikapisteistä. Mukana myös oletetut variantit. Variantit on arvioitu sen mukaan, milloin kyseinen variantti on ilmestynyt Suomeen.

- Näyte 1: Helmikuu 2021, villityyppi
- Näyte 2: Huhtikuu 2021, alfavariantti
- Näyte 3: Syyskuu 2021, deltavariantti
- Näyte 4: Helmikuu 2022, omikron BA1
- Näyte 5: Maaliskuu 2022, omikron BA2
- Näyte 6: Maaliskuu 2022, omikron BA2

Näytteiden variantit on valittu seuraavin perustein: Helmikuun näyte on todennäköisesti villityyppi, sillä SARS-CoV-2 saapui Suomeen THL:n mukaan helmikuussa (Onnettomuustutkintakeskus 2021). Alfavariantti kulkeutui Suomeen 2021 keväällä, deltavariantti 2021 kesällä ja omikronvariantti joulukuussa (Terveyskirjasto 2022). Näistä voidaan päätellä, että näyte 2 on

alfavariantti ja näyte 3 deltavariantti. Puolestaan näytteet 4, 5 ja 6 ovat omikronvarianttia. Omikron BA.1 yleistyi Suomessa tammikuussa, joten näyte 4 on BA.1-alavariantti. Näytteet 5 ja 6 ovat todennäköisesti omikronin alavarianttia BA.2, sillä se yleistyi valtavariantiksi Suomessa maaliskuussa (THL 2022c). Omikronin alavariantteja BA.4 ja BA.5 ei näytteissä todennäköisesti ollut, sillä ne ilmestyivät Suomeen vasta toukokuussa 2022 (THL 2022d).

2.2 Jätevesinäytteiden sekvensointi Ion Torrent -menetelmällä

Tutkimuksessa oli tavoitteena sekvensoida SARS-CoV-2-viruksen piikkiproteiinista noin 400 bp pituisia alueita, joiden perusteella voitaisiin tunnistaa kyseessä oleva virusvariantti. Jätevesinäytteet otettiin tulevasta jätevedestä Nenäinniemen jätevedenpuhdistamolla seuraavista aikapisteistä:

1A, 1B: 17.3.2022
2A, 2B: 13.4.2022
3A: 19.4.2022
4A: 28.4.2022
5A: 4.5.2022
6A, 6B, 6C: 10.5.2022
7A: 31.5.2022

Jätevesinäytteet (100 ml) otettiin putkiin, joihin oli lisätty 50 ml desinfiointi- ja säilytysainetta DNA/RNA Shield (Zymo Research, USA). Näytteet kuljetettiin kertänäytteenoton jälkeen nopeasti laitokselle, jossa ne pantiin pakasteeseen odottamaan käsittelyä. Näytteistä 1A, 1B ja 2B tehtiin vain SARS-CoV-2-PCR-osoitus. Näitä näytteitä ei otettu mukaan sekvensointiin. Kaikki muut näytteet otettiin mukaan sekvensointiin. Sekvensointi tehtiin Ion Torrent -sekvensoinnilla Jyväskylän yliopiston tiloissa ja sen suoritti Elina Virtanen.

Ennen sekvensointiin päätyneiden näytteiden käsittelyä, jouduttiin tekemään useita kokeita ja tutkimaan, mitkä alukkeet toimivat parhaiten sekvensoinnissa ja mitä reagentteja kannattaa käyttää. Aluksi käytettiin muun muassa RT-reaktiossa sellaista kittiä, jossa oli mukana random-alukkeita ja tämä luultavasti sotki PCR-tuloksia. Lisäksi huomattiin, että tilattu aluke 77r ei sopinut hyvin omikronin sekvenssiin, koska sekvensoinnissa syntyi vain lyhyitä primeridimeerejä.

2.2.1 Näytteiden konsentroidi

Ennen RT-reaktiota jätevesinäytteet konsentroidiin sentrifugoimalla. Tässä käytettiin Amicon Ultra-15 Centrifugal Filter unit -putkia (Merck KGaA, Saksa). Putkiin laitettiin ruiskusuodattamalla jätevesinäytteitä 2A, 3A, 4A, 5A, 6A, 6B, 6C ja 7A. Näytettä laitettiin 15 ml ja tämän jälkeen sentrifugoitiin 4000 g 30 minuuttia (sentrifugi Thermo Scientific SL 16R, USA). Konsentroidinnissa näytteet saatiin pienempään tilavuuteen. Konsentroidut (310–490 µl) näytteet siirrettiin 96-kuoppalevyille RNA:n eristystä varten. Näytteiden numerointi ja tilavuudet

on esitetty taulukossa 2. 96-kuoppalevyllä näytteet vietiin BiopSense Oy:lle RNA:n eristystä varten. Eristys tehtiin cheMagic Viral300 360 H96 prefilling 45 min ICP VD210201.che -ohjelmalla valmistajan ohjeiden mukaan. Tavallista suurempien näytetilavuuksien vuoksi guanidiini-tiosyanaatti-pitoista lyysis-puskuria annosteltiin näytteisiin vain 100 µl eikä 300 µl, kuten kitin ohje on. Automaatissa lyysis-puskurin määrä lisääntyi edelleen. Näytteistä eristettiin RNA ja tehtiin COVID-19-PCR-osoitus (mukana positiivinen kontrolli). COVID-19-PCR-osoitus tehtiin näytteille 1A-6A ilman konsentrointia.

TAULUKKO 2 Jätevesinäytteiden konsentroidut tilavuudet ja näytteiden numerointi.

Numero	Näyte	Tilavuus (µl)
1	2A	430
2	3A	440
3	4A	440
4	5A	480
5	6A	480
6	6B	490
7	6C	400
8	7A	310

2.2.2 RT-reaktio

Eristetyt RNA:t käännettiin cDNA:ksi alukkeen 77R avulla. RT-reaktiossa käytettiin kittiä Maxima H Minus First Strand cDNA synthesis Kit with dsDNAase (Thermo Scientific, USA). Reaktiota varten valmistettiin kolme mixiä. Mix 1 sisälsi kitin sisältämää 10 x puskuria 6 µl, dsDNAase-entsyymiä 6 µl ja vettä 18 µl. Mix 2 sisälsi 6 µl aluketta 77R, dNTP mix 6 µl ja vettä 18 µl. Mix 3 sisälsi reaktiopuskuria 5x 12 µl, RT-entsyymiä 6 µl ja vettä 12 µl. PCR-stripille lisättiin 2,5 µl Mix 1 ja 2,5 µl eristettyä RNA:ta. Tämän jälkeen inkuboitiin 30 minuuttia +37 °C. Inkuboinnin jälkeen lisättiin näytteisiin 2,5 µl Mix 2 ja 2,5 µl Mix 3. Näytteitä oli 8 ja näyte 9 oli negatiivinen kontrolli, johon lisättiin Mix 3 tilalta 2,5 µl vettä. Mix 2 ja 3 lisäyksen jälkeen inkuboitiin +50 °C 30 minuuttia ja +85 °C 5 minuuttia.

2.2.3 Ensimmäinen PCR

Ensimmäisellä PCR:llä monistettiin haluttua aluetta piikkiproteiinista. Reaktiossa käytettiin alukepareja NTD forward + NTD reverse ja RBD forward + 76 alt1 reverse. Valmistettiin kaksi Master Mixiä. Master Mix sisälsi 2x Maxima Mastermix 163 µl (Maxima SybrGreen/Fluorecein qPCR Master Mix 2x, Thermo Scientific, USA), 13 µl alukkeita (10 µM) ja 124 µl vettä. PCR-stripille laitettiin 23 µl Master Mixiä ja 2 µl RT-reaktion tuotetta. PCR-ajo tehtiin seuraavan protokollan mukaan:

1. 95 °C 10 minuuttia
2. 95 °C 30 sekuntia
3. 55 °C 30 sekuntia
4. 72 °C 1 minuuttia (kuoppalevyn luku)
5. 72 °C 5 minuuttia
6. 4 °C jatkuvasti

Vaiheet 2–4 toistettiin 40 kertaa.

2.2.4 Toinen PCR

Ensimmäisiin PCR tuotteisiin liitettiin barkoodit ja M13- ja P1-alukkeet. Käytetyt alukeparit olivat M13 NTD forward + P1 NTD reverse ja M13 RBD forward + P1 76alt reverse. Alukepareille tehtiin kaksi Master Mixiä. Yksi Master Mix sisälsi 100 µl Master Mix (sama kuin ensimmäisessä PCR:ssä), 8 µl molempia alukkeita (10 µM) ja 68 µl vettä. PCR-stripille laitettiin 24 µl Master Mixiä, 1 µl qPCR-tuotetta ja 1 µl barkoodia. Näytteet ja barkoodit on esitetty taulukossa 3. Ensimmäisellä rivillä kuoppalevyllä oli myös BiopSense Oy:llä alukeseoksen kanssa eristetyt virus-RNA näytteet. Nämä käännettiin cDNA:ksi 77R-alukkeella ja monistettiin NTD-alukkeilla. Tarkoituksena näistä oli tutkia, vaikuttiko alukkeiden lisääminen RNA-eristyksen qPCR-tulokseen. Kaikkia näytteitä ei laitettu ensimmäisestä PCR:stä toiseen PCR:ään, koska huomattiin, että osa näytteistä ei monistunut ollenkaan tai monistuminen tapahtui huonosti. Tämän perusteella valittiin toiseen PCR:ään näytteet 1, 2, 3, 4, 7 ja 8. PCR-ajoprotokolla oli seuraava:

1. 95 °C 10 minuuttia
2. 94 °C 30 sekuntia
3. 50 °C 30 sekuntia
4. 72 °C 1 minuuttia
5. 12 °C 5 minuuttia

Vaiheet 2–4 toistettiin 15 kertaa.

TAULUKKO 3 Toisessa PCR:ssä käytettyjen näytteiden näyttenumerot, alukkeet sekä barkoodit ylhäältä alas soluissa.

RT7	RT8	7	8	7	8		
NTD	NTD	NTD	NTD	RBD	RBD		
1	2	3	4	5	6		
1	2	3	4	1	2	3	4
NTD	NTD	NTD	NTD	RBD	RBD	RBD	RBD
9	10	11	12	13	14	15	16

2.2.5 PCR-tuotteiden puhdistus

Toisen PCR:n tuotteet puhdistettiin AMPure-puhdistuksella käyttäen Agencourt AMPure XP bufferia (Beckman, USA). PCR-tuotteista otettiin 20 µl Eppendorf-

putkeen ja lisättiin 36 µl AMPure bufferia. Tämän jälkeen sekoitettiin pipetillä 10 kertaa ja inkuboitiin huoneenlämmössä 5 minuuttia. Inkuboinnin jälkeen putket laitettiin magneettiseen putkitelineeseen, jossa inkuboitiin 2 minuuttia. Tämän jälkeen poistettiin syntynyt supernatantti ja lisättiin 200 µl etanolia. Etanolin lisäämisen jälkeen inkuboitiin 30 sekuntia huoneenlämmössä ja käännettiin telinettä kaksi kertaa tänä aikana. Inkuboinnin jälkeen poistettiin supernatantti ja lisättiin uudelleen 200 µl etanolia ja inkubointi toistettiin. Supernatantti poistettiin ja tämän jälkeen putkia kuivatettiin huoneenlämmössä 5 minuuttia. Kuivatuksen jälkeen putket poistettiin magneettisesta putkitelineestä ja lisättiin 40 µl vettä jokaiseen putkeen. Tämän jälkeen putket laitettiin takaisin magneettiseen putkitelineeseen ja inkuboitiin 2 minuuttia, jonka jälkeen puhdistettu tuote siirrettiin uusiin puhtaisiin putkiin. Tässä vaiheessa näytteet nimettiin 2R, 3R, 4R, 7R, 8R ja 8N käytetyn alukkeen ja näytteen numeron perusteella.

2.2.6 Qubit-mittaus

Puhdistettujen PCR-tuotteiden konsentraatio mitattiin Qubit-laitteella. Laitteeseen valmistettiin näytteet lisäämällä 195 µl Working Solution (Invitrogen, USA) ja 5 µl näytettä. Tämän jälkeen vortexoitiin ja inkuboitiin huoneenlämmössä 2 minuuttia. Mittaus suoritettiin ja sen perusteella laskettiin tuotteiden poolaukseen eri näytteiden tilavuudet.

2.2.7 Poolaus ja Tapestation

Qubit-mittauksen perusteella haluttiin 5 ng jokaista näytettä pooliin. Tämän jälkeen tehtiin Tapestation-mittaus ja huomattiin, että näytteet 8N ja 8R olivat ainoat näytteet, joissa toinen PCR oli onnistunut ja oli syntynyt noin 400 bp pituisia DNA-pätkiä. Muissa näytteissä oli syntynyt vain noin 100 bp pitkiä fragmentteja. Tämän tiedon myötä päätettiin tehdä uusi pooli, jossa oli 20 ng näytteitä 8R ja 8N ja 0,2 ng näytettä 7R.

2.2.8 Ion Torrent

Pooli sekvensointiin Ion Torrent -menetelmällä (Life Technologies). Sekvensoinnin suoritti laboratoriomestari Elina Virtanen Jyväskylän yliopistosta, ja valmistajan ohjeen mukaiset protokollan pääpiirteet kerrotaan seuraavassa lyhyesti: Sekvensoinnissa haluttiin sekvensoida 400 bp pituisia fragmentteja. Emulsio-PCR tehdään Ion OneTouch instrumentilla. Kirjaston määrä on ratkaiseva, koska kirjastofragmenttien määrän tulee olla huomattavasti pienempi kuin ISP-partikkelien määrä, jotta kutakin partikkelia kohti seoksessa on enintään yksi templaatti. Emulsio-PCR:n aikana biotinyloidut alukkeet monistavat kirjaston fragmentteja ja lopulta syntyy miljoonia ISP-partikkeleja sekä monoklonalisella PCR-tuotteella, että ilman tuotetta. Tämän jälkeen tehdään rikastus, jolloin poistetaan ne ISP-partikkelit, joissa ei ole templaattia. Tässä käytetään apuna streptavidiinilla päällystettyjä magneettipartikkeleita. Streptavidiinillä päällystetyt partikkelit sitoutuvat templaatin sisältäviin ISP-

partikkeleihin biotiinin avulla. Magneettinen ydin streptavidiini partikkeleissa mahdollistaa niiden keräämisen magneetin avulla ja tällöin voidaan pesemällä poistaa partikkelit, joissa ei ole templaattia. Rikastamisen viimeisessä vaiheessa poistetaan kaksijuosteiset DNA-fragmentit hyvin alhaisella MELT-liuoksella. Toinen juoste säilyy streptavidiini-partikkelissa kiinni ja toinen komplementaarinen juoste, joka on kiinni ISP:ssä, siirtyy supernatanttiin ja se voidaan käyttää sekvensointiin.

Varsinainen sekvensointi tehdään Ion Torrent PGM -sekvensointilaitteella. Sekvensointilaitteeseen täytetään tarvittavilla nukleotideilla (dATP, dGTP, dCTP ja dTTP). Näytteeseen lisätään 3 µl Ion PGM Hi-Q sekvensointipolymeraasia ja se yhdistetään IonA-sekvensointialukkeeseen kanssa 27 µl tilavuudessa. Näyte ladataan sekvensointisirulle ja sekvensointipartikkelit sentrifugoidaan sirun kuoppiin. Laitteen valmistelun (valmistajan ohjeen mukaan) jälkeen tehdään suunniteltu ajo ja ladataan siru laitteeseen.

Näytteiden sekvensointia valmistellessa huomattiin, että näytteet eivät monistuneet halutulla tavalla. Tämä johtui virheellisesti alukkeesta. Tämän myötä tilattiin uusi aluke ja sekvensoitiin vain näyte 7A toistaen kaikki edellä mainitut työvaiheet. Sekvensointidata käsiteltiin CLC-ohjelmalla (<https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/>). Sekvensseistä poistettiin kaikki sekvenssit, jotka olivat alle 180 bp ja yli 250 bp pitkiä. Sekvenssien poistamisen jälkeen jäljelle jäi 25738 sekvenssiä, jotka koottiin OTU:ksi (*operational taxonomic unit*). OTU:t luotiin 98 % samankaltaisuuden perusteella. OTU:ja syntyi 12 ja näistä luotiin fylogeneettinen puu (ML-metodi) MEGA-ohjelmalla (samoin kuin Sanger-sekvenssien käsittelyssä) mallisekvenssien kera. Linjatut sekvenssit analysoitiin myös Nextclade-ohjelmistossa (Aksamentov ym. 2021), jotta saadaan selville mutaatioiden aiheuttamat muutokset aminohapposekvenssiin.

2.3 Tilastollinen analyysi

Tilastolliset analyysit tehtiin RStudiolla (versio 2022.7.1.554 Spotted Wakerobin, RStudio Team). Lineaariset sekamallit rakennettiin paketin *nlme* avulla ja aineistoa kuvattiin *ggplot2*-paketilla. Aineiston perusteella rakennettiin SARS-CoV-2-viruksen määrää jätevedessä ennustava malli. Tätä mallia varten Cq-arvot muutettiin virusmääräksi. Cq-arvo tarkoittaa PCR-syklien määrää ennen kuin *threshold*-arvo ohitetaan (Engelmann ym. 2021). qPCR-osoituksessa käytetyn cheMagic-laitteen mukana saatiin valmistajan tieto virusten kappalemäärästä koko eluutiosta eri Cq-arvoilla (liite 5). Näihin arvoihin sovitettiin eksponentiaalinen malli ja saatiin yhtälö (yhtälö 2), jolla laskettiin jätevesinäytteissä oleva virusmäärä (kappaletta) koko eluutiosta. Tämän jälkeen voitiin laskea reaktioissa käytettyjen tilavuuksien avulla, paljonko viruksia on yhdessä millilitrassa ja kuutiosta (m³) jätevetä.

$$\text{virusta eluutiosta} = 3 \times 10^{13} \times e^{-0,693 \times (Cq - arvo)} \quad (2)$$

Virusten määrää per millilitra jätevettä käytettiin vastemuuttujana lineaarisessa sekamallissa. Selittäjinä mallissa käytettiin Kuopion yliopistollisen sairaalan erityisvastuualueella tehohoidossa olevien potilaiden määrää COVID-19-taudin takia tai kanssa jätevesinäytteenottamisen päivänä, Jyväskylässä todettujen COVID-19-tapausten määrää näytteenottamisen viikolla, mitattua fluoresenssia, jätevedenpuhdistamon tunti- ja vuorokausivirtaamaa. Virtaamatietoja saatiin J-S-Puhdistamolta ja muut selittäjät saatiin THL:n tuottamasta COVID-19-pandemia datasta (THL kuutiokäyttöliittymä 2022). Sovitetut mallit olivat lineaarisia sekamalleja, joiden sovituksessa käytettiin funktiota *lme*. Osassa malliehdokkaista otettiin huomioon satunnaismuuttujana fluoresenssi, koska se vaihtelee jokaisen näytteen Cq-arvossa. Fluoresenssi vaikuttaa suuresti näytteen Cq-arvoon ja edelleen myös virusmäärään. Estimointi tehtiin ML-metodilla, jotta saadaan vertailukelpoisia informaatiokriteereitä. Paras malli valittiin AIC- ja BIC-kriteerien avulla. Mallidiagnostiikkaa tehtiin funktioilla *qqnorm*- ja *r2beta*-funktioilla paketista *r2glmm*. R²-arvo laskettiin käyttäen Nakagawa ja Schielzeth (2013) menetelmää.

Mallien rakentamisen lisäksi tutkittiin, eroavatko tämän tutkimuksen tulokset virus-RNA:n määrästä jätevedessä THL:n tuottamista arvoista. Virtaamakorjaus RNA:n määrälle tehtiin mukaillen Tiwari ym. (2022) menetelmää. Virtaamakorjatun RNA:n yksikkö oli miljoona RNA-kopiota/1000 henkilöä/vuorokausi. Nenäinniemen jätevedenpuhdistamon alueella asiakkaita on 154600. Virtaamakorjauksen jälkeen THL:n tuottamasta jätevesiseurannan datasta otettiin mitatut virtaamakorjatut RNA-määrät samalta viikolta kuin tämän tutkimuksen jätevesinäytteet oli kerätty. Näytteille tehtiin Wilcoxonin testi funktiolla *wilcox.test*. Lisäksi laskettiin THL:n virtaamakorjauksen ja itse tehdyn virtaamakorjauksen välinen suhde. Näissä laskuissa käytettiin vain FAM-fluoresenssin mittauksia ja rinnakkaisista näytteistä käytettiin vain ensimmäistä mittausta.

Käyttäen THL:n tuottamaa dataa rakennettiin myös toinen lineaarinen sekamalli. Siinä vastemuuttujana oli virus-RNA:n määrä jätevesinäytteessä, joka oli otettu Nenäinniemen jätevedenpuhdistamolta. Virus-RNA:n määrän määrittämisen teki THL (yksikkö miljoona virus-RNA kopiota/1000 henkilöä/vuorokausi). Kiinteinä selittäjinä mallissa käytettiin näytteenottopäivänä tehohoidossa olleita potilaita COVID-19 kanssa tai takia Kuopion yliopistollisen sairaalan erityisvastuualueella, näytteenottoviikolla tehtyjä COVID-19-PCR-testejä Keski-Suomen sairaanhoitopiirin alueella, jätevedenpuhdistamon vuorokausivirtaamaa sekä Jyväskylässä todettuja COVID-19-tapauksia viikolla, jolloin jätevesinäyte oli otettu. Satunnaisefektinä mallissa on jätevesinäytteen ottamisen vuodenaika, joka määriteltiin näytteenottamisen ajankohdan perusteella. Lineaarisia sekamalleja sovitettiin erilaisilla kiinteiden mallien yhdistelmillä ja paras malli valittiin AIC-informaatiokriteerin avulla (sovitetut mallit liitteessä 1). THL:n data oli aikaväliltä 3.1.2022–7.1.2022.

2.4 Systemaattinen katsaus kirjallisuudesta

Systemaattiseen kirjallisuuskatsaukseen (myöhemmin SKK) valittiin englanninkielisiä tutkimuksia seuraavilla perusteilla: tutkimuksissa tutkittiin SARS-CoV-2-viruksen esiintymistä jätevesissä tai arvioitiin toisten tutkimuksia kyseisestä aiheesta, niissä esiintyi määritellyt hakusanat "coronavirus", "correlation" ja "wastewater", niistä oli tehty vertaisarvioita, sekä tutkimuksia, joissa näytteiden määrä oli vähintään yhdeksän kappaletta tai vaihtoehtoisesti näytteet oli otettu vähintään kuuden kuukauden aikaväliltä. Katsaukseen kelpuutettiin myös kirjallisuuskatsauksia, joissa aineistoja tutkittiin vähintään kymmenestä tutkimuksesta.

Edellä mainittujen kriteerien avulla katsaukseen sisällytettiin 40 tutkimusta ja ne edustivat erilaisia tutkimustapoja sekä -paikkoja. Nämä artikkelit toivat esille myös eritasoisia tutkimuksia ja tutkimusryhmiä mahdollisimman laajasti maailmalta. Näiden 40 artikkelin joukossa oli 28 tutkimusta, 10 kirjallisuuskatsausta, yksi protokolla-artikkeli, sekä yksi tutkimus, joka keskittyi enemmän tilastolliseen analyysiin ja parhaan epidemiamallin löytämiseen.

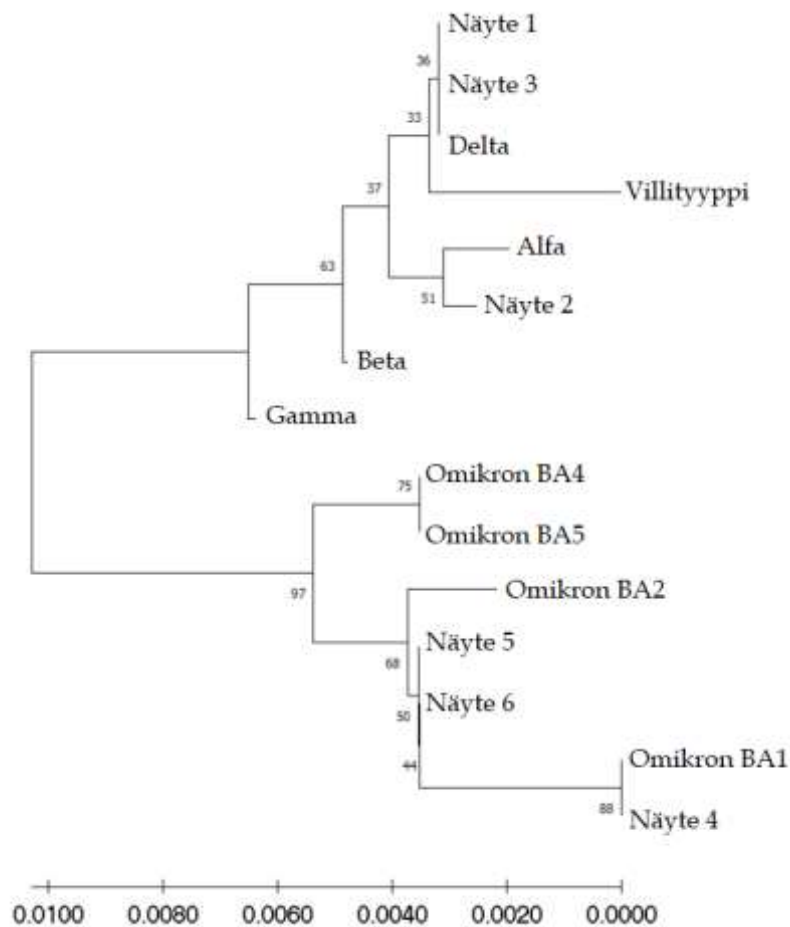
SKK:ssa keskityttiin eri tutkimusten näytteiden määrään, tutkimuksen toteutuspaikkaan ja tutkimusmenetelmiin. Nämä tulokset taulukoitiin ja visualisoitiin. Kirjallisuuskatsauksen tavoitteena oli vahvistaa omia tuloksia ja vertailla yleisesti käytettyjä menetelmiä tässä tutkimuksessa käytettyihin.

3 TULOKSET

Sanger-sekvensoinnin tuottamat sekvenssit löytyvät liitteestä 2. Näistä sekvensseistä selvitettiin, mitä virusvarianttia ne edustavat käyttämällä apuna fylogeneettistä puuta. Tuloksia liittyen jätevesinäytteisiin saatiin RNA-eristyksestä, RT-reaktiosta, ensimmäisestä ja toisesta PCR:stä sekä Qubit-mittauksesta. Lisäksi jokaisesta jätevesinäytteestä testattiin SARS-CoV-2-positiivisuus qPCR-testauksella. Tilastollinen analyysi antoi tuloksia muuttujien välisistä korrelaatioista, lineaaristen mallien estimaateista sekä mallien jäännöstarkasteluista. SKK:n tulosten avulla tämän tutkielman menetelmiä ja tuloksia voitiin verrata aiempiin tutkimuksiin ja täten selvittää tutkielman luotettavuutta ja toistettavuutta.

3.1 Sanger-sekvenssien tulokset

Nenänielunäytteiden sekvenssien geneettiset erot olivat hyvin pieniä. Sekvenssit ovat 571 bp pituisia. Fylogeneettisessä puussa käytetty linjaus löytyy liitteestä 3.



KUVA 6 Sanger-sekvensoinnin tuottamien sekvenssien fylogeneettinen puu. Puun sekvenssit ovat 571 emäksen pituisia. Kuvassa oleva asteikko kuvaa emästen eroavaisuutta tuhannesosina.

Fylogeneettisestä puusta (kuva 6) ja sekvenssien linjauksesta (liite 3) nähdään, että näytesekvenssit olivat hyvin samanlaisia niin mallisekvenssien kanssa kuin keskenäänkin. Esimerkiksi näytteet 1, 3 ja delta olivat linjauksessa täsmälleen samanlaisia keskenään. Muidenkin sekvenssien erot olivat hyvin pieniä: näytteen 4 ja omikron BA.1 sekvenssien eroavaisuus näytteisiin 5 ja 6, sekä omikron BA.2 mallisekvenssiin oli vain kaksi emäksen eroavaisuutta. Fylogeneettisestä puusta nähdään lisäksi, että omikronin haara erottuu selkeästi muiden SARS-CoV-2-varianttien haarasta. Yhteensä linjatulla alueella oli kuitenkin vain 13 emästä (liite 3), jotka määrittivät varianttien eroja, ja linjattu alue oli lyhyt (571 bp pituinen), mikä saattaa lisätä virheiden määrää tulkinnassa.

TAULUKKO 4 Nenänielunäytteiden variantit. Variantit on arvioitu vertailemalla fylogeneettista puuta ja varianttien saapumista Suomeen.

Näytteen numero	Variantti
1	Villityyppi
2	Alfa
3	Delta

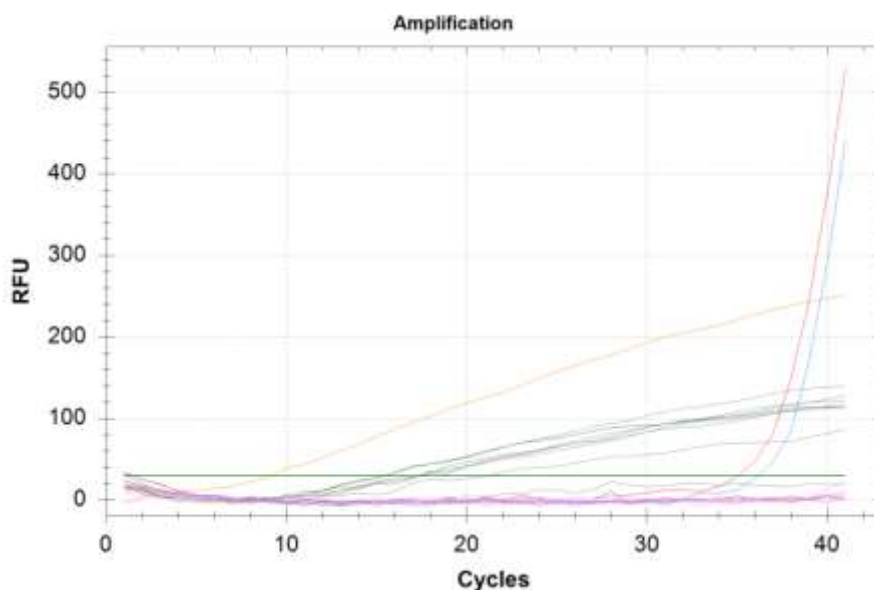
4	Omikron BA.1
5	Omikron BA.2
6	Omikron BA.2

3.2 Jätevesinäytteet

3.2.1 SARS-CoV-2-osoitus qPCR-reaktiolla

SARS-CoV-2-osoitus tehtiin cheMagic-laitteella. Tämä laite mittaa neljää eri fluoresenssia: Cy5, FAM, HEX ja ROX. Cy5 on kontrolli ihmisen RNA:lle (RPP30-geeni). FAM mittaa N- ja E-geeniä ja HEX mittaa ORF1ab aluetta. ROX mittaa deletiota Δ H69/V70. Kaikki näytteet tuottivat fluoresenssia Cy5, FAM ja HEX fluoresensseilla. Ainoa näyte, joka tuotti tuloksen ROX-kanavalla, oli näyte 7A. qPCR:n tulokset ovat liitteessä 1.

3.2.2 Ensimmäinen PCR



KUVA 7 Jätevesinäytteiden ensimmäisen PCR:n amplifikaatiokäyrät.

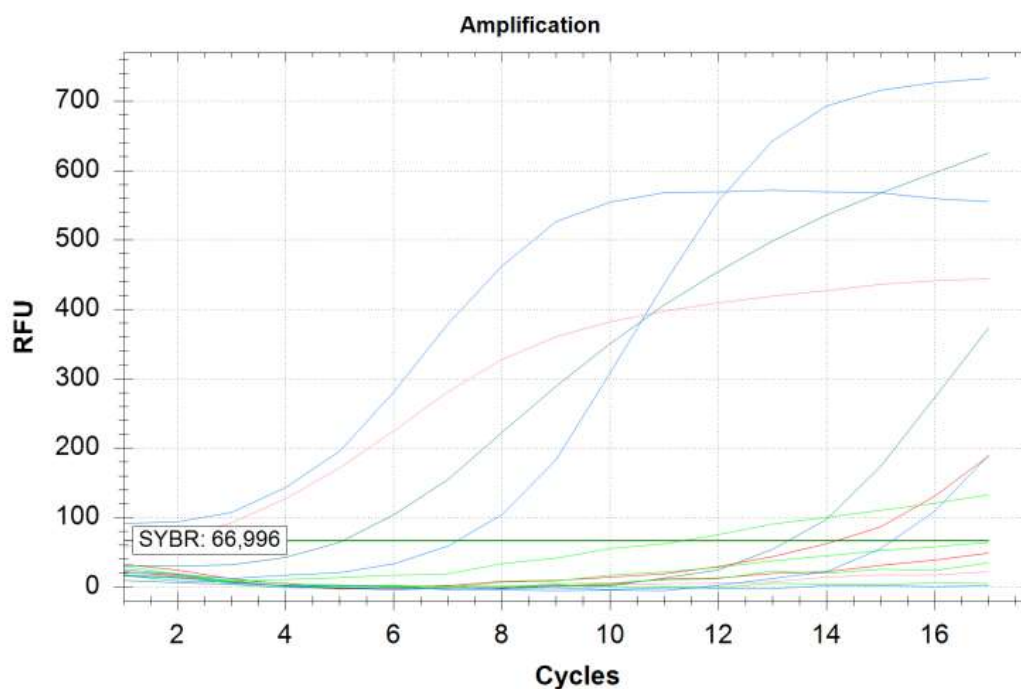
Ensimmäisen PCR:n tuottamat Cq-arvot ovat taulukossa 5 ja kuvassa 7. Tuloksissa näkyy, että RBD-alukkeet eivät tuottaneet kuin näytteessä 7A tuotetta. NTD-alukkeet tuottivat oikeanlaisen tuotteen vain näytteessä 7A, muissa näytteissä *threshold*-arvo ylitettiin liian aikaisin. Tämä näkyy kuvaajassa lineaarisina suorina, kun taas eksponentiaalinen käyrä kertoo oikeanlaisesta monistumisesta.

TAULUKKO 5 Jätevesinäytteiden ensimmäisen PCR:n tulokset numeerisessa muodossa.

Näyte	NTD Cq	RBD Cq
2A	17,32	N/A
3A	17,27	N/A
4A	15,41	N/A
5A	N/A	N/A
6A	17,27	N/A
6B	17,38	N/A
6C	21,14	N/A
7A	35,04	36,17

3.2.3 Toinen PCR

Jätevesinäytteiden toisen PCR:n tuloksissa (kuva 8, taulukko 6) huomataan, että NTD-alukkeet eivät tuota tuotetta ja vain osa RBD-alue näytteistä tuotti tulosta. NTD-alukkeista näyte 8 tuotti vähän tuotetta, joten se otetaan mukaan pooliin, samoin kuin kaikki RBD-alueen tuottamat tuotteet eli näytteet 2R, 3R, 4R, 7R ja 8R.



KUVA 8 Jätevesinäytteiden toisen PCR:n tuottamat amplifikaatiokäyrät.

TAULUKKO 6 Jätevesinäytteiden toisen PCR:n tulokset numeerisessa muodossa.

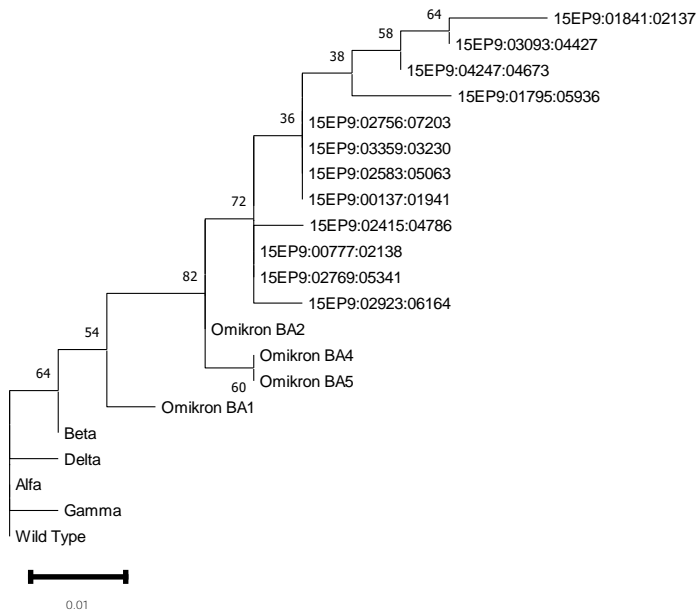
Näyte	NTD C _q	RBD C _q
RT 7	N/A	Ei tutkittu
RT 8	N/A	Ei tutkittu
7	N/A	13,29
8	1,71	5,06
1	N/A	N/A
2	N/A	15,22
3	N/A	7,17
4	N/A	N/A

3.2.4 Qubit-mittaus poolia varten ja sekvensoinnin tulokset

Toisen PCR:n jälkeen mitattiin Qubitilla näytteiden DNA-konsentraatiot poolauksista varten. Nämä tulokset ovat taulukossa 7. Tuotteiden poolauksen jälkeen pooli yritettiin sekvensoida. Tästä ei kuitenkaan saatu mitään tuloksia, koska sekvensointi ei onnistunut (barkoodit olivat tyhjiä). Myöhemmissä tutkimuksissa selvisi, että toisessa PCR:ssä käytetty aluke P1 oli molemmilla alueilla väärä, koska sekvenssi oli väärä. Tämän myötä PCR ei onnistunut ja barkoodeilta ei löytynyt mitään tuotetta. Virheen huomaamisen jälkeen tilattiin uudet alukkeet ja sekvensointi toistettiin RBD-alueelle näytteestä 7A. Fylogeneettinen puu (kuva 9) näyttää sekvensoidun jätevesinäytteen kuuluvan omikron BA.2 haaraan.

TAULUKKO 7 Jätevesinäytteiden toisen PCR:n jälkeisten tuotteiden Qubit-mittaukset. R ja N viittaavat alukkeiden sijoittumiseen SARS-CoV-2-piikkiproteiinissa.

Näyte	ng/μl
2R	0,35
3R	0,27
4R	0,39
7R	0,02
8R	5,64
8N	4,48



KUVA 9 NGS-sekvensoinnin tuottama fylogeneettinen puu Sanger-sekvensoinnin kanssa käytettyjen mallisekvenssien kanssa.

3.2.5 Tilastolliset mallit ja tunnusluvut

Aineistosta lasketut Pearsonin korrelaatiokertoimet ovat taulukossa 8. Korrelaatioista huomataan, että virusmäärä korreloi kaikkien muuttujien kanssa positiivisesti. Malleissa käytettyjen muuttujien keskiarvot ja keskihajonnat ovat esitetty taulukossa 9.

TAULUKKO 8 Tutkimuksessa käytettyjen muuttujien korrelaatiot.

Muuttuja	Tuntivirtaama (m ³ /h)	Vrk.virtaama (m ³ /d)	Tapausmäärät (näytteenotto viikolla)	Tehohoidossa (näytteenotto päivänä)	Virusmäärä (virusta/ml)	Testausmäärät (näytteenottamisen viikolla)
Tuntivirtaama (m ³ /h)	1					
Vrk.virtaama (m ³ /d)	0,8930	1				
Tapausmäärä (näytteenotto viikolla)	-0,0194	0,1184	1			
Tehohoidossa (näytteenotto päivänä)	-0,0616	0,5034	0,9351	1		
Virusmäärä (virusta/ml)	0,0954	0,1179	0,3867	0,4105	1	
Testausmäärä t (näytteenottamisen viikolla)	-0,0653	-0,0557	0,8263	0,9334	0,3808	1

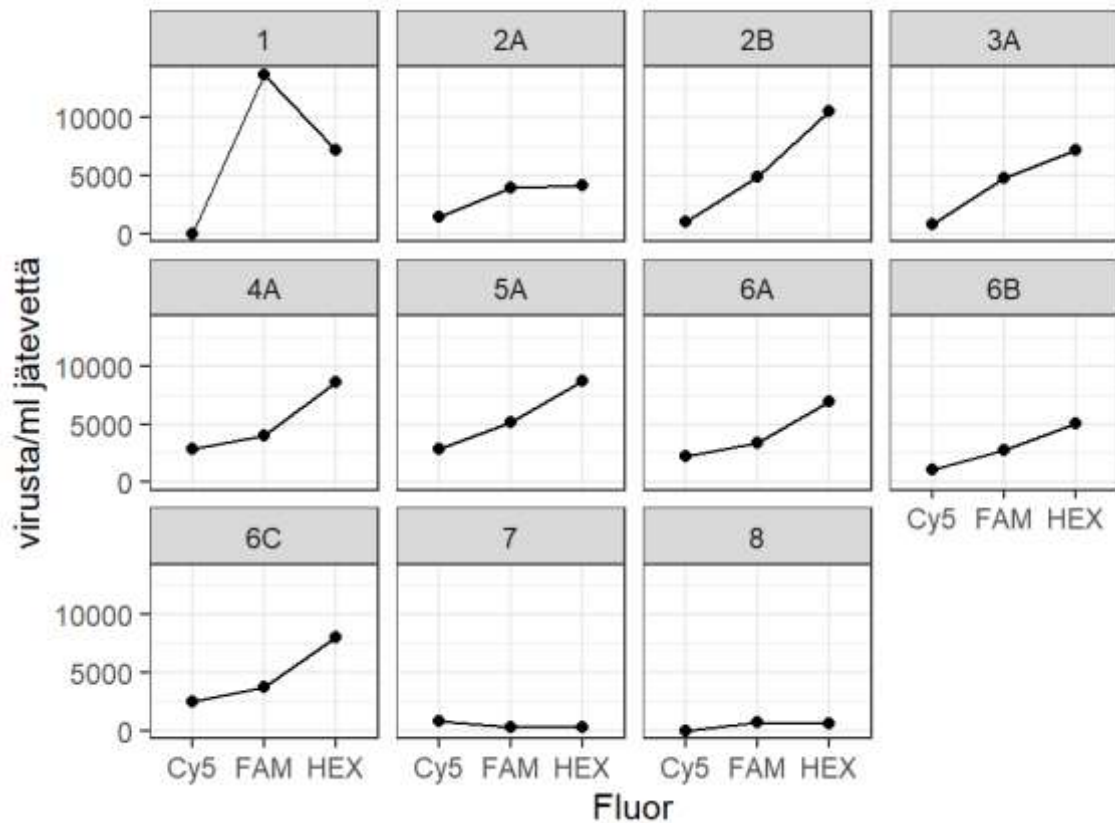
TAULUKKO 9 Jätevesidatassa käytettyjen parametrien keskiarvot ja keskihajonnat (σ).

Parametri	Keskiarvo	σ
Tuntivirtaama (m ³ /h)	2372,818	367,453
Vrk. virtaama (m ³ /d)	49321,82	10449,5
Tapausmäärä (kpl näytteenotto viikolla)	448,364	231,687
Tehohoidossa (kpl näytteenotto päivänä)	5,818	1,928
Virusmäärä (virus/ml)	3958,106	3374,675
Virtaamakorjattu RNA-määrä (miljoonaa virus-RNA kopiota/1000 hlöä/vrk)	1238316	1089036

Virusmäärää mallinnettaessa aineisto kuvattiin *ggplot*-funktioilla (kuva 10). Mallinnettaessa virusmäärää jätevedessä sovitettiin yhteensä 45 mallia, jotka on esitelty liitteessä 1. Näistä malleista valittiin AIC- ja BIC-arvojen perusteella paras. Lisäksi otettiin huomioon selittävien muuttujien määrä. Valittu malli oli lopulta malli vv15. Tämän mallin kaava on

$$y_{ij} = b_0 + b_1 \times Fluor + b_2 \times Tapausmäärät + b_3 \times Tehohoidossa olevat + u_{i0} + u_{i1} \times Fluor_{ij} + \varepsilon_{ij} \quad (3)$$

missä vaste y_{ij} kuvaa näytteen i virusmäärää, ε_{ij} on satunnaisvirheiden vektori ja u_{i0} on yksilöefekti. Tämän mallin regressiokertoimet ovat taulukossa 10.

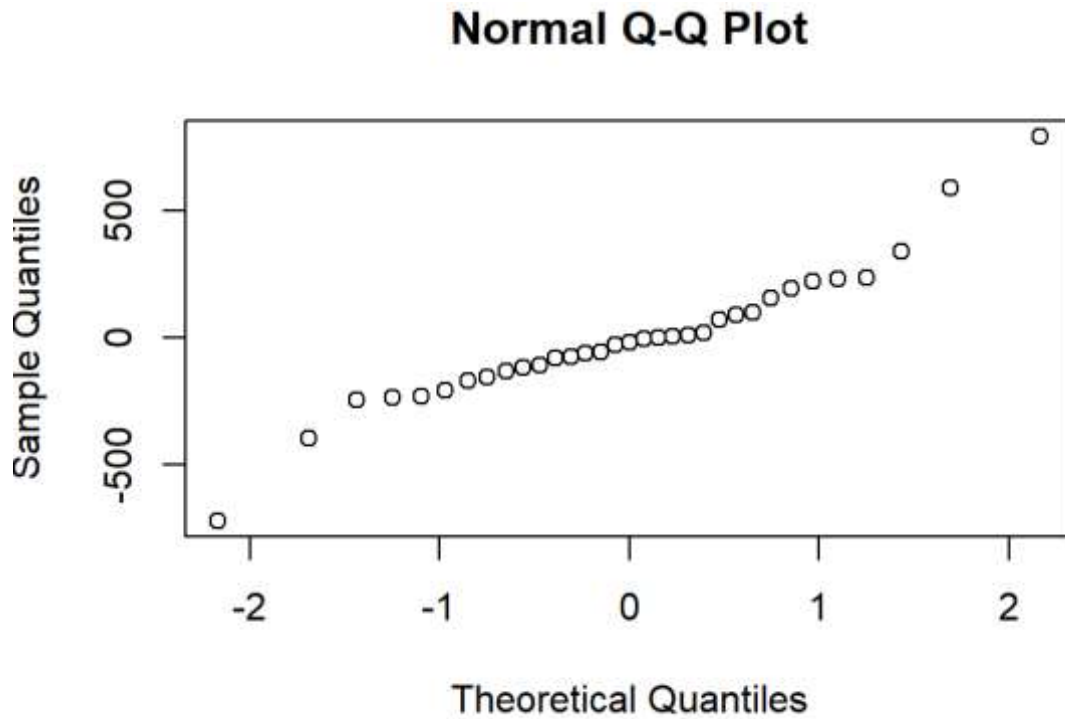


KUVA 10 Virus-RNA:ta per ml jätevedettä eri näytteissä fluoresenssi luokkaa vastaan. Näyte 7 on näyte 6C konsentroituna ja näyte 8 on näyte 7A.

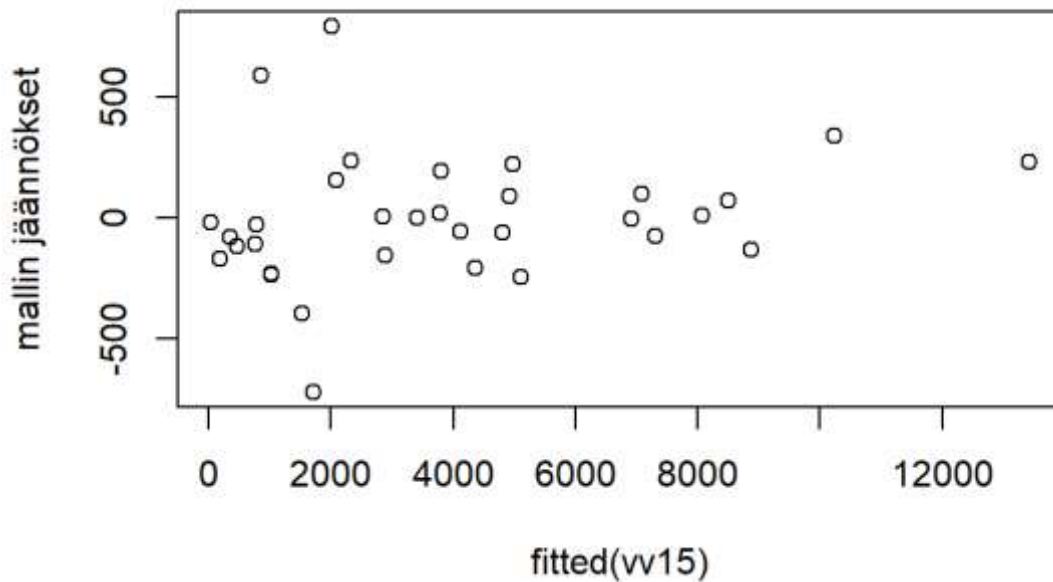
TAULUKKO 10 Virusta/ml jätevedessä mallintavan mallin kertoimet ja niiden *p*-arvot.

Muuttuja	Kerroin	<i>p</i> -arvo
<i>Intercept</i>	-768,85	0,3638
FluorFAM	2902,89	0,02036
FluorHEX	47074,86	<0,0001
Tapausmäärät (kpl)	-4,18	0,0372
Tehohoidossa olevat (kpl)	698,87	0,1104

Virusmäärää kuvaavan mallin perusteella virusmäärä per millilitra jätevedettä on -768,85, kun fluoresenssi on Cy5, tehohoidossa ei ole potilaita ja SARS-CoV-2-tapauksia on 0. Tämä tulos ei ole tilastollisesti merkitsevä. Kun fluoresenssi muuttuu FAM:ksi ja muut taustamuuttujat pysyvät samoina virusmäärä kasvaa 2902,89. Fluoresenssin muuttuessa HEX:ksi ja taustamuuttujien pysyessä samoina, virusmäärä kasvaa 4704,86. Fluoresenssien kertoimista molemmat ovat tilastollisesti merkitseviä. Tapausmäärän kasvaessa yhdellä ja muiden taustamuuttujien pysyessä samana virusmäärä laskee 4,18. Tämä tulos on tilastollisesti merkitsevä. Kun tehohoitoon tulee yksi potilas lisää, niin virusmäärä lisääntyy 698,87. Mallin vv15 normaaliutta on tarkasteltu kuvassa 11 ja vv5 mallin jäännösten ja sovitettujen arvojen yhteisjakauma näytteittäin on kuvassa 12. Mallille laskettu R^2 -arvo oli 0,40 (95 % -CI [0,21-0,64]).



KUVA 11 vv15 mallin normaaliuden tutkiminen Q-Q-plotin avulla.



KUVA 12 vv15 mallin jäännösten ja sovitettujen arvojen yhteisjakauma.

Virtaamakorjatuille RNA-määrille tehdyn Wilcoxonin testin tulos p -arvoksi oli 0,0039. Tämän perusteella siis tämän tutkimuksen tuloksilla ja THL:n tekemillä määrittäyksillä on tilastollisesti merkitsevä ero ($p < 0,05$). Laskettaessa THL:n

virtaamakorjauksen ja itse tehdyn virtaamakorjauksen suhdetta, tämän tutkimuksen tulokset olivat 0,2–49-kertaisia (keskiarvo 8,2) THL:n tuloksiin verrattuna.

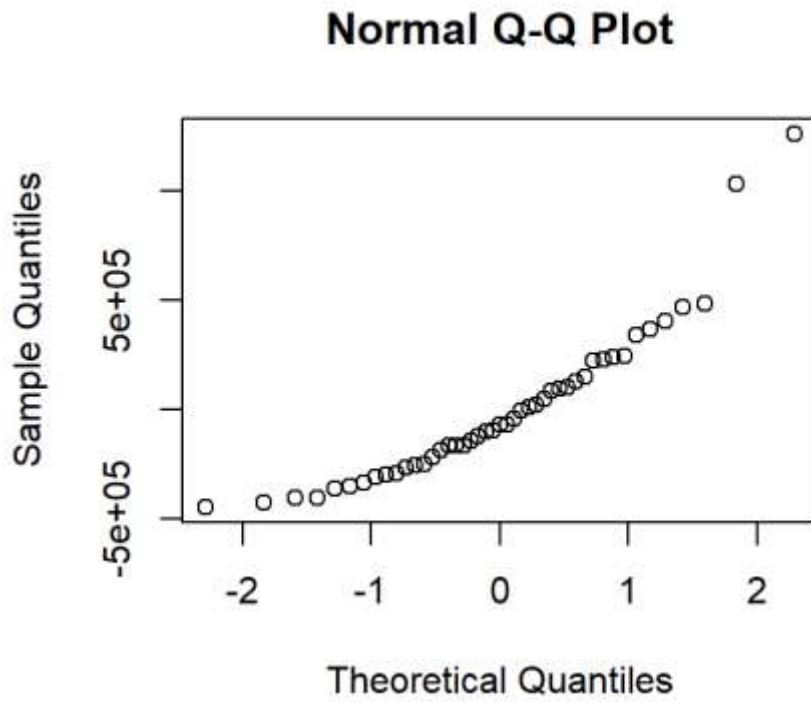
Mallinnettaessa THL:n tuottamaa dataa paras malli mallintamaan virus-RNA:n määrää jätevedessä oli

$$y_{ij} = b_0 + b_1 \times Fluor + b_2 \times Tapausmäärät + b_3 \times COVID - 19 - testit + u_i + \varepsilon_{ij} \quad (4)$$

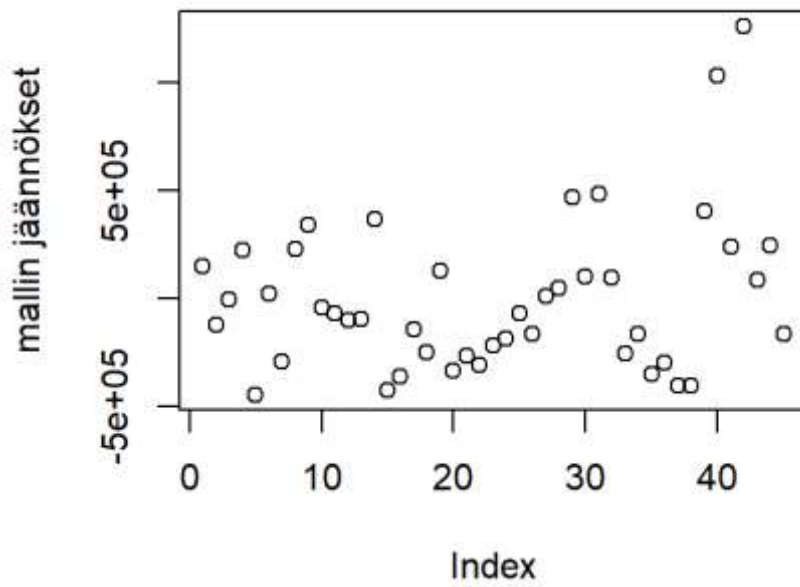
missä vaste y_{ij} kuvaa näytteen virus-RNA:n määrää, ε_{ij} on satunnaisvirheiden vektori ja u_i on vuodenajan luoma satunnaisefekti. Kaikki sovitetut mallit ovat liitteessä 1. Taulukossa 11 on esitetty valitun mallin regressiokertoimet. Kun COVID-19-tapauksia on 0 ja COVID-19-testejä on 0, niin virus-RNA:ta on jätevedessä 580151,4 (miljoonaa RNA-kopiota/1000 henkilöä/vrk). COVID-19 tapauksien lisääntyessä yhdellä, virusmäärä lisääntyy 703,4. Jos COVID-19-testejä tehdään 1 ja taustamuuttujat pysyvät samoina, niin virusmäärä alenee 173. *Intercept* regressiokerroin on tilastollisesti merkitsevä ($p < 0,0001$) ja muut regressiokertoimet eivät ole. Kuvissa 13 ja 14 on esitetty sovitetun mallin Q-Q-plot sekä standardoitujen jäännösten ja sovitettujen arvojen yhteisjakaumat. Sovitetun mallin R^2 -arvo oli 0,065 (95 % CI [0,005–0,276]).

TAULUKKO 11 THL:n datan perusteella tehdyn lineaarisen sekamallin regressiokertoimet ja niiden p -arvot.

Muuttuja	Kerroin	p -arvo
<i>Intercept</i>	580151,4	<0,0001
Tapausmäärät (kpl)	703,4	0,1090
Testausmäärät (kpl)	-173,0	0,1038



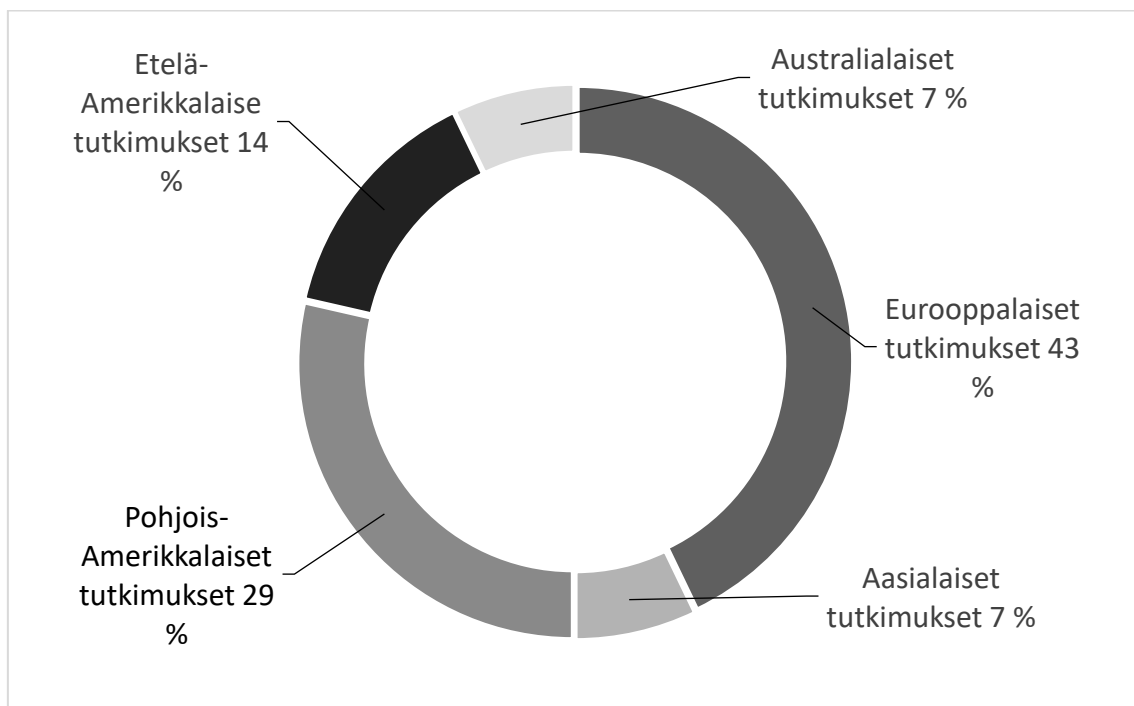
KUVA 13 Ainoastaan THL:n dataa mallintavan mallin Q-Q-plot.



KUVA 14 Standardoidut jäännökset sovitettuja arvoja vastaan THL:n dataa mallintavassa sekamallissa.

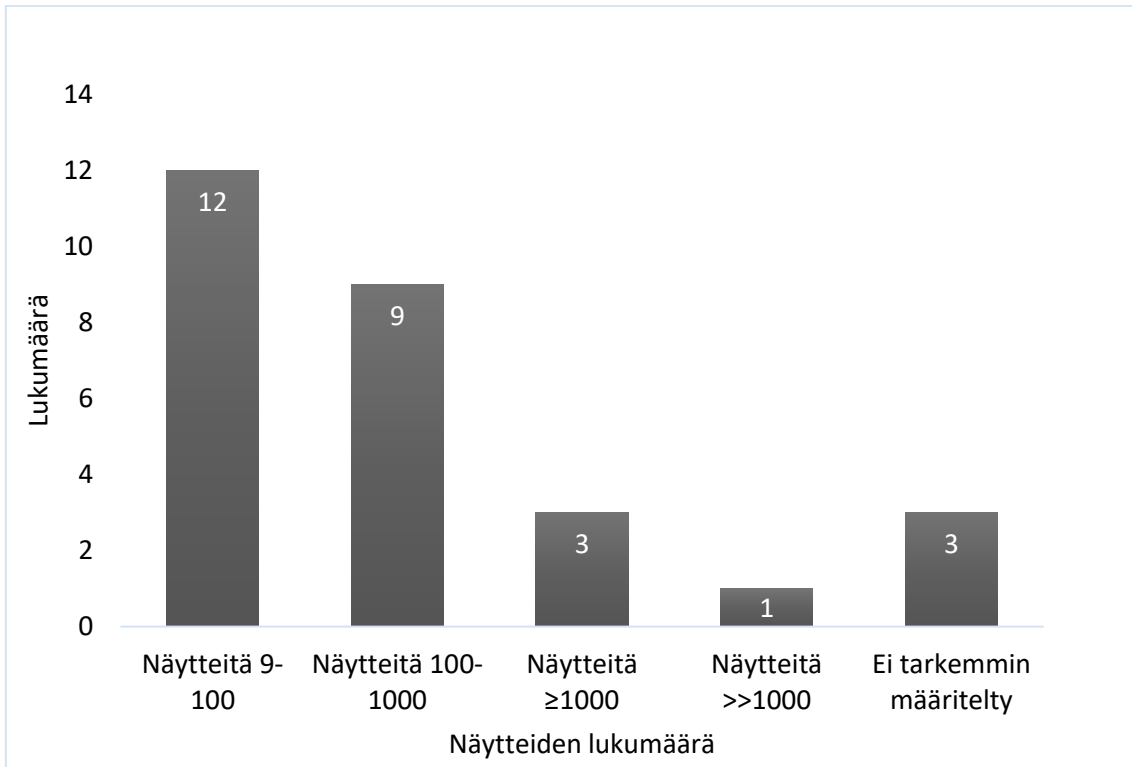
3.3 Systemaattisen kirjallisuuskatsauksen tulokset

Systemaattisen kirjallisuuskatsauksen artikkeleita (myöhemmin numeroituina, liite 6) oli kirjoitettu eri puolilla maailmaa, pääosin artikkelit olivat kuitenkin Pohjois-Amerikasta tai Euroopasta. Kuvassa 15 on esitelty tarkemmin tutkimuspaikat maanosittain. SKK:n kahdessa artikkelissa neljästäkymmenestä ei todettu korrelaatiota jätevesistä löytyvien SARS-CoV-2-konsentraatioiden ja jonkin COVID-19-pandemiaan liittyvän muuttujan välillä. Lisäksi yhden artikkelin korrelaatio oli epätäydellinen ja toinen artikkeli ei tutkinut korrelaatiota. Kuitenkin katsauksen artikkeleista 92,5 % totesi jonkinlaista korrelaatiota muuttujien välillä.

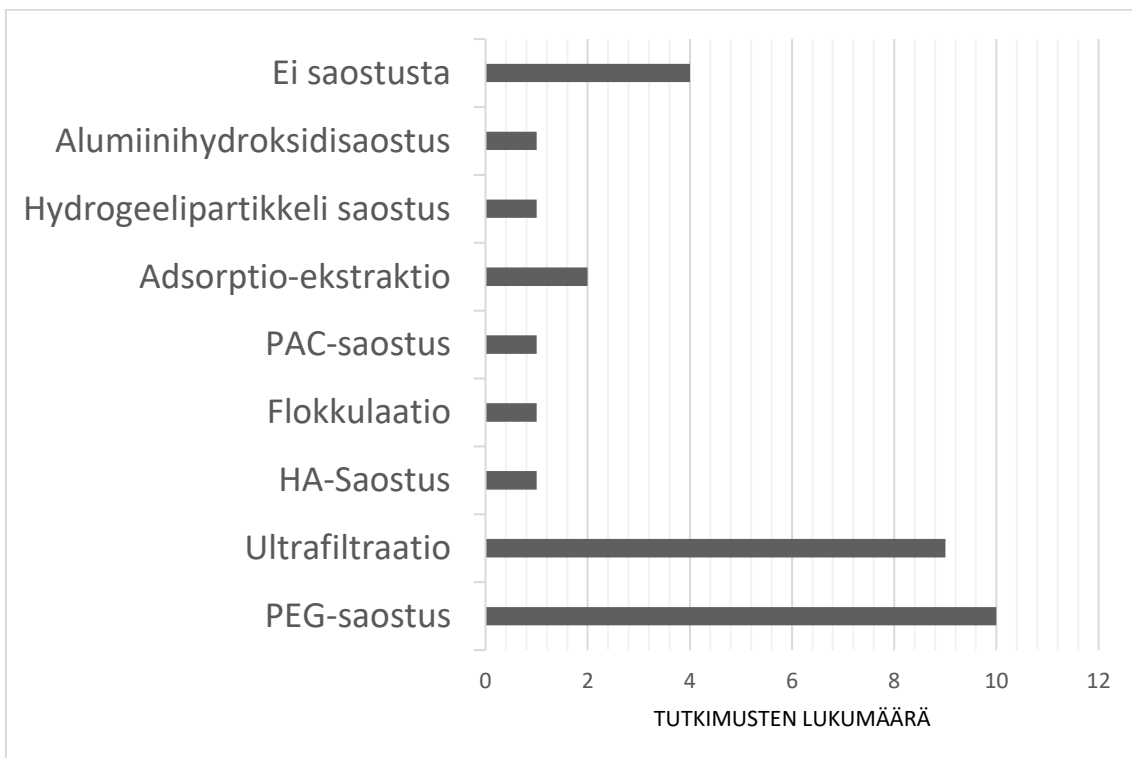


KUVA 15 SKK:n tutkimusten toteutuspaikka maanosittain

SKK:n tutkimuksissa näytteiden määrä vaihteli huomattavasti. Kuvasta 16 on nähtävissä, että suurimmassa osassa tutkimuksia näytteiden määrät sijoittuivat yhdeksän ja tuhannen välille. Saostusmenetelmistä yleisimmät olivat polyetyleeniglykolisaostus (PEG) ja ultrafiltraatio, jota käytettiin tässä LuK-tutkimuksessa. Vähiten tutkimuksissa käytettyjen saostusmenetelmien joukossa olivat esimerkiksi polyammoniumkloridisaostus (PAC) ja hydroksiapatiittisaostus (HA). Osa tutkimuksista ei saostanut jätevesinäytteitään tai saostusmenetelmiä ei ollut raportoitu (katso kuva 17).



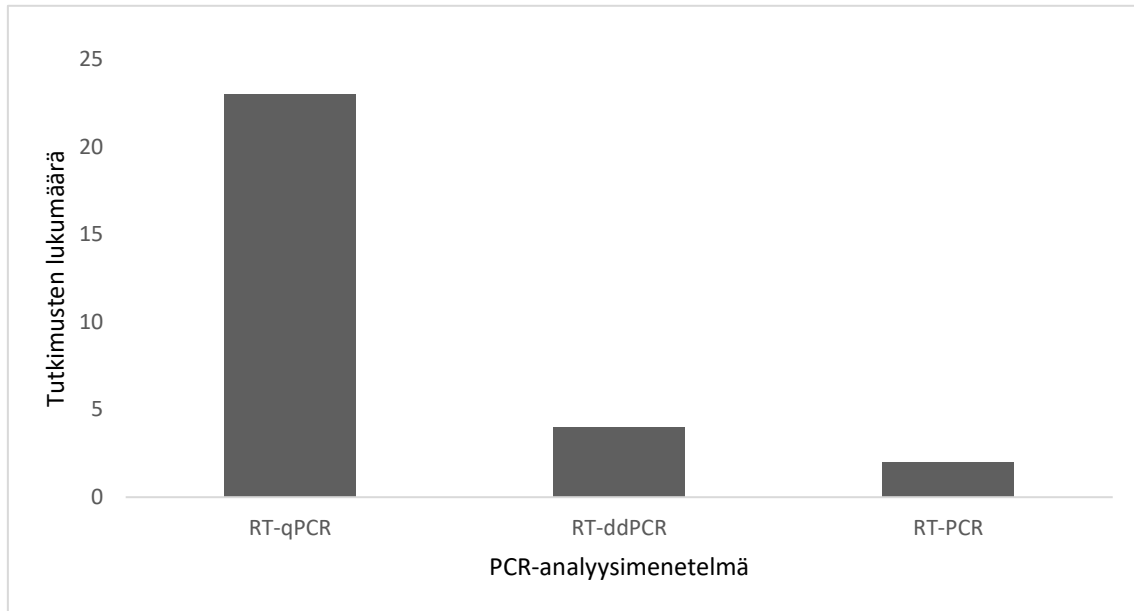
KUVA 16 SKK:n tutkimuksissa käytettyjen näytteiden lukumäärä



KUVA 17 SKK:n tutkimuksissa käytetyt saostusmenetelmät

PCR-analyysimenetelmiä käytettiin monipuolisesti eri tutkimuksissa, mutta kuitenkin huomattava osa tutkimuksista tukeutui RT-qPCR-menetelmään. Tässä

tutkimuksessa käytettiin myös RT-qPCR-menetelmää, mutta RT-ddPCR (digitaalinen PCR, jossa yhden fluoresenssimittauksen sijaan niitä on monia ja ei paikassa näytettä) oli myös vaihtoehtona. Kuvasta 18 on nähtävissä, kuinka epätavallinen RT-ddPCR käytettynä menetelmänä oli.



KUVA 18 Tutkimuksissa käytetyt PCR-analyysimenetelmät

TULOSTEN TARKASTELU

Tämän tutkimuksen hypoteesien mukaan Sanger-sekvensointi tuottaa ajallisen kuvan Jyväskylässä vallinneista varianteista eri aikoina. Saatujen tulosten perusteella voidaan päätellä, että ajallinen kuva varianteista vastaa pääosin THL:n tekemiä määrittelyjä. Jätevesinäytteistä sekvensoitiin vain yksi, toisin kuin alkuperäinen tarkoitus oli. Sekvensoidun näytteen qPCR-tuloksen ja fylogeneettisen puun perusteella näytteen varianttia ei voida varmasti sanoa. Systemaattinen kirjallisuuskatsaus antoi tukea tämän tuloksen menetelmille ja tilastolliselle tarkastelulle. Tämän tutkielman tuottamissa regressiomalleissa huomattavaa on alhaiset R^2 -arvot, jotka viittaavat regressiomallin vaikeaan estimointiin COVID-19-datan tapauksessa.

4.1 Sanger-sekvensoinnin tuloksista

Fylogeneettisen puun (kuva 6) perusteella voidaan päätellä, että näytteet 1 ja 3 ovat todennäköisimmin joko villityyppiä tai deltavarianttia, näyte 2 on alfavariantti, näyte 4 omikronin alavariantti BA.1 ja näytteet 5 ja 6 puolestaan omikronin alavarianttia BA.2. Näytesekvensseistä ei fylogeneettisen puun

mukaan löytynyt omikron BA.4 ja BA.5 varianttia. Näitä tuloksia vertaamalla näytteiden aikapisteisiin voidaan huomata, että fylogeneettisen puun tulokset olivat oikeaa suuntaa antavia. Kun näitä tietoja yhdisteltiin, saatiin taulukon 4 mukaiset tulokset. Fylogeneettisen puun kanssa yhteneviä tuloksia olivat näytteille 2, 4, 5 ja 6 saadut variantit. Näytteiden 1 ja 3 kohdalla fylogeneettisen puun mukaan oli näytteelle kaksi vaihtoehtoa: delta tai villityyppi. Näytteiden aikapisteiden mukaan, näyte 1 osoittautui villityypiksi ja näyte 3 puolestaan deltavariantiksi. Tämä päätelmä tehtiin siksi, koska näyte 1 oli otettu jo helmikuussa 2021, eikä se tämän perusteella voinut olla deltavariantti, sillä deltavariantti saapui Suomeen vasta vuoden 2021 kesällä (Terveyskirjasto 2022).

Linjauksesta (liite 3) voidaan lisäksi todeta, että jos tarvittaisiin Sanger-sekvensoitava lyhyt alue näytteen variantin selvittämiseen, voisi alue 23010–23080 bp toimia tässä hyvin. Tältä alueelta löytyi kahdeksan yksittäistä emäsparin eroavaisuutta (yhteensä 571 pituisella alueella oli 12, joista 8 jo tällä 70 bp alueella), ja tässä tutkielmassa suoritettua sekvensointia sekä fylogeneettisen puun sovittamisen perusteella variantit saatiin selvitettyä melko suurella tarkkuudella. Lisäksi sekvensoinnin epäonnistuessa suurimmissa osissa alueista, kannattaa jatkossa sekvensointi kohdistaa tälle 70 emäsparin alueelle. Fylogeneettisen puun (kuva 6) sovittaminen todisti sen, että näinkin pienellä alueella saadaan suuntaa siihen, mikä variantti näytesekvenssissä on kyseessä.

4.2 Jätevesinäytteiden SARS-CoV-2-osoitus ja PCR-tulokset

Jätevesinäytteiden SARS-CoV-2-osoituksen myötä saatiin tietää, että kaikki näytteet olivat SARS-CoV-2-positiivisia. Tärkeä huomio on, että ainoastaan näyte 7A tuottaa ROX-fluoresenssia. ROX fluoresoi, jos näytteessä on deletio Δ H69/V70. Näyte 7A on toukokuun 2022 loppupuolelta. Tällöin omikronvarianteista Jyväskylässä BA.2 oli valtavariantti ja BA.5 oli jo havaittavissa (THL 2022). Aiemmissa tutkimuksissa on huomattu, että Δ H69/V70-deletio esiintyy omikron alalinjoissa BA.4 ja BA.5 (katso kuva 1) (Wang ym. 2022). Tämän perusteella voidaan päätellä, että näyte 7A sisältää omikron BA.4 tai BA.5 linjaa.

Jätevesinäytteiden ensimmäisestä PCR:stä huomataan, että monistuminen ei tapahtunut toivotusti. Kuvaajista osa on lineaarisia ja *threshold*-arvo ohitetaan aikaisessa vaiheessa (kuva 8 ja taulukko 4). Kun *threshold*-arvo ohitetaan aikaisessa vaiheessa, qPCR tuottaa vain vähän templaattia (Bivins ym. 2021). Tämä puolestaan johtaa seuraavissa vaiheissa entistä vähäisempään määrään templaattia, kuten huomattiin Qubit-mittauksessa. Jätevedessä on SARS-CoV-2-RNA:n lisäksi paljon muita materiaaleja kuten esimerkiksi muiden virusten RNA:ta (Hellmér ym. 2014). Tämän myötä SARS-CoV-2-viruksen havaitsemiseen jätevedessä voi vaikuttaa monet fyysiset ja kemialliset tekijät kuten esimerkiksi PCR-reaktiota inhiboivat aineet (Tiwari ym. 2022). Lisäksi jäteveden puhdistamolle tuleva virtaama vaikuttaa jäteveden konsentraatioon. SARS-CoV-2-virus voi liittyä veteen liuenneisiin kiinteisiin aineisiin. Tällöin SARS-CoV-2:n RNA:ta voidaan eristää sitä enemmän mitä vähemmän on

virtaamaa, koska vähäisessä virtaamassa on suurempi konsentraatio kiinteitä liuenneita aineita (D'Aoust ym. 2021, Tiwari ym. 2022).

PCR-tuloksista huomataan, että ensimmäisessä PCR:ssä RBD-alukkeet monistivat ainoastaan näytettä 7A ja NTD-alukkeet monistivat näytteitä 2A-6B heikosti. Näiden alukkeiden istuvuudessa oli selkeästi jokin vialla. Näyte 7A oli viimeisin jätevesinäyte ja se on ainoa, joka tuotti molemmilla alukkeilla halutun lopputuloksen. Toisessa PCR:ssä mikään näyte ei onnistunut täydellisesti. Näyte 7A monistuu hieman molemmilla alukkeilla, mutta tuote ei ollut toivotunlaista. Tässäkään tapauksessa alukkeet eivät istuneet kunnolla. Myöhemmissä tutkimuksissa huomattiin, että toisessa PCR:ssä käytetyt P1-alukkeet eivät sisältäneet oikeaa P1-adapterisekvenssiä ja osassa alukkeista sekvenssi oli väärä. Tämän myötä tuotetta ei syntynyt. Virheen huomaamisen jälkeen näyte 7A sekvensointiin uudelleen RBD-alueelta. Fylogeneettisestä puusta huomataan, että kaikki jätevesinäytteet ovat omikron BA.2 haaraa. THL:n (2022b) määritysten mukaan tämän tutkimuksen jätevesinäytteiden ottamisen aikaan omikron BA.2 oli valtavariantti jätevesissä. Tämä tieto on kuitenkin ristiriidassa näytteen tuottaman qPCR:n ROX-fluoresenssin kanssa. ROX-kanava kertoo deleetion Δ H69/V70 olemassaolosta, jota ei kuitenkaan ole omikron BA.2 linjassa (Wang ym. 2022).

Nextclade-analyysissä huomataan, että kaikki jätevesinäytteen sekvenssit sisältävät omikronille tyypillisiä mutaatioita muun muassa K417N ja N440K. Omikron alalinjalla BA.2 tyypillisiä mutaatioita ovat D405N ja R408S (Chen ja Wei 2022). Nämä mutaatiot löytyvät sekvensoidun jätevesinäytteen sekvensseistä ja näin tukevat päätelmää, että sekvenssit edustavat omikron BA.2 linjaa. Sekvensoitu alue oli viruksen piikkiproteiinin RBD-alueella. Wang ym. (2022) mukaan varianteilla BA.4/5 ja BA.2 on hyvin samankaltaiset RBD-alueet. Tällöin on mahdollista, että jätevesinäyte sisälsi BA.4/5 varianttia, mutta sitä ei voida eritellä erikseen, koska sekvensointi ei ollut NTD-alueella ja eri varianttien RBD-alueet ovat hyvin samankaltaisia. THL:n määritysten perusteella omikron BA.2 ja BA.4/5 variantteja oli näytteen 7A ottamisen aikaan liikkeellä (THL 2022b). Jätevesinäyte luultavasti sisälsi molempia omikron BA.2 ja BA.4/5 linjoja ja sekvensoimalla sekä RBD- että NTD-alueen variantista saataisiin varmuus. Näiden tulosten perusteella johtopäätöksiä ei voida tehdä.

4.3 Tilastolliset mallit ja tunnusluvut

Virusten määrää mallintavassa mallissa fluoresenssit ja tapausmäärät ovat tilastollisesti merkitseviä kertoimia. Tapausmäärät vaikuttavat virusten määrään jätevedessä mallin mukaan negatiivisesti. Tarkastellessa virusmäärän ja COVID-19-tapausten välistä korrelaatiota huomataan, että korrelaatio on positiivinen. Lineaarinen sekamalli estimoitii ML-menetelmällä, jonka myötä estimaatit saattavat olla harhaisempia verrattuna REML-menetelmän estimaatteihin (Bolker ym. 2009). Lisäksi kevään 2022 virallisesti todettujen tapausten määrä väheni Suomen koronastrategian muuttumisen myötä ja tämän myötä tapausten määrä ei kuvaa yhtä tehokkaasti tartuntojen todellista tilannetta. Tehohoidossa

olevien potilaiden määrä vaikuttaa positiivisesti virus-RNA:n määrään jätevedessä. Tämä tulos ei kuitenkaan ole tilastollisesti merkitsevä. Tehohoidossa olevien potilaiden määrä kuvaa Kuopion yliopistollisen sairaalan erityisvastuualueen potilaita (THL kuutiokäyttöliittymä 2022). Tämän myötä data ei liity suoraan Jyväskylän tilanteeseen ja tämä heikentää selittäjän tehokkuutta. Tulevaisuudessa mallin rakentamisessa kannattaisi käyttää Jyväskylän alueen tehohoitopotilaiden määrää, jos eettinen näkökulma sallii sen.

Jäännöstarkastelussa huomataan, että Q-Q-plot on lähes lineaarinen, joten aineisto noudattaa melko hyvin normaalijakaumaa (Zubir ym. 2018). Yhteisjakaumasta huomataan että, että mallin sovitetuissa arvoissa ja jäännöksissä ei ole selkeää rakennetta. Tällöin malli sopii dataan hyvin, koska jäännösten virheet noudattavat normaalijakaumaa (Moffat ym. 2022). Mallin R²-arvon perusteella voidaan sanoa, että malli sopii dataan kohtalaisesti (Kvålseth 1985).

Tarkastellessa korrelaatioita muuttujien välillä huomataan, että virusmäärä jätevedessä korreloi positiivisesti vuorokausi- ja tuntivirtaaman, tapausmäärien, tehooidossa olevien potilaiden ja testausmäärien kanssa. Positiivinen korrelaatio positiivisten tapausten ja virusmäärän kanssa on todettu myös aiemmissa tutkimuksissa esimerkiksi Mantilla-Calderon ym. (2022) mukaan päivittäiset uudet SARS-CoV-2-tapaukset korreloivat virus-RNA:n määrän kanssa kertoimella 0,28 (95 %-CI [0,01–0,51]). Tämän tutkimuksen korrelaatio osuu Mantilla-Calderon ym. (2022) määrittämälle luottamusvälille, joka antaa tulokselle luotettavuutta. Positiivinen korrelaatio tehooidossa olevien potilaiden kanssa on 0,41. Tulos on melko matala verrattuna aiempiin tutkimuksiin, joissa on saatu suurempia korrelaatioita virus-RNA:n ja tehooidossa olevien potilaiden määrään (Perez-Zabaleta ym. 2022, Hopkins ym. 2023). Tähän saattaa vaikuttaa mallinnuksessa käytetyt Kuopion yliopistollisen sairaalan erityisvastuualueen tehohoitopotilaiden määrät, koska nämä potilaat eivät kuvaa täsmällisesti Jyväskylän tilannetta.

Tunti- ja vuorokausivirtaaman kasvaessa virusmäärä jätevedessä nousee laskettujen korrelaatioiden mukaan. Jätevesinäytteet ovat kevään 2022 ajalta, jolloin jätevedenpuhdistamolle ajautui sulamisvesiä. Tämä kasvattaa virtaamaa ja laimentaa jätevettä (Tiwari ym. 2022). Tähän perustuen tämän tutkimuksen laskettu korrelaatio on hyvin poikkeava. Lisäksi on huomattavaa, että THL:n data, jonka avulla rakennettiin erillinen lineaarinen sekamalli (yhtälö 4), sisältää huomattavasti pienemmän vuorokausivirtaaman keskiarvon. Näihin tietoihin pohjautuen normaalia suuremmat muuttujan arvot saattavat vääristää estimaatteja. Tulevaisuudessa analyysissa voisi huomioida sääolosuhteet ja jätevedenpuhdistamon geospaatialiset tiedot entistä tarkemmin, kuten esimerkiksi Fitzgerald ym. (2021) teki tutkimuksessaan. Tämä luultavasti tarkentaisi saatuja korrelaatioita, koska tässä tutkimuksessa luultavasti sulamisvedet vääristivät tuloksia.

Wilcoxonin testin tulosten perusteella tämän tutkimuksen ja THL:n tekemän tutkimuksen välillä on tilastollisesti merkittävä ero. Tähän eroon on monia syitä. Tämän tutkimuksen otos on pieni. THL (2022a) ottaa

jätevesinäytteet kokoomänäytteinä ja tämä tutkimus otti kertänäytteitä, jotka kuvasivat näytteenottohetkellä vallinneita olosuhteita. Tässä tutkimuksessa käytetyt menetelmät eivät olleet täysin samat muun muassa tässä tutkimuksessa näytteiden SARS-CoV-2-osoitus tehtiin cheMagic-laitteella ja THL:n laitteesta ei ole tietoa.

Tämän tutkimuksen virtaamakorjaus oli keskimäärin 10-kertaa suurempi kuin THL:n vastaavat. Tässä tutkimuksessa käytetyssä qPCR-osoitus menetelmässä Cq-arvo 42 vastasi yhtä virusta. Lisäanalyysi virtaamadatalle tehtiin kokeilemalla, jos yhtä virusta vastaa Cq-arvo 39. Tämä analyysi antoi virtaamadatalle 0,02–7-kertaisen arvon verrattuna THL:n tuottamaan dataan. Tämän perusteella voidaan päätellä, että Cq-raja-arvon määrittäminen tuottaa merkittävän virhelähteen tämän tutkimuksen ja THL:n tulosten välille. Menetelmään riippuva virhe ei luultavasti ole, koska jätevesinäytteistä osa oli rinnakkaisia ja nämä tuottivat samankaltaisia tuloksia. Käytetyn cheMagic-laitteen ja kitin valmistajan mukaan kaikki näytteet, joiden Cq-arvo on alle 40, ovat COVID-19-positiivisia. Monissa muissa menetelmissä positiivisen näytteen määrittävä Cq-arvo on matalampi (Garg ym. 2021). Tähän perustuen on mahdollista, että cheMagic-laitteen valmistajan ilmoittama positiivisen näytteen raja-arvo vääristää tuloksia ja todellisuudessa THL:n ja tämän tutkimuksen tulosten välillä ei ole tilastollisesti merkitsevää eroa. Kitit muuttuvat kehityksen kulkiessa eteenpäin ja tämän myötä on mahdollista, että THL käyttää kittiä, jonka entsyymit ovat tehokkaampia kuin tässä tutkimuksessa käytetyt. Lisäksi on huomattavaa, miten tämän tutkimuksen muuttujien keskiarvot eroavat THL:n datasta, jolla rakennettiin malli 4 (yhtälö 4). Virusmäärä on keskimääräisesti suurempi tässä tutkimuksessa kuin pidempää aikaväliä kuvaavassa THL:n datassa. Tämä tukee entisestään tulkintaa qPCR positiivisen näytteen raja-arvon tärkeydestä ja kuinka sen asettaminen saattaa antaa todellisia arvoja suurempia arvoja. Virhettä virusmäärään saattaa aiheuttaa myös qPCR:n *threshold*-arvon määrittäminen. Käyttäjä voi itse manuaalisesti määrittää qPCR:n määrittää *threshold*-arvon. Tällöin Cq-arvot ovat myös datan lukijasta ja lukijan valitsemasta *threshold*-arvosta riippuvaisia.

THL:n dataa mallintaessa huomataan, että paras malli AIC-arvoon perustuen on malli, joka sisältää selittäjinään tapausmäärät sekä tehohoidossa olevien potilaiden määrän. Q-Q-plottia tarkastellessa huomataan, että se ei ole täysin lineaarinen. Tähän perustuen jätesinäytteiden data ei ole täysin normalijakautunutta (Zubir ym. 2018). Kun tutkitaan jäännösten ja sovitettujen arvojen jakaumaa, siinä ei näkyä hieman rakennetta. Tällöin malli sopii hieman puutteellisesti aineistoon, koska jäännösten virheiden pitäisi noudattaa normaalijakaumaa (Moffat ym. 2022). Mallin R²-arvo on alhainen, joten malli ei selitä dataa kovin hyvin. Mallia olisi voinut tarkentaa lisäämällä mahdollisia selittäjiä tai interaktiotermejä. Lisäksi vuoden 2022 aikana kaikkia COVID-19-tapauksia ei virallisesti tilastoitu THL:n toimesta. Tämä vähentää näennäisesti tapauksia ja saattaa aiheuttaa epävarmuutta estimaatteihin. Regressiomallit ovat usein antaneet alhaisia R²-arvoja, kun niitä on käytetty mallintamaan SARS-CoV-2-virus-RNA:n määrää jätevedessä. Tämä saattaa johtua esimerkiksi jätevedessä

olevista tuntemattomista tekijöistä, käytetyistä menetelmistä tai esimerkiksi tilastoitujen tapausten viiveestä (Fitzgerald ym. 2021, Mantilla-Calderon ym. 2022). Molemmissa tässä tutkimuksessa tuotetuissa regressiomalleissa jokin edellä mainituista tekijöistä saattaa vaikuttaa tuloksiin ja aiheuttaa vääristymää estimaatteihin. Aiemmissa tutkimuksissa on tarkasteltu viranomaisten COVID-19-pandemia datan sisältävän viiveen vaikutusta tilastollisiin tarkasteluihin (Weidhaas ym. 2021, Mantilla-Calderon ym. 2022). Tässä tutkimuksessa THL:n dataa mallintavan mallin rakentamisen jälkeen tehohoitopotilaiden dataan tehtiin kahden viikon viive. Tämä ei parantanut mallin estimaatteja tai R^2 -arvoa. Viiveen määrää viranomaisten tuottamassa datassa on vaikea estimoida ja saatujen tulosten perusteella viiveen estimointi käsin ei paranna estimaatteja.

4.4 Systemaattisesta kirjallisuuskatsauksesta

Systemaattisessa kirjallisuuskatsauksessa käytetyt tutkimukset käsittivät laajalti eri saostus- ja PCR-analyysimenetelmiä, jotka vastasivat tässä tutkielmassa käytettyjä. Suurin osa tutkimuksista oli toteutettu Pohjois-Amerikassa (1, 12, 16, 18, 19, 20, 21 ja 27) tai Euroopassa (2, 7, 8, 10, 11, 14, 17, 22, 23, 24 ja 26). Tutkimuksia oli myös toteutettu Aasiassa (5 ja 6), Australiassa (3 ja 4) ja Etelä-Amerikassa (9, 13, 25 ja 28). Huomattavaa on, ettei yksikään tutkimuksista ole toteutettu Afrikassa mikä voi liittyä puutteelliseen infrastruktuuriin ja SARS-CoV-2-diagnostiikkaan (Wang ym. 2012, Tegally ym. 2022)

Tässä tutkimuksessa käytetyt näytemäärät vastasivat läheisesti SKK:n tutkimusten näytemääriä. SKK:ssa vain neljässä tutkimuksessa ei ollut määritelty tarkemmin näytemääriä (11, 13, 15 ja 24). Yleisimmin tutkimuksissa käytettiin 9–100 näytettä (1, 2, 3, 4, 5, 7, 12, 14, 20, 25, 26, 28) tai 100–1000 näytettä (6, 8, 9, 10, 13, 17, 21, 22, 27) (kuva 16). Tämän tutkimuksen näytemäärä osui myös tälle vaihteluvälille (11 jätevesinäytettä), jolloin menetelmämme ja tuloksemme ovat vertailukelpoisia SKK:n tutkimusten kanssa. Pienestä näytemäärästä huolimatta tuloksemme ovat toistettavissa perustuen, koska näytteiden lisääminen ei vaikuta parantavan tuloksia perustuen SKK:hon.

Yleistä tutkimuksissa oli saostaa näytteet jollakin menetelmällä, mutta osa SKK:n tutkimuksista ei saostanut näytteitään (6, 16, 19 ja 27). Saostusmenetelmällä ei ole suurta vaikutusta tuloksiin, mutta tähän tutkimukseen ehdotetut saostusmenetelmät (PEG ja ultrafiltraatio) näyttävät olevan yleisiä. PEG:tä käyttivät tutkimukset 5, 7, 10, 12, 13, 14, 21, 23 ja 28; kun taas ultrafiltraatiota käyttivät tutkimukset 2, 7, 11, 15, 20, 24, 25 ja 26. Adsorptio-metodi vaikuttaa olevan harvemmin käytetty, vaikka sen talteenottoaste vaikuttaa olevan parhain SARS-CoV-2-viruksen jätevesiseurannassa (Ahmed ym. 2020).

PCR-analyysimetodeista RT-qPCR kattaa suurimman osan tutkimuksia (1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25 ja 27), mikä ei kuitenkaan ole tarkin saatavilla olevista menetelmistä eikä RT-qPCR tuota parhaiten toistettavia tuloksia. RT-ddPCR vaikuttaa tuottavan tuloksia pienemmällä tulosten vaihtelevuudella ja paremmalla tulosten toistettavuudella,

minkä takia kyseistä menetelmää olisi voinut harkita myös tässä tutkimuksessa (Taylor ym. 2017). LuK-tutkielmassa päätettiin kuitenkin käyttää RT-qPCR-menetelmää, sillä RT-ddPCR-menetelmä oli vaihtoehtona kalliimpi, ja kyseiseen menetelmään ei ollut valmiina kaupallista kittiä ja se olisi pitänyt tilata. Yksittäisen tutkimuksen tapauksessa RT-qPCR vaikuttaa kuitenkin olevan tarpeeksi tarkka perustuen tämän tutkimuksen tuloksiin jätevesinäytteistä.

5 PÄÄTELMÄT

Tämän tutkimuksen ja systemaattisen kirjallisuuskatsauksen perusteella voidaan todeta, että jätevesiseuranta pandemian seurannassa on perusteltua, mutta siinä on monia tekijöitä, jotka voivat vaikuttaa tulosten vertailukelpoisuuteen (esimerkiksi tulosten standardointi ja näytteiden käsittelymenetelmät). Itse mitattujen jätevesinäytteiden ja THL:n datan avulla voidaan päätellä, että virusmäärien mallintaminen on vaikeaa. Tämä johtuu monista syistä kuten esimerkiksi havaitsemattomista muuttujista, joiden vaikutusta vasteeseen ei tunneta ja tilastoidun datan viiveestä verrattuna jätevesinäytteisiin. Rakennettujen mallien mukaan virallisesti tilastoidut tapaukset ennustavat jätevedessä olevan virus-RNA:n määrää. Estimoidun virusmäärän perusteella voitaisiin päätellä tapausten kokonaismäärä (virallisesti tilastoidut ja tilastoimattomat), jos tiedetään, paljonko yksi sairastunut henkilö erittää virus-RNA:ta. Tieto tapausten kokonaismäärästä auttaa viranomaisia seuraamaan pandemian suuntaa ja suunnittelemaan mahdollisia toimenpiteitä.

Sanger-sekvenssien BLAST-hakua ei ole käsitelty tässä tutkielmassa, koska haun tulokset eivät tuottaneet yksiselitteistä tulosta. Näiden tulosten perusteella ei onnistuttu luomaan ajallista kuvaa Jyväskylässä vallinneista varianteista. Sen sijaan näytteiden variantit selvitettiin linjaamalla ne ja luomalla fylogeneettinen puu. Fylogeneettisen puun ja näytteiden aikapisteiden avulla selvitettyjen näytteiden varianttien perusteella voidaan todeta, että jo 571 bp pituisella sekvenssillä voidaan selvittää näytteen variantti. Tutkimuksen tulokset myös osoittivat, että alueella 23010-23080 bp olisi lyhyt, mutta tehokas alue SARS-CoV-2-variantin selvittämiseen. Tältä alueelta löytyi tulosten mukaan 8 emäksen eroavaisuutta, mikä yksistään jo riitti variantin toteamiseen. Tuloksia olisi silti tarkasteltava näytteenoton ajankohdan mukaan, jotta alueelta hävinneet variantit voidaan poissulkea.

Tulevaisuudessa tämän tutkimuksen tulosten perusteella voidaan päätellä, että jätevesiseuranta COVID-19-pandemian ja muiden epidemioiden seurannassa on kannattavaa. Tuloksiin voi kuitenkin vaikuttaa monet tekijät ja tämän myötä menetelmiä kannattaa vielä kehittää. Menetelmissä kannattaisi käyttää mahdollisimman samanlaisia menetelmiä kuin THL, jotta data olisi täysin vertailukelpoista. Tutkimusta voisi laajentaa hankkimalla dataa Jyväskylän COVID-19-tilanteesta esimerkiksi tietoa Jyväskylän tehohoitopotilaista, joilla on COVID-19. Tämä luultavasti parantaisi tilastollisen mallinnuksen estimaatteja ja R^2 -arvoja.

KIITOKSET

Haluamme kiittää ohjaajiamme Marja Tirolaa ja Elina Ruuhilehtoa kaikesta avusta ja tuesta tutkielman kirjoittamisessa. Kiitämme BiopSense Oy:tä nenänielunäytteiden jakamisesta meille sekä kaikesta käytännön avusta virus-RNA:n eristyksessä. Kiitämme Sonja Pyykköstä J-S-puhdistamolta jätevesinäytteiden kanssa auttamisesta ja virtaamadatan antamisesta. Kiitämme myös THL:n Soile Blomqvistia Sanger-sekvensoinnin aluketiedoista.

KIRJALLISUUSLUETTELO

- Acer P.T. 2022. Quantifying the Relationship between SARS-CoV-2 Wastewater Concentrations and Building-Level COVID-19 Prevalence at an Isolation Residence: A Passive Sampling Approach. *International journal of environmental research and public health* 19: 11245-.
- Agrawal S. 2021. Long-term monitoring of SARS-CoV-2 RNA in wastewater of the Frankfurt metropolitan area in Southern Germany. *Scientific reports* 11: 5372–5377.
- Ahmed W. 2021. SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. *The Science of the total environment* 761: 144216–144216.
- Ahmed W., Angel N., Edson J., Bibby K., Bivins A., O'Brien J.W., Choi P.M., Kitajima M., Simpson S.L., Li J., Tschärke B., Verhagen R., Smith W.J.M., Zaugg J., Dierens L., Hugenholz P., Thomas K.V. & Mueller J.F. 2020. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Science of The Total Environment* 728: 138764. 1
- Ahmed W., Bertsch P.M., Bivins A., Bibby K., Farkas K., Gathercole A., Haramoto E., Gyawali P., Korajkic A., McMinn B.R., Mueller J.F., Simpson S., Smith W.J.M., Symonds E.M., Thomas K. v., Verhagen R. & Kitajima M. 2020. Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. *Science of The Total Environment* 739: 139960.
- Aksamentov I., Roemer C., Hodcroft E. & Neher R. 2021. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software* 6: 3773.
- Alhama J. 2022. Monitoring COVID - 19 through SARS - CoV - 2 quantification in wastewater: progress, challenges and prospects. *Microbial Biotechnology* 15: 1719.
- Alkhatib M., Svicher V., Salpini R., Ambrosio F.A., Bellocchi M.C., Carioti L., Piermatteo L., Scutari R., Costa G., Artese A., Alcaro S., Shafer R. & Ceccherini-Silberstein F. 2021. SARS-CoV-2 Variants and Their Relevant Mutational Profiles: Update Summer 2021. *Microbiology Spectrum* 9.

- Amahmid O. 2022. Occurrence of SARS-CoV-2 in excreta, sewage, and environment: epidemiological significance and potential risks. *International journal of environmental health research* 32: 1686–1706.
- Arora S. 2020. Sewage surveillance for the presence of SARS-CoV-2 genome as a useful wastewater based epidemiology (WBE) tracking tool in India. *Water science and technology* 82: 2823–2836.
- Arora S. 2022. Successful application of wastewater-based epidemiology in prediction and monitoring of the second wave of COVID-19 with fragmented sewerage systems—a case study of Jaipur (India). *Environmental monitoring and assessment* 194: 342–342.
- Basavaraju S, Aswathanarayan JB, Basavegowda M, Somanathan B. 2021. Coronavirus: occurrence, surveillance, and persistence in wastewater. *Environ Monit Assess.* 193(8):508.
- Boeraş I. 2022. Anthropogenic Sewage Water Circuit as Vector for SARS-CoV-2 Viral ARN Transport and Public Health Assessment, Monitoring and Forecasting—Sibiu Metropolitan Area (Transylvania/Romania) Study Case. *International journal of environmental research and public health* 19: 11725–.
- Bolker B.M., Brooks M.E., Clark C.J., Geange S.W., Poulsen J.R., Stevens M.H.H. & White J.-S.S. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution* 24: 127–135.
- Bivins A., Kaya D., Bibby K., Simpson S.L., Bustin S.A., Shanks O.C. & Ahmed W. 2021. Variability in RT-qPCR assay parameters indicates unreliable SARS-CoV-2 RNA quantification for wastewater surveillance. *Water Research* 203: 117516.
- Calabria de Araujo J. 2021. SARS-CoV-2 sewage surveillance in low-income countries: potential and challenges. *Journal of water and health* 19: 1–19.
- Cao Y., Yisimayi A., Jian F., Song W., Xiao T., Wang L., Du S., Wang J., Li Q., Chen X., Yu Y., Wang P., Zhang Z., Liu P., An R., Hao X., Wang Y., Wang J., Feng R., Sun H., Zhao L., Zhang W., Zhao D., Zheng J., Yu L., Li C., Zhang N., Wang R., Niu X., Yang S., Song X., Chai Y., Hu Y., Shi Y., Zheng L., Li Z., Gu Q., Shao F., Huang W., Jin R., Shen Z., Wang Y., Wang X., Xiao J. & Xie X.S. 2022. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature* 608: 593–602.
- Carcereny A. 2021. Monitoring Emergence of the SARS-CoV 2 B.1.1.7 Variant through the Spanish National SARS-CoV 2 Wastewater Surveillance System (VATar COVID-19). *Environmental science & technology* 55: 11756–11766.
- Cascella M., Rajnik M., Aleem A., Dulebohn S.C. & Di Napoli R. 2022. Features, Evaluation, and Treatment of Coronavirus (COVID-19). Teoksessa: StatPearls, StatPearls Publishing, Treasure Island (FL).
- Cavanagh D. 1995. The Coronavirus Surface Glycoprotein. Teoksessa: Siddell S.G. (toim.), *The Coronaviridae*, Springer US, Boston, MA, ss. 73–113.
- Centers for Disease Control and Prevention (CDC). 2013. CDC SARS Response Timeline <https://www.cdc.gov/about/history/sars/timeline.html> (luettu 3.11.2022)

- Centers for Disease Control and Prevention (CDC). Middle East Respiratory Syndrome (MERS). 2019. <https://www.cdc.gov/coronavirus/mers/index.html> (luettu 3.11.2022).
- Chen J. & Wei G.W. 2022. Omicron BA.2 (B.1.1.529.2): High Potential for Becoming the Next Dominant Variant. *Journal of Physical Chemistry Letters* 13: 3840–3849.
- Cherian S., Potdar V., Jadhav S., Yadav P., Gupta N., Das M., Rakshit P., Singh S., Abraham P., Panda S. & Team N. 2021. SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms* 9: 1542.
- Claro I.C.M., Cabral A.D., Augusto M.R., Duran A.F.A., Graciosa M.C.P., Fonseca F.L.A., Speranca M.A. & Bueno R. de F. 2021. Long-term monitoring of SARS-COV-2 RNA in wastewater in Brazil: A more responsive and economical approach. *Water Research* 203: 117534.
- D'Aoust P.M., Mercier E., Montpetit D., Jia J.J., Alexandrov I., Neault N., Baig A.T., Mayne J., Zhang X., Alain T., Langlois M.A., Servos M.R., MacKenzie M., Figeys D., MacKenzie A.E., Graber T.E. & Delatolla R. 2021. Quantitative analysis of SARS-CoV-2 RNA from wastewater solids in communities with low COVID-19 incidence and prevalence. *Water Research* 188: 116560.
- Daughton C.G. 2020. Wastewater surveillance for population-wide Covid-19: The present and future. *The Science of the Total Environment* 736: 139631.
- De Giglio O. 2021. Potential Use of Untreated Wastewater for Assessing COVID-19 Trends in Southern Italy. *International journal of environmental research and public health* 18: 10278-.
- Dzinamarira T. 2022. Utilization of SARS-CoV-2 Wastewater Surveillance in Africa-A Rapid Review. *International journal of environmental research and public health* 19: 969-.
- Engelmann I., Alidjinou E.K., Ogiez J., Pagneux Q., Miloudi S., Benhalima I., Ouafi M., Sane F., Hober D., Roussel A., Cambillau C., Devos D., Boukherroub R. & Szunerits S. 2021. Preanalytical Issues and Cycle Threshold Values in SARS-CoV-2 Real-Time RT-PCR Testing: Should Test Results Include These? *ACS Omega* 6: 6528–6536.
- Fahrenfeld N.L. 2022. Year - long wastewater monitoring for SARS - CoV - 2 signals in combined and separate sanitary sewers. *Water environment research* 94: e10768-n/a.
- Farkas K. 2021. Concentration and Quantification of SARS-CoV-2 RNA in Wastewater Using Polyethylene Glycol-Based Concentration and qRT-PCR. *Methods and protocols* 4: 17-.
- Fitzgerald S.F., Rossi G., Low A.S., McAteer S.P., O'Keefe B., Findlay D., Cameron G.J., Pollard P., Singleton P.T.R., Ponton G., Singer A.C., Farkas K., Jones D., Graham D.W., Quintela-Baluja M., Tait-Burkard C., Gally D.L., Kao R. & Corbishley A. 2021. Site Specific Relationships between COVID-19 Cases and SARS-CoV-2 Viral Load in Wastewater Treatment Plant Influent. *Environmental Science & Technology* 55: 15276–15286.

- Flood M.T. 2021. Methods Evaluation for Rapid Concentration and Quantification of SARS-CoV-2 in Raw Wastewater Using Droplet Digital and Quantitative RT-PCR. *Food and environmental virology* 13: 303–315.
- Garg A., Ghoshal U., Patel S.S., Singh D. v., Arya A.K., Vasanth S., Pandey A. & Srivastava N. 2021. Evaluation of seven commercial RT-PCR kits for COVID-19 testing in pooled clinical specimens. *Journal of Medical Virology* 93: 2281–2286.
- Giraud-Billoud M. 2021. Monitoring of SARS-CoV-2 RNA in wastewater as an epidemiological surveillance tool in Mendoza, Argentina. *The Science of the total environment* 796: 148887–148887.
- Girón-Navarro R. 2021. The impact of coronavirus SARS-CoV-2 (COVID-19) in water: potential risks. *Environmental science and pollution research international* 28: 52651–52674.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41:95-98.
- Hellmér M., Paxéus N., Magnus L., Enache L., Arnholm B., Johansson A., Bergström T. & Norder H. 2014. Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Applied and Environmental Microbiology* 80: 6771–6781.
- Ho J. 2022. SARS-CoV-2 wastewater surveillance in Germany: Long-term RT-digital droplet PCR monitoring, suitability of primer/probe combinations and biomarker stability. *Water research (Oxford)* 210: 117977–117977.
- Hopkins L., Persse D., Caton K., Ensor K., Schneider R., McCall C. & Stadler L.B. 2023. Citywide wastewater SARS-CoV-2 levels strongly correlated with multiple disease surveillance indicators and outcomes over three COVID-19 waves. *Science of The Total Environment* 855: 158967.
- Izquierdo-Lara R. 2021. Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing, the netherlands and belgium. *Emerging infectious diseases* 27: 1405–1415.
- Jordan P., Alessandro Z., Brackney D.E., Grubaugh N.D., Kaplan E.H., Arnau C.-M., Ko A.I., Malik A.A., Wang D., Wang M., Warren J.L., Weinberger D.M., Wyatt A., Omer S.B. 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology* 38: 1164–1167.
- Joseph-Duran B. 2022. Assessing wastewater-based epidemiology for the prediction of SARS-CoV-2 incidence in Catalonia. *Scientific reports* 12: 15073–15073.
- Kaplan E.H. 2021. Aligning SARS-CoV-2 indicators via an epidemic model: application to hospital admissions and RNA detection in sewage sludge. *Health care management science* 24: 320–329.
- Karthikeyan S. 2022. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature (London)* 609: 101–104.

- Kaya D. 2022. Correlation between Clinical and Wastewater SARS-CoV-2 Genomic Surveillance, Oregon, USA. *Emerging infectious diseases* 28: 1906–1908.
- Kuzniewski S. 2021. Prevalence, environmental fate, treatment strategies, and future challenges for wastewater contaminated with SARS - CoV - 2. *Remediation (New York, N.Y.)* 31: 97.
- Kvålseth T.O. 1985. Cautionary Note about R 2. *The American Statistician* 39: 279–285.
- Li L. 2022a. Detecting SARS-CoV-2 variants in wastewater and their correlation with circulating variants in the communities. *Scientific reports* 12: 16141–16141.
- Li L. 2022b. Longitudinal monitoring of SARS-CoV-2 in wastewater using viral genetic markers and the estimation of unconfirmed COVID-19 cases. *The Science of the total environment* 817: 152958–152958.
- Lindsey J. K. & Jones B. 1998. Choosing among generalized linear models applied to medical data - Lindsey - 1998 - *Statistics in Medicine* - Wiley Online Library. *Statistics in medicine*: 59–68.
- Liu Z., VanBlargan L.A., Bloyet L.-M., Rothlauf P.W., Chen R.E., Stumpf S., Zhao H., Errico J.M., Theel E.S., Liebeskind M.J., Alford B., Buchser W.J., Ellebedy A.H., Fremont D.H., Diamond M.S. & Whelan S.P.J. 2021. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host & Microbe* 29: 477-488.e4.
- Mandal P. 2020. A review on presence, survival, disinfection/removal methods of coronavirus in wastewater and progress of wastewater-based epidemiology. *Journal of environmental chemical engineering* 8: 104317–104317.
- Mantilla-Calderon D., Huang K., Li A., Chibwe K., Yu X., Ye Y., Liu L. & Ling F. 2022. Emerging investigator series: meta-analyses on SARS-CoV-2 viral RNA levels in wastewater and their correlations to epidemiological indicators. *Environmental Science: Water Research & Technology* 8: 1391–1407.
- McCarthy K.R., Rennick L.J., Nambulli S., Robinson-McCarthy L.R., Bain W.G., Haidar G. & Paul Duprex W. 2021. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* 371: 1139–1142.
- Meng B., Kemp S.A., Papa G., Datir R., Ferreira I.A.T.M., Marelli S., Harvey W.T., Lytras S., Mohamed A., Gallo G., Thakur N., Collier D.A., Mlcochova P., Robson S.C., Loman N.J., Connor T.R., Golubchik T., Martinez Nunez R.T., Ludden C., Corden S., Johnston I., Bonsall D., Smith C.P., Awan A.R., Bucca G., Torok M.E., Saeed K., Prieto J.A., Jackson D.K., Hamilton W.L., Snell L.B., Moore C., Harrison E.M., Goncalves S., Fairley D.J., Loose M.W., Watkins J., Livett R., Moses S., Amato R., Nicholls S., Bull M., Smith D.L., Barrett J., Aanensen D.M., Curran M.D., Parmar S., Aggarwal D., Shepherd J.G., Parker M.D., Glaysher S., Bashton M., Underwood A.P., Pacchiarini N., Loveson K.F., Templeton K.E., Langford C.F., Sillitoe J., Silva T.I. de, Wang D., Kwiatkowski D., Rambaut A., O’Grady J., Cottrell S., Holden M.T.G.,

- Thomson E.C., Osman H., Andersson M., Chauhan A.J., Hassan-Ibrahim M.O., Lawniczak M., Alderton A., Chand M., Constantinidou C., Unnikrishnan M., Darby A.C., Hiscox J.A., Paterson S., Martincorena I., Volz E.M., Page A.J., Pybus O.G., Bassett A.R., Ariani C. v., Chapman M.H.S., Li K.K., Shah R.N., Jesudason N.G., Taha Y., McHugh M.P., Dewar R., Jahun A.S., McMurray C., Pandey S., McKenna J.P., Nelson A., Young G.R., McCann C.M., et al. 2021. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Reports* 35.
- Mistry P., Barmania F., Mellet J., Peta K., Strydom A., Viljoen I.M., James W., Gordon S. & Pepper M.S. 2021. SARS-CoV-2 Variants, Vaccines, and Host Immunity. *Frontiers in Immunology* 12.
- Mlejnkova H. 2020. Preliminary Study of Sars-Cov-2 Occurrence in Wastewater in the Czech Republic. *International journal of environmental research and public health* 17: 5508-.
- Moffat I.U., Akpan E.A. & Dashed L.M. Linear Mixed Effects Regression Modeling of SARS-CoV-2 Spread.
- Morvan M. 2022. An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence. *Nature communications* 13: 4313–4313.
- Nakagawa S. & Schielzeth H. 2013. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4: 133–142.
- Onnetomuustutkintakeskus. 2021. Koronaepidemian ensimmäinen vaihe Suomessa vuonna 2020. (P2020-1) <https://www.turvallisuustutkinta.fi/fi/index/tutkintaselostukset/poikkueksellisettapahtumat/p2020-01koronaepidemianensimmainenvaihesuomessavuonna2020.html#>
- O'Toole Á., Hill V., Pybus O.G., Watts A., Bogoch I.I., Khan K., Messina J.P., Tegally H., Lessells R.R., Giandhari J., Pillay S., Tumedi K.A., Nyepetsi G., Kebabonye M., Matsheka M., Mine M., Tokajian S., Hassan H., Salloum T., Merhi G., Koweyes J., Geoghegan J.L., Ligt J. de, Ren X., Storey M., Freed N.E., Pattabiraman C., Prasad P., Desai A.S., Vasanthapuram R., Schulz T.F., Steinbrück L., Stadler T., Parisi A., Bianco A., García de Viedma D., Buenestado-Serrano S., Borges V., Isidro J., Duarte S., Gomes J.P., Zuckerman N.S., Mandelboim M., Mor O., Seemann T., Arnott A., Draper J., Gall M., Rawlinson W., Deveson I., Schlegelbusch S., McMahon J., Leong L., Lim C.K., Chironna M., Loconsole D., Bal A., Josset L., Holmes E., George K. st., Lasek-Nesselquist E., Sikkema R.S., Oude Munnink B., Koopmans M., Brytting M., Sudha rani V., Pavani S., Smura T., Heim A., Kurkela S., Umair M., Salman M., Bartolini B., Rueca M., Drosten C., Wolff T., Silander O., Eggink D., Reusken C., Vennema H., Park A., Carrington C., Sahadeo N., Carr M., Gonzalez G., Oliveira T. de, Faria N., Rambaut A. & Kraemer M.U.G. 2021. Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2. *Wellcome Open Research* 6: 121.

- Peinado B. 2022. Improved methods for the detection and quantification of SARS-CoV-2 RNA in wastewater. *Scientific reports* 12: 7201–7201.
- Peng H. & Lu Y. 2012. Model selection in linear mixed effect models. *Journal of Multivariate Analysis* 109: 109–129.
- Perez-Zabaleta M., Archer A., Khatami K., Jafferli M.H., Nandy P., Atasoy M., Birgersson M., Williams C. & Cetecioglu Z. 2022. Long-term SARS-CoV-2 surveillance in the wastewater of Stockholm: What lessons can be learned from the Swedish perspective? *Science of The Total Environment*: 160023.
- Planas D., Bruel T., Grzelak L., Guivel-Benhassine F., Staropoli I., Porrot F., Planchais C., Buchrieser J., Rajah M.M., Bishop E., Albert M., Donati F., Prot M., Behillil S., Enouf V., Maquart M., Smati-Lafarge M., Varon E., Schortgen F., Yahyaoui L., Gonzalez M., Sèze J. de, Péré H., Veyer D., Sève A., Simon-Lorière E., Fafi-Kremer S., Stefic K., Mouquet H., Hocqueloux L., Werf S. van der, Prazuck T. & Schwartz O. 2021. Sensitivity of infectious SARS-CoV-2 B.1.1.7 and B.1.351 variants to neutralizing antibodies. *Nature Medicine* 2021 27:5 27: 917–924.
- Reno U. 2022. Wastewater-Based Epidemiology: Detection of SARS-CoV-2 RNA in Different Stages of Domestic Wastewater Treatment in Santa Fe, Argentina. *Water, air, and soil pollution* 233: 372–372.
- Saini G. 2021. Wastewater-based epidemiology for novel Coronavirus detection in wastewater. *Global journal of environmental science and management* 7: 643–658.
- Simmons G., Zmora P., Gierer S., Heurich A. & Pöhlmann S. 2013. Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. *Antiviral Research* 100: 605.
- Singh D. & Yi S.V. 2021. On the origin and evolution of SARS-CoV-2. *Exp Mol Med* 53: 537–547.
- SnapGene. 2022. SARS-CoV-2 (COVID-19) Genome. [https://www.snapgene.com/resources/coronavirus-resources/?resource=SARS-CoV-2_\(COVID-19\)_Genome](https://www.snapgene.com/resources/coronavirus-resources/?resource=SARS-CoV-2_(COVID-19)_Genome) (luettu 6.11.2022)
- Swiss Institute of Bioinformatics. 2022. SARS-CoV-2 circulating variants <https://viralzone.expasy.org/9556> (luettu 14.11.2022)
- Tamura K, Stecher G, and Kumar S. (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution* 38:3022-3027.
- Taylor S.C., Laperriere G. & Germain H. 2017. Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: from variable nonsense to publication quality data. *Scientific Reports* 7: 2409.
- Tegally H., San J.E., Cotten M., Moir M., Tegomoh B., Mboowa G., Martin D.P., Baxter C., Lambisia A.W., Diallo A., Amoako D.G., Diagne M.M., Sisay A., Zekri A.R.N., Gueye A.S., Sangare A.K., Ouedraogo A.S., Sow A., Musa A.O., Sesay A.K., Abias A.G., Elzagheid A.I., Lagare A., Kemi A.S., Abar A.E., Johnson A.A., Fowotade A., Oluwapelumi A.O., Amuri A.A., Juru A., Kandeil A., Mostafa A., Rebai A., Sayed A., Akano K., Balde A., Christoffels A., Trotter A.J., Campbell A., Keita A.K., Kone A., Bouzid A., Souissi A.,

- Agweyu A., Naguib A., Gutierrez A. v., Nkeshimana A., Page A.J., Yadouleton A., Vinze A., Happi A.N., Chouikha A., Iranzadeh A., Maharaj A., Batchi-Bouyou A.L., Ismail A., Sylverken A.A., Goba A., Femi A., Sijuwola A.E., Marycelin B., Salako B.L., Oderinde B.S., Bolajoko B., Diarra B., Herring B.L., Tsofa B., Lekana-Douki B., Mvula B., Njanpop-Lafourcade B.M., Marondera B.T., Khaireh B.A., Kouriba B., Adu B., Pool B., McInnis B., Brook C., Williamson C., Nduwimana C., Anscombe C., Pratt C.B., Scheepers C., Akoua-Koffi C.G., Agoti C.N., Mapanguy C.M., Loucoubar C., Onwuamah C.K., Ihekweazu C., Malaka C.N., Peyrefitte C., Chukwa G.C., Omoruyi C.E., Rafaï C.D., Morang'a C.M., Erameh C., Lule D.B., Bridges D.J., Mukadi-Bamuleka D., et al. 2022. The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly expanding genomic surveillance. *Science* 378.
- Terveyden ja hyvinvoinnin laitos (THL). 2022a. Koronaviruksen jätevesiseuranta <https://thl.fi/fi/tutkimus-ja-kehittaminen/tutkimukset-ja-hankkeet/sars-cov-2-jatevedenpuhdistamoilla/koronaviruksen-jatevesiseuranta> (luettu 23.9.2022)
- Terveyden ja hyvinvoinnin laitos (THL). 2022b. Koronavirusmuunnokset jätevedessä. <https://thl.fi/fi/tutkimus-ja-kehittaminen/tutkimukset-ja-hankkeet/sars-cov-2-jatevedenpuhdistamoilla/koronaviruksen-jatevesiseuranta/koronavirusmuunnokset-jatevedessa> (luettu 31.10.2022)
- Terveyden ja hyvinvoinnin laitos (THL) kuutiokäyttöliittymä. 2022. Raportit. <https://sampo.thl.fi/pivot/prod/fi/epirapo> (luettu 29.11.2022)
- Terveyden ja hyvinvoinnin laitos (THL). 2022c. Omikron BA2 on nyt yleisin Suomen jätevesistä havaittu koronavirusmuunnos <https://thl.fi/fi/-/omikron-ba.2-on-nyt-yleisin-suomen-jatevesista-havaittu-koronavirusmuunnos> (luettu 20.11.2022)
- Terveyden ja hyvinvoinnin laitos (THL). 2022d. Muuntuneet koronavirukset <https://thl.fi/fi/web/infektiotaudit-ja-rokotukset/ajankohtaista/ajankohtaista-koronaviruksesta-covid-19/tarttuminen-ja-suojautuminen-koronavirus/muuntuneet-koronavirukset> (luettu 20.11.2022)
- Terveyskirjasto. 2022. Koronavirus (SARS-CoV-2, COVID-19) <https://www.terveyskirjasto.fi/dlk01257> (luettu 20.11.2022)
- Tiwari A., Lipponen A., Hokajärvi A.-M., Luomala O., Sarekoski A., Rytönen A., Österlund P., Al-Hello H., Juutinen A., Miettinen I.T., Savolainen-Kopra C. & Pitkänen T. 2022. Detection and quantification of SARS-CoV-2 RNA in wastewater influent in relation to reported COVID-19 incidence in Finland. *Water Research* 215: 118220.
- Viana R., Moyo S., Amoako D.G., Tegally H., Scheepers C., Althaus C.L., Anyaneji U.J., Bester P.A., Boni M.F., Chand M., Choga W.T., Colquhoun R., Davids M., Deforche K., Doolabh D., Plessis L. du, Engelbrecht S., Everatt J., Giandhari J., Giovanetti M., Hardie D., Hill V., Hsiao N.Y., Iranzadeh A., Ismail A., Joseph C., Joseph R., Koopile L., Kosakovsky Pond S.L., Kraemer M.U.G., Kuate-Lere L., Laguda-Akingba O., Lesetedi-Mafoko O., Lessells R.J., Lockman S., Lucaci A.G., Maharaj A., Mahlangu B., Maponga T.,

- Mahlakwane K., Makatini Z., Marais G., Maruapula D., Masupu K., Matshaba M., Mayaphi S., Mbhele N., Mbulawa M.B., Mendes A., Mlisana K., Mnguni A., Mohale T., Moir M., Moruisi K., Mosepele M., Motsatsi G., Motswaledi M.S., Mphoyakgosi T., Msomi N., Mwangi P.N., Naidoo Y., Ntuli N., Nyaga M., Olubayo L., Pillay S., Radibe B., Ramphal Y., Ramphal U., San J.E., Scott L., Shapiro R., Singh L., Smith-Lawrence P., Stevens W., Strydom A., Subramoney K., Tebeila N., Tshiabuila D., Tsui J., Wyk S. van, Weaver S., Wibmer C.K., Wilkinson E., Wolter N., Zarebski A.E., Zuze B., Goedhals D., Preiser W., Treurnicht F., Venter M., Williamson C., Pybus O.G., Bhiman J., Glass A., Martin D.P., Rambaut A., Gaseitsiwe S., Gottberg A. von, et al. 2022. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 2022 603:7902 603: 679–686.
- Viveros M.L. 2022. Wild type and variants of SARS-COV-2 in Parisian sewage: presence in raw water and through processes in wastewater treatment plants. *Environmental science and pollution research international* 29: 67442–67449.
- Wang H., Omosa I.B., Keller A.A. & Li F. 2012. Ecosystem Protection, Integrated Management and Infrastructure are Vital for Improving Water Quality in Africa. *Environmental Science & Technology* 46: 4699–4700.
- Wang P., Casner R.G., Nair M.S., Wang M., Yu J., Cerutti G., Liu L., Kwong P.D., Huang Y., Shapiro L. & Ho D.D. 2021. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. *Cell Host & Microbe* 29: 747-751.e4.
- Wang Q., Guo Y., Iketani S., Nair M.S., Li Z., Mohri H., Wang M., Yu J., Bowen A.D., Chang J.Y., Shah J.G., Nguyen N., Chen Z., Meyers K., Yin M.T., Sobieszczyk M.E., Sheng Z., Huang Y., Liu L. & Ho D.D. 2022. Antibody evasion by SARS-CoV-2 Omicron subvariants BA.2.12.1, BA.4 and BA.5. *Nature* 2022 608:7923 608: 603–608.
- Weidhaas J., Aanderud Z.T., Roper D.K., VanDerslice J., Gaddis E.B., Ostermiller J., Hoffman K., Jamal R., Heck P., Zhang Y., Torgersen K., Laan J.V. & LaCross N. 2021. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Science of The Total Environment* 775: 145790.
- World Health Organization (WHO). Severe Acute Respiratory Syndrome (SARS) https://www.who.int/health-topics/severe-acute-respiratory-syndrome#tab=tab_1 (luettu 3.11.2022)
- World Health Organization (WHO). Timeline: WHO's COVID-19 response <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline#!> (luettu 3.11.2022)
- Xiao F., Tang M., Zheng X., Liu Y., Li X. & Shan H. 2020. Evidence for Gastrointestinal Infection of SARS-CoV-2. *Gastroenterology* 158: 1831-1833.e3.
- Zarza E. 2022. Monitoring SARS-CoV-2 in the Wastewater and Rivers of Tapachula, a Migratory Hub in Southern Mexico. *Food and environmental virology* 14: 199–211.

Zubir N.S.A., Abas M.A., Ismail N., Rahiman M.H.F., Tajuddin S.N. & Taib M.N.
2018. Statistical analysis of agarwood oil compounds in discriminating the
quality of agarwood oil. *Journal of Fundamental and Applied Sciences* 9: 45.

LIITE 1

Sovitetut virusmäärää kuvaavat mallit omaan dataan.

Malli	Fluor	Vuorokausivir- taama (m ³ /h)	Tuntivir- taama (m ³ /h)	Tapaus- määrät	Tehohoi- dossa	Fluore- sensi sa- tunnais- taso	AIC	BIC
v1	x	x	x	x	x		622,36	635,83
v2	x		x	x	x		620,36	632,33
v3	x	x		x	x		620,49	632,46
v4	x	x	x		x		620,36	632,33
v5	x	x	x	x			620,36	632,95
v6	x				x		617,3	626,28
v7	x			x			618,41	627,38
v8	x	x					624,46	633,44
v9	x		x				624,3	633,28
v10	x	x		x			620,31	630,79
v11	x	x			x		618,6	629,08
v12	x		x	x			619,8	630,28
v13	x		x		x		618,36	628,84
v14	x	x	x				626,24	636,71
v15	x			x	x		619,3	629,77
vv1	x	x	x	x	x	x	612,58	633,53
vv2	x		x	x	x	x	610,58	630,03
vv3	x	x		x	x	x	610,59	630,04
vv4	x	x	x		x	x	610,9	630,36
vv5	x	x	x	x		x	612,06	631,51
vv6	x				x	x	608,87	625,32
vv7	x			x		x	612,05	628,51
vv8	x	x				x	611,54	628
vv9	x		x			x	612,06	628,52
vv10	x	x		x		x	612,8	630,76
vv11	x	x			x	x	609,59	627,55
vv12	x		x	x		x	614,05	632,01
vv13	x		x		x	x	610,62	628,58
vv14	x	x	x			x	611,91	629,86
vv15	x			x	x	x	608,59	626,55
ve1		x	x	x	x		635,93	646,4

ve2		x	x	x	634,15	643,12
ve3	x		x	x	633,96	642,94
ve4	x	x		x	634,06	643,04
ve5	x	x	x		634,95	643,93
ve6				x	630,73	636,82
ve7			x		631,48	637,46
ve8	x				636,36	642,34
ve9		x			636,52	642,5
ve10	x		x		633,28	640,77
ve11	x			x	632,07	639,55
ve12		x	x		633,06	640,54
ve13		x		x	632,15	639,63
ve14	x	x			638,34	645,83
ve15			x	x	632,73	640,21

Alla olevassa taulukossa pelkästään THL:n dataa mallintavat mallit ja niiden AIC- ja BIC-arvot. Satunnaisefektinä vuodenaika kaikissa malleissa.

Malli	Vuorokausivirtaama (m ³ /h)	Testausmäärät	Tapausmäärät	Tehohoidossa (kpl)	AIC	BIC
1	x	x	x	x	1289,42	1302,06
2	x	x		x	1290,62	1301,46
3	x		x	x	1290,02	1300,86
4		x	x	x	1287,65	1298,49
5	x	x	x		1288,23	1299,07
6	x	x			1289,20	1298,23
7	x		x		1289,30	1298,33
8		x	x		1287,03	1296,06
9			x	x	1288,12	1297,15
10		x	x	x	1288,82	1297,85
11	x			x	1288,62	1297,66

Alla olevassa taulukossa on esitetty kaikki data, jota käytettiin mallintamaan THL:n itse tuottamia virus-RNA:n määriä. Muuttujat:

- vko: viikko, jolloin näyte otettiin
- season: vuodenaika, jolloin näyte otettiin
- tapaus: koko Suomen tapaukset näytteenottamisen päivän
- testaus: COVID-19-PCR-testit, jotka tehtiin Keski-Suomen sairaanhoitopiirin alueella näytteenottamisen viikolla
- rna: vastemuuttuja
- populaatio: jätevedenpuhdistamon piirissä olevien ihmisten määrä
- virtaama: jätevedenpuhdistamon kokonaisvirtaama (m³/d)

- tehokpo+2: Kuopion yliopistollisen sairaalan eritysvastuualueella olevat COVID-19 takia tai kanssa olevat tehohoitopotilaat näytteenottamisen päivänä ja keinotekoinen kahden viikon viive
- tehokpo: Kuopion yliopistollisen sairaalan eritysvastuualueella olevat COVID-19-positiiviset tehohoitopotilaat näytteenottamisen päivänä
- tehohoidossa.olevat: koko Suomessa COVID-19-positiiviset tehohoitopotilaat näytteenottamisen päivänä

tehohoi- dossa.ole- vat	53	60	62	50	50	39	38	35	31	35	43
tehohoit- o- lag+2	62	58	53	60	62	50	50	39	38	35	31
tehokpo	6	6	10	8	3	5	12	4	4	8	12
tehokpo+2	10	5	6	6	10	8	3	5	12	4	4
virtaama	30668	31739	33371	31294	31997	30377	30026	32821	31783	31701	32927
populaatio	154600	154600	154600	154600	154600	154600	154600	154600	154600	154600	154600
rma	373015	247100	377550	900737	254813	610916	238961	854185	940785	614934	634345
testaus	5309	5864	5387	4457	4370	4627	4084	3512	3300	3635	3503
tapaus	802	1145	1041	1234	1244	1150	936	929	837	1002	1033
season	talvi	talvi	talvi	talvi	talvi	talvi	talvi	talvi	kevät	kevät	kevät
vko	1	2	3	4	5	6	7	8	9	10	11

45	42	43	31	33	26	26	23	20	22	16	13
35	43	45	42	43	31	33	26	26	23	20	22
12	14	9	7	6	5	4	4	5	4	2	3
8	12	12	14	9	7	6	5	4	4	5	4
38930	42530	37987	42156	57356	69595	47731	46839	42263	35594	34322	34326
154600	154600	154600	154600	154600	154600	154600	154600	154600	154600	154600	154600
599421	562345	1053722	177099	276103	397916	298634	599879	215692	307552	246985	407300
3414	3285	2549	2205	2028	2166	2035	2088	1331	981	945	661
1011	915	777	574	582	480	452	362	282	228	199	226
kevät	kevät	kevät	kevät	kevät	kevät	kevät	kevät	kevät	kevät	kesä	kesä
12	13	14	15	16	17	18	19	20	21	22	23

9	13	11	15	8	13	14	13	14	8	15	16
16	13	9	13	11	15	8	13	14	13	14	8
2	2	0	1	0	0	2	3	3	4	7	6
2	3	2	2	0	1	0	0	2	3	3	4
31734	37561	29821	29302	29257	36341	33144	31157	29934	29177	30008	31391
154600	154600	154600	154600	154600	154600	154600	154600	154600	154600	154600	154600
436193	537815	459548	654740	720449	1129345	739298	1090485	675378	311624	341503	127050
967	790	923	839	1063	1450	1432	1363	1271	1122	1097	1176
296	233	292	298	392	475	437	374	309	252	167	142
kesä	kesä	kesä	kesä	kesä	kesä	kesä	kesä	kesä	kesä	kesä	syksy
24	25	26	27	28	29	30	31	32	33	34	35

14	10	10	11	10	12	22	12	15	10
15	16	14	10	10	11	10	12	22	12
4	4	5	3	3	5	5	2	2	2
7	6	4	4	5	3	3	5	5	2
30130	29803	32922	32491	31940	32857	36209	34272	34063	33740
154600	154600	154600	154600	154600	154600	154600	154600	154600	154600
202530	140419	147536	964546	1585441	797199	1792346	603074	762170	366716
1125	981	997	1142	1336	1565	1453	1375	1227	1160
163	189	203	252	294	353	289	252	211	212
syksy	syksy	syksy	syksy	syksy	syksy	syksy	syksy	syksy	syksy
36	37	38	39	40	41	42	43	44	45

Alla olevassa taulukossa on tämän tutkimuksen tuottama data. Datan sisältämät muuttujat:

- id: tunnistenumero mittaukselle
- Fluor: mitattu fluoresenssi qPCR-osoituksessa
- Virusta PCR-putkessa: valmistajan tietojen mukaan määritetty virusten määrä PCR-putkessa
- virtaus: virtaamakorjatun RNA:n määrä yksikössä miljoonaa virus-RNA-kopiota/1000 henkilöä/vuorokausi
- virtaamakorjaus39: virtaamakorjatun RNA:n määrä yksikössä miljoonaa virus-RNA-kopiota/1000 henkilöä/vuorokausi, kun Cq-arvo 39 vastaa yhtä virusta
- virtaamakorjaus: virtaamakorjatun RNA:n määrä jaettuna, kuinka monta tuhatta asiakasta jätevedenpuhdistamon alueella on
- virus_paiva39: montako viruskopiota per kuutiojäteveettä saapuu päivässä jätevedenpuhdistamolle, kun Cq-arvo 39 vastaa yhtä virusta
- virus_paiva: montako viruskopiota per kuutio jäteveettä saapuu päivässä jätevedenpuhdistamolle
- virus/m3 39: montako viruskopiota per kuutio jäteveettä, kun Cq-arvo 39 vastaa yhtä virusta
- virus /m3: montako viruskopiota per kuutio jäteveettä
- virus/ml39: montako viruskopiota per millilitra jäteveettä, kun Cq-arvo 39 vastaa yhtä virusta
- virus/ml: montako viruskopiota per millilitra jäteveettä
- virusta_eluutio cq 39: montako viruskopiota oli qPCR-osoituksen eluutiossa, kun Cq-arvo 39 vastaa yhtä virusta
- virusta_eluutiossa: montako viruskopiota oli qPCR-osoituksen eluutiossa
- Cq: mitattu Cq-arvo qPCR-osoituksessa
- nayte: mitattu näyte
- Vrk.virtaama: jätevedenpuhdistamon vuorokausivirtaama
- tehokpo: Kuopion yliopistollisen sairaalan erityisvastuualueella olevat COVID-19 kanssa tai takia tehohoitopotilaat näytteenottamisen päivänä
- tapausjkl: virallisesti todetut COVID-19-tapaukset näytteenottamisen viikolla
- Testausmäärä: tehdyt COVID-19-PCR-testit jätevesinäytteenottamisen viikolla Keski-Suomen sairaanhoitopiirin alueella
- Tuntivirtaama.m3.h: jätevedenpuhdistamon virtaama tuntina, jolloin näyte otettiin
- thl.virtaamakorjattu: THL:n virtaamakorjatun RNA:n määrä jätevesinäytteenottamisen viikolla

thl virtaamakorjattu	64345	177099	177099	276103	397916	298634	599876	599879	599879	599879
Tuntivirtaama.m3.h	1890	2306	2306	3274	2688	2477	2326	2326	2326	2326
Testausmäärä	3503	2205	2205	2028	2166	2035	2088	2088	2088	2088
tapausjkl	1033	574	574	582	480	362	282	282	282	282
tehokpo	11	6	6	6	7	5	5	5	5	5
Vrk-virtaama	35637	56360	56360	73218	55184	48447	46139	46139	46139	36044
nayte	1	2A	2B	3A	4A	5A	6A	6B	6C	7
Cq	42,43846	36,015454	36,562867	36,88392	35,034464	35,058899	35,381399	36,379377	35,193583	31,240625
Virusta eluutiossa	5,065058864434,2015494297,1246569237,8551251856,91119352842,5237213673,7837874337,4138258767,44398511878,26687									
virusta_eluutio cq39	0,6753412	57,89354	39,616621	31,714017	114,25492	112,3365	89,837838	44,98851	102,32586	1583,7689
Virusta/ml	16,88353	1447,3385	990,41552	792,85042	2856,3731	2808,4124	2245,946	1124,7128	2558,1466	791,88446
virus/ml39	2,2511373	192,97847	132,0554	105,71339	380,84975	374,45499	299,45946	149,9617	341,08622	105,58459
virus/m3	16883530	1,447E+09	990415523	792850417	2,856E+09	2,808E+09	2,246E+09	1,125E+09	2,558E+09	791884458
virus/m3_39	2251137,3	192978466	132055403	105713389	380849749	374454987	299459461	149961700	341086216	105584594
virus_paiva	6,017E+11	8,157E+13	5,582E+13	5,805E+13	1,576E+14	1,361E+14	1,036E+14	5,189E+13	1,18E+14	2,854E+13
virus_paiva39	8,022E+10	1,088E+13	7,443E+12	7,74E+12	2,102E+13	1,814E+13	1,382E+13	6,919E+12	1,574E+13	3,806E+12
virtaamakorjaus	3,892E+09	5,276E+11	3,611E+11	3,755E+11	1,02E+12	8,801E+11	6,703E+11	3,357E+11	7,635E+11	1,846E+11
virtaamakorjaus39	518,9119	70351,011	48141,284	50065,478	135943,16	117342,95	89371,022	44754,741	101794,16	24616,372
virtaus	3891,8392	527632,59	361059,63	375491,09	1019573,7	880072,16	670282,67	335660,55	763456,19	184622,79
virtus	16,88353	1447,3385	990,41552	792,85042	2856,3731	2808,4124	2245,946	1124,7128	2558,1466	791,88446
Virusta PCR-putkessa	1	64	32	32	128	128	128	32	128	54,613333
Fluor	Cy5	Cy5	Cy5	Cy5	Cy5	Cy5	Cy5	Cy5	Cy5	Cy5
id	1	2	3	4	5	6	7	8	9	10

246985	64345	177099	177099	276103	397916	298634	599879	599879	599879	599879	246985
1856	1890	2306	2306	3274	2688	2477	2326	2326	2326	2326	1856
945	3503	2205	2205	2028	2166	2035	2088	2088	2088	2088	945
199	1033	574	574	582	480	362	282	282	282	282	199
3	11	6	6	6	7	5	5	5	5	5	3
36044	35637	56360	56360	73218	55184	48447	46139	46139	46139	46139	36044
8	1	2A	2B	3A	4A	5A	6A	6B	6C	7	8
37,394198	32,776323	34,550536	34,266113	34,29866	34,527052	34,168404	34,776938	35,098947	34,622546	32,466335	31,318268
167,00759594097,887347	1198,345598	1459,431116	1426,882003	1218,007732	1561,675521	1024,336626	819,4624131	1140,01195	5079,908699	11256,02519	
22,267679	546,38498	159,77941	194,59082	190,25093	162,40103	208,2234	136,57822	109,26166	152,00159	677,32116	1500,8034
11,13384	13659,624	3994,4853	4864,7704	4756,2733	4060,0258	5205,5851	3414,4554	2731,5414	3800,0398	338,66058	750,40168
1,484512	1821,2833	532,59804	648,63605	634,16978	541,33677	694,07801	455,26072	364,20552	506,67198	45,154744	100,05356
11133840	1,366E+10	3,994E+09	4,865E+09	4,756E+09	4,06E+09	5,206E+09	3,414E+09	2,732E+09	3,8E+09	338660580	750401680
1484512	1,821E+09	532598044	648636052	634169779	541336770	694078009	455260723	364205517	506671978	45154744	100053557
4,013E+11	4,868E+14	2,251E+14	2,742E+14	3,482E+14	2,24E+14	2,522E+14	1,575E+14	1,26E+14	1,753E+14	1,563E+13	2,705E+13
5,351E+10	6,491E+13	3,002E+13	3,656E+13	4,643E+13	2,987E+13	3,363E+13	2,101E+13	1,68E+13	2,338E+13	2,083E+12	3,606E+12
2,596E+09	3,149E+12	1,456E+12	1,773E+12	2,253E+12	1,449E+12	1,631E+12	1,019E+12	8,152E+11	1,134E+12	1,011E+11	1,75E+11
346,10446	419825,82	194160,58	236462,66	300340,51	193228,51	217503,22	135868,53	108693,91	151211,76	13476,033	23326,846
2595,7834	3148693,6	1456204,4	1773470	2252553,8	1449213,9	1631274,1	1019014	815204,32	1134088,2	101070,25	174951,35
11,13384	13659,624	3994,4853	4864,7704	4756,2733	4060,0258	5205,5851	3414,4554	2731,5414	3800,0398	338,66058	750,40168
0,8533333	1024	128	256	256	128	256	128	128	128	81,92	163,84
Cy5	FAM	FAM	FAM	FAM	FAM	FAM	FAM	FAM	FAM	FAM	FAM
11	12	13	14	15	16	17	18	19	20	21	22

LIITE 2

Tuotetut Sanger-sekvenssit. Sekvenssit eivät ole yhtenäisiä, mutta ne on yhdistetty yhtenäiseksi, sillä tyhjät alueet poistettiin linjauksen vuoksi. Näytesekvenssit ovat SARS-CoV-2-piikkiproteiinin alueelta.

Näyte 1

```
ATGGTTATGTCATGCATGCAAATTACATATTTTGGAGGAATACAAATCCAATTCAGTTGTCTTCCTA
TTCTTTATTTGACATGAGTAAATTTCCCTTAAATTAAGGGGTTACTGCTGTTATGCTTTAAAATTTA
TCTCTTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGIGATGTT
CTTGTTAACTAAACGAACAATGTTTGTGTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGIG
TTAATCTTACAACCAGAACTCAATTACCCCTGCATACACTAATTCTTCGCACGTAGTGTAGCTA
GTCAATCACAAGTGCCTTTTTGGTGAAGTTTTTAAACGCCACCAGATTTTGCATCTGTTTTATGCTT
GGAACAGGAAGAGAATCAGCAACTGTGTGCTGATTATTCTGTCCTATATAATTCCGCATCATTTT
CCACTTTAAGTGTTATGGAGTGTCTCTACTAAATTAATGATCTCTGCTTTATAATGCTATGCA
GATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAAATCGCTCCAGGGCAAACCTGGAAAGATTGC
TGATATAATTATAAATTACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAAATTCTAACAACTTG
ATTTAAGGTGGTGGTAATTATAATTCCGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAG
AGAGATATTTCAACTGAAATCTATCAGGCCGGTAGCACCTGTAAATGGTGTGAAGGTTTTAATTGT
TACTTTCTTTACAATCATATGGTTTTCCAACCCACTAATGGTGTGGTTACCAACCATACAGAGTAG
TAGTACTTTCTTTGAACTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTACTAATTT
GGTAAAAACAATGTGTCAATTTCAACTTCAATGGTTAACAGGCACAGGTGTTCTTACTGAGTC
TAACAAAAAGTTTCTGCCTTTCCAACAATTTGGCAGAGACATTGCTGACACTACTGATGCTGTCCG
TGATCCACAGACACTTGAGATTCTTGACATTACCCATGTTCTTTTGGTGGTGTGAGTGTATAACA
CCAGGAACAAATACTTCTAACCCAGGTGCTGTTCTTTATCAGGGTGTAACTGCACAGAAGTCCCT
GTTGCTATTCATGCAGATCAACTTACTCCTACTTGGCGTGTATTCTACAGGTTCTAATGTTTTTCA
AACACGTGCAGGCTGTTAATAGGGGCTGAACATGTCAACAACATCATATGAGTGTGACATACCCT
TGGTGCAGGTATATGCGCTAGTTATCAGACTCAGACTAATCTCTTTGCACAAGTCAAACAAATT
TACAAAACACCACCATTAAAGATTTTGGTGGTTTTAATTTTTACAAAATATTACCAGATCCTCAA
ACCAAGAGGTCATTTATTGAAGATCTACTTTTCAACAAAGTGACACTTGCAGATGCTGGCTTCATC
AAACAATATGGTATTGCCTTGGTGTATTTGCTGCTAGAGACCTCATTGTGCACAAAAGTTTTAAC
GGCCTTACTGTTTTGCCACCTTTGCTCACAGATGAAATGATTGCTCAATACACTTCTGCACTGTTAG
CGGGTACAATCACTTCTGGTGGACCTTTGGTGCAGGTGCTCCGTGCTTTAACGGWATAGCTGTTG
ACAAGACAAAAACCCCAAGAAGTTTTGCACAAGTCAAACAAATTTACAAAACACCACCAAT
TAAAGATTTTGGTGGTTAATTTACAAA
```

Näyte 2

```
AGATGGTTATGTCATGCATGCAAATTACATATTTTGGAGGAATACAAATCCAATTCAGTTGTCTTC
CTATTCCTTATTTGACATGAGTAAATTTCCCTTAAATTAAGGGGTTACTGCTGTTATGCTTTAAA
GAAGGGTTTTATCTCTTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTT
CTAGTGATGTTCTTGTTAACTAAACGAACAATGTTTGTGTTTTCTTGTTTTATTGCCACTAGTCTC
TAGTCAGTGTGTTAATCTTACAACCAGAACTCAATTACCCCTGCATACACTAATTCTTCGCACGT
AGTGTAGCTAGTCAATCATGCTCACAAAATCTTAGGGAGATTTTGGCGTTCAGGAATATATGATC
GGTTATTATAAAATTTTCTAAGCACACAGGCCTATAATTTAGGCGGATTTCTTCAAGGATTTTGG
CTTAGAGACCATTTGGTAGATTTGCCAATAGGTATTACATCACTAGGTTTCAAACCTTACTCTTACAT
AGAAGTACTGACTTCTGGGATTTCTTTCAGGGTGGGACAGCTGTTGCAGCTTATTATGGGTTATC
TTCAACCTGGACTTTTCTATTAATAATATAATGAAAATGGAACCATTACAGATGCTGTAGAAGTGT
GCACTTGACCCTCTCTCAGAAAACAAGTGTACGTTTGAATCCTTCACTGTAGAAAAAGGAATCT
ATCAAAACTTCTAACTTTTAGAGTCCAACCAACAGAATCTATTGTTAGATTTCTAATATTACAAA
CTTGTCCTTTTGGTGAAGTTTTTAAACGCCACCAGATTTGCATCTGTTATGCTTGGAAACAGGAAG
AGAATCAGCAACTGTGTGCTGATTATTCTGTCCTATATAATTCCGCATCATTTTCCACTTTTAAAGT
GTTATGGAGTGTCTCTACTAAATTAATGATCTCTGCTTACTAATGTCTATGCAGATTCATTTGT
AATTAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAAACCTGGAAAGATTGCTGATTATAATT
ATAAATTACCAGATGATTTTACAGGCTGCGTTATAGCTTGGAAATCTAACAACTTGTATTCTAAGG
```

TTGGTGGTAATTATAATTACCTGTATAGATTGTTTAGGAAGTCTAATCTCAAACCTTTTGAGAGAG
ATATTTCAACTGAAATCTATCAGGCCGGTAGCACACCTTGTAAATGGTGTGAAGGTTTTAAATTGTT
ACTTTCCTTTACAATCATATGGTTTCCAACCCACTATGGTGTGGTTACCAACCATAACAGAGTAGTA
GTACTTTCCTTTGAACCTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAAGTCTACTAATTTGG
TTAAAAACAAATGTGTCAATTTCAACTTCAATGGTTTAAACAGGCACAGGTGTTCTTACTGAGTCTA
ACAAAAAGTTTTCTGCCTTTCCAACAATTTGGCAGAGACATTGATGACACTACTGATGCTGTCCGTG
ATCCACAGACACTTGAGATTCTTGACATTACCCATGTTCTTTTGGTGGTGTGAGTGTATAACACC
AGGAACAAATACTTCTAACCAGGTGCTGTTCTTTATCAGGGTGTAACTGCACAGAAGTCCCTGT
TGCTATTCATGCAGATCAACTTACTCCTACTTGGCGTGTATTCTACAGGTTCTAATGTTTTTCAA
CACGTGCAGGCTGTTAATAGGGGCTGAACATGTCAACAACACTCATATGAGTGTGACATACCCATT
GGTGCAGGTATATGCGCTAGTTATCAGACTCAGACTAATTCATCCGTAGTGTAGAACGGAATA
GTGTGAACAAGACAAAAACCCCAAGAAGTTTTTGCCGTGCTTAACTGGAATAGCTGTGAACA
AGACAAAAACCCCAAGAAGTTTTTGACAAAGTCAAAACAAATTTACAAAAACACCACCAATTTAA
GATTTTGGTGGTTTTAAATTTTTCACAAATATTACCAGATCCATCAAACCAAGCAAGAGGTCATTT
ATTGAAGATCTACTTTTCAACAAAGTGACACTTGCAGATGCTGGCTTTCATCAAACAATATGGTGT
TGCTTGGTGATATTGCTGCTAGAGACCTCATTTGTGCACAAAAGTTAACGGCCTTACTGTTTTGC
CACCTTGTCTACAGATGAAATGATTGCTCAATACACTTCTGCACGTTAGCGGGTACAATCACTT
CTGGTGGACCTTTGGTGCAGGTGCTGCATTACAAATACCATTTGCTATGCAAATGGCTTATAGGTT
TAATGGTATTGGAGTTACACAGAATGTTCTCTATGAGAACCAAAAATTGATTGCCAACCAATTTAA
TAGTGCTATTGGCAAAAATTCAAGACTCACTTCTTCCACAGCAAGTGCACCTTGGAAAACCTTCAAGA
TGTTGTCACCAAAAATGCACAAGCTTTAAACACGCTTGTAAACAACCTTAGCTCCAATTTTGGTGC
AATTTCAAGTGTTTTAAATGATATCCTTGCACGTCTTGACAAAGTTGAGGCTGAAGTGCAAATTGA
TAGGTTGATCACAGGCAGACTTCAAAGTTTGCAGACATATGTGACTCAACAATTAATTAGAGCTG
CAGAAATCAGAGCTTCTGCTAATCTTGTGCTACTAAAATGTCAGAGTGTACTTGGACAATCAA
AAAGAGTTGATTTTTGTGAAAGGGCTATCATCTTATGTCCTTCCCTCAGTCAGCACCTCATGGTGT
AGTCTTCTGCAATGTGACTTATGTCCTGCACAAGAAAAGAACTGTCCTGCCATTTGTCATGATG
GAAAAGCACACTTTCCTCGTGAAGGTGTCTTTGTTTCAAATGGCACACACTGGTTTTGTAACACAAA
GGAATTTTTATGAACCACAAATCATTACTACACACAACACATTTGTGTCTGGTAACTGTGATGTTG
TAATAGGAATTGTCAACAACACAGTTTATGATCCTTTGCAACCTGAATTAGACTCATTCAAGGAGG
AGTTAGATAAATATTTAAGAATCATAACATCACCAGATGTTGATTTAGGTGACATCTCTGGCATT
ATGCTTCAAGTGTAAACATTCAAAAAGAAATTGACCGCTCAATGAGGTTGCCAAGAATTTAAAT
GAATCTCTCATCGATCTCCAAGAACTTGGAAAGTATGAGCAGTATATAAAATGGCCATGGTACAT
TTGGCTAGGTTTTATAGCTGGCTGATTGCCATAGTAATGGTGACAATTATGCTTGTGATGACC
AGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTTGTGGATCCTGCTGCAAATTTGATGAAGACGACT
CTGAGCCAGTGCTCAAAGGAGTCAAATTACATTACACATAAACGAACCTTATGGATTTGTTTATGA
GAATCTTCACAATTGGAACGTGAACCTTTGAAGCAAGGIGAATCAAGGATGCTA

Näyte 3

TAGATGGTTATGTCATGCATGCAAATTACATATTTTGGAGGAATACAAATCCAATTCAGTTGTCTT
CCTATTCCTTATTTGACATGAGTAAATTTCCCCTTAAATTAAGGGGTAAGTGTGTTATGCTTTTAAA
AGAAGTTATCTCTTCTTAGTAAAGGTAGACTTATAAATTAGAGAAAACAACAGAGTTGTTATTTCTA
GTGATGTTCTTGTAAACAATAACGAACAATGTTGTTTTTCTTGTTTTTATGCCACTAGCTCTAG
TCAGTGTGTTAATCTTAGAACCCAGAACTCAATTACCCCTGCATACACTAATTTCTTTCGCAGTAT
GTTAGCTAGTCAATACGTTGAGATCCTCACGTGTAATAAAGGAATCTATCAACTTCTAACTTTAGAGT
CCAACCAACAGAATCTATTGTTAATTTCCATAATTACAAACTTGTGCCCTTTGGTGAAGTTTTTA
ACGCCACCAATTTGCATCTGTTTATGCTTGGAACAGGAAGAACAGGGTTGCTGATTATCTGTCTCT
ATATAATCCCATCATTTTCCACTTTTAAAGTGTATGGAGTGTCTCTACTAAATTAATGATCTCT
GCTTACTAATGTCTATGCAGATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAG
GGCAAACCTGGAAAGATTGCTGATTATAAATTATAAATTACCAGATGATTTTACAGGCTGCGTTATAG
CTTGGAAATCTAACAATCTGATTCTAAGGTTGGTGGTAATTATAATTACCGGTATAATTGTTTAGG
AAGTCTAATCTCAAACCTTTTGAGAGAATATTTCAACTGAAATCTATCAGGCCGGTAGCAAACCTGT
ATGGGTTAAGGTTTTAATTGTTACTTTCTTTACAATCATATGGTTTCCAACCCACTAATGGTGTG
GTTACCAACCATAACAGAGTAGTAGTACTTTCTTTGAACTTCTACATGCACCAGCAACTGTTTGTG
GACCTATAGTCTACTAATTTGGTAAAAACAAATGTGTCAATTTCAACTTCAATGGTTTAAACAGGC
ACAGGTGTTCTTACTGAGTCTAACAAAAAGTTTCTGCTTTCCAACAATTTGGCAGAGACATTTCT
GACACTACTGATGTTCCGTGATCCACAGACACTTGAGATTCTTGACATTACACCATGTTCTTTTGT
GTGGTGTGAGTGTATAACACCAGGAACAAATACTTCTAACCAGGTTGCTGTTCTTTATCAGGGTG
TTAACTGCACAGAAGTCCCTGTTGCTATTATGCAGATCAACTTACTCCTACTTGGCGTGTATTTC
TACAGGTTCTAATGTTTTTCAAACACGTGCAGGCTGTTAATAGGGGCTGAACATGTCAACAACCT
ATATGAGTGTGACATACCCATTGGTGCAGGTATATGCGCTAGTTATCAGACTCAGACTAATTCTCG

TCGGACGGGCACGTAGTGAGCTAGTCAATCCATCATTGCCTACACTATGTCACTTGGTGCAGAAA
ATTGAGTTGCTTACTCTAATAACTCTATTGCCATACCCACAAATTTTACTATTAGTGTTACCACAGA
AATTCTACCAGTGTCTATGACCAAGACATCAGTAGATTGTACAATGTACATTTGTGGTGATTCAAC
TGAATGCAGCAATCTTTTGTGCAATATGGCAGTTTTTGTACACAATTAACCGTGCITTAACCTGGA
ATAGCTGTTGAACAAGACAAAAACCCCAAGAAGTTTTTGCACAAGTCAAAACAAATTTACAAAA
CACCACCAATTAAGATTTTGGTGGTTTTAATTTTTACAAATATTACCAGATCCATCAAAACCAA
GCAAGGTCATTTATTGAAGATCTACTTTTCAACAAAGTGACACTTGCAGATGCTGGCTTCATCAA
CAATATGGTGATTGCCTTGGTGATATTGCTGCTAGAGACCTCATTGTGCACAAAAGTTTAAACGGC
CTTACTGTTTTGCCACTTTGCTCACAGATGAAATGATGTCTCAATACACTTCTGCACCTGTAGCGG
GTACAATCACTTCTGGTTGGACCTTTGGTGCAGGTGCTGCATTACAAATACCATTGTCTATGCAA
TGGCTTATAGGTTAATGGTATGGAGTTACACAGAATGTTCTCTATGAGAACCACAAAATTTGATTG
CCAACCAATTTAATAGTGCTATTGGCAAAAATTCAAGACTCACTTTCTCCACAGCAACAGCAAGTG
CACTTGGAAAACTTCAAAATGTGGTCAACCAAAAATGCACAAGCTTTAAACACGCTTGTAAACAA
CTTAGCTCCAATTTTGGTGCAATTTCAAGTGTTTAATGATATCCTTTCACGCTTGACAAAAGTTG
AGGCTGAAGTGCAAATTTGATAGGTTGATCACAGGCAGACTTCAAAGTTTGCAGACATATGTGACT
CAACAATTAATTAGAGCTGCAGAAATCAGAGCTTCTGCTTATCTGCTGCTACTAAAATGTCAGAG
TGTGACTTGGACAATCAAAAAGAGTTGATTTTTGTGGAAAGGGCTATCATCTTATGTCCTTCCCTC
AGTCAGCACCTCATGGTGTAGTCTTCTGCTGCTGACTTATGTCCTGCACAAGAAAAGAAGTTCA
CAACTGCTCCTGCCATTTGTCATGATGGAAAAGCACACTTTCCTCGTGAAGGTGTCTTTGTTTCAA
TGGCACACACTGGTTTGTAAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACA
CATTGTGCTGGTAACTGTGATGTTGTAATAGGAATGTCAACAACACAGTTTATGATCCTTTGCA
ACCTGAATTAGACTCATTCAAGGAGGAGTTAGATAAATATTTAAGAATCATAACATCACCAGATG
TTGATTTAGGTGACATCTCTGGCATTAAATGCTTCAGTTGTAAACATTCAAAAAGAAATGACCGCC
TCAATGAGGTTGCCAAGAATTTAAATGAATCTCTCATCGATCTCCAAGAAGTTGAAAAGTATGAG
CAGTATATAAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTTATTGCCATAGTAATG
GTGACAATTATGCTTTGCTGTATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTTGTGGAT
CCTGCTGCAAATTTGATGAAGACGACTCTGAGCCAGTGCTCAAGGAGTCAATTACATTTACACAT
AAACGAAGTTATGGATTTGTTTATGAGAATCTTCAATTGGAAGTGAAGTTTGAAGCAAGGTGAAC
AAGGATTTAAC

Näyte 4

GGTTATGTCATGCATGCAAATTACATATTTTTTGGAGGATACAAATCCATTGAGTTGTCTTCTTCTATT
CTTTATTTGACATGAGTAAATTTCCCTTAAATTAAGGGGACTGCTGTTATGCTTTAAAAGAAGG
CAAATATCTCTTCTTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGT
GATGTTCTTGTTAACAATAAAGCAACAATGTTTGTGTTTTCTTGTTTTATTGCCACTAGTCTCTAGTC
AGTGTGTTAATCTTACAACCAGAAGTCAATTACCCCTGCATACACTAATTTTTCGCACGTAGTG
TATCAATCCACGTTTATGCTTGGAACAGGAAGAGGATCTCCAGAGTGTGCTGATTATTCTGCTCT
ATACATCTCGCACCATTTTTCTTTTAAATGTTATGGAGTGCTCCTACTAAATTAATGATCTCTGC
TTTACTATGCTATGCGATTCTTTGTAATTAGAGGTGATGAAGTCAGACAATCCTCCGGGCAAACCTG
GAAATATTGCTGATTATAATTATAAATTACCAATGATTTTACAGGCTGCGTTATAGTTGGAATTTGGT
GTTGCAGGTTTTAATTGTTACTTTCTTTACGATCATATAGTTTCCGACCCACTTATGGTGTGGTCA
CCAACCATACAGATAGTAGTACTTTCTTTGAACTTCTACATGCACCAGCAATCTACTAATTTGG
TTAAAAACAAATGTGTCAATTTCAACTTCAATTGGTTTTAAAGGCACAGGTGTTCTTACTGAGTCTA
ACAAAAAGTTTCTGCCTTTCCAACAATTTGGCAGAGACATTGCTGACACTACTGATGCTGTCCGTG
ATCCACAGACACTTGGAGATTCTTGACATTACCCATGTTCTTTTGGTGGTGTGAGTGTATAACACC
AGGAACAAATACTTCTAACCAGGTGCTGTTCTTTATCAGGGTGTAACTGCACAGAAGTCCCTGT
TGCTATTCATGCAGATCAACTTACTCCTACTTGGCGTGTATTCTACAGGTTCTAATGTTTTTCAA
CACGTGCAGGCTGTTAATAGGGGCTGAATATGTCAACAACCTCATATGAGTGTGACATACCCATTG
GTGCAGGTATATGCGCTAGTTATCAGACTCAGACTAAGTCTCATCGGCGGCACGTAGTGTGTTGAA
CAAGACAAAACACCCAAGAAGTTTTTGCACAAGTCAAAACAAATTTACAAAACACCACCAATTA
AATATTTTGGTGGTTTTAATTTTTTACAAATATTACCAGATCCATCAAAACCAAGCAAGGTGAAAT
CAAGATGACTTTTCAACAAAGTGACACTTGCAGATGCTGGCTTCATCAAAACAAATATGGTGATTGCC
TTGGTGATATTGCTGCTAGAGACCTCATTGTGCACAAAAGTTTAAAGGCCTTACTGTTTTGCCACC
TTTGTCTCACAGATGAAATGATTGCTCAATACACTTCTGCACCTGTTAGCGGGTACAATCACTTCTGG
TTGGACCTTTGGTGCAGGTGCTGCATTACAAATACCATTGCTATGCAAATGGCTTATAGGTTAAT
GGTATTGGAGTTACACAGAATGTTCTCTATGAGAACCACAAAATTTGATTGCCAACCAATTTAATAGT
GCTATTGGCAAAAATCAAGACTCACTTCTTCCACAGCAAGGTGAAATCAAGGATGCTAATATCTT
TTCACGTCTTGACAAAAGTTGAGGCTGAAGTGCAAATTTGATAGGTTGATCACAGGCAGACTTCAA
GTTTTGCAGACATATGTGACTCAACAATTAATTAGAGCTGCAGAAATCAGAGCTTCTGGTATKTTGC
TGCTACTAAAATGTCAGAGTGTGTACTTGGACAATCAAAAAGAGTTGATTTTTGTGGAAAGGGCT

ATCATCTTATGTCCTTCCCTCAGTCAGCACCTCATGGGTAGTCTTCTTGCATGTGACTTATGTCCT
GCACAAGAAAAGAAGTTCACAAGTCTCTGCCATTTGTCATGATGGAAAAGCACACTTTCCTCG
TGAAGGTGTCTTTGTTTCAATGGCACACACTGGTTTGTAAACACAAGGAATTTTTATGAACCACAAA
TCATTAACAGACAACACATTTGTGTCTGGTAAGTGTGATGTTGTAATAGGAATTGTCAACAACA
CAGTTTATGATCCTTGAACCTGAATTAGATTCATTCAAGGAGGAGTTAGATAAATATTTAAGAA
TCATACATCACCAGATGTTGATTTAGGTGACATCTCTGGCATTAAATGCTTCAGTTGTAAACATTCA
AAAAGAAATTGACCGCCTCAATGAGGTTGCCAAGAATTTAAATGAATCTCTCATCGATCTCCAAG
AACTTGGAAAGTATGAGCAGTATATAAAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGC
TTTGATTGCCATAGTAATGGTGACAATTATGCTTTTGTGTATGACCAGTTGCTGTAGTTGTCTCAA
GGGCTGTTGTTTCTTGTGGATCCTGCTGCAAATTTTGTGATGAAGACGACTCTGGACCAGTGTCTAAA
GGAGTCAAATTCATTACACCATAAAACAACCTTATGGATTGTTTATGAGAATCTTCAATTGAACTGT
AACTTTGAACAAAGGTGGAAAAGGA

Näyte 5

ATTACATATTTGGAGGATACAAATCCATTCAGTTGTCTTCTTATTCTTTATTGACATGAGTAAAT
TTCCCTTAAATTAAGGGGTACTGCTGTTATGTCTTTAAAAGAAGGCAATCAATGAATGAATTTAT
CTCTTCTAGTAAAGGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGTGATGTTT
TTGTTAAACAATAAACGAACAATGTTTGTCTTTTCTTGTCTTTATTGCCACTAGTCTCTAGTCAGTGT
TAATCTTATAACCAGAACTCAATCATACTAATTCTTTCGCACGTAGTGTAGCTAGTCAATCATA
TGACCAACTCTGGGAGTTTGGGTCTTAGAGGAGGATTACGGGGAAGAAGGAAATCAAACAGTGA
TTGGCCTCTTTAAGACTAAGAGCATCCCCCTTAATTTAAGGGGAAATTTACTCTGTCAAATAAA
GAAAGGAAGACAACGAATTGGATTTGTATTCCTCTAAAAATGTAATTTGTATGAGACATAACTCT
ATTTGTTTCGCGGGGTTTGGCAGACAACAAACCTTGAATGGTGTTCAGGTTTTAATTGTTACTTTC
CTTTACGATCATATGGTTTCCGACCCACTTATGGTGTGGTCACCAACCATAACAGAGTAGTAGTAC
TTTCTTTTGAACCTTCTACATGCACCAGCAACTGTTTGTGGACCTAAAAGTCTACTAATTTGGTTAA
AAACAAATGTGTCAATTTCAACTTCAATGGTTAACAGGCACAGGTGTTCTTACTGAGTCTAACAA
AAAGTTTCTGCTTTCCAACAATTTGGCAGAGACTTGTGACACTACTGATGCTGTCCGTGATCC
ACAGACTTGTGATTTCTGACATTACACATGTTCTTTTGGTGGTGTGAGTGTATAACACCAGG
AACAAATACTTCTAACCAGGTTGCTGTTCTTTATCAGGGTGTAACTGCACAGAAGTCCCTGTTGC
TATTCATGCAGATCAACTTACTCTACTTGGCGTGTATTCTACAGGTTCTAATGTTTTTCAAACA
CGTGCAGGCTGTTAATAGGGGCTGAATATGTCAACAACCTCATATGAGTGTGACATACCCATTGGT
GCAGGTATATGCGCTAGTTATCAGACTCAGACTAAGTCTCATCGGCGGGCACGTAGTGTCTCAG
ATATCCGTGCTTTAACTGGAATAGGTGTTGAACAAGACAAAACACCCAAGAAGTTTTTGCACA
AGTCAAACAAATTTACAAAACACCACCAATTAATATTTTGGTGGTTTTAATTTTTCACAAATATT
ACCAGATCCATCAAACCAAGCAAGAGGTCATTTATTGAAGATCTACTTTTCAACAAAGTGACAC
TTGCAGATGCTGGCTTCATCAAACAATATGGTGAATGCTTGGTGATATTGCTGCTAGAGACCTCA
TTTGTGCACAAAAGTTTAAACGGCCTTACTGTTTTGCCACCTTGTCTCACAGATGAAATGATTGCTCA
ATACACTTCTGCACGTTAGCGGGTACAATCACTTCTGGTTGGACCTTTGGTGCAGGTGCTGCATTA
CAAATACCATTTGCTATGCAAATGGCTTATAGGTTTAAATGGTATTGGAGTTACACAGAATGTTCTC
TATGAGAACCAAAAATTGATTGCCAACCAATTTAATAGTGCTATTGGCAAAAATTCAAGACTCACTT
CTTCCACAGCAAGGTGAAACAAAGATGCAAAGATGTGGTCAACCATAATGCACAAGCTTTAAA
CAGCTTGTTAAACAACCTTAGCTCCAAATTTGGTGAATTTCAAGTGTTTTAAATGATATCCTTTCA
CGCTTGACAAAAGTTGAGGCTGAAAGTGCAAATTTGATAGGTTGATCACAGGCAGACTTCAAAGTT
TGCAGACATATGTGACTCAACAATTAATTAGAGCTGCGAAATAAAAATGTCAGAGTGTGTACTTGG
ACAATCAAAAAGAGTTGATTTTTGTGGAAAGGGCTATCATCTTATGTCCTTCCCTCAGTCAGCACC
TCATGGTGTAGTCTTCTTGCATGTGACTTATGTCCCTGCACAAGAAAAGAAGTTCACAACTGCTCC
TGCCATTTGTCATGATGGAAAAGCACACTTTCCTCGTGAAGGTGCTTTTGTTCAAATGGCACACA
CTGGTTTGTAAACACAAAGGAATTTTTATGAACCACAAATCATTACTACAGACAACACATTTGTGTC
TGGTAACTGTGATGTTGTAATAGGAATTGTCAACAACACAGTTTATGATCCTTGAACCTGAATT
AGATTCATTCAAGGAGGAGTTAGATAAATATTTAAGAATCATAACATCACCAGATGTTGATTTAGG
TGACATCTCTGGCATTAAATGCTTCAGTTGTAAACATTCAAAAAGAAATTGACCGCCTCAATGAGGT
TGCCAAGAATTTAAATGAATCTCTCATCGATCTCAAGAACTTGGAAAGTATGAGCAGTATATAA
AATGGCCATGGTACATTTGGCTAGGTTTTATAGCTGGCTTGAATGCCATAGTAATGGTGACAATTA
TGCTTTGCTGTATGACCAGTTGCTGTAGTTGTCTCAAGGGCTGTTGTTCTTGTGGATCCTGCTGCAA
ATTTGATGAAGACGACTCTGAGCCAGTGTCTCAAAGGAGTCAAATACATTACACATAAACGAAC
TATGGATTGTTTATGAGAATCTTCACAATTGGAAGTGAACCTTGAACAAGGTGAAATAGGATGC
TA

Näyte 6

TGGAGGATACAAATCCAATTCAGTTGTCCTTCTATTCTTTATTTGACATGAGTAAATTTCCCCTTAA
ATTAAGGGGTACTGCTGTTATGTCTTTAAAGAGGTGAATCAGATTTATTTTATCTCTTCTTAGTAAA
GGTAGACTTATAATTAGAGAAAACAACAGAGTTGTTATTTCTAGTGATGTTCTTGTTAACAACCTAA
ACGAACAATGTTTGTCTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGTTAATCTTATAACCA
GAACTCAATCATACACTAATTTCTTTTCGCACGTAGTGTAGCTAGTCAATCAACCTTGTAAATGGTGT
GCAGGTTTTAATTGTTACTTTCTTTACGATCATATGGTTTCCGACCCACTTATGGTGTGGTCACCA
ACCATACAGAGTAGTAGTACTTTCTTTTGAACTTTTACATGCACCAGCAAGGTGAAATCAAGGATG
CAAATCTACTAATTTGGTTAAAAACAATGTGTCAATTTCAACTTCAATGGTTTAAACAGGCACAGG
TGTTCTTACTGAGTCTAACA AAAAGTTTCTGCCTTTCCAACAATTTGGCAGAGACATTGCTGACAC
TACTGATGCTGTCCGTGATCCACAGACACTTGAGATTCTTGACATTACACCATGTTCTTTTGGTGGT
GTCAGTGTATAACACCAGGAACAAATACTTCTAACCAGGTTGCTGTTCTTTATCAGGGTGTAAAC
TGCACAGAAGTCCCTGTTGCTATTTCATGCAGATCAACTTACTCCTACTTGGCGTGTTTATTCTACAG
GTTCTAATGTTTTTCAAACACGTGCAGGCTGTTTAAATAGGGGCTGAATATGTCAACAACCTCATATG
AGTGTGACATACCCATTGGTGCAGGTATATGCGCTAGTTATCAGACTCAGACTAAGTCTCATCGGC
GGGCACGTAGTGTAGCTAGTCAATCCCGTGCTTTAACTGGAATAGCTGTTGAACAAGACAAAAAC
ACCCAAGAATTTTAGCACAAGTCAAACAAATTTACAAACACCACACACCAATTAATATTTTGGT
GGTTTTAATTTTCACAAATATTACCAGATCCATCAAACCAAGCAAGGTGAAATCAAGGATGCT
AA

LIITE 3

Näytteiden ja mallisekvenssien linjaus

Linjauksessa alue 23010–23605 bp. Linjauksen pituus on 571 bp.

Alueella 13 emäksen eroavaisuutta, joista 8 sijaitsee ensimmäisen 70 emäksen alueella. Eroavaisuudet on lihavoitu ja korostettu alfavariantin mallisekvenssiin nähden.

Alueelta puuttuvat emäkset ovat 23013, 23065, 23115, 23131–23150, 23271 ja 23559.

```
      .....|.....| .....|.....| .....|.....| .....|.....| .....|.....|
      10      20      30      40      50
Alfa      GGTAAGGTTT TAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCAACCC
Beta      GGTAAGGTTT TAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCAACCC
Delta      GGTAAGGTTT TAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCAACCC
Gamma      GGTAAGGTTT TAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCAACCC
Omikron BA.1 GGTCAGGTTT TAATTGTTAC TTTCCCTTTAC GATCATATAG TTTCCGACCC
Omikron BA.4 GGTCAGGTTAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCGACCC
Omikron BA.5 GGTCAGGTTAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCGACCC
Näyte 1      GGTAAGGTTT TAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCAACCC
Näyte 2      GGTAAGGTTT TAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCAACCC
Näyte 3      GGTAAGGTTT TAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCAACCC
Näyte 4      GGTCAGGTTT TAATTGTTAC TTTCCCTTTAC GATCATATAG TTTCCGACCC
Näyte 5      GGTCAGGTTT TAATTGTTAC TTTCCCTTTAC GATCATATGG TTTCCGACCC
Näyte 6      GGTCAGGTTT TAATTGTTAC TTTCCCTTTAC GATCATATGG TTTCCGACCC
Omikron BA.2 GTTCAGGTTT TAATTGTTAC TTTCCCTTTAC GATCATATGG TTTCCGACCC
Wild type    GTTAAGGTTT TAATTGTTAC TTTCCCTTTAC AATCATATGG TTTCCAACCC
```

```
      .....|.....| .....|.....| .....|.....| .....|.....|
      60      70      80      90     100
Alfa      ACTTTGGTGT TGGTTACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Beta      ACTTTGGTGT TGGTTACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Delta      ACTATGGTGT TGGTTACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Gamma      ACTTTGGTGT TGGTTACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Omikron BA.1 ACTTTGGTGT TGGTCACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Omikron BA.4 ACTTTGGTGT TGGTCACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Omikron BA.5 ACTTTGGTGT TGGTCACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Näyte 1      ACTATGGTGT TGGTTACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Näyte 2      ACTATGGTGT TGGTTACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Näyte 3      ACTATGGTGT TGGTTACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Näyte 4      ACTTTGGTGT TGGTCACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Näyte 5      ACTTTGGTGT TGGTCACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Näyte 6      ACTTTGGTGT TGGTCACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Omikron BA.2 ACTTTGGTGT TGGTCACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
Wild type    ACTATGGTGT TGGTTACCAA CCATACAGAG TAGTAGTACT TTCTTTTGAA
```

```
      .....|.....| .....|.....| .....|.....| .....|.....|
      110     120     130     140     150
Alfa      CTTTACATGC ACCAGCAATC TACTAATTTG GTTAAAAACA AATGTGTCAA
Beta      CTTTACATGC ACCAGCAATC TACTAATTTG GTTAAAAACA AATGTGTCAA
Delta      CTTTACATGC ACCAGCAATC TACTAATTTG GTTAAAAACA AATGTGTCAA
Gamma      CTTTACATGC ACCAGCAATC TACTAATTTG GTTAAAAACA AATGTGTCAA
Omikron BA.1 CTTTACATGC ACCAGCAATC TACTAATTTG GTTAAAAACA AATGTGTCAA
Omikron BA.4 CTTTACATGC ACCAGCAATC TACTAATTTG GTTAAAAACA AATGTGTCAA
Omikron BA.5 CTTTACATGC ACCAGCAATC TACTAATTTG GTTAAAAACA AATGTGTCAA
```

Näyte 1	CTTTACATGC	ACCAGCAATC	TACTAATTTG	GTTAAAAACA	AATGTGTCAA
Näyte 2	CTTTACATGC	ACCAGCAATC	TACTAATTTG	GTTAAAAACA	AATGTGTCAA
Näyte 3	CTTTACATGC	ACCAGCAATC	TACTAATTTG	GTTAAAAACA	AATGTGTCAA
Näyte 4	CTTTACATGC	ACCAGCAATC	TACTAATTTG	GTTAAAAACA	AATGTGTCAA
Näyte 5	CTTTACATGC	ACCAGCAATC	TACTAATTTG	GTTAAAAACA	AATGTGTCAA
Näyte 6	CTTTACATGC	ACCAGCAATC	TACTAATTTG	GTTAAAAACA	AATGTGTCAA
Omikron BA.2	CTTTACATGC	ACCAGCAATC	TACTAATTTG	GTTAAAAACA	AATGTGTCAA
Wild type	CTTTACATGC	ACCAGCAATC	TACTAATTTG	GTTAAAAACA	AATGTGTCAA

	
		160	170	180	190	200
Alfa	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Beta	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Delta	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Gamma	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Omikron BA.1	TTTCAACTTC	AATGGTTTAA	A AGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Omikron BA.4	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Omikron BA.5	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Näyte 1	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Näyte 2	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Näyte 3	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Näyte 4	TTTCAACTTC	AATGGTTTAA	A AGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Näyte 5	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Näyte 6	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Omikron BA.2	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	
Wild type	TTTCAACTTC	AATGGTTTAA	CAGGCACAGG	TGTTCTTACT	GAGTCTAACA	

	
		210	220	230	240	250
Alfa	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATTAT	GACACTACTG	
Beta	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Delta	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Gamma	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Omikron BA.1	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Omikron BA.4	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Omikron BA.5	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Näyte 1	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Näyte 2	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATTAT	GACACTACTG	
Näyte 3	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Näyte 4	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Näyte 5	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Näyte 6	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Omikron BA.2	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	
Wild type	AAAAGTTTCT	GCCTTTCCAA	CAATTTGGCA	GAGACATT CT	GACACTACTG	

	
		260	270	280	290	300
Alfa	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Beta	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Delta	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Gamma	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Omikron BA.1	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Omikron BA.4	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Omikron BA.5	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Näyte 1	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Näyte 2	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Näyte 3	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Näyte 4	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Näyte 5	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Näyte 6	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	
Omikron BA.2	ATGCTGTCCG	TGATCCACAG	ACACTTGAGA	TTCTTGACAT	TACACCATGT	

Wild type ATGCTGTCCG TGATCCACAG ACACTTGAGA TTCTTGACAT TACACCATGT

.....|.....||.....||.....||.....||.....|
310 320 330 340 350

Alfa TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Beta TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Delta TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Gamma TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Omikron BA.1 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Omikron BA.4 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Omikron BA.5 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Näyte 1 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Näyte 2 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Näyte 3 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Näyte 4 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Näyte 5 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Näyte 6 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Omikron BA.2 TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA
Wild type TCTTTTGGTG GTGTCAGTGT TATAACACCA GGAACAAATA CTTCTAACCA

.....|.....||.....||.....||.....||.....|
360 370 380 390 400

Alfa GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Beta GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Delta GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Gamma GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Omikron BA.1 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Omikron BA.4 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Omikron BA.5 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Näyte 1 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Näyte 2 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Näyte 3 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Näyte 4 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Näyte 5 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Näyte 6 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Omikron BA.2 GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA
Wild type GGTGCTGTT CTTTATCAGG GTGTTAACTG CACAGAAGTC CCTGTTGCTA

.....|.....||.....||.....||.....||.....|
410 420 430 440 450

Alfa TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Beta TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Delta TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Gamma TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Omikron BA.1 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Omikron BA.4 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Omikron BA.5 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Näyte 1 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Näyte 2 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Näyte 3 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Näyte 4 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Näyte 5 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Näyte 6 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Omikron BA.2 TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT
Wild type TTCATGCAGA TCAACTTACT CCTACTTGGC GTGTTTATTC TACAGTTCT

.....|.....||.....||.....||.....||.....|
460 470 480 490 500

Alfa AATGTTTTTC AAACACGTGC AGGCTGTTTA ATAGGGGCTG AACATGTCAA

Beta	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AACATGTCAA
Delta	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AACATGTCAA
Gamma	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AA T ATGTCAA
Omikron BA.1	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AA T ATGTCAA
Omikron BA.4	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AA T ATGTCAA
Omikron BA.5	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AA T ATGTCAA
Näyte 1	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AACATGTCAA
Näyte 2	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AACATGTCAA
Näyte 3	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AACATGTCAA
Näyte 4	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AA T ATGTCAA
Näyte 5	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AA T ATGTCAA
Näyte 6	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AA T ATGTCAA
Omikron BA.2	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AA T ATGTCAA
Wild type	AATGTTTTTC	AAACACGTGC	AGGCTGTTTA	ATAGGGGCTG	AACATGTCAA

	510	520	530	540	550
Alfa	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Beta	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Delta	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Gamma	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Omikron BA.1	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Omikron BA.4	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Omikron BA.5	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Näyte 1	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Näyte 2	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Näyte 3	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Näyte 4	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Näyte 5	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Näyte 6	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Omikron BA.2	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA
Wild type	CAACTCATAT	GAGTGTGACA	TACCCTTGGT	GCAGGTATAT	GCGCTAGTTA

	560	570	
Alfa	TCAGACTCAG	ACTAATTCTC	G
Beta	TCAGACTCAG	ACTAATTCTC	G
Delta	TCAGACTCAG	ACTAATTCTC	G
Gamma	TCAGACTCAG	ACTAATTCTC	G
Omikron BA.1	TCAGACTCAG	ACTA A GTCTC	T
Omikron BA.4	TCAGACTCAG	ACTA A GTCTC	T
Omikron BA.5	TCAGACTCAG	ACTA A GTCTC	T
Näyte 1	TCAGACTCAG	ACTAATTCTC	G
Näyte 2	TCAGACTCAG	ACTAATTCTC	G
Näyte 3	TCAGACTCAG	ACTAATTCTC	G
Näyte 4	TCAGACTCAG	ACTA A GTCTC	T
Näyte 5	TCAGACTCAG	ACTA A GTCTC	T
Näyte 6	TCAGACTCAG	ACTA A GTCTC	T
Omikron BA.2	TCAGACTCAG	ACTA A GTCTC	T
Wild type	TCAGACTCAG	ACTAATTCTC	C

LIITE 4

Sanger-sekvensoinnin alukkeet

SEKVENSSI; PAIKKA GENOMISSA

SARS-CoV-2_71_LEFT GGCAAACCACGCGAACAAATAG; 21294-21316
SARS-CoV-2_76_RIGHT_alt1 GTCCACAAACAGTTGCTGGTG; 2312-23141
SARS-CoV-2_74_LEFT TGGACCTTGAAGGAAAACAGGG; 22091-22113
SARS-CoV-2_78_RIGHT GGATTGACTAGCTACACTACGTGC; 23611-23635
SARS-CoV-2_77_LEFT CAAACCTTTTGAGAGAGATATTTCAACTGA;
22944-22974
SARS-CoV-2_81_RIGHT TGAAGTCTGCCTGTGATCAACC; 24545-24567
SARS-CoV-2_80_LEFT CCGTGCTTTAACTGGAATAGCTG; 23853-23876
SARS-CoV-2_84_RIGHT AGCATCCTTGATTTACCTTGCT; 25438-25461
SARS-CoV-2_72_LEFT GTGATGTTCTTGTTAACAACACTAAACGAAC; 21532-
21561
SARS-CoV-2_73_RIGHT ACCTAGTGATGTTAATACCTATTGGCA; 22247-
22274
SARS-CoV-2_75_LEFT GAAAATGGAACCATTACAGATGCTGT; 22402-
22428
SARS-CoV-2_77_RIGHT CACTGACACCACCAAAGAACATG; 23327-23351
SARS-CoV-2_79_LEFT ACCCATTGGTGCAGGTATATGC
SARS-CoV-2_79_RIGHT AATTGGTGGTGTGTTTGTAAATTTGTTTGAC; 23914-
23944
SARS-CoV-2_82_LEFT TGCACAAGCTTTAAACACGCTT; 24426-24448
SARS-CoV-2_82_RIGHT CACGAGGAAAGTGTGCTTTTCC; 24814-24836

Ion-Torrent -sekvensoinnissa käytetyt alukkeet ja niiden paikat SARS-CoV-2-viruksen genomissa. Osa käytetyistä alukkeista oli samoja kuin Sanger-sekvensoinnissa.

TGTAACACGACGGCCAGT TAGTCTCTAGTCAGTGTGTTAATC 23617-
23641

P1_NTDrev

CCACTACGCCTCCGCTTT CTTTCCATCCAACCTTTTGTTG 23244-23265

M13-RBDfor

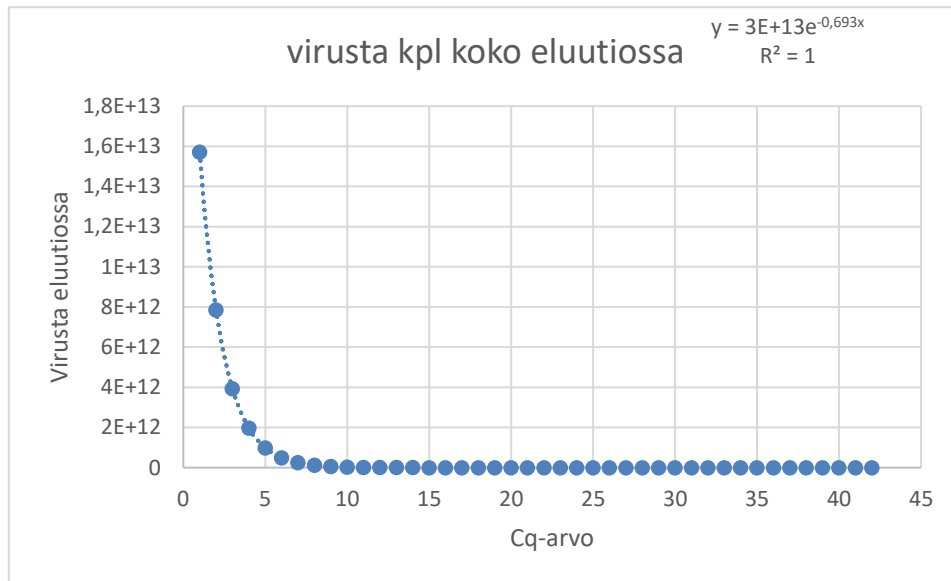
TGTAACACGACGGCCAGT CTGCTTTACTAATGTCTATGCAGATTC 22712-
22758

P1-76right

CCACTACGCCTCCGCTTT ACCTAGTGATGTTAATACCTATTGGCA 23120-
23141

LIITE 5

Cq-arvo muunnettiin seuraavan kuvaajan avulla virusten kappalemääräksi koko qPCR-eluutiosta.



LIITE 6

Systemaattisen kirjallisuuskatsauksen artikkelit ja niitä vastaavat järjestysluvut

1. Acer P.T. 2022. Quantifying the Relationship between SARS-CoV-2 Wastewater Concentrations and Building-Level COVID-19 Prevalence at an Isolation Residence: A Passive Sampling Approach. *International journal of environmental research and public health* 19: 11245-.
2. Agrawal S. 2021. Long-term monitoring of SARS-CoV-2 RNA in wastewater of the Frankfurt metropolitan area in Southern Germany. *Scientific reports* 11: 5372–5377.
3. Ahmed W. 2021. SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. *The Science of the total environment* 761: 144216–144216.
4. Ahmed W., Angel N., Edson J., Bibby K., Bivins A., O'Brien J.W., Choi P.M., Kitajima M., Simpson S.L., Li J., Tschärke B., Verhagen R., Smith W.J.M., Zaugg J., Dierens L., Hugenholtz P., Thomas K.V. & Mueller J.F. 2020. First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Science of The Total Environment* 728: 138764.
5. Arora S. 2020. Sewage surveillance for the presence of SARS-CoV-2 genome as a useful wastewater based epidemiology (WBE) tracking tool in India. *Water science and technology* 82: 2823–2836.
6. Arora S. 2022. Successful application of wastewater-based epidemiology in prediction and monitoring of the second wave of COVID-19 with fragmented sewerage systems—a case study of Jaipur (India). *Environmental monitoring and assessment* 194: 342–342.
7. Boeraş I. 2022. Anthropogenic Sewage Water Circuit as Vector for SARS-CoV-2 Viral ARN Transport and Public Health Assessment, Monitoring and Forecasting—Sibiu Metropolitan Area (Transylvania/Romania) Study Case. *International journal of environmental research and public health* 19: 11725-.
8. Carcereny A. 2021. Monitoring Emergence of the SARS-CoV-2 B.1.1.7 Variant through the Spanish National SARS-CoV-2 Wastewater Surveillance System (VATar COVID-19). *Environmental science & technology* 55: 11756–11766.
9. Claro I.C.M., Cabral A.D., Augusto M.R., Duran A.F.A., Graciosa M.C.P., Fonseca F.L.A., Speranca M.A. & Bueno R. de F. 2021. Long-term monitoring of SARS-COV-2 RNA in wastewater in Brazil: A more responsive and economical approach. *Water Research* 203: 117534.

10. De Giglio O. 2021. Potential Use of Untreated Wastewater for Assessing COVID-19 Trends in Southern Italy. *International journal of environmental research and public health* 18: 10278-.
11. Fitzgerald S.F. 2021. Site Specific Relationships between COVID-19 Cases and SARS-CoV-2 Viral Load in Wastewater Treatment Plant Influent. *Environmental science & technology* 55: 15276–15286.
12. Flood M.T. 2021. Methods Evaluation for Rapid Concentration and Quantification of SARS-CoV-2 in Raw Wastewater Using Droplet Digital and Quantitative RT-PCR. *Food and environmental virology* 13: 303–315.
13. Giraud-Billoud M. 2021. Monitoring of SARS-CoV-2 RNA in wastewater as an epidemiological surveillance tool in Mendoza, Argentina. *The Science of the total environment* 796: 148887–148887.
14. Ho J. 2022. SARS-CoV-2 wastewater surveillance in Germany: Long-term RT-digital droplet PCR monitoring, suitability of primer/probe combinations and biomarker stability. *Water research (Oxford)* 210: 117977–117977.
15. Izquierdo-Lara R. 2021. Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing, the netherlands and belgium. *Emerging infectious diseases* 27: 1405–1415.
16. Jordan P., Alessandro Z., , Brackney D.E., , Grubaugh N.D., Kaplan E.H., Arnau C.-M., Ko A.I., Malik A.A., Wang D., Wang M., Warren J.L., Weinberger D.M., Wyatt A., Omer S.B. 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nature Biotechnology* 38: 1164–1167.
17. Joseph-Duran B. 2022. Assessing wastewater-based epidemiology for the prediction of SARS-CoV-2 incidence in Catalonia. *Scientific reports* 12: 15073–15073.
18. Karthikeyan S. 2022. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature (London)* 609: 101–104.
19. Kaya D. 2022. Correlation between Clinical and Wastewater SARS-CoV-2 Genomic Surveillance, Oregon, USA. *Emerging infectious diseases* 28: 1906–1908.
20. Li L. 2022a. Detecting SARS-CoV-2 variants in wastewater and their correlation with circulating variants in the communities. *Scientific reports* 12: 16141–16141.
21. Li L. 2022b. Longitudinal monitoring of SARS-CoV-2 in wastewater using viral genetic markers and the estimation of unconfirmed COVID-19 cases. *The Science of the total environment* 817: 152958–152958.
22. Mlejnkova H. 2020. Preliminary Study of Sars-Cov-2 Occurrence in Wastewater in the Czech Republic. *International journal of environmental research and public health* 17: 5508-.
23. Morvan M. 2022. An analysis of 45 large-scale wastewater sites in England to estimate SARS-CoV-2 community prevalence. *Nature communications* 13: 4313–4313.

24. Peinado B. 2022. Improved methods for the detection and quantification of SARS-CoV-2 RNA in wastewater. *Scientific reports* 12: 7201–7201.
25. Reno U. 2022. Wastewater-Based Epidemiology: Detection of SARS-CoV-2 RNA in Different Stages of Domestic Wastewater Treatment in Santa Fe, Argentina. *Water, air, and soil pollution* 233: 372–372.
26. Viveros M.L. 2022. Wild type and variants of SARS-COV-2 in Parisian sewage: presence in raw water and through processes in wastewater treatment plants. *Environmental science and pollution research international* 29: 67442–67449.
27. Weidhaas J., Aanderud Z.T., Roper D.K., VanDerslice J., Gaddis E.B., Ostermiller J., Hoffman K., Jamal R., Heck P., Zhang Y., Torgersen K., Laan J.V. & LaCross N. 2021. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Science of The Total Environment* 775: 145790.
28. Zarza E. 2022. Monitoring SARS-CoV-2 in the Wastewater and Rivers of Tapachula, a Migratory Hub in Southern Mexico. *Food and environmental virology* 14: 199–211.

LIITE 7

Sekvensoidun jätevesinäytteen tuottamat sekvenssit ja fylogeneettisessa puussa käytetyt mallisekvenssit eri virusvarianteista.

	10	20	30	40	50
15EP9:0292	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0013	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0077	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0241	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0258	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0275	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0276	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0184	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0309	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0424	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0179	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
15EP9:0335	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATTGCTCCA	GGGCAAAC TG
Wild_Type	ATTTGTAATT	AGAGGTGATG	AAGTCAGACA	AATCGCTCCA	GGGCAAAC TG
Alfa	ATTTGTAATT	AGAGGTGATG	AAGTCAGACA	AATCGCTCCA	GGGCAAAC TG
Beta	ATTTGTAATT	AGAGGTGATG	AAGTCAGACA	AATCGCTCCA	GGGCAAAC TG
Delta	ATTTGTAATT	AGAGGTGATG	AAGTCAGACA	AATCGCTCCA	GGGCAAAC TG
Gamma	ATTTGTAATT	AGAGGTGATG	AAGTCAGACA	AATCGCTCCA	GGGCAAAC TG
Omikron_BA1	ATTTGTAATT	AGAGGTGATG	AAGTCAGACA	AATCGCTCCA	GGGCAAAC TG
Omikron_BA2	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATCGCTCCA	GGGCAAAC TG
Omikron_BA4	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATCGCTCCA	GGGCAAAC TG
Omikron_BA5	ATTTGTAATT	AGAGGTAATG	AAGTCAGCCA	AATCGCTCCA	GGGCAAAC TG

	60	70	80	90	100
15EP9:0292	GAAATATTGC	TGATTATGAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0013	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0077	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0241	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0258	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0275	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0276	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0184	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0309	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0424	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0179	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
15EP9:0335	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
Wild_Type	GAAAGATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
Alfa	GAAAGATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
Beta	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
Delta	GAAAGATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
Gamma	GAACGATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
Omikron_BA1	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
Omikron_BA2	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
Omikron_BA4	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC
Omikron_BA5	GAAATATTGC	TGATTATAAT	TATAAATTAC	CAGATGATTT	TACAGGCTGC

	110	120	130	140	150
15EP9:0292	GTTATAGCTT	GGAATCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0013	GTTATAGCTT	GGAATCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0077	GTTATAGCTT	GGAATCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA

15EP9:0241	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0258	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0275	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0276	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0184	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0309	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0424	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0179	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
15EP9:0335	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
Wild_Type	GTTATAGCTT	GGAATTCTAA	CAATCTTGAT	TCTAAGGTTG	GTGGTAATTA
Alfa	GTTATAGCTT	GGAATTCTAA	CAATCTTGAT	TCTAAGGTTG	GTGGTAATTA
Beta	GTTATAGCTT	GGAATTCTAA	CAATCTTGAT	TCTAAGGTTG	GTGGTAATTA
Delta	GTTATAGCTT	GGAATTCTAA	CAATCTTGAT	TCTAAGGTTG	GTGGTAATTA
Gamma	GTTATAGCTT	GGAATTCTAA	CAATCTTGAT	TCTAAGGTTG	GTGGTAATTA
Omikron_BA1	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
Omikron_BA2	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
Omikron_BA4	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA
Omikron_BA5	GTTATAGCTT	GGAATTCTAA	CAAGCTTGAT	TCTAAGGTTG	GTGGTAATTA

	160	170	180	190	
15EP9:0292	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
15EP9:0013	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATATCAA	CCTTT-
15EP9:0077	TAATTAC-TG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
15EP9:0241	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATCTAAA-	CCTTT-
15EP9:0258	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATATCAA	CCTTT-
15EP9:0275	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATA-CAA-	-CTTT-
15EP9:0276	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA-	-CTT--
15EP9:0184	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATACTCA	CCTTT-
15EP9:0309	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATACTCAA	-CTTT-
15EP9:0424	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATACCCAA	CCTTT-
15EP9:0179	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATTTACAA	CCTTT-
15EP9:0335	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATATCAA	CCTTT-
Wild_Type	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
Alfa	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
Beta	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
Delta	TAATTACCGG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
Gamma	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
Omikron_BA1	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
Omikron_BA2	TAATTACCTG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
Omikron_BA4	TAATTACCGG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-
Omikron_BA5	TAATTACCGG	TATAGATTGT	TTAGGAAGTC	TAATCTCAA	CCTTT-