

HELSINGIN YLIOPISTO

Mikrobit virtavesien ympäristön tilan kuvaamisessa:

Harvinaisten lajien poiston vaikutus yhteisö-ympäristösuhteisiin

Tiina Laamanen
Pro gradu -tutkielma
Perinnöllisyystiede
Biotieteiden laitos
Helsingin yliopisto
Huhtikuu 2019

Tiedekunta – Fakultet – Faculty Bio- ja ympäristötieteellinen tiedekunta		Laitos – Institution– Department Biotieteiden laitos	
Tekijä – Författare – Author Tiina Laamanen (o.s. Nokela)			
Työn nimi – Arbetets titel – Title Mikrobit virtavesien ympäristön tilan kuvaamisessa: Harvinaisten lajien poiston vaikutus yhteisö- ympäristösuhteisiin			
Oppiaine – Läroämne – Subject Perinnöllisyystiede			
Työn laji – Arbetets art – Level Pro gradu		Aika – Datum – Month and year 3.4.2019	Sivumäärä – Sidoantal – Number of pages 93
Tiivistelmä – Referat – Abstract			
<p>Ihmistoiminnan vaikutukset ovat muuttaneet voimakkaasti ekosysteemien rakennetta ja toimintaa maailmanlaajuisesti viimeisten vuosikymmenten aikana. Ekologisen tilan heikentyminen aiheuttaa vakavan uhan virtavesien eliöyhteisöjen monimuotoisuudelle.</p> <p>Tulevaisuudessa tarve kustannustehokkaille, luotettaville ja nopeille keinoille arvioida vesistöjen ekologista tilaa kasvaa. Biomonitorointi, eli eliöiden käyttö ympäristömuutosten seurantaan ja mittaamiseen, pyrkii tunnistamaan, seuraamaan ja arvioimaan ekosysteemien toimintaa uhkaavia ympäristön stressitekijöitä. Tulevaisuudessa yksi mahdollinen biomonitoroinnin työkalu ja ekologisen tilan mittari yhdessä muiden biologisten (esim. piilevät, pohjaeläimet, vesikasvit ja kalat) ja fysikaalis-kemiallisten mittarien kanssa voi olla mikrobit. Uuden polven sekvensointimenetelmien kehitys ja sekvensointikustannusten aleneminen viime vuosina on mahdollistanut yhä useamman ympäristömikrobiaineiston analysoinnin uuden polven sekvensointimenetelmillä.</p> <p>Tämän pro gradu -tutkielman tarkoitus on tuoda lisätietoa siitä, mitkä eri ympäristötekijät vaikuttavat virtaavien vesien biofilmbakteerien yhteisörakenteeseen, yhteisö-ympäristösuhteen voimakkuuteen, yhteisöjen koostumusta selittävien muuttujien identiteettiin ja miten harvinaisten bakteerisukujen poisto aineistosta vaikuttaa erilaisiin ympäristömuuttujiin. Tutkimusaineistoa analysoidaan redundanssianalyysin (redundancy analysis, RDA) ja monimuuttujaregressiopuiden (Multivariate regression trees, MRT) avulla. Tutkielman tavoitteena on myös kuvata länsisuomalaisten virtavesien koskipaikkojen kivien pinnalla elävien biofilmbakteerien ominaispiirteitä pääjako-, luokka-, laho-, heimo- ja sukutasoille tehtävien BLAST-analyysien (basic local alignment search tool) avulla.</p> <p>Pro gradu tutkielmani bakteeriaineiston näytteet on otettu 21 suomalaiselta kolmannen jakovaiheen jokivaluma-alueelta, joilta eristettiin erillisten purojen koskipaikoista bakteerinäytteitä yhteensä 84 kappaletta. Samaan aikaan bakteerinäytteiden ottamisen kanssa tutkimuskohteilla toteutettiin vesinäytteenotto, ja kohteista määritettiin myös piilevä- ja pohjaeläinlajisto. Näytteet sekvensoitiin Ion Torrent Semiconductor -sekvensaattoria käyttäen ja valmis sekvenssiaineisto analysoitiin QIIME (Quantitative Insights Into Microbial Ecology) –ohjelmalla. Sekvenssit klusteroitiin toiminnallisiksi taksonomisiksi yksiköiksi (OTUiksi) 97 % samankaltaisuusastetta käyttäen. OTUille haettiin BLASTia (basic local alignment search tool) käyttäen paras mahdollinen pääjako-, luokka-, laho-, heimo- ja sukutason vastaavuus. Yleisimmät havaitut luokat olivat <i>Alphaproteobacteria</i> (28.9%), <i>Betaproteobacteria</i> (27.7%), <i>Gammaproteobacteria</i> (7.5%) <i>Cytophagia</i> (7.3%), <i>Saprospirae</i> (6.5%) ja <i>Flavobacteriia</i> (5.1%).</p> <p>Tuloksista havaitaan, että yhteisön ympäristösuhteisiin vaikuttavat erityisesti yleiset lajit. Harvinaisten lajien poistolla ei tässä tutkimuksessa siten havaita olevan suurta merkitystä, vaan merkitsevät muuttujat pysyvät lähes samoina huolimatta siitä, poistetaanko aineistosta ne lajit, jotka esiintyvät vain yhdellä paikalla vai kaikki ne lajit, jotka esiintyvät vain ≤ 40 paikalla. Myös selitysasteet jäävät alhaisiksi ja satunnaisten eli stokastisten tekijöiden vaikutus yhteisön ympäristöolosuhteisiin luultavasti korostuu.</p> <p>Jatkossa eri bakteeritaksonien roolista muuttuvissa ympäristöolosuhteissa tarvitaan tarkempaa tietoa ennen kuin virtavesien biofilmbakteereja voidaan laajamittaisesti käyttää biomonitoroinnin täysimittaisena työkaluna ja ympäristön tilan kuvaajana. Tämän tutkielman redundanssianalyysin, vaihtelun osituksen ja monimuuttujaregressiopuiden avulla saatujen tulosten perusteella voidaan todeta virtaavien vesien mikrobiyhteisöjen yhteisö-ympäristösuhteiden olevan heikkoja. Ympäristön tilan seurannassa tulisikin jatkossa tulosten perusteella käyttää mikrobien osalta muita vastemuuttujia kuin yhteisörakennetta.</p>			
Avainsanat – Nyckelord – Keywords mikrobit, virtavedet, ekologisen tilan arviointi			
Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors Jani Heino ja Hannu Saarilahti			
Säilytyspaikka – Förvaringställe – Where deposited Helsingin yliopisto, Viikin kampuskirjasto			
Muita tietoja – Övriga uppgifter – Additional information -			

Sisällysluettelo

Harvinaisten lajien poiston vaikutus yhteisö-ympäristösuhteisiin.....	1
1 Johdanto.....	7
1.1 Tutkielman tavoitteet ja tutkimuskysymykset.....	9
1.2 Virtavesien mikrobiyhteisöt	10
1.2.1 Mikrobien tutkiminen sekvensointimenetelmien avulla	10
1.2.2 Biofilmin rakenne ja toiminta	11
1.2.3 Biofilmien rakenteeseen vaikuttavat tekijät	13
1.3 Mikrobiyhteisöihin vaikuttavat tärkeimmät tekijät: happamuus ja ravinnekuormitus.....	14
1.3.1 Happamuuteen vaikuttavat tekijät, happamuus ja mikrobit.....	14
1.3.2 Vesistöjen ravinnekuormitus.....	16
2 Materiaalit ja menetelmät	17
2.1 Tutkimusalue	17
2.2 Maastomenetelmät.....	20
2.2.1 Bakteerinäytteenotto	20
2.2.2 Vesikemia- ja habitaattianalyysit	21
2.3 Valuma-alueanalyysit	21
2.4 Laboratoriomenetelmät	22
2.5 Bioinformatiikan menetelmät.....	24
2.6 Tilastomenetelmät	25
2.6.1 Selittävät muuttujat	25
2.6.2 Redundanssianalyysi ja vaihtelun ositus.....	27
2.6.3 Monimuuttujaregressiopuut	29
2.6.4 OTU -aineisto.....	29

2.6.5	Harvinaisten OTUjen poisto	30
3	Tulokset	31
3.1	Näytteenottoajankohdan sää.....	31
3.2	Ympäristömuuttujien tunnuslukuja	31
3.2.1	pH.....	33
3.2.2	Kokonaisfosfori.....	33
3.2.3	Kokonaistyyppi	33
3.2.4	Sähkönjohtavuus	33
3.2.5	Väri.....	34
3.2.6	Valuma-alueuuttajat.....	34
3.2.7	Habitaattimuuttajat.....	34
3.3	Harvinaisten lajien poiston merkitys	35
3.3.1	Erilaisten OTU-taulukkojen muodostaminen sukujen esiintyvyyden perusteella	35
3.3.2	Selittävien muuttujien valinta	36
3.3.3	Millaisia vaikutuksia harvinaisten lajien poistolla havaitaan?.....	36
3.3.4	Redundanssianalyysi ja vaihtelun ositus.....	39
3.4	Monimuuttujaregressiopuuanalyysit	44
3.5	BLAST-tulokset	45
4	Tulosten tarkastelu.....	49
4.1	Tutkimuspaikkojen ekologinen tila	49
4.2	Mikrobiyhteisöjen rakenteeseen vaikuttavat tekijät	51
4.3	Harvinaisten lajien poiston merkitys	53
4.4	Länsisuomalaisten virtavesien biofilmibakteerit verrattuna muissa tutkimuksissa havaittuihin biofilmibakteereihin	54
4.5	Sovelтуuko mikrobien yhteisötutkimus biomonitoroinnin työkaluksi? .	57
4.6	Ajatuksia jatkotutkimuksille.....	59

4.6.1	Taksonomiset, funktionaaliset ja fylogeneettiset diversiteetti- indeksit	60
4.6.2	Fosmidikirjastoperusteiset menetelmät	61
4.6.3	RNA –perusteiset menetelmät.....	63
4.6.4	Metagenomi- ja metatranskriptomianalyysjä yhdistävät menetelmät 63	
4.6.5	Metagenomi- ja metaproteomitutkimusta yhdistävät menetelmät ..	64
5	Kiitokset.....	66
	Lähteet.....	68
LIITTEET		
	Liite 1 Tutkimuksessa käytetyt alukkeet.....	81
	Liite 2. Tutkimuspaikkojen sijaintia kuvaava pienin puu	85
	Liite 3 Mallikuva vaihtelun ositusten tulosten tulkintaan	86
	Liite 4 Tutkimuspaikkojen vesikemiatulokset	87
	Liite 5 Tutkimuspaikkojen yksilölliset a-alukkeet.....	89
	Liite 6. Luokkatasolle BLAST-analysoidut bakteerit	90
	Liite 7. Lahkotasolle BLAST-analysoidut bakteerit.	91
	Liite 8. Heimotasolle BLAST-analysoidut bakteerit.	92
	Liite 9. Sukutasolle BLAST-analysoidut bakteerit.	93

Käytettyjä lyhenteitä ja sanastoa

BLAST = työkalu jonka avulla sekvenssitietokannasta haetaan OTUille vastaavuuksia (basic local alignment search tool, Altschul ym. 1990)

cDNA = komplementaarinen DNA (complementary DNA)

dNTP = deoksiribonukleosiditriposfaatit (deoxynucleoside triphosphates). PCR -reaktiossa polymeerasientsyymien avulla rakennetaan uusi DNA-juoste yksittäisistä dNTP-molekyyleistä, joita ovat dATP, dTTP, dCTP ja dGTP

GC-puskuri = kaupallinen puskuriliuos, joka parantaa PCR -reaktiossa DNA-polymeerasientsyymien toimintaa (GC-buffer)

MRT = monimuuttujaregressiopuut (Multivariate regression trees, De' Ath, 2002)

NGS = Uuden polven sekvensointimenetelmät (Next generation sequencing)

OTU = yleensä 97 % samankaltaisuusasteella määritetty toiminnallinen taksonominen yksikkö (operational taxonomic unit)

PCR = polymeerasiketjureaktio (Polymerase chain reaction)

QIIME = Sekvenssiaineistojen analysointiin käytettävä ohjelma (Quantitative Insights Into Microbial Ecology, Caporaso ym. 2010)

RDA = redundanssianalyysi (redundancy analysis, Rao 1964)

SRB = Sulfaattia tuottavat bakteerit (sulfate-reducing bacteria)

VPD = Vesipuitedirektiivi (Water Framework Directive, WFD). Vuonna 2000 voimaan tullut direktiivi (2000/60/EY), joka määrittää EU:n vesipolitiikan suuntaviivat

1 Johdanto

Ihmistoiminnan vaikutukset ovat muuttaneet voimakkaasti ekosysteemien rakennetta ja toimintaa maailmanlaajuisesti viimeisten vuosikymmenten aikana (Millennium Ecosystem Assessment. 2005). Virtavesiekosysteemeissä ekologisen tilan heikentyminen aiheuttaa vakavan uhan niiden eliöyhteisöjen monimuotoisuudelle (Dudgeon ym. 2006). Maailman virtavesistä korkean tai keskinkertaisen uhan alla katsotaan olevan tällä hetkellä 65 prosenttia (Vörösmarty ym. 2010). Virtaavia vesiä uhkaavat muun muassa ilmastonmuutos, habitaattien heikkeneminen ja vedenlaadun huonontuminen. Ihmistoiminnan vaikutukset, kuten virtavesiin päätyvän kuormituksen ja kiintoaineen määrää lisäävä maankäyttö (esimerkiksi maatalous ja metsätalous), kaupungistuminen ja teollistuminen, uomien perkaukset ja kanavoinnit sekä keinotekoiset rakenteet, kuten padot ja kastelu- ja vedenjohtamisjärjestelmät heikentävät virtavesien ekologista tilaa ja eliöyhteisöjen monimuotoisuutta (Vörösmarty ym. 2010).

Tulevaisuudessa tarve kustannustehokkaille, luotettaville ja nopeille keinoille arvioida vesistöjen ekologista tilaa kasvaa. EU:n vesipuitedirektiivin (2000/60/EY) ja siihen pohjautuvan Suomen vesienhoitolain (1299/2004) tavoitteina on saada kaikki vesistöt vähintään hyvään tilaan. Alun perin tavoitteena oli saavuttaa hyvä tila vuoden 2015 aikana, mutta monien vesistöjen tavoitetila on tällä hetkellä asetettu vuosien 2021 ja 2027 välille. Tämänhetkisen arvion mukaan 65 prosenttia Suomen jokivesistä on hyvässä tai erinomaisessa tilassa, mutta erityisesti rannikko-alueen joet ovat huonossa tai välttävässä tilassa (Pintavesien ekologinen ja kemiallinen tila. 2016). Suomessa vesien tilaa on laajalaisesti heikentänyt erityisesti maa- ja metsätalouden hajakuormitus (Aroviita ym. 2014).

Biomonitorointi, eli eliöiden käyttö ympäristömuutosten seurantaan ja mittaamiseen, pyrkii tunnistamaan, seuraamaan ja arvioimaan ekosysteemien toimintaa uhkaavia ympäristön stressitekijöitä (Friberg ym. 2011; Bonada ym. 2006; Gerhardt 2000). Tulevaisuudessa yksi mahdollinen biomonitoroinnin työkalu ja ekologisen tilan mittari yhdessä muiden biologisten (esim. piilevät,

pohjaeläimet, vesikasvit ja kalat) ja fysikaalis-kemiallisten mittarien kanssa voi olla mikrobit (Friberg ym. 2011). Uuden polven sekvensointimenetelmien kehitys ja sekvensointikustannusten aleneminen viime vuosina on mahdollistanut yhä useamman ympäristömikrobiaineiston analysoinnin uuden polven sekvensointimenetelmillä (ks. esim. Schlötterer ym. 2014).

Baas Becking totesi vuonna 1934: ”*Kaikki on kaikkialla, mutta ympäristö valitsee*” (Baas Becking 1934). Hänen lausahduksensa on hyvin laajalti siteerattu ympäristömikrobiologian ja ekologian julkaisuissa (De Wit & Bouvier 2006). Mikrobeja on siis perinteisesti pidetty kosmopoliitteina. Viimeaikaiset tutkimukset vahvistavat Baas Beckingin käsitystä siitä, että ”*ympäristö valitsee*”, ja on osaltaan vaikuttamassa spatiaaliseen vaihteluun mikrobien diversiteetissä. Toisaalta viimeaikaiset tutkimukset ovat kyseenalaistaneet väitteen ”*kaikki on kaikkialla*” osoittamalla, että myös maantieteelliseen sijaintiin liittyvät tekijät voivat olla merkityksellisiä bakteeriyhteisöjen muodostumisessa (Martiny ym. 2006).

Rabinowitz (1981) suositteli, että lajien harvinaisuutta tulisi tarkastella paikallisen yhteisön koon, levinneisyyden ja elinympäristön kautta (Rabinowitz 1981). Harvinaisuuden määritelmä on biologian alan kirjallisuudessa vaihdellut vuosien ja vuosikymmenten saatossa lähestymistavasta riippuen (ks. esim. Mayr 1963; Hanski 1991; Batianoff & Burgess 1993). Yhteistä kaikille määritelmille on se, että harvinaisuus kuvataan esimerkiksi lajien runsauden, levinneisyysalueen koon tai elinympäristön ominaisuuksiin liittyvien tekijöiden kautta. Harvinaisuutta on määriteltä myös erilaisten tarkastelun kohteena olevaan lajiin liittyvien uhkien, geneettisen monimuotoisuuden ja taksonomisen erottuvuuden kautta (Gaston & Kunin 1997).

Harvinaisten lajien on perinteisesti katsottu tuovan aineistoihin vain vähän tulkinnallista lisäarvoa, mutta aiheuttavan tilasto-analyyseissä ongelmia (esim. Webb ym. 1967, Austin & Greig-Smith 1968; Gauch 1982). Harvinaisten lajien käsittely on kuitenkin yksi keskeinen kysymys biomonitoroinnissa, sillä niiden aineistoon sisällyttäminen tai siitä poisjättäminen voi vaikuttaa suoraan kykyyn havaita ja mitata ekologisia vaikutuksia (Cao ym. 2001).

Mikrobiyhteisöjen lajien yleisyyden ja runsauden jakaumaa tutkittaessa on havaittu, että hyvin usein aineistoissa havaitaan pitkä ”häntä” alhaisen yleisyyden ja runsauden omaavia, harvinaisia lajeja. Suuri osa mikrobien monimuotoisuudesta onkin havaittu juuri harvinaisten lajien joukossa. Nämä harvinaiset mikrobit voivat erota merkittävästi yleisemmistä mikrobilajeista muun muassa elinympäristönsä ekologisiin ja biogeografisiin tekijöihin liittyvissä asioissa (Lynch & Neufeld 2015).

Tutkittaessa harvinaisten mikrobien merkitystä ympäristö-yhteisösuhteeseen on hyvä huomioida ja tunnistaa mahdolliset virhelähteet esimerkiksi aineiston sekvensointiin (sequencing noise), bioinformatiikan työkalujen valintoihin (esim. raakasekvenssien käsittelyssä käytettävät laatuksiteerit) tai tilastoanalyysien valintoihin ja tulosten tulkintaan liittyen (ks. esim. Lynch & Neufeld (2015)).

Tässä tutkielmassa harvinaisten bakteerisukujen vaikutusta yhteisön ympäristösuhteisiin tutkitaan poistamalla aineistosta harvinaisia bakteerisukuja. Harvinaisuudella tarkoitetaan tässä tutkielmassa bakteerisukujen esiintymispaikkojen lukumäärää. Sukujen poistoja aineistosta tehdään eri tasoilla alkaen niistä suvuista, jotka esiintyvät vain yhdellä tutkimuspaikalla, ja päättyen niihin, jotka esiintyvät ≤ 40 , eli maksimissaan noin puolella, tutkimuspaikoista.

1.1 Tutkielman tavoitteet ja tutkimuskysymykset

Tämän pro gradu -tutkielman tarkoitus on tuoda lisätietoa siitä, mitkä eri ympäristötekijät vaikuttavat virtaavien vesien biofilmibakteerien yhteisörakenteeseen, yhteisö-ympäristösuhteen voimakkuuteen ja yhteisöjen koostumusta selittävien muuttujien identiteettiin. Tutkielmassa tarkastellaan erityisesti harvinaisten bakteerisukujen poiston vaikutusta yhteisön ympäristösuhteisiin. Tutkimusaineistoa analysoidaan redundanssianalyysin (redundancy analysis, RDA) (Rao 1964) ja monimuuttujaregressiopuiden (Multivariate regression trees, MRT) (De’Ath, 2002) avulla. Tutkielman tavoitteena on myös kuvata länsisuomalaisten virtavesien koskipaikkojen kivien pinnalla elävien biofilmibakteerien ominaispiirteitä pääjakso-, luokka-, lahko-,

heimo- ja sukutasoille tehtävien BLAST -analyysien (basic local alignment search tool) (Altschul ym. 1990) avulla.

Tutkielman päätutkimuskysymykset ovat:

1. Mitkä ympäristömuuttajat selittävät akvaattisten biofilmibakteerien esiintyvyyttä Länsi-Suomen jokivaluma-alueilla?
2. Parantaako harvinaisten bakteerisukujen poisto ympäristö-yhteisösuhteen selityskykyä tilastollisissa analyyseissä?
3. Millaisia bakteeritaksoneiden pääjaksoja Länsi-Suomen jokivaluma-alueiden virtavesien kivien pintojen biofilmiyhteisöissä havaitaan?

Ensimmäistä tutkimuskysymystä tarkastellaan monimuuttujaregressiopuiden (MRT, Multivariate regression trees, De'Ath, 2002) ja redundanssianalyysin (RDA, redundancy analysis, Rao 1964) avulla, toiseen tutkimuskysymykseen haetaan vastausta redundanssianalyysillä ja kolmannen tutkimuskysymyksen tarkastelussa hyödynnetään BLAST -analyysien (basic local alignment search tool, Altschul ym. 1990) tuloksia.

1.2 Virtavesien mikrobiyhteisöt

1.2.1 Mikrobien tutkiminen sekvensointimenetelmien avulla

Sekvensointimenetelmät ovat kehittyneet viime vuosina nopeasti ja sekvensointiin pohjautuvat lähestymistavat ovat korvanneet perinteisiä viljelymenetelmiin perustuneita tutkimuksia ja näin ollen mahdollistaneet ympäristönäytteiden entistä kattavammat mikrobiekologiset tutkimukset (Auvinen & Hultman 2010; Staley & Sadowsky 2016). Ympäristömikrobien sekvensointi voidaan toteuttaa niin kutsutulla amplikonisekvensointimenetelmällä, jolloin mikrobin tiettyä

amplikonien eli DNA- tai RNA- jaksoa monistetaan polymeerasiketjureaktiolla. Bakteereilla käytetään hyvin yleisesti tähän tarkoitukseen 16S rRNA geenin (Franzosa ym. 2015). Uuden polven sekvensointimenetelmät (Next generation sequencing, NGS) ovat mahdollistaneet yhä laajempien ympäristömikrobiaineistojen kustannustehokkaan analysoinnin (Staley & Sadowsky 2016).

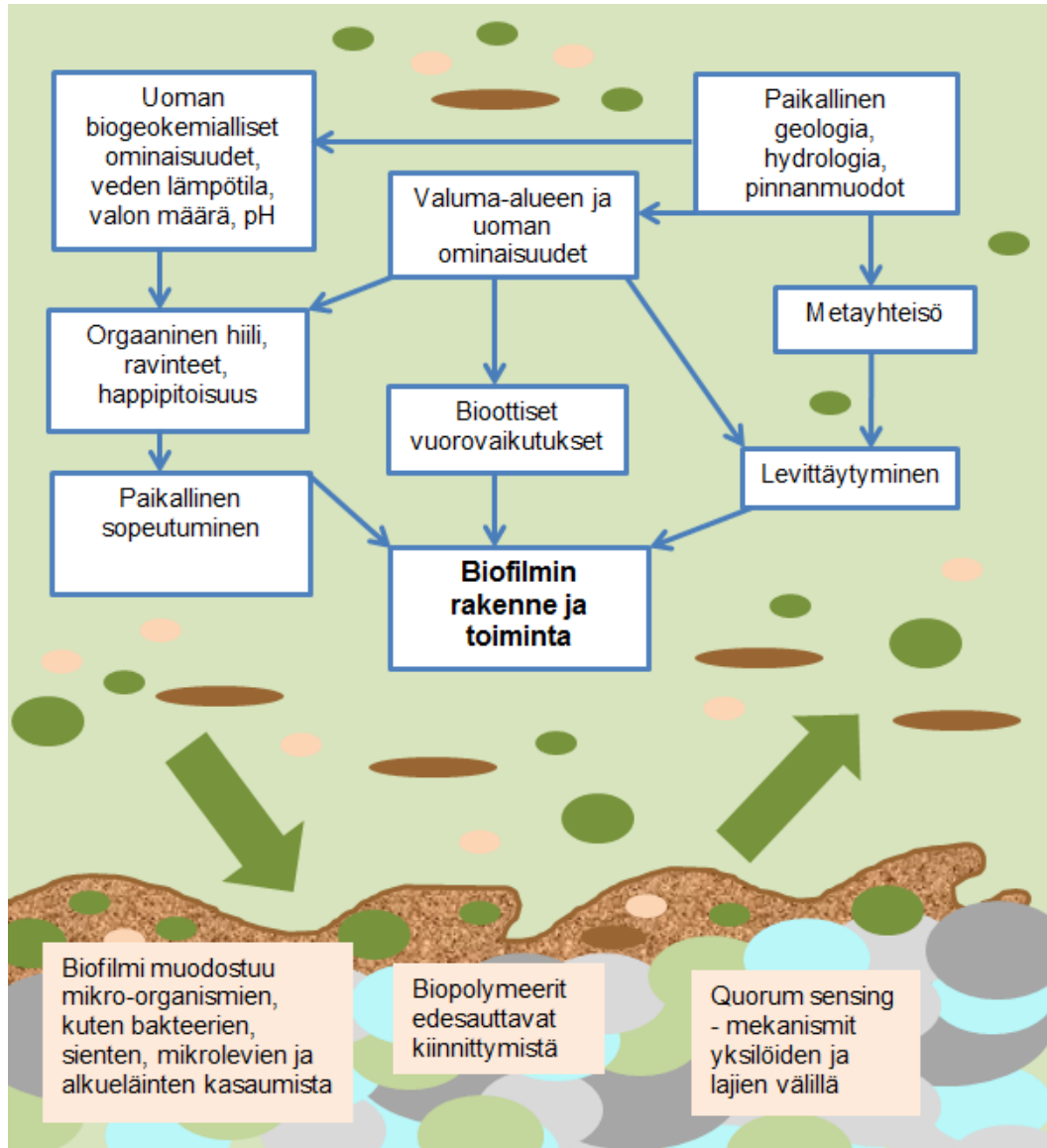
1.2.2 Biofilmin rakenne ja toiminta

Mikrobien yhteistoiminnallisista eli mutualistisista verkostoista tärkeimpiä ovat biofilmit, jotka muodostuvat pinnoilla kasvavien mikro-organismien, kuten bakteerien, sienten ja mikrolevien kasaumista (Karimi ym. 2017). Liikkuvassa nesteessä, kuten esimerkiksi virtavedessä, liikkuvat mikrobit kiinnittyvät asteittain alustaansa, esimerkiksi kiven pintaan. Kiinnittymistä edesauttavat bakteerien valmistamat biopolymeerit, jotka ovat muun muassa sokereita ja proteiineja. Biofilmin muodostuksessa eri mikrobilajit toimivat yhteistyössä muodostaen biofilmin pinnalle suojaavan kerroksen (kuva 1) (Karimi ym. 2017; Davey & O'toole 2000).

Biofilmin muodostumiseen vaikuttavat muun muassa paikallinen geologia ja hydrologia sekä valuma-alueen ja uoman ominaisuudet. Biofilmiyhteisöjen rakentumiseen vaikuttavia uomatason ominaisuuksia ovat muun muassa valon määrä, pH, veden lämpötila, ravinteet, orgaanisen hiilen määrä ja happipitoisuus (kuva 1). (Besemer 2016; Karimi ym. 2017; Davey & O'toole 2000; Kalia ym. 2013).

Biofilmien muodostumisessa tapahtuvia mikrobilajien välisiä kommunikaatiomekanismeja ei vielä tänä päivänä täysin tunneta (Karimi ym. 2017). Tiedetään kuitenkin, että bakteerilajit hyödyntävät keskinäisessä viestinnässään niin kutsuttuja ”*quorum sensing*” -kommunikaatiomekanismeja, joiden avulla bakteerisolut tuottavat ja monitoroivat signaalimolekyylejä (Cornforth ym. 2014). Solut myös pystyvät muokkaamaan käyttäytymisensä näiden signaalien perusteella. Bakteerit käyttävät ”*quorum sensing*” -signalointia sekä lajien sisäiseen, että myös lajien väliseen viestintään. Ympäristötekijät voivat

vaikuttaa ”quorum sensing” -signaalimolekyylien kautta ja aikaansaada spesifisten bakteerifenotyyppien tuotannon (Kalia ym. 2013).



Kuva 1. Biofilmin rakenteeseen ja toimintaan vaikuttavia paikallisia ja alueellisia tekijöitä..

Kuva on muokattu Besemer (2016); Karimi ym. (2017); Davey & O'toole (2000) ja Kalia ym. (2013) pohjalta. Kuvassa vihreällä, tummanruskealla ja vaaleanruskealla kuvatut pallot esittävät uomassa vapaana liikkuvia mikro-organismeja, jotka kasautuvat ja kiinnittyvät esimerkiksi uoman pohjalla olevien kivien pinnoille. Tällaisia uoman pohjan kivien pinnoille syntyneitä biofilmejä kuvastavat kuvan alaosan vaaleansiniset, vaaleanvihreät ja harmaat muodot. Vihreät nuolet kuvastavat biofilmien ja uomassa vapaana liikkuvien mikro-organismien kumpaankin suuntaan tapahtuvaa liikettä. Vuokaavio pyrkii kokoamaan yhteen kappaleissa 1.2.2 ja 1.2.3 tarkemmin käsitellyt biofilmin rakenteeseen ja toimintaan liittyvät tekijät ja käsitteet.

1.2.3 Biofilmien rakenteeseen vaikuttavat tekijät

Viime aikoina ekologian alan tutkimuksissa on yhä enemmän tarkasteltu monilajisia yhteisöjä ja yhteisöjen yhteistoimintaan vaikuttavia tekijöitä. *Metayhteisöillä* tarkoitetaan ryhmää paikallisyhteisöjä, joita yhdistää yhden tai useamman lajin yhteisöjenvälinen liikkuminen (mm. Gilpin & Hanski 1991; Hanski 1998; Leibold ym. 2004). Metayhteisöekologia yhdistää toisiinsa pienen mittakaavan populaatio- ja yhteisöekologisia kysymyksiä, joita ovat muun muassa lajien väliset vuorovaikutukset, valinta (selektio) ja stokastiset, eli sattumanvaraiset prosessit, sekä laajan mittakaavan kysymyksiä, joita ovat muun muassa leviäminen ja elinympäristön monimuotoisuus (Leibold & Chase 2017). Biologisten yhteisöjen rakenteeseen ja monimuotoisuuteen vaikuttavat niin alueelliset (esim. leviäminen) kuin paikallisetkin (elinympäristö, bioottiset vuorovaikutukset) tekijät (Leibold ym. 2004; Leibold & Chase 2017).

Alueellisten prosessien merkitys mikrobisyhteisöjen rakenteeseen on ollut viime aikoina laajan kiinnostuksen, tutkimuksen ja jopa väittelyn kohteena. Korkean levittäytyneisyyden asteen ja lyhytikäisten bakteerisukupolvien on todettu saavan aikaan sen, että ympäristötekijöiden on katsottu olevan mikrobisyhteisöjen monimuotoisuudelle tärkeämpiä kuin maantieteellisten etäisyyksien (Fierer ym. 2007).

Besemer ja kumppanit (2012) havaitsivat, että stokastinen eli sattumanvarainen levittäytyminen alkuperäisestä yläpuolisesta yhteisöstä virtavedessä vesipatsaan mukana uoman biofilmiyhteisöihin kulkeutuen ei pystynyt selittämään mikrobisyhteisön rakennetta pohjoisissa virtavesissä. Todennäköisesti paikalliset ympäristötekijät ja bioottiset, eli elolliset, vuorovaikutukset vaikuttivat uoman biofilmin muodostumiseen levittäytymistä enemmän (Besemer ym. 2012). Toisaalta Lear ja kumppanit (2013) havaitsivat levittäytymisrajoittuneisuuden ja lajin paikkaan sekä toimintoihin ekosysteemissä perustuvien prosessien vaikuttavan yhdessä uusiseelantilaisten virtavesien biofilmi bakteereiden yhteisörakenteeseen ja monimuotoisuuteen. Ympäristötekijöiden vaikutuksen nähtiin tässäkin tutkimuksessa olevan

merkittävä (Lear ym. 2013). Vastaavia havaintoa on tehty myös suomalaisissa virtaavissa vesissä (Heino ym. 2014).

Länsisuomalaisiin, tässäkin pro gradussa analysoituihin biofilmibakteereihin perustuvan tutkimuksemme tulokset puolestaan osoittavat, että mikrobiyhteisöjen alueelliseen vaihteluun vaikuttavat yhdessä paikalliset ja valuma-alueen ympäristötekijät sekä ilmastotekijät ja mahdollisesti levittäytymisrajoittuneisuus (Jyrkänkallio-Mikkola ym. 2017). Koillismaalla sijaitsevan Kitkajärven kivikkorantojen biofilmibakteeriyhteisöillä tekemämme tutkimus puolestaan osoitti järven sisäisen mikrobiyhteisön spatiaalisen vaihtelun muodostuvan niin sattumanvaraisista tekijöistä, lajin paikkaan ja toimintoihin ekosysteemissä perustuvista prosesseista kuin myös mahdollisesti levittäytymisrajoittuneisuudesta (Langenheder ym. 2017). Abioottiset, eli elottomat kemialliset ja fysikaaliset ympäristötekijät selittivät bioottisia, eli elollisia ympäristötekijöitä paremmin järven kivikkorantojen biofilmibakteereiden taksonomista ja fylogeneettistä yhteisörakennetta (Langenheder ym. 2017). Biofilmien rakenteeseen ja toimintaan vaikuttavat siten monet eri alueelliset ja paikalliset tekijät yhdessä (Besemer 2016). Näitä alueellisia ja paikallisia prosesseja on kuvattu kuvassa 1.

1.3 Mikrobiyhteisöihin vaikuttavat tärkeimmät tekijät: happamuus ja ravinnekuormitus

1.3.1 Happamuuteen vaikuttavat tekijät, happamuus ja mikrobit

Suomen rannikkoalueilla on paljon happamia sulfaattimaita, jotka ovat muodostuneet noin 7500–4000 vuotta sitten Litorinameri-vaiheen aikana. Merenpohjaan pelkistyi mikrobitoiminnan vaikutuksesta sulfideja, jotka pohjalle sitoutuneen pelkistyneen raudan kanssa reagoidessaan saostuivat pohjan sedimenttiin rautasulfideina, eli pyriittinä (FeS₂) ja monosulfidina (FeS). (Hadzic ym. 2014; Nyberg ym. 2012). Maankohoamisen seurauksena näitä vanhoja merenpohjan sedimenttejä esiintyy nykyisin Suomen rannikolla noin 100 metrin

korkeuteen asti (Kivinen 1944). Tämän tutkielman tutkimuspaikat sijaitsevat Suomen länsirannikolla happamien sulfaattimaiden vaikutuspiirissä.

Happamien sulfaattimaiden lisäksi myös mustaliuske voi aiheuttaa happamuushaittoja. Mustaliuske on eloperäistä hiiltä, rikkiä ja erilaisia metalleja sisältävä vanhasta merenpohjasta muodostunut sedimenttikivilaji, jota esiintyy esimerkiksi Sotkamon alueella. (Loukola-Ruskeeniemi & Heino 1996; Hadzic ym. 2014).

Maankäytön, kuten esimerkiksi ojitusten, seurauksena rautasulfidipitoinen maaperä voi päätyä kosketuksiin hapen kanssa ja muodostaa veteen liuetessaan rikkihappoa. Vesistöihin päätyessään happamien sulfaattimaiden aiheuttamat alhaiset pH:t ja alhaisissa pH-lukemissa liukoisessa muodossa olevat metallit voivat aiheuttaa laajamittaista haittaa vesieliöstölle. Suomessa koettiin viimeksi vuonna 2006 runsaasti happamien sulfaattimaiden aiheuttamia kalakuolemia. Osa kalalajeista voi kuolla pH:n laskiessa alle 5,5 ja niiden lisääntyminen voi häiriintyä jo pH:n ollessa alle 5,75 (Sutela ym. 2012). Happamien sulfaattimaiden olemassa olon mahdollisuus tulisikin entisen Litorinameren alueella huomioida kaikessa maankäytössä jo suunnitteluvaiheessa. Geologian tutkimuskeskus on laajamittaisesti kartoittanut happamien sulfaattimaiden esiintymisen todennäköisyyttä (Geologian tutkimuskeskus. 2018). Kansalaisille avoin karttapalvelu on saatavissa internet-osoitteesta: <http://gtkdata.gtk.fi/Hasu/index.html>

Sulfaattia tuottavat bakteerit (SRB) ovat tärkeässä roolissa erityisesti kaivosten vesien ja sedimenttien mikrobiyhteisöissä. Näissä olosuhteissa on usein alhainen pH ja paljon metalleja. SRB:t voivat osallistua metallien detoksifikaatioon muuttamalla niitä sulfidiksi (Kaksonen ym. 2008; Kovaliova ym. 2017). Suurin osa tunnetuista sulfaattia tuottavista bakteereista kuuluu *Firmicutes*-pääjakson *Desulfosporosinus*, *Desulfotomaculum* ja *Deltaproteobacteria*-sukuihin (Kovaliova ym. 2017).

1.3.2 Vesistöjen ravinnekuormitus

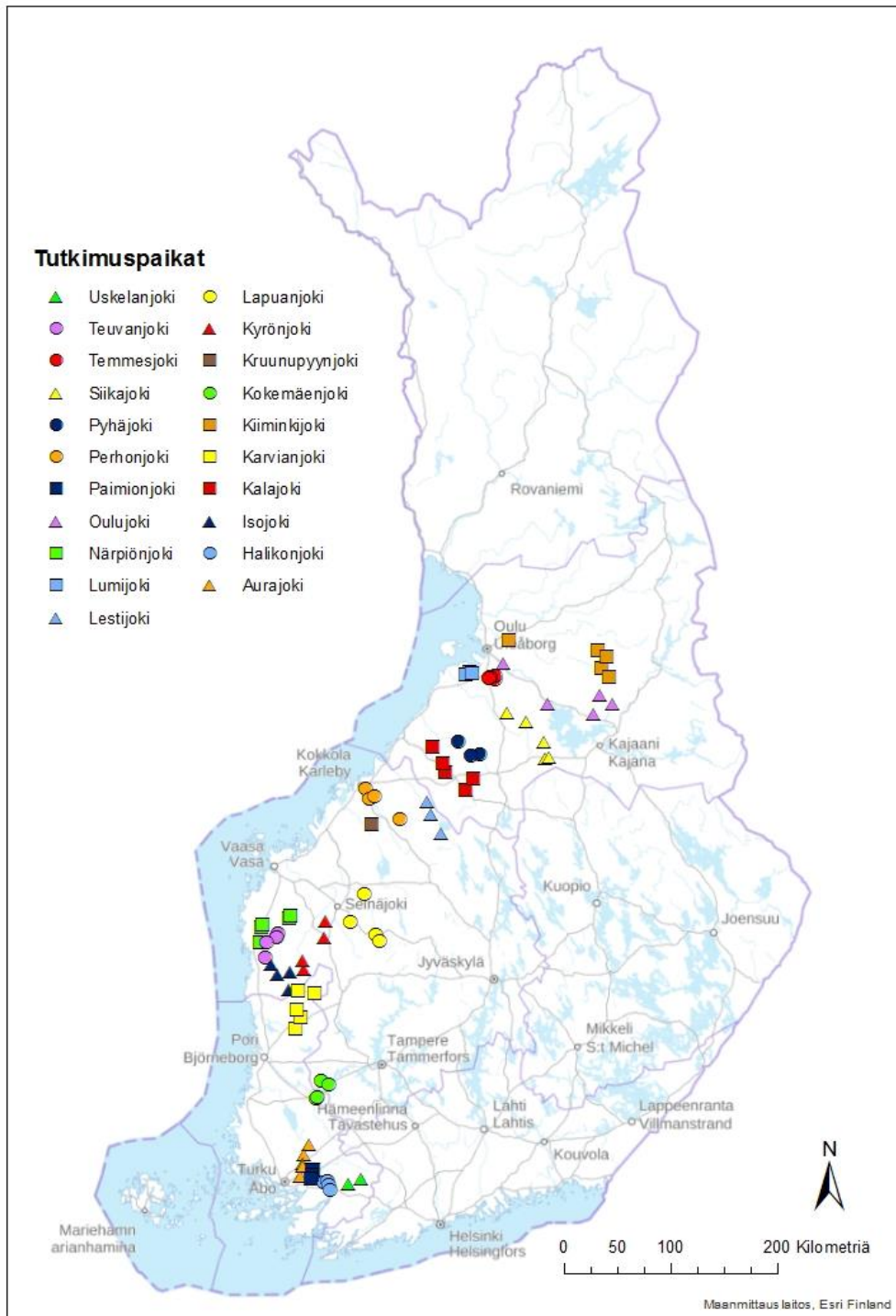
Fosfori- ja typpikuormitus heikentävät rehevöittämällä voimakkaasti vesistöjen tilaa (Thornton ym. 1999). Ravinteiden määrä vaikuttaa levätuotannon lisäksi niin vesikasvien, päällyslievien, pohjaeläimien, kalaston kuin mikrobienkin runsauteen ja lajikoostumukseen (Mitikka 2013). Jokivesien ekologisen tilan luokittelussa käytetään kokonaisfosforin ja -typen vuosikeskiarvoja. Jokityypistä riippuen erinomaisen ja hyvän ekologisen tilan raja-arvona käytetään kokonaistypen osalta 335–450 µg/l pitoisuuksia ja kokonaisfosforin osalta pitoisuuksia 15–40 µg/l (Aroviita ym. 2012). Suomen sisävedet ovat yleensä fosforirajoitteisia, eli fosforipitoisuudet rajoittavat niiden levätuotantoa. Erittäin rehevöityneet ja humuspitoiset vedet voivat olla myös typpirajoitteisia (Mitikka 2013).

2 Materiaalit ja menetelmät

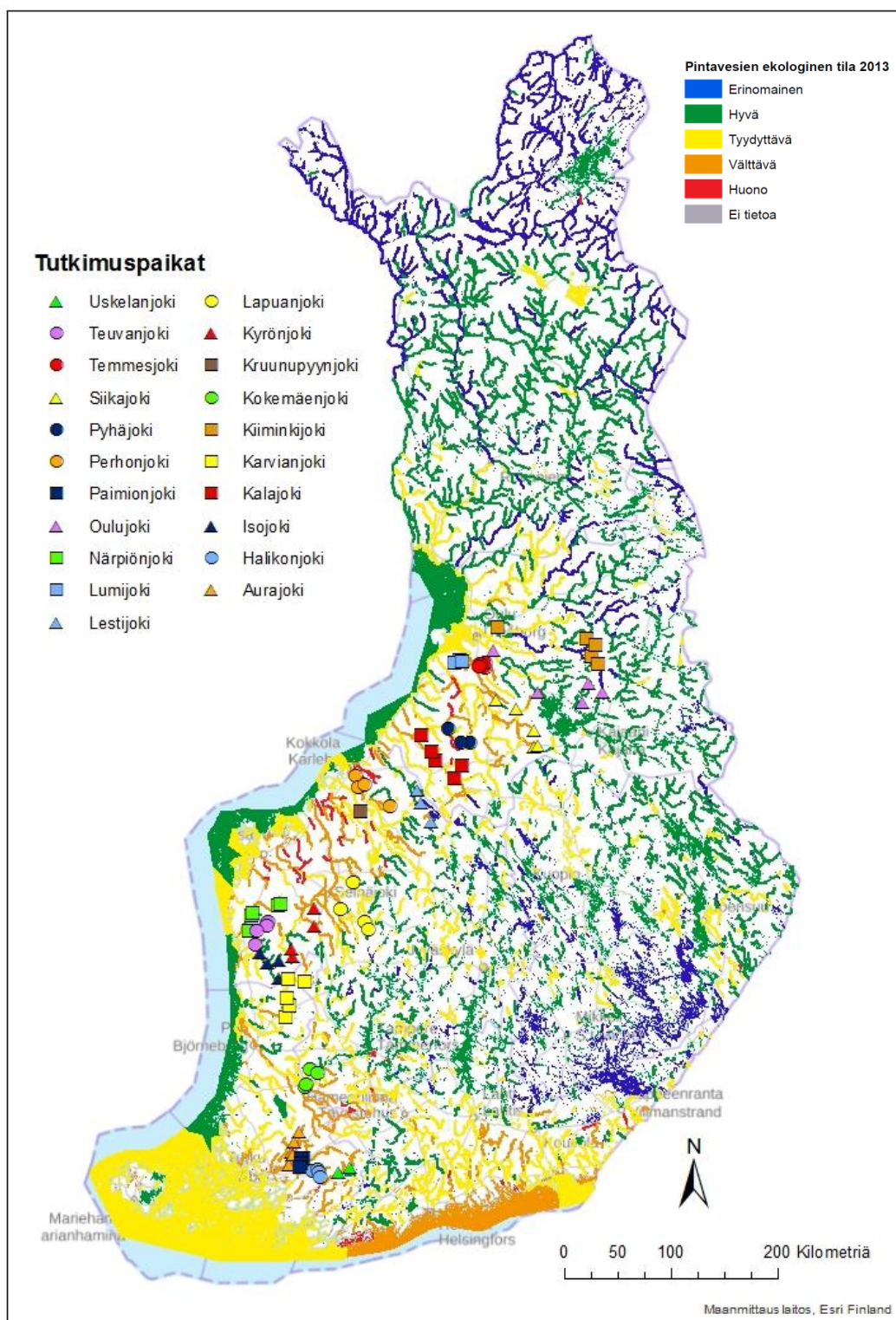
2.1 Tutkimusalue

Pro gradu tutkielmani bakteeriaineiston näytteet on otettu 21 suomalaiselta kolmannen jakovaiheen jokivaluma-alueelta, joilta eristettiin erillisistä koskipaikoista bakteerinäytteitä yhteensä 84 kappaletta. Ajallisen vaihtelun minimoimiseksi näytteenotto suoritettiin kaikissa kohteissa kahden viikon sisällä syyskuussa 2014. Tutkimuspaikat sijoittuvat Suomen länsirannikolle Turun eteläpuolelta hieman Oulun pohjoispuolelle ulottuvalla alueella. Tutkimusalue on pohjois-eteläsuunnassa 520 kilometriä ja itä-länsi -suunnassa 330 kilometriä. (Jyrkänkallio-Mikkola ym. 2017). Tutkimuspaikkojen sijainti on kuvattu kuvassa 3. Tutkimuspaikkojen sijainti suhteessa pintavesien ekologiseen tilaan on puolestaan kuvattu kuvassa 4.

Samaan aikaan bakteerinäytteiden ottamisen kanssa tutkimuskohteilla toteutettiin vesinäytteenotto, ja kohteista määritettiin myös piilevä- ja pohjaeläinlajisto. Tutkimuspaikkojen lähivaluma-alueet edustavat erilaisia maankäyttömuotoja ja esimerkiksi peltojen määrä lähivaluma-alueella vaihtelee suuresti (0-60%). Laajat, samanaikaisesti otetut vesikemia-analyysit mahdollistavat bakteerien toiminnallisten taksonomisten yksikköjen (operational taxonomic unit, OTU) ja muun muassa ravinteiden suhteen tarkastelun.



Kuva 3. Tutkimuspaikat kolmannen jakovaiheen jokivaluma-alueittain jaoteltuina. Samanväriset ja muotoiset symbolit edustava jokia, jotka kuuluvat samaan kolmannen jakovaiheen jokivaluma-alueeseen. Näytepisteet sijaitsivat latvapuroissa kullakin jokivaluma-alueella. $N = 84$.



Kuva 4. Tutkimuspaikat ja pintavesien ekologinen tila Suomessa vuonna 2013. Sininen väri edustaa erinomaista, vihreä hyvää, keltainen tyydyttävää, oranssi välttävää ja punainen huonoa pintavesien ekologista tilaa. Harmailta alueilta tiedot puuttuvat. $N = 84$ tutkimuspistettä.

2.2 Maastomenetelmät

Näytteenotto toteutettiin ajallisen vaihtelun minimoimiseksi kaikilla tutkimuspaikoilla kahden viikon aikana syyskuussa 2014. Bakterinäytteenotto, vesinäytteenotto ja habitaattianalyysit tehtiin tutkimuspaikoilla samanaikaisesti. Koska tutkimuspaikkoja oli runsaasti ja näytteet haluttiin ottaa mahdollisimman samanaikaisesti, oli maastossa yhtä aikaa kaksi erillistä maastotyöryhmää, joista toiset aloittivat näytteenoton eteläisimmistä ja toiset pohjoisimmista tutkimuspaikoista edeten kahden viikon aikana toisiaan kohti.

2.2.1 Bakterinäytteenotto

Bakterinäytteet otettiin jokaiselta tutkimuspaikalta 10 noin nyrkin kokoisen kiven yläpinnalta. Kivien keräyssyvyys oli noin 40 cm, ja ne kerättiin mahdollisimman keskeltä uomaa mahdollisen varjostusvaikutuksen välttämiseksi. Jokaiseen näytteenottoon otettiin uudet, valmiiksi pakatut steriilit näytteenottovälineet, jotka sisälsivät pienen minigrip-pussin sisään pakatut kertakäyttökäsineet, keltakorkkisen näytteenottoastian (valmistaja: Mekalasi) ja puhtaan, noin 4 cm x 4 cm x 4 cm kokoisen keittiösienen palan. Lisäksi käytettiin puhdasta pakastusrasiaa näytteenottoa varten.

Keltakorkkista astiaa avuksi käyttäen kerättiin aluksi kertakäyttöhansikkaat kädessä noin 1,5 x keltakorkkisen näyteastian tilavuuden verran paikan vettä pakasterasiaan. Seuraavaksi sientä kostutettiin ensin pakasterasiassa ja kiven pinnalta siirrettiin biofilmimateriaali pakasterasiaan niin, että kosteaa sientä hierottiin noin 6 cm x 6 cm kokoisella kiven yläpinnan alueella ja irronnut biofilmimateriaali puristettiin pakasterasiassa olevaan näytteenottopaikan veteen (Vilmi ym. 2016). Kiven pinnan sienellä harjaamista toistettiin niin kauan kunnes silmämääräisesti nähtiin kiven pinnalla olleen materiaalin siirtyneen näyteastiaan. Yllä kuvattu toistettiin kaikkien kymmenen kerätyn kiven kohdalla ja puristettiin kaikki biofilmimateriaali samaan näyteastiaan (puhdas pakastusrasia) kymmenen kiven kokoomänäytteeksi. Lopuksi kaadettiin biofilmimateriaalin sisältävä näyte keltakorkkiseen näyteastiaan ja merkittiin siihen vedenkestävällä tussilla näytteenottoaika, paikka

ja näytteenottajan nimikirjaimet. Näytteet pakastettiin välittömästi näytteenoton jälkeen autossa toimivaan kannettavaan Waeco Cool Freeze CFX35 –pakastimeen (valmistaja: Dometic WAECO).

2.2.2 Vesikemia- ja habitaattianalyysit

Vesinäytteet otettiin tutkimuspaikoilta samanaikaisesti bakteerinäytteiden kanssa. Vesinäytteistä määritettiin kokonaistyyppi, kokonaisfosfori ja väri standardien SFS-EN ISO 6878:en, SFS-EN ISO 11905-1 ja SFS-EN ISO 7887:en mukaisesti (SFS-käsikirja 147-3. 2011; Water quality. Examination and determination of colour. 2011).

Lisäksi tutkimuspaikalta mitattiin sähkönjohtavuus ja pH YSI-Professional Plus kenttämittarilla (valmistaja: YSI Incorporated). Virrannopeus mitattiin jokaiselta tutkimuspaikalta 30 eri paikasta Schiltknecht MiniAir 2 -virrannopeusmittarilla (valmistaja: Schiltknecht). Uoman leveys puolestaan mitattiin jokaiselta tutkimuspaikalta kymmenestä eri kohdasta ja uoman syvyys 30 eri kohdasta (Jyrkänkallio-Mikkola ym. 2017).

Uoman raekoko arvioitiin muokattua Wentworthin asteikkoa käyttäen. Partikkelikooltaan 0.25 mm–2 mm oleva aines luokiteltiin hiekaksi, 2 mm–16 mm kokoinen aines soraksi, 16 mm–64 mm kokoinen aines pikkukiviksi, 64 mm–256 mm kokoinen aines mukulakiviksi ja yli 256 mm kokoinen kiviaines lohkareiksi (Wentworth 1922). Lisäksi arvioitiin sammalpeittävyys uomassa ja uoman molemmilta puolilta sekä lehtipuiden määrä että latvuston peittävyys. Tarkempi kuvaus vesikemia- ja habitaattianalyyseista on luettavissa julkaisusta Jyrkänkallio-Mikkola ym. (2017).

2.3 Valuma-alueanalyysit

FT Virpi Pajunen rajasi jokaiselle tutkimuspaikalle valuma-alueen ArcGIS 10.3.1 -ohjelmalla hyödyntäen GTK:n virtaamasuunta- ja korkeusaineistoja. Valuma-alueelle määritettiin koko ja kaltevuus. Valuma-alueen maankäyttö ja -peitto määritettiin CORINE 2012 aineiston (Suomen ympäristökeskus 2013) ja

maaperäkartan (GTK 2012) avulla. Tarkempi kuvaus valuma-alueanalyyseistä on luettavissa julkaisusta Jyrkänkallio-Mikkola ym. (2017).

2.4 Laboratoriomenetelmät

Kaikessa laboratoriotyöskentelyssä kiinnitettiin huomiota aseptisiin työtapoihin. Pöytäpinnat pyyhittiin aina laboratoriotöitä ennen ja niiden jälkeen 96 % etanolilla kontaminaatoriskin pienentämiseksi. Kaikki pipetoinnit tehtiin filterillisiä pipetinkärkiä käyttäen ja suojakäsineitä käytettiin kaikissa työvaiheissa. DNA-työt suositellaan tehtäväksi laminaarivirtauskaapissa, mutta sitä ei valitettavasti tilojen puolesta ollut tässä työssä mahdollista toteuttaa.

Laboratoriossa näytteet siirrettiin välittömästi näytteenottajien maastosta palattua kannettavasta pakastimesta laboratorion pakastimeen odottamaan jatkokäsittelyä. Näytteitä pakastekylmäkuivattiin ALPHA 1-4 LD plus-pakastekylmäkuivurilla (valmistaja: Martin Christ) vuorokauden ajan ja pakastekylmäkuivatusta näytemateriaalista eristettiin PowerSoil DNA isolation -kitillä (valmistaja: MOBIO Laboratories) DNA. Eristyksen onnistuminen tarkistettiin määrittämällä DNA:n saanto SimpliNano-spektrofotometrillä (valmistaja: Biochrom Ltd.).

Näytteiden DNA monistettiin käyttäen universaaleja bakteerialukkeita 519F (5-CAGCMGCCGCGTAATWC-3) ja 926R (5-CCGTCAATTCCTTTRAGTTT-3) (Lane 1985; Lane 1991). Jokaiselle näytteelle käytettiin erilaista yksilöllisen ”viivakoodin” sisältävää 519F-aluketta (nk. a-aluke), jotta sekvensointia varten yhdistetyt näytteet voitiin aineiston jatkoanalyseissa erottaa toisistaan. 519F alukkeet sisälsivät Ion Torrent pyrosekvensointi-adapterisekvenssin A (Lifescience Technologies, USA), yhdeksän emäsparin mittaisen yksilöllisen ”viivakoodisekvenssin” ja yhden nukleotidilinkkerin. 926R aluke (nk. b-aluke) puolestaan sisälsi Ion Torrentin adapterisekvenssin trP1 (Tejesvi ym. 2016). Käytetyt alukkeet on kuvattu liitteessä 1. Testasimme alukkeet etukäteen SILVA-tietokannassa (Quast ym. 2013) ja saimme tulokseksi noin 94 % kattavuuden kaikille bakteeritaksoneille. Tarkempi kuvaus on luettavissa julkaisusta Heino ym. (2015).

PCR (Polymerase chain reaction) -reaktioita varten valmistettiin niin kutsuttu mastermix-liuos, johon pipetoitiin 5,0 µl GC-puskuria (Thermo Scientific), 0,5 µl dNTP:tä (Thermo Scientific), 1 µl b-aluketta, 0,5 µl Phusion hot start polymeraasia (Thermo Scientific) ja 13 µl molekyylibiologisiin töihin soveltuvaa steriiliä, nukleaasi-vapaata vettä (VWR International).

Mastermix-liuosta valmistettiin aina sen päivän tarvetta vastaava määrä. Jokaiseen kuppalevyn kuoppaan pipetoitiin 20 µl mastermix-liuosta, 1 µl näytespesifistä a-aluketta ja 4 µl eristettyä DNA-näytettä. Jokaisen kuoppalevyn kuopan kokonaistilavuudeksi tuli näin ollen 25 µl. Kaikki PCR-vaiheen pipetoinnit suoritettiin laakeassa astiassa johon oli lisätty jäitä ja pipetointiin käytettiin erillisiä vain PCR-töihin tarkoitettuja pipettejä kontaminaatoriskin vähentämiseksi.

Bakteerit monistettiin Veriti Thermal Cycler -PCR-laitteella (Life technologies). Kustakin näytteestä pipetoitiin 96-kuoppalevylle pipetointiepätarkkuuksien minimoimiseksi triplikaatit. PCR-ohjelman jälkeen triplikaatit yhdistettiin yhdeksi näytteeksi.

Bakteerien monistamisessa käytin seuraavaa PCR-ohjelmaa: Vaihe 1, tehtiin kerran (askel 1, 98°C, 3:00 min). Vaihe 2, tehtiin 30 kertaa (askel 1, 98°C, 0:10 min, askel 2, 64°C, 0:30 min, askel 3, 72°C, 0:20 min). Vaihe 3, tehtiin kerran (askel 1, 72°C, 5:00 min, askel 2, 4°C, ∞).

Monistamisen jälkeen PCR:ien onnistuminen tarkistettiin Shimadzun valmistamalla MultiNA (MCE-202 Microchip Electrophoresis System for DNA/RNA Analysis) -laitteella. MultiNA on uuden polven mikrosiruelektroforeesilaitte, jolla voidaan korvata perinteinen agarosigeelielektroforeesiajo näytteiden tarkistamiseksi (MultiNA 2017).

Näytteet puhdistettiin, yhdistettiin sekvensointia varten ja sekvensoitiin Oulun yliopiston molekyylibiologian (BioSer) -laboratoriossa Ion Torrent Semiconductor -sekvenssaattoria käyttäen. Aineisto sekvensoitiin Ion 316™ Chip -sirua käyttäen. Siru sisältää 6 miljoonaa kaivoa ja sen ulosanto on 100 Mb (Ion Torrent 2012).

Ion Torrent sekvenssaattorin toiminta perustuu DNA:n polymerisaation yhteydessä vapautuvien vetyionien havaitsemiseen. Sekvensointiprosessi

aloitetaan emulsio-PCR:lla, jossa DNA-kirjaston fragmentit kiinnittyvät pienen mikrohelmiä pinnalle. Tämän jälkeen mikrohelmet pipetoidaan sekvensointisirulle, jossa sijaitsee miljoonia mikrokaivoja. Mikrokaivoihin puolestaan lisätään sekvensoinnin aikana vuorotellen neljää DNA -nukleotidia. Sopivien komplementaaristen emästen löytyminen aiheuttaa vetyionin vapautuessa pienen pH -muutoksen, jonka sekvenssaattori havaitsee. Jos DNA-juosteessa on peräkkäin kaksi identtistä emästä, on jännite kaksinkertainen ja sekvenssaattori tunnistaa ne kahdeksi peräkkäiseksi samaksi emäkseksi. Ion Torrentin teknologia siirtää kemiallisen informaation (A,C,G,T) suoraan digitaaliseen (0,1) muotoon. Yllä kuvattu reaktio tapahtuu yhtä aikaa miljoonissa mikrokokoisissa kaivoissa. Tästä johtuen Ion Torrent Semiconductor -sekvensointi on huomattavasti perinteisiä sekvensointimenetelmiä nopeampi ja sekvensointi kestää vain muutaman tunnin (Glenn 2011; Ion Torrent 2012; Rothberg ym. 2011).

2.5 Bioinformatiikan menetelmät

Sekvensointiaineistossa raakasekvenssien määrä oli 2 708 611. Sekvenssiaineisto analysoitiin QIIME (Quantitative Insights Into Microbial Ecology) -ohjelmalla (Caporaso ym. 2010). Sekvensointikirjasto jaettiin näytteisiin ja huonolaatuiset, laatuksittomat läpäisemättömät sekvenssit poistettiin. Esimerkiksi kaikki sekvenssit, jotka alittivat asetetun laatuarvon 25 tai olivat lyhempiä kuin 200 emäsparia tai pidempiä kuin 1000 emäsparia poistettiin. Laatuksittomien sekvenssien minimipituus oli 200 emäsparia, maksimipituus 501,0 emäsparia ja pituuden keskiarvo 341,4 emäsparia. Laatuksittomien sekvenssien määrä oli yhteensä 549548. Näytteiden välinen sekvenssimäärä vaihteli huomattavan paljon, 20505 ja 2219 välillä, joten aineisto rarefoitiin alimman sekvenssimäärän (2219) sisältäneen näytteen perusteella. Tämän jälkeen sekvenssit klusteroitiin toiminnallisiksi taksonomisiksi yksiköiksi (OTUiksi) 97 % samankaltaisuusastetta käyttäen usearch61-algoritmilla (Edgar 2010), ja OTUille haettiin BLASTia (basic local alignment search tool) (Altschul ym. 1990) käyttäen paras mahdollinen sukutason vastaavuus. BLAST-analyysit tehtiin myös

pääjakso-, luokka-, lahko-, ja heimotasoisille, mutta kaikissa tilastanalyysissä käytettiin sukutason vastaavuudelle analysoituja toiminnallisia taksonomisia yksiköitä (OTUja).

2.6 Tilastomenetelmät

Kaikki tilastanalyysit tehtiin R -ympäristössä (R Core Team 2013) käyttäen R:n versioita 2.15.3 ja 3.3.2 ja R:n paketteja vegan 2.0-6 (Oksanen ym. 2013), PCNM (Legendre ym. 2012), Rcmdr (Fox 2005), BiodiversityR (Kindt 2013), labdsv (Roberts 2016), mvpart (Therneau ym. 2004), MVPARTwrap (Ouellette, M. H., & Legendre, P. 2013)

2.6.1 Selittävät muuttujat

Aineiston selittävät muuttujat jaettiin neljään ryhmään: vesikemiamuuttujiin, habitaattimuuttujiin, valuma-alueuuttujiin ja spatiaalimuuttujiin.

Vesikemiamuuttujina analyysissä olivat kokonaistyyppi, kokonaisfosfori, väri, pH ja sähkönjohtavuus. Valuma-alueuuttujina puolestaan olivat alueen koko, kaltevuus, rakennetut alueet, maatalousmaat, harvapuustoiset metsät, havupuumetsät, sekametsät, pensaikot, avoimet alueet, kosteikot, vesimuodostumat, kallioid, karkearakeinen maa-aines, moreeni, hienojakoinen maa-aines, savi, lieju ja turve. Habitaattimuuttujina olivat uoman leveys, varjostus, virrannopeus, uoman syvyys, sammalpeittävyys, hiekka, sora, pikkukivet, mukulakivet ja lohkarit.

Perinteisesti kahden etäisyysmatrisiin (esimerkiksi geneettinen ja maantieteellinen etäisyys) välisen korrelaation mittaamiseen on käytetty Mantelin testiä (Mantel 1967). Viime aikoina tutkimukset ovat kuitenkin osoittaneet, ettei Mantelin testiä tulisi käyttää spatiaalisen rakenteen tai korrelaation mittaamiseen monimuuttujaisten tutkimusaineistojen (yhteisörakenne ja abiottiset, eli elottomat, parametrit) välillä (Guillot & Rousset 2013), vaan tällaisten aineistojen kohdalla tulisi käyttää etäisyysperusteisia Moranin ominaisvektorikarttoja (Distance-based Moran's Eigenvector Maps, db-MEM) (Dray ym. 2006).

Lajien havaittu spatiaalinen levinneisyys voi kohota lukuisten sisä- ja ulkosyntyisten prosessien, kuten lajien välisten vuorovaikutusten, kasvun, yhteisöjen dynamiikan tai fysikaalisten tekijöiden vuoksi yhtä aikaa useilla eri spatiaalisilla tasoilla (Václavík ym. 2012). Tällaisten yhtä aikaa eri skaaloilla tapahtuvien prosessien ymmärtämiseksi tarvitaan matemaattinen työkalu (Dray ym. 2012). Tilastotieteen menetelmistä tähän tarkoitukseen hyvin soveltuvat Drayn ja kumppaneiden 2006 kehittämät etäisyysperusteiset Moranin ominaisvektorikartat (Moran Eigenvector Maps, db-MEM) ja niiden alkuperäinen muoto, Borcardin ja Legendren (2002) kehittämä PCNM (Principal Coordinates of Neighbor Matrices) (Dray ym. 2006; Borcard & Legendre 2002). Etäisyysperusteiset Moranin ominaisvektorikartat muodostetaan spektrisen graafiteorian avulla, ja ne kuvaavat tutkimuspaikkojen välistä etäisyyttä (Dray ym. 2006). Spatiaalimuuttujat luotiin käyttämällä etäisyysperusteisia Moranin ominaisvektorikarttoja (db-MEM, Dray ym. 2006) R-ohjelman PCNM-paketin (Legendre ym. 2012) avulla.

Pienin puu (minimum spanning tree), joka yhdistää kaikki tutkimuspisteet, on suuntaamaton verkko, jonka kahden pisteen (tässä tapauksessa tutkimuspaikkojen) välillä on yksi polku. Suuntaamattoman verkon $G = (V, E)$ puu on mikä tahansa puu (V, T) , missä $T \subseteq E$. Pienin puu muodostetaan joukosta E käyttäen mahdollisimman vähän kaaria T , kuitenkin siten että syntyvä verkko pysyy yhtenäisenä ja syntyvän puun paino on pienin mahdollinen (Gower & Ross 1969; Cormen ym. 2009) Pienimmän puun laskemiseksi on esitetty useita erilaisia algoritmeja, esimerkiksi Primin algoritmi (Jarník, 1931) ja Kruskalin algoritmi (Kruskal, 1956). Tutkimuspaikkojen välistä etäisyyttä kuvaava, R-ohjelman PCNM-paketilla (Legendre ym. 2012) luotu pienin puu (minimum spanning tree) on kuvattu liitteessä 2.

Spatiaalimuuttujat voidaan jakaa laajan mittakaavan spatiaalisia rakenteita kuvastaviin (esimerkiksi V1 ja V2) ja pienen mittakaavan spatiaalisia rakenteita kuvastaviin (esimerkiksi V14 ja V15) muuttujiin. Spatiaalimuuttujat muodostettiin Moranin etäisyyteen perustuvien ominaisvektorikarttojen avulla R-ohjelman PCNM-paketilla (Legendre ym. 2012). Vain merkitsevää positiivista spatiaalista

autokorrelaatiota kuvaavat PCNM-muuttujat otettiin jatkotarkasteluun mukaan (Borcard ym. 2011). Positiivisia PCNM-muuttujia olivat muuttujat V1–V15.

2.6.2 Redundanssianalyysi ja vaihtelun ositus

Tarkastelin mikrobiyhteisön rakenteeseen vaikuttavia tekijöitä redundanssianalyysin (redundancy analysis, RDA) (Rao 1964) avulla. RDA:ta käytetään paljon erilaisiin yhteisöekologisiin analyysihin. Esimerkiksi Wang ja kumppanit käyttivät RDA-analyysia tutkiessaan mikrobien phylogeneettisten klusterien ja ympäristön korkeusgradientin suhdetta (Wang ym. 2012). Sun kumppaneineen puolestaan käytti RDA-analyysia tutkiessaan jokisuualueen mikrobien yhteisörakennetta (Sun ym. 2014).

RDA on pääkomponenttianalyysin (PCA) muoto, joka tutkii samaan aikaan sekä ympäristömuuttujia että lajiaineistoja. Legendre & Gallagher (2001) toteavat, että yhteisöekologiassa analyysijä suositellaan usein tehtäväksi kanonisella korrespondenssianalyysillä (canonical correspondence analysis, CCA) (ter Braak 1986) redundanssianalyysin sijaan, mutta molemmissa menetelmissä on omat heikkoutensa. RDA:ssa lajit voivat 'korvautua' toisilla lajeilla, josta voi aiheutua virheellisiä nolla-arvoja aineistoon. CCA:n heikkoutena puolestaan on se, että harvinaiset lajit voivat siinä saada liian suuren painoarvon (Legendre & Gallagher 2001). Aineistolle tehtiin yllä kuvatun ongelman välttämiseksi Legendre & Gallagherin (2001) kuvaama Hellingerin muunnos R-ohjelman decostand-toiminnolla. Hellingerin muunnos muuntaa unimodaalin aineiston RDA-analyysille ja vaihtelun ositukselle sopivaan muotoon (Legendre & Gallagher 2001).

Lopullisiin tilastoanalyysihin valittiin selittävät muuttujat muokatun eteenpäin askeltavan (forward selection) mallin avulla, jonka ovat kehittäneet Blanchet ym. (2008). Eteenpäin askeltavaa mallia käytetään usein ekologian alan tutkimuksissa muuttujien valinnassa. Menetelmän suurimpia etuja on sen kyky toimia myös sellaisissa tilanteissa jolloin selittäviä muuttujia on tutkimuspaikkoja enemmän (Blanchet ym. 2008). Menetelmän heikkous on kuitenkin sen taipumus yliarvioida selitettävää vaihtelua.

Tavallisesta eteenpäin askeltavasta mallista poiketen muokatussa eteenpäin askeltavassa mallissa (Blanchet ym. 2008) tehdään muuttujien valinnassa aluksi niin kutsuttu globaalimalli, jossa on mukana kaikkien muuttujaryhmien muuttujat mukana. Globaalimallin ollessa merkitsevä, jatketaan muuttujien valintaan suoraan. Muuttujien valinta aloitetaan tyhjästä mallista, johon lisätään selitysastetta eniten parantava muuttuja. Muuttujien valinta tehdään kahden kriteerin, p-arvon ($\alpha = 0.05$) ja korjatun selitysasteen (adjusted R²) avulla. Muuttuja valikoituu mukaan malliin, jos sen lisäämisen jälkeen mallin p-arvo on pienempi kuin 0.05 ja korjattu selitysaste on pienempi kuin globaalimallin selitysaste. Näihin kriteereihin sopimattomat muuttujat hylätään muokatussa eteenpäin askeltavassa mallista ja estetään näin selitettävän vaihtelun yliarviointi (Blanchet ym. 2008).

Yhteisöaineistolle tehtiin vaihtelun ositus (variation partitioning) (Borcard ym. 1992). Vaihtelun ositusta käyttämällä saadaan kokonaiskuva siitä, kuinka hyvin valittujen muuttujaryhmien avulla voi selittää vastemuuttujia. Vaihtelun ositus tarkastelee sitä, millainen osuus selitettävän muuttujan vaihtelusta selittyy eri selittävien muuttujaryhmien yksittäisillä ja yhteisillä suhteellisilla osuuksilla ja mikä selittämättömän vaihtelun määrä on (Borcard ym. 1992).

Vaihtelun osituksella yhteisön rakennetta arvioitaessa on erittäin tärkeää suhteuttaa malliin lisättyjen muuttujien määrä havaintojen määrään. Tämä voidaan tehdä käyttämällä korjattuja selitysasteita (adjusted R²) (Ezekiel 1930), jotka huomioivat mallissa olevien muuttujien määrän ja antavat näin ollen luotettavampia tuloksia. Eroavaisuudet korjaamattomien ja korjattujen selitysasteiden välillä riippuvat aina malliin lisättyjen muuttujien määrästä (Peres-Neto ym. 2006). RDA-analyysi ja vaihtelun ositus tehtiin R-ohjelman vegan-paketilla (Oksanen ym. 2013).

2.6.3 Monimuuttujaregressiopuut

Monimuuttujaregressiopuiden (Multivariate regression trees, MRT) (De'Ath, 2002) avulla voidaan tutkia, kuvailla ja ennustaa yhteisön ja ympäristön suhdetta. MRT muodostaa klustereita tutkimuspaikoista toistamalla aineiston jakamista klustereihin uudelleen ja uudelleen. Jokainen näin muodostettava klusteri perustuu yksinkertaiseen ympäristömuuttujista valittuun sääntöön. Muodostuvan klusterin sisällä paikkojen väliset erot pyritään pitämään mahdollisimman pieninä. Analyysin tuloksena syntyy graafinen ”puu”, joka kuvaa tutkimuspaikkojen ja ympäristömuuttujien välistä suhdetta. Jokainen aineiston onnistunut jako klustereihin on määritelty jonkin selittävän muuttujan kynnysarvolla

MRT:n avulla voidaan analysoida myös sellaisia monimutkaisia ekologisia aineistoja, joista esimerkiksi puuttuu joitain arvoja, tai joiden muuttujien väliset suhteet eivät ole lineaarisia. MRT-analyysien avulla voidaan myös ennustaa yhteisökoostumusta sellaisilla tutkimuspaikoilla, joilta on saatavissa vain ympäristömittauksia ((De'Ath, 2002; Legendre & Legendre 2012)

Myös MRT -analyseissa käytettiin R-ohjelman decostand-toiminnon avulla Hellinger-muunnettua (Legendre & Gallagherin 2001) OTU-aineistoa. Hellinger-muunnos on kuvattu tarkemmin kappaleessa 2.6.1. MRT-analyysit tehtiin R-ympäristössä (R Core Team 2013) mvpart (Therneau ym. 2004) ja MVPARTwrap (Ouellette & Legendre 2013) paketteja käyttäen.

2.6.4 OTU -aineisto

Sekvensoiduista, alimman sekvenssimäärän (2219) sisältäneen näytteen perusteella rarefoidusta bakteeriaineistosta tilasto-analyyseissä käytettiin L6-tasolle eli tarkimmalle mahdolliselle tasolle (nk. sukutaso) BLAST-ohjelman avulla 97 % samankaltaisuusasteella määritellyjä bakteeri-OTUja. OTU-aineistoille tehtiin Hellingerin muunnos R-ohjelman decostand-toiminnolla, jotta aineisto saatiin muunnettua jatkoanalyysille sopivaksi (Legendre & Gallagher 2001).

2.6.5 Harvinaisten OTUjen poisto

Harvinaisten OTUjen merkitystä tutkittiin poistamalla aineistosta OTUja, jotka esiintyivät vain yhdellä, kahdella, kolmella, neljällä, viidellä, 10, 15, 20, 25, 30, 35 tai 40 paikalla. OTU-aineiston muokkaus tehtiin R-ohjelmann labdsv-paketin (Roberts 2016) dropspc-funktiolla. Erilaisia OTU-aineistoja muodostettiin näin kaksitoista erilaista (taulukko 1).

Kaikille edellä kuvatulla tavalla muodostetuille OTU-aineistoille tehtiin tämän jälkeen selittävien muuttujien (vesikemia-, habitaatti-, valuma-alue ja spatiaalimuuttujat) valinta R-ohjelmann vegan-paketin ordiR2step -toiminnolla (Oksanen 2013). Tämän jälkeen kaikille OTU-aineistoille tehtiin vaihtelun ositus käyttämällä vegan-paketin varpart-toimintoa (Borcard ym. 1992).

3 Tulokset

3.1 Näytteenottoajankohdan sää

Syyskuu 2014 oli hieman tavanomaista lämpimämpi. Kuukauden ylin havaittu lämpötila oli 23,1 astetta ja se mitattiin Kokemäellä 8.9.2014. Osassa maata terminen syksy alkoi myöhemmin kuin tavallisesti. Kuukauden sademäärät jäivät monin paikoin keskimääräistä pienemmiksi (Mäntykangas 2014). Pienillä sademäärillä ja keskimääräistä korkeammilla lämpötiloilla voi olla hieman vaikutusta tutkielman tulosten yleiseen edustavuuteen.

3.2 Ympäristömuuttujien tunnuslukuja

Tilasto-analyyseissa ympäristömuuttujina käytettiin tutkimuspaikoilta yhtä aikaa bakteerinäytteiden kanssa otettuja vesikemia- ja habitaattianalyysejä sekä ArcGIS 10.3.1 -ohjelmalla rajattujen tutkimuspaikkojen valuma-alueiden tietoja (ks. tarkemmin kappaleet 2.2.2 ja 2.3). Tilastoanalyyseissa käytettyjen ympäristömuuttujien tunnuslukuja on koottu taulukkoon 1 ja tarkasteltu seuraavissa kappaleissa. Kaikkien tutkimuspaikkojen vesikemiatulokset puolestaan on kuvattu liitteessä 4.

Taulukko 1. Ympäristömuuttujien tunnuslukuja

Muuttuja	Minimi	Maksimi	Keskiarvo	Keskihajonta
Vesikemiamuuttujat				
kokonaistyyppi µg/l	150	2763	897	549
kokonaisfosfori µg/l	0,07	357	74,8	60
väri (mgPt/l)	15	375	130	81
pH	5,95	8,07	7,20	0,51
sähkönjohtavuus µS/cm	13	619	91	84
Habitaattimuuttujat				
uoman leveys (m)	0,7	15,5	3,8	2,5
varjostus (%)	0,00	98,0	39,5	24,2
virrannopeus (m/s)	0,04	0,5	0,2	0,1
uoman syvyys (m)	0,06	0,4	0,2	0,1
sammalpeittävyys (%)	0,00	89,5	25,7	25,9
hiekkä (raekoko 0.25 mm–2 mm, %)	0,00	90,0	11,1	17,9
sora (2 mm–16 mm, %)	0,00	39,0	8,2	8,9
pikkukivet (16 mm–64 mm, %)	0,00	65,5	14,4	14,0
mukulakivet (64 mm–256 mm, %)	0,00	64,5	25,7	16,9
lohkareet (yli 256 mm, %)	0,00	78,0	26,7	20,3
Valuma-alueuuttujat				
alueen koko (km ²)	2,5	731,8	57,7	95,8
kaltevuus (% valuma-alueen keskiarvo)	3,4	36,8	13,6	7,4
rakennetut alueet (% valuma-alueesta)	0,00	7,18	2,00	1,67
maatalousmaat (% valuma-alueesta)	0,00	60,9	16,0	15,0
harvapuustoiset metsät (% valuma-alueesta)	0,112	8,85	1,75	1,57
havupuumetsät (% valuma-alueesta)	19,1	83,6	45,9	11,8
sekametsät (% valuma-alueesta)	3,5	31,2	14,2	6,2
pensaikot (% valuma-alueesta)	5,6	29,0	14,0	4,2
avoimet alueet (% valuma-alueesta)	0,00	0,51	0,05	0,05
kosteikot (% valuma-alueesta)	0,00	28,3	4,9	5,5
vesimuodostumat (% valuma-alueesta)	0,00	13,6	1,2	2,5
kalliot (% valuma-alueen maaperästä)	0,00	78,0	12,4	15,5
karkearakeinen maa-aines (% valuma-alueen maaperästä)	0,00	86,9	8,7	14,1
moreeni (% valuma-alueen maaperästä)	0,00	94,2	42,8	23,8
hienojakoinen maa-aines (% valuma-alueen maaperästä)	0,00	63,3	5,9	12,5
savi (% valuma-alueen maaperästä)	0,00	81,2	11,2	21,0
lieju (% valuma-alueen maaperästä)	0,00	18,2	0,5	2,2
turve (% valuma-alueen maaperästä)	0,00	58,5	16,9	12,5

3.2.1 pH

pH vaihteli tutkimuspaikoilla 5,95–8,07 välillä. Keskiarvo-pH oli 7,20 (taulukko 1).

3.2.2 Kokonaisfosfori

Tutkimuspaikkojen kokonaisfosfori vaihteli 0,07–357 µg/l välillä ja kokonaisfosforitulosten keskiarvo oli 74,8 µg/l (taulukko 1). Korkeimmat fosforipitoisuudet havaittiin Uskelanjoen (208 µg/l), Kyrönjoen (242 µg/l) ja Halikonjoen (357 µg/l) valuma-alueella sijaitsevilla yksittäisillä tutkimuspaikoilla. Alhaisimmat kokonaisfosforipitoisuudet tavattiin neljällä Kiiminkijoen valuma-alueella (0,07 µg/l, 0,6 µg/l, 3,3 µg/l ja 7,0 µg/l) ja kahdella Oulujoen (2,7 µg/l, 6,1 µg/l) valuma-alueella sijaitsevalla tutkimuspaikalla.

3.2.3 Kokonaistyyppi

Tutkimuspaikkojen kokonaistyyppi vaihteli Kiiminkijoen valuma-alueella sijaitsevalla tutkimuspaikalla havaitusta 150 µg/l pitoisuudesta Temmesjoen valuma-alueella havaittuun 2763 µg/l pitoisuuteen. Kokonaistyyppimääritysten keskiarvo oli 897 µg/l (taulukko 1).

3.2.4 Sähkönjohtavuus

Sähkönjohtavuus vaihteli tutkimuspaikoilla Siikajoen valuma-alueella havaitusta 13,2 µS/cm Temmesjoen valuma-alueella havaittuun 619 µS/cm. Havaittujen sähkönjohtavuusarvojen keskiarvo oli 91 µS/cm (taulukko 1).

3.2.5 Väri

Tutkimuspaikkojen väriarvot vaihtelivat Kiiminkijoen valuma-alueella havaitusta 15 mgPt/l Aurajoen valuma-alueella havaittuun 375 mgPt/l arvoon. Väriarvojen keskiarvo oli 130 mgPt/l (taulukko 1).

3.2.6 Valuma-alueuuttajat

Tutkimuspaikkojen yläpuolisten valuma-alueiden koko vaihteli Halikonjoen tutkimuspaikan 2,5 km² ja Kyrönjoen tutkimuspaikan 732 km² välillä. Tutkimuspaikkojen yläpuolisten valuma-alueiden keskiarvo oli 57,7 km². Esimerkiksi havupuumetsien prosenttiosuus tutkimuspaikkojen yläpuolisilla valuma-alueilla vaihteli 19,1 % ja 83,6 % välillä ja sekametsien 3,5 % ja 31,2 % välillä. Yksityiskohtaisemmin valuma-alueuuttajien tunnuslukuja on kuvattu taulukossa 1.

3.2.7 Habitaattimuuttajat

Uoman leveys vaihteli tutkimuspaikkojen välillä 0,7 metrin ja 15,5 metrin välillä. Sammalpeittävyys puolestaan oli suurimmillaan 89,5 % ja pienimmillään sammalia ei havaittu uoman pohjalla ollenkaan. Varjostus vaihteli tutkimuspaikkojen välillä paljon, suurimmillaan se oli 98,0 % ja pienimmillään varjostusta ei havaittu lainkaan. Yksityiskohtaisemmin habitaattimuuttajien tunnuslukuja on kuvattu taulukossa 1.

3.3 Harvinaisten lajien poiston merkitys

3.3.1 Erilaisten OTU-taulukkojen muodostaminen sukujen esiintyvyyden perusteella

Harvinaisten lajien poiston merkityksen selvittämiseksi R-ohjelman labdsv-paketin (Roberts 2016) dropspc-funktiolla muodostettiin erilaiset OTU-taulukot. Muodostetut taulukot, niille tehty käsittely ja OTUjen määrä on kuvattu taulukossa 2. Yhteensä erilaisia OTU –taulukoita muodostettiin 13 kappaletta: Taulukko, jossa kaikki OTUt olivat mukana, sekä taulukot, joista oli poistettu OTUt, jotka esiintyvät vain 1- 5 paikalla, 10 paikalla, 15 paikalla, 20 paikalla, 25 paikalla, 30 paikalla, 35 paikalla ja 40 paikalla. OTUjen määrä analyyseissa pieneni alkutilanteen 753 OTUsta harvinaisten sukujen poistojen myötä tasaisesti ollen 116 OTUa dropspc- funktiolla muodostetusta taulukossa, josta oli poistettu kaikki sellaiset OTUt, jotka esiintyvät vain ≤ 40 paikalla, eli noin puolella tutkimuspaikoista (taulukko 2).

Taulukko 2. Analyyseissä käytetyt OTU-taulukot. Hellinger-muunnos tehtiin kullekin osaineistolle erikseen.

OTU- taulukon nimi	Käsittely	OTUjen määrä analyyseissa
species.hel	-	753
species.hel.drop1	Poistettu OTUt, jotka esiintyvät vain yhdellä paikalla	585
species.hel.drop2	Poistettu OTUt, jotka esiintyvät vain \leq kahdella paikalla	498
species.hel.drop3	Poistettu OTUt, jotka esiintyvät vain \leq kolmella paikalla	433
species.hel.drop4	Poistettu OTUt, jotka esiintyvät vain \leq neljällä paikalla	393
species.hel.drop5	Poistettu OTUt, jotka esiintyvät vain \leq viidellä paikalla	371
species.hel.drop10	Poistettu OTUt, jotka esiintyvät vain ≤ 10 paikalla	282
species.hel.drop15	Poistettu OTUt, jotka esiintyvät vain ≤ 15 paikalla	229
species.hel.drop20	Poistettu OTUt, jotka esiintyvät vain ≤ 20 paikalla	196
species.hel.drop25	Poistettu OTUt, jotka esiintyvät vain ≤ 25 paikalla	164
species.hel.drop30	Poistettu OTUt, jotka esiintyvät vain ≤ 30 paikalla	143
species.hel.drop35	Poistettu OTUt, jotka esiintyvät vain ≤ 35 paikalla	131
species.hel.drop40	Poistettu OTUt, jotka esiintyvät vain ≤ 40 paikalla	116

3.3.2 Selittävien muuttujien valinta

Kaikille OTU-aineistoille tehtiin tämän jälkeen selittävien muuttujien (vesikemia, habitaatti-, valuma-alue- ja spatiaalimuuttujat) valinta R:n vegan paketin ordiR2step -toiminnolla (Oksanen 2013).

3.3.3 Millaisia vaikutuksia harvinaisten lajien poistolla havaitaan?

Tuloksista havaitaan, että PCNM-spatiaalimuuttujista malliin valikoituvat aina V2, V8, V4 ja V1 huolimatta siitä ovatko mukana kaikki bakteerisuvut vai ovatko kaikki ne, jotka esiintyvät vain 40 paikalla poistettu.

Myös valuma-aluemuuttujat pysyvät lähes samoina. Lähes kaikilla bakteerisukujen poistokäsittelyillä malliin valikoituvat valuma-aluemuuttujista sekametsät, karkearakeinen maa-aines, lieju, kosteikot, kaltevuus ja hienojakoinen maa-aines. Hienojakoinen maa-aines valikoituu malliin vielä poistettaessa ne suvut, jotka esiintyvät vain ≤ 35 paikalla, mutta ei enää poistettaessa ne suvut, jotka esiintyvät vain ≤ 40 paikalla.

Valikoituvat vesikemiamuuttujat vaihtelivat erilaisten harvinaisten sukujen poistokäsittelyjen välillä jonkin verran. pH oli ainoa malliin aina sukujen poiston määrästä riippumatta valikoitunut vesikemiamuuttuja. Kun harvinaisia sukuja ei ollut poistettu, valikoituivat malliin vesikemiamuuttujista pH, sähkönjohtavuus ja kokonaistyyppi. Kun kaikki ne suvut, jotka esiintyvät vain yhdellä paikalla poistettiin, valikoitui mukaan äsken mainittujen lisäksi myös kokonaisfosfori, mutta poistettaessa ne suvut jotka esiintyivät vain kahdella tai kolmella paikalla valikoitui malliin jälleen vain pH, sähkönjohtavuus ja kokonaistyyppi. Kun aineistosta poistettiin ne suvut, jotka esiintyvät vain 25–35 paikalla valikoitui malliin mukaan pH:n ja sähkönjohtavuuden lisäksi väri (taulukko 3).

Taulukko 3a. Merkitsevien ($p < 0.05$) muuttujien valintojen tulokset OTU –taulukoille, joissa kaikki lajit ovat mukana (taulukko nimeltään species.hel) ja taulukoille joista on poistettu sellaiset lajit, jotka esiintyvät vain 1-5 paikalla, ≤ 10 paikalla, ≤ 15 paikalla ja ≤ 20 paikalla. Ne OTU-taulukot, joiden osalta merkitsevien muuttujien valintojen tulokset ovat keskenään täysin samanlaiset, on taulukon luettavuuden parantamiseksi yhdistetty samalle taulukon riville.

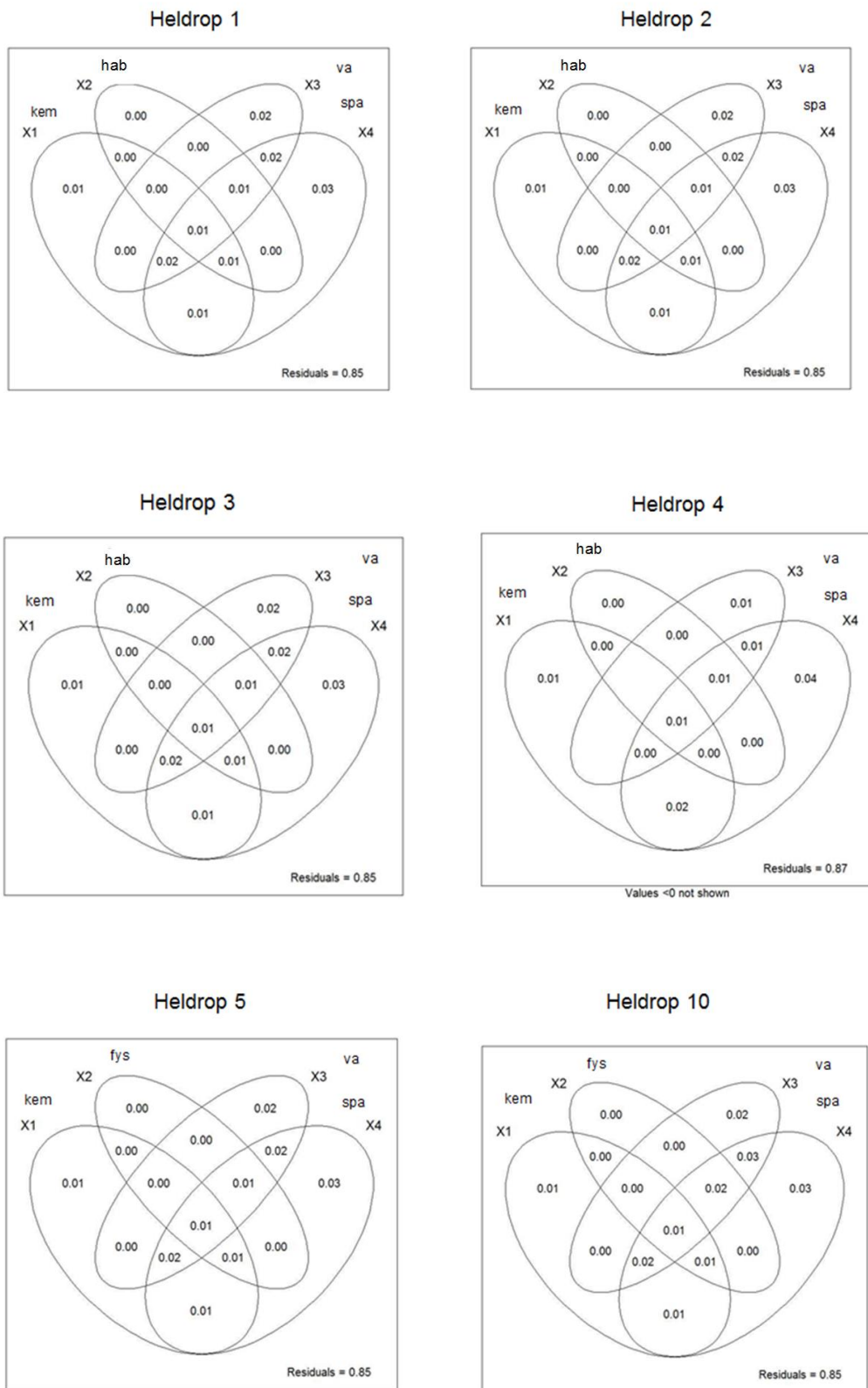
OTU- taulukon nimi	Spatiaali-muuttujat (PCNM.pos)	Valuma- aluemuuttujat	Vesikemia-muuttujat	Habitaatti-muuttujat
species.hel	V2, V8, V4,	sekametsät,	pH,	mukulakivet,
species.hel.drop2	V1	karkearakeinen	sähkönjohtavuus,	pikkukivet
species.hel.drop3		maa-aines, lieju,	kokonaistyyppi	
species.hel.drop5		kosteikot, kaltevuus, hienojakoinen maa-aines		
species.hel.drop1	V2, V8, V4, V1	sekametsät, karkearakeinen maa-aines, lieju, kosteikot, kaltevuus, hienojakoinen maa-aines	pH, sähkönjohtavuus, kokonaistyyppi, kokonaisfosfori	mukulakivet, pikkukivet
species.hel.drop4	V2, V8, V4, V1	sekametsät, karkearakeinen maa-aines	pH	mukulakivet, pikkukivet
species.hel.drop10	V2, V8, V4,	sekametsät,	pH	mukulakivet,
species.hel.drop15	V1	karkearakeinen maa-aines, lieju, kosteikot, kaltevuus, hienojakoinen maa-aines		pikkukivet
species.hel.drop20	V2, V8, V4, V1	sekametsät, karkearakeinen maa-aines, lieju, kosteikot, kaltevuus, hienojakoinen maa-aines	pH, sähkönjohtavuus, kokonaistyyppi, kokonaisfosfori	mukulakivet

Taulukko 3b. Merkitsevien ($p < 0.05$) muuttujien valintojen tulokset OTU –taulukkoille, joista on poistettu sellaiset lajit, jotka esiintyvät vain ≤ 25 paikalla, ≤ 30 paikalla, ≤ 35 paikalla ja ≤ 40 paikalla. Ne OTU-taulukot, joiden osalta merkitsevien muuttujien valintojen tulokset ovat keskenään täysin samanlaiset, on taulukon luettavuuden parantamiseksi yhdistetty samalle taulukon riville.

OTU- taulukon nimi	Spatiaali- muuttujat (PCNM.pos)	Valuma- aluemuuttujat	Vesikemia- muuttujat	Habitaatti- muuttujat
species.hel.drop25 species.hel.drop30	V2, V8, V4, V1	sekametsät, karkearakeinen maa-aines, lieju, kosteikot, kaltevuus, hienojakoinen maa-aines	pH, väri, sähkönjohtavuus	mukulakivet
species.hel.drop35	V2, V8, V4, V1	sekametsät, karkearakeinen maa-aines, lieju, kosteikot, kaltevuus, hienojakoinen maa-aines	pH, väri, sähkönjohtavuus	mukulakivet
species.hel.drop40	V2, V8, V4, V1	sekametsät, karkearakeinen maa-aines, lieju, kosteikot, kaltevuus	pH, väri	mukulakivet

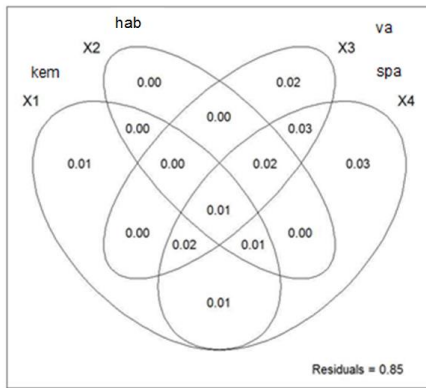
3.3.4 Redundanssianalyysi ja vaihtelun ositus

Kaikille OTU-aineistoille tehtiin redundanssianalyysi ((Rao 1964) ja vaihtelun ositus (Borcard ym. 1992). Korjatut kokonaisselityssasteet paranivat harvinaisten OTUjen poiston myötä hieman. Kun aineistossa olivat mukana kaikki OTUt, oli RDA-analyysin kokonaisselityssaste 14,6 % ja kun aineistosta poistettiin ne suvut, jotka esiintyivät vain yhdellä paikalla, oli kokonaisselityssaste 15,0 %. Harvinaisten OTUjen poistaminen paransi tasaisesti hieman kokonaisselityssastetta ja korkeimmillaan kokonaisselityssaste oli 17,1 % kun aineistosta poistettiin kaikki ne suvut, jotka esiintyivät vain ≤ 35 paikalla. Alle 40 paikalla esiintyvien OTUjen poistaminen puolestaan laski hieman kokonaisselityssastetta, joka oli tuolloin 16,2 % (taulukko 3a). Tuloksissa korostuu selittämättömän vaihtelun suuri osuus, joka vaihteli 87–83 prosentin välillä (kuvat 9a ja 9b; taulukko 3a ja 3b). Kaikkien OTUjen mukana ollessa vesikemiamuuttajat selittivät 6 % vaihtelusta, habitattimuuttajat 3 % vaihtelusta, valuma-alueuuttajat 8 % vaihtelusta ja spatiaalimuuttajat 10 % vaihtelusta (kuva 9a; taulukot 3a ja 3 b). Redundanssianalyysin ja vaihtelun osituksen tulokset eri harvinaisten OTUjen poiston tasoilla on kuvattu tarkemmin taulukoissa 3a ja 3b. Kun aineistosta poistettiin kaikki ne OTUt, jotka esiintyivät vain ≤ 35 paikalla, vesikemiamuuttajat selittivät 7 % vaihtelusta, habitaattimuuttajat 4 % vaihtelusta, valuma-alueuuttajat 10 % vaihtelusta ja spatiaalimuuttajat 11 % vaihtelusta. Eri muuttujaryhmien puhtaat selitysosuudet olivat todella pieniä (0–4%). (kuva 9b; taulukot 3a ja 3 b).

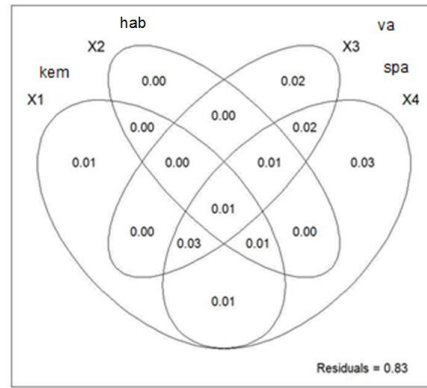


Kuva 9a. OTU-aineistoille tehtyjen vaihtelun ositusten tulokset. Kuvassa lyhenteet spa = PCNM-spatiaalimuuttujat, va= valuma-aluemuuttujat, kem = vesikemia-muuttujat, hab = habitaattimuuttujat. Kuvien tulkinnan mallikuva on liitteessä 3.

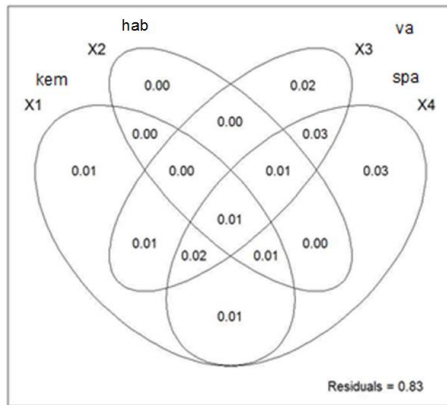
Heldrop 15



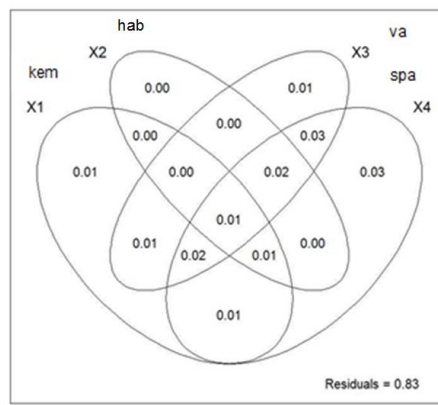
Heldrop 20



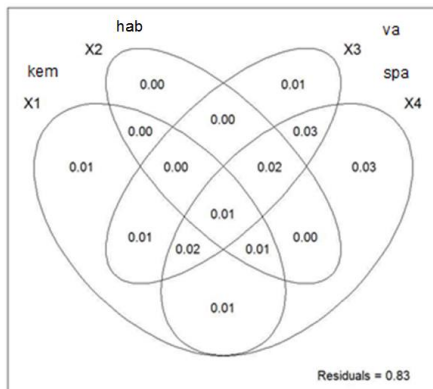
Heldrop 25



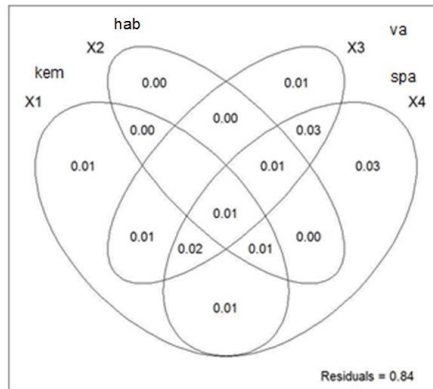
Heldrop 30



Heldrop 35



Heldrop 40



Values <0 not shown

Kuva 9b. OTU-aineistoille tehtyjen vaihtelun ositusten tulokset. Kuvassa lyhenteet spa = PCNM- spatiaalimuuttujat, va= valuma-alueuuttujat, kem = vesikemia-muuttujat, hab = habitaatti-muuttujat. Kuvien tulkinnan mallikuva on liitteessä 3.

Taulukko 3a. Redundanssianalyysin ja vaihtelun osituksen tulokset eri harvinaisten OTUjen poiston tasoilla. Arvot perustuvat korjattuihin R^2 arvoihin (adj. R^2)

	species. hel adj. R^2	species. hel.drop 1 adj. R^2	species. hel.drop 2 adj. R^2	species. hel.drop 3 adj. R^2	species. hel.drop 4 adj. R^2	species. hel.drop 5 adj. R^2	species. hel.drop 10 adj. R^2	species. hel.drop 15 adj. R^2	species. hel.drop 20 adj. R^2	species. hel.drop 25 adj. R^2	species. hel.drop 30 adj. R^2	species. hel.drop 35 adj. R^2	species. hel.drop 40 adj. R^2
[aeghklno] = X1, Vesikemia-muuttujat = kem	0,062	0,069	0,063	0,064	0,046	0,064	0,047	0,048	0,076	0,068	0,069	0,071	0,064
[befiklmo] = X2, habitaatti-muuttujat = fys	0,043	0,043	0,043	0,044	0,044	0,044	0,045	0,046	0,040	0,041	0,042	0,042	0,043
[cfgilmno] = X3, valuma-aluemuuttujat = va	0,091	0,092	0,093	0,093	0,050	0,094	0,095	0,099	0,102	0,103	0,104	0,105	0,098
[dhijkmno] = X4, spatiaalimuuttujat = spa	0,108	0,108	0,109	0,110	0,111	0,111	0,113	0,117	0,122	0,124	0,127	0,129	0,131
[abefghijklmno] = kem+fys	0,083	0,089	0,084	0,085	0,072	0,085	0,073	0,075	0,096	0,088	0,090	0,092	0,084
[acefghijklmno] = kem+va	0,116	0,120	0,118	0,119	0,083	0,119	0,117	0,120	0,134	0,131	0,134	0,136	0,124
[adefghijklmno] = kem+spa	0,123	0,128	0,125	0,126	0,120	0,126	0,122	0,126	0,144	0,146	0,149	0,152	0,148
[bcefghijklmno] = fys+va	0,104	0,104	0,105	0,106	0,064	0,106	0,108	0,112	0,115	0,116	0,117	0,119	0,113
[bdefghijklmno] = fys+spa	0,117	0,118	0,119	0,120	0,120	0,120	0,122	0,126	0,130	0,132	0,135	0,137	0,139
[cdfghijklmno] = va+spa	0,133	0,133	0,135	0,136	0,118	0,136	0,138	0,144	0,149	0,151	0,153	0,155	0,149
[abcefghijklmno] = kem+fys+va	0,121	0,124	0,123	0,124	0,092	0,124	0,121	0,125	0,138	0,134	0,137	0,139	0,128
[abdefghijklmno] = kem+fys+spa	0,130	0,135	0,131	0,133	0,127	0,133	0,130	0,134	0,149	0,151	0,154	0,157	0,153
[acdefghijklmno] = kem+va+spa	0,143	0,147	0,145	0,146	0,128	0,147	0,146	0,151	0,165	0,164	0,167	0,169	0,160
[bcdefghijklmno] = fys+va+spa	0,137	0,138	0,139	0,140	0,124	0,141	0,143	0,148	0,153	0,155	0,158	0,160	0,155
[abcdefghijklmno] = Kokonaisselitysaste	0,147	0,150	0,149	0,150	0,133	0,150	0,149	0,155	0,167	0,166	0,169	0,171	0,162

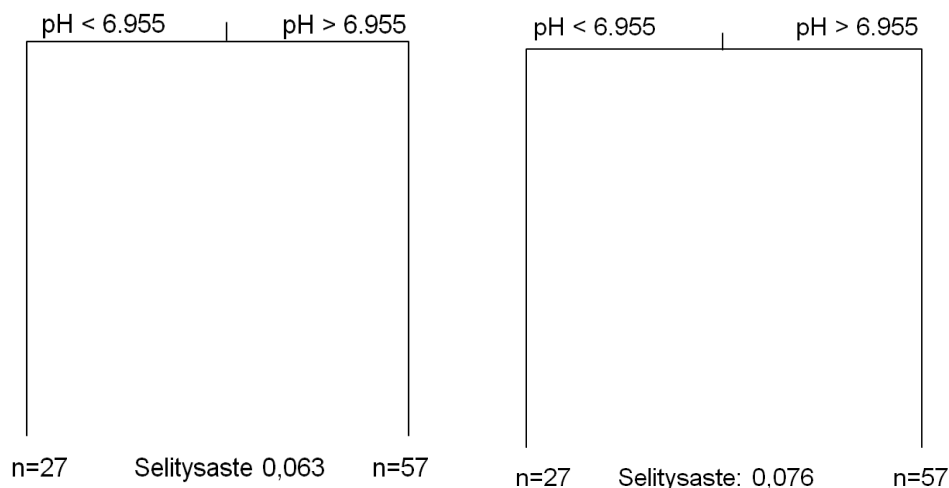
Taulukko 3b. Redundanssianalyysin ja vaihtelun osituksen tulokset eri harvinaisten OTUjen poiston tasoilla. Arvot perustuvat korjattuihin R^2 arvoihin (adj. R^2)

	species. hel adj. R^2	species.h el.drop 1 adj. R^2	species. hel.drop 2 adj. R^2	species. hel.drop 3 adj. R^2	species. hel.drop 4 adj. R^2	species. hel.drop 5 adj. R^2	species. hel.drop 10 adj. R^2	species. hel.drop 15 adj. R^2	species. hel.drop 20 adj. R^2	species. hel.drop 25 adj. R^2	species. hel.drop 30 adj. R^2	species. hel.drop 35 adj. R^2	species. hel.drop 40 adj. R^2
[a] = kem fys+va+spa	0,009	0,012	0,010	0,010	0,008	0,010	0,006	0,006	0,014	0,011	0,011	0,012	0,007
[b] = fys kem+va+spa	0,004	0,003	0,004	0,004	0,005	0,004	0,003	0,003	0,002	0,002	0,002	0,002	0,002
[c] = va kem+fys+spa	0,017	0,015	0,017	0,017	0,005	0,017	0,020	0,021	0,019	0,015	0,015	0,015	0,009
[d] = spa kem+fys+va	0,026	0,026	0,026	0,026	0,041	0,026	0,029	0,030	0,030	0,032	0,032	0,032	0,035
[e]	0,001	0,002	0,001	0,001	0,002	0,001	0,001	0,001	0,002	0,003	0,003	0,003	0,004
[f]	0,003	0,004	0,003	0,003	0,003	0,003	0,004	0,004	0,003	0,003	0,003	0,003	0,003
[g]	0,003	0,005	0,003	0,003	-0,001	0,003	0,001	0,001	0,005	0,008	0,008	0,008	0,007
[h]	0,008	0,008	0,008	0,008	0,019	0,008	0,006	0,006	0,009	0,007	0,008	0,008	0,007
[i]	0,001	0,001	0,001	0,001	0,004	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
[j]	0,021	0,020	0,021	0,021	0,015	0,022	0,028	0,029	0,023	0,031	0,031	0,032	0,035
[k]	0,007	0,007	0,007	0,007	0,004	0,007	0,008	0,008	0,007	0,007	0,008	0,008	0,007
[l]	0,002	0,001	0,002	0,002	0,000	0,002	0,000	0,000	0,001	0,001	0,000	0,000	-0,002
[m]	0,013	0,013	0,013	0,013	0,014	0,014	0,018	0,019	0,014	0,015	0,015	0,016	0,013
[n]	0,020	0,021	0,020	0,020	0,001	0,020	0,015	0,015	0,028	0,021	0,021	0,022	0,019
[o]	0,013	0,013	0,013	0,013	0,012	0,013	0,009	0,009	0,010	0,009	0,010	0,010	0,014
[p] = Selittämätön vaihtelu	0,853	0,850	0,851	0,850	0,867	0,850	0,851	0,845	0,833	0,834	0,831	0,829	0,838

3.4 Monimuuttujaregressiopuuanalyysit

MRT-analyysin tuloksena syntyy graafinen ”puu”, joka kuvaa tutkimuspaikkojen ja ympäristömuuttujien välistä suhdetta. Jokainen aineiston onnistunut jako klustereihin on määritelty jonkin selittävän muuttujan kynnsarvolla, tässä tapauksessa pH -arvolla 6,955 (kuva 10).

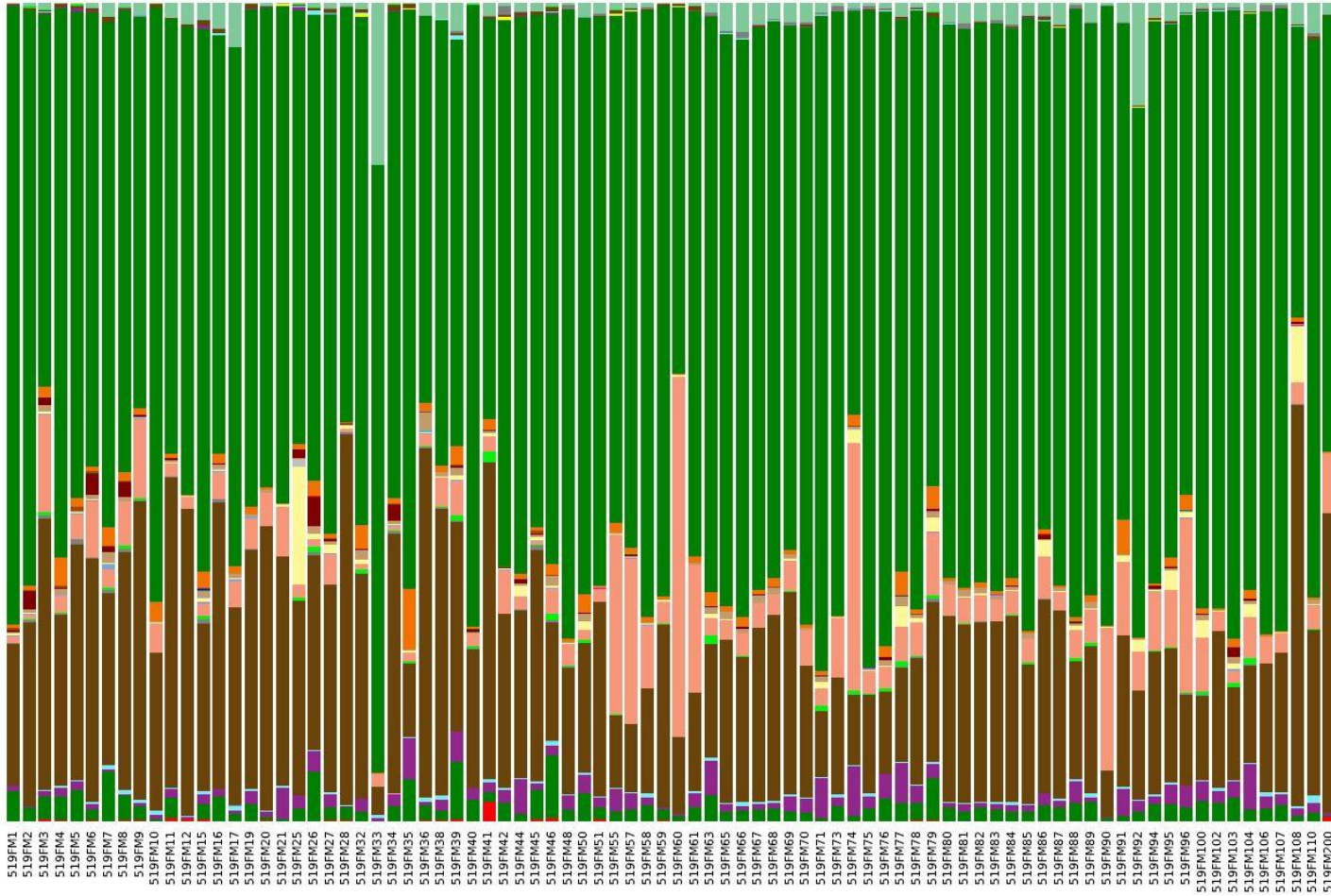
MRT-tuloksista havaitaan hyvin samansuuntaiset tulokset kuin redundanssianalyysistä ja vaihtelun osituksestakin; yhteisö-ympäristösuhteet virtavesien mikrobeilla ovat heikkoja. MRT-analyyseissa kandidaatti-ympäristömuuttujia oli mukana yli 30. Toisin kuin RDA:ssa, analyyseja ei toistettu eri harvinaisten lajien poistojen tasoilla, vaan tarkasteltiin ainoastaan kaikki OTUt sisältävää aineistoa sekä aineistoa, jossa olivat mukana vain ne OTUt, jotka esiintyvät ≥ 40 tutkimuspaikalla. Selitysasteet jäivät alhaisiksi. Kaikki OTUt sisältävän aineiston MRT-analyysin selitysaste oli 6,3 %. Selitysaste parani harvinaisten lajien poiston myötä hieman ollen ≥ 40 paikalla esiintyvien OTUjen MRT -analyysissä 7,6 % (kuva 10).



Kuva 10. MRT -analyysien tuloksena syntyneet ”puut”. Vasemmanpuoleisessa ”puussa” kaikki OTUt mukana ja oikeanpuoleisessa vain ne, jotka esiintyvät ≥ 40 paikalla.

3.5 BLAST-tulokset

BLAST -analyysit (Altschul ym. 1990) tehtiin QIIME-ympäristössä (Caporaso ym. 2010) pääjaksot-, luokka-, lahko-, heimo-, ja sukutasoille. Yleisimmät havaitut pääjaksot ovat *Proteobacteria* (64,9%), *Bacteroidetes* (21,3%) ja *Cyanobacteria* (5,2%) (kuva 11). Yleisimmät havaitut luokat puolestaan ovat *Alphaproteobacteria* (28,9%), *Betaproteobacteria* (27,7%), *Gammaproteobacteria* (7,5%), *Cytophagia* (7,3%), *Saprospirae* (6,5%) ja *Flavobacteriia* (5,1%) (liite 6). Lahkoista yleisimmät ovat *Burkholderiales* (24,0%), *Rhizobiales* (11,4%), *Sphingomonadales* (9,3%), *Cytophagales* (7,3%), *Saprospirales* (6,5%), *Xanthomonadales* (6,5%) ja *Flavobacteriales* (5,1%) (liite 7). Yleisimmät havaitut heimot ovat *Comamonadaceae* (21,6%), *Sphingomonadaceae* (7,9%), *Cytophagaceae* (7,2%), *Chitinophagaceae* (5,8%), *Flavobacteriaceae* (4,8%), lahkotasolla *Rhizobiales*:iksi määrittäneet, mutta heimotasolla ilman tarkennusta jääneet (4,7%) ja *Hyphomicrobiaceae* (4,5%) (liite 8). Yleisimmät havaitut suvut puolestaan ovat heimotasolla *Comamonadaceae*:ksi blast-analyysin perusteella määrittäneet, mutta sukutasolla ilman tarkempaa määrettä jääneet (18,7%), *Flectobacillus* (4,6%), heimotasolla *Sphingomonadaceae*:ksi BLAST-analyysin perusteella sijoittuneet, mutta sukutasolla ilman tarkempaa määrettä jääneet (6,1%), heimotasolla *Chitinophagaceae*:ksi määrittäneet, mutta sukutasolla ilman tarkempaa määrettä jääneet (5,2%), *Flavobacterium* (4,8%), lahkotasolla *Rhizobiales*:iksi määrittäneet, mutta heimo- ja sukutasolla ilman tarkennusta jääneet (4,7%) (liite 9). Kuvassa 11 ja liitteissä 6–9 on tutkimuspaikkojen nimien sijaan käytetty paikan yksilöllisen a-alukkeen nimeä. Alukkeet ja tutkimuspaikat toisiinsa yhdistävä taulukko on liitteessä 5.



Kuva 11. Pääjaksotasolle BLAST-analysoidut bakteerit. Yleisimmät pääjaksot ovat Proteobakteria 64,9% (kuvassa vihreällä), Bacteroidetes 21,3% (kuvassa ruskealla) ja Cyanobacteria 5,2% (kuvassa oranssinruskealla värillä). Alukkeet ja tutkimuspaikat toisiinsa yhdistävä taulukko on liitteessä 5.

Taulukko 4a. Yhteenveto eri taksonomisilla tasoilla havaituista yleisimmistä bakteereista.

BLAST- analyysin taksonominen taso	Pääjakso	Luokka	Lahko	Heimo	Suku	Yleisyys (% kyseisellä taksonomisella tasolla havaituista)
Pääjakso	<i>Proteobacteria</i>	-	-	-	-	64,9
Pääjakso	<i>Bacteroidetes</i>	-	-	-	-	21,3
Pääjakso	<i>Cyanobacteria</i>	-	-	-	-	5,2
Luokka	<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	-	-	-	28,9
Luokka	<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	-	-	-	27,7
Luokka	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	-	-	-	7,5
Luokka	<i>Bacteroidetes</i>	<i>Cytophagia</i>	-	-	-	7,3
Luokka	<i>Bacteroidetes</i>	<i>Saprospirae</i>	-	-	-	6,5
Luokka	<i>Bacteroidetes</i>	<i>Flavobacteriia</i>	-	-	-	5,1
Lahko	<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	<i>Burkholderiales</i>	-	-	24,0
Lahko	<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Rhizobiales</i>	-	-	11,4
Lahko	<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Sphingomonadales</i>	-	-	9,3
Lahko	<i>Bacteroidetes</i>	<i>Cytophagia</i>	<i>Cytophagales</i>	-	-	7,3
Lahko	<i>Bacteroidetes</i>	<i>Saprospirae</i>	<i>Saprospirales</i>	-	-	6,5

Taulukko 4b. Yhteenveto eri taksonomisilla tasoilla havaituista yleisimmistä bakteereista (jatkuu)

BLAST- analyysin taksonominen taso	Pääjakso	Luokka	Lahko	Heimo	Suku	Yleisyys (% kyseisellä taksonomisella tasolla havaituista havaituista)
Lahko	<i>Proteobacteria</i>	<i>Gammaproteobacteria</i>	<i>Xanthomonadales</i>	-	-	6,5
Lahko	<i>Bacteroidetes</i>	<i>Flavobacteriia</i>	<i>Flavobacteriales</i>	-	-	5,1
Heimo	<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	<i>Burkholderiales</i>	<i>Comamonadaceae</i>	-	21,6
Heimo	<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Sphingomonadales</i>	<i>Sphingomonadaceae</i>	-	7,9
Heimo	<i>Bacteroidetes</i>	<i>Cytophagia</i>	<i>Cytophagales</i>	<i>Cytophagaceae</i>	-	7,2
Heimo	<i>Bacteroidetes</i>	<i>Saprospirae</i>	<i>Saprospirales</i>	<i>Chitinophagaceae</i>	-	5,8
Heimo	<i>Bacteroidetes</i>	<i>Flavobacteriia</i>	<i>Flavobacteriales</i>	<i>Flavobacteriaceae</i>	-	4,8
Heimo	<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Rhizobiales</i>	ei tarkennusta	-	4,7
Heimo	<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Rhizobiales</i>	<i>Hyphomicrobiaceae</i>	-	4,5
Suku	<i>Proteobacteria</i>	<i>Betaproteobacteria</i>	<i>Burkholderiales</i>	<i>Comamonadaceae</i>	ei tarkennusta	18,7
Suku	<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Sphingomonadales</i>	<i>Sphingomonadaceae</i>	ei tarkennusta	6,1
Suku	<i>Bacteroidetes</i>	<i>Saprospirae</i>	<i>Saprospirales</i>	<i>Chitinophagaceae</i>	ei tarkennusta	5,2
Suku	<i>Bacteroidetes</i>	<i>Flavobacteriia</i>	<i>Flavobacteriales</i>	<i>Flavobacteriaceae</i>	<i>Flavobacterium</i>	4,8
Suku	<i>Proteobacteria</i>	<i>Alphaproteobacteria</i>	<i>Rhizobiales</i>	ei tarkennusta	ei tarkennusta	4,7

4 Tulosten tarkastelu

4.1 Tutkimuspaikkojen ekologinen tila

Tämän tutkielman tutkimuspaikat (N=84) sijaittivat 21 länsisuomalaisella kolmannen jakovaiheen jokivaluma-alueella. Tutkimuspaikat pyrittiin valitsemaan niin, että ne edustaisivat kattavasti maatalouden ja metsätalouden vaikutuspiiriä. Mukaan valittiin kuitenkin myös muutamia hyvässä tai erinomaisessa ekologisessa tilassa olevia jokia (esim. Kiiminkijoki, Isojoki). Länsisuomen rannikolla erityisesti monet maatalousvaikutteiset joet ovat välttävässä tai huonossa ekologisessa tilassa (Pintavesien ekologinen ja kemiallinen tila 2016; kuva 4).

Pintavesien ekologisen tilan luokittelussa käytetään Suomessa pH:n osalta vuosittaista minimiarvoa. Jokivesien ekologisen luokittelun vertailuolojen laskennallinen minimiarvo vaihtelee jokityypistä riippuen pH-arvojen 5,6–5,8 välillä (Aroviita ym. 2012). pH vaihteli tutkimuspaikoilla 5,95–8,07 välillä. Keskiarvo-pH oli 7,20 (taulukko 1).

Jokityypistä riippuen erinomaisen ja hyvän ekologisen tilan raja-arvona käytetään Suomessa kokonaisfosforin osalta pitoisuuksia 15–40 µg/l, hyvän ja tyydyttävän tilan raja-arvona pitoisuuksia 40–60 µg/l, tyydyttävän ja välttävän tilan raja-arvona pitoisuuksia 60–100 µg/l ja välttävän ja huonon ekologisen tilan raja-arvona pitoisuuksia 90–130 µg/l (Aroviita ym. 2012). Kokonaisfosforin perusteella huonoon ekologiseen tilaan luokituu tässä tutkimuksessa esimerkiksi kaksi Paimionjoen (133 µg/l ja 146 µg/l), yksi Aurajoen (136 µg/l), yksi Kokemäenjoen (147 µg/l), kaksi Teuvanjoen (150 µg/l ja 197 µg/l), yksi Kruunupyynjoen (167 µg/l), yksi Temmesjoen (167 µg/l) ja yksi Oulujoen (183 µg/l) valuma-alueiden tutkimuspaikoista.

Jokityypistä riippuen erinomaisen ja hyvän ekologisen tilan raja-arvona käytetään Suomessa kokonaistypen osalta pitoisuuksia 335–450 µg litrassa, hyvän ja tyydyttävän tilan raja-arvona pitoisuuksia 800–900 µg litrassa, tyydyttävän ja välttävän tilan raja-arvona pitoisuuksia 1400–1500 µg litrassa ja välttävän ja

huonon ekologisen tilan raja-arvona pitoisuuksia 2400–2500 µg litrassa (Aroviita ym. 2012). Kokonaistypen perusteella huonoon ekologiseen tilaan luokituu vain yksi Temmesjoen valuma-alueen tutkimuspaikka (2763 µg/l). Erinomaiseen ekologiseen tilaan sen sijaan luokituu neljä Kiiminkijoen valuma-alueen tutkimuspaikkaa ((havaitut kokonaistyyppipitoisuudet 150 µg/l, 163 µg/l, 219 µg/l, 313µg/l) neljä Oulujoen valuma-alueen tutkimuspaikkaa (havaitut kokonaistyyppipitoisuudet 200 µg/l, 213 µg/l, 275 µg/l, 313 µg/l), kaksi Isojoen (275 µg/l ja 313 µg/l) ja yksi Karvianjoen valuma-alueen tutkimuspaikka (294 µg/l).

Sähkönjohtavuus ei kuulu pintavesien ekologisen tilan fysikaalis-kemiallisiin laatutekijöihin. Korkeaa sähköjohtokykyä voidaan kuitenkin asiantuntija-arvioinnissa käyttää lisäperusteluna ekologisen tilaluokan määräytymiselle luokittelumuuttujien ja vesiin kohdistuvien ihmistoiminnan paineiden kokonaistilannetta arvioitaessa (Aroviita ym. 2012). Sähkönjohtokyky kertoo veteen liuenneiden suolojen (esimerkiksi natrium, kalium, kalsium, magnesium, kloridi ja sulfaatti) määrästä ja sitä nostavat muun muassa jätevedet ja lannoitteet. Virtavesien sähköjohtavuus on yleensä 100–200 µS/cm. (Suomen ympäristökeskus 2017). Sähkönjohtavuus vaihteli tutkimuspaikoilla Siikajoen valuma-alueella havaitusta 13,2 µS/cm Temmesjoen valuma-alueella havaittuun 619 µS/cm. Havaittujen sähköjohtavuusarvojen keskiarvo oli 91 µS/cm (taulukko 1).

Veden väriin vaikuttaa valuma-alueen maaperä ja humuksen määrä. Suomessa humuksesta peräisin oleva rusehtava väri on hyvin tyypillinen vesistöissä. Humusaineet ovat orgaanisia yhdisteitä, joiden huuhtoutumista vesistöihin maanpinnan käsittelyt ja ojitukset lisäävät. Väriarvojen vuosi- ja vuodenaikaisvaihtelut voivat sadannasta johtuvien valumaolojen muutosten vuoksi olla merkittäviä. Lievästi humuspitoisten vesien väriarvo on 20–40 mgPt/l, humuspitoisen 40–100 mgPt/l ja erittäin humuspitoisen yli 100 mgPt/l. Luonnontilaisissa humusvesissä väriarvo saattaa olla kuitenkin jopa yli 200 mgPt/l. (Mitikka 2013; Suomen ympäristökeskus 2017). Tutkimuspaikkojen väriarvot vaihtelivat Kiiminkijoen valuma-alueella havaitusta 15 mgPt/l Aurajoen

valuma-alueella havaittuun 375 mgPt/l arvoon. Väriarvojen keskiarvo oli 130 mgPt/l (taulukko 1).

4.2 Mikrobiyhteisöjen rakenteeseen vaikuttavat tekijät

Redundanssianalyysin ja vaihtelun osituksen perusteella parhaiten mikrobiyhteisöjen rakennetta selittivät spatiaalimuuttujat (10 % vaihtelusta), ja valuma-aluemuuttujat (8 % vaihtelusta). Vesikemiamuuttujat selittivät 6 % vaihtelusta ja habitattimuuttujat vain 3 %. Eri muuttujaryhmien puhtaasta selitysosuudet olivat todella pieniä (0–4%). (kuva 9a; taulukot 3a ja 3 b). RDA-analyysin, vaihtelun osituksen ja monimuuttujaregressiopuiden avulla saatujen tulosten perusteella voidaan todeta virtaavien vesien mikrobiyhteisöjen yhteisöympäristösuhteiden olevan heikkoja. Ympäristön tilan seurannassa tulisikin jatkossa tulosten perusteella käyttää mikrobien osalta muita vastemuuttujia kuin OTU-perusteista yhteisörakennetta. Tuloksissa korostuu selittämättömän vaihtelun suuri osuus, joka RDA:ssa vaihteli 87–83 prosentin välillä. Myös MRT-analyysit osoittivat yhteisöympäristösuhteiden olevan heikkoja.

Selittämättömän vaihtelun suuri osuus on hyvin tyypillistä tämän kaltaisille ekologian alan tutkimuksille (Cottenie 2005). Tämä voi johtua ainakin osittain tilastoanalyysissä käytettyjen selittävien muuttujien määrän huomioivien korjattujen R²-arvojen käytöstä. Samaa tilastollista lähestymistapaa käyttävissä tutkimuksissa on havaittu piilevillä, pohjaeläimillä, vesikasveilla ja päällyksillä samansuuntaisia tuloksia (ks. esim. Manninen 2014; Heino ym. 2012). Myös yhteisörakenteen sattumanvarainen eli stokastinen vaihtelu tai tärkeiden ympäristötekijöiden puuttuminen mallista voivat osaltaan selittää havaittuja tuloksia (Pandit ym. 2009).

Tutkimme aiemmin samalla länsisuomalaisen virtavesimikrobien aineistolla taksonomista monimuotoisuutta, jossa OTUt olivat BLAST-analyysien sijaan analysoitu OTU-runsauksina. Havaitimme, että mikrobiyhteisöjen alueelliseen vaihteluun vaikuttaa yhdessä paikalliset ja valuma-alueen ympäristötekijät sekä ilmastotekijät ja mahdollisesti levittäytymisrajoittuneisuus. Valuma-alueominaisuuksia, erityisesti maatalouden määrää, voidaan

tutkimuksemme perusteella käyttää epäsuorana mittarina muutokselle vesistöjen mikrobiyhteisöissä. (Jyrkänkallio-Mikkola ym. 2017). Heino ja kumppanit (2017) puolestaan havaitsivat RDA-analyysitarkastelussa alueellisten ja paikallisten ympäristömuuttujien olevan merkitseviä yhteisörakenteeseen vaikuttavia tekijöitä niin virtavesissä tavatuilla piilevillä, hyönteisillä ja sammalilla kuin järviympäristössä havaituilla vesikasveilla, kasviplanktonilla ja kaloilla. RDA:n selityssasteet jäivät kuitenkin myös tässä tutkimuksessa alhaisiksi vaihdellen eri eliöryhmien välillä 10-20 prosenttiin. (Heino ym. 2017)

Virtavesien bakteeriyhteisöjen rakenteeseen paikallisesti ja alueellisesti vaikuttavien tekijöiden on aiemmissa tutkimuksissa havaittu olevan vaikeasti selitettäviä (Staley ym. 2014). Fysikaalis-kemialliset ja ravinnepitoisuuden muutokset voivat vaikuttaa paikallisesti suoraan vain pieneen määrään bakteerilahkoja, mutta vaikutukset voivat kuitenkin mahdollisesti ylettyä koko yhteisöön bakteerien välisten vuorovaikutusmekanismien kautta (Staley ym. 2014). Wang kumppaneineen (2012) havaitsi RDA-analyysin avulla korkeuden olevan parhaiten mikrobien yhteisörakennetta selittävä tekijä ($P < 0.01$) kiinalaisissa 1820–4050 metrin korkeudessa sijaitsevilla virtavesillä. He havaitsivat myös kokonaistypen ja leveyspiirin olevan tilastollisesti merkitseviä ($p = 0.056$ ja $p = 0.058$) mikrobien yhteisörakenteeseen vaikuttavia tekijöitä, mutta mitkään muut muuttujat eivät nousseet tilastollisesti merkitseviksi silloin, kun korkeus oli sisällytetty mukaan analyysiin (Wang ym. 2012).

Virtavesien biofilmeistä, järvien kivikkorannoilta ja virtavesien lehtikarikkeesta eristettyjen bakteerien ja vesikemiamuuttujien välillä on todettu aiemmissa tutkimuksissa olevan selkeä yhteys (esim. Fierer ym. 2007; Heino ym. 2014; Vilmi ym. 2016). Virtavesien pH-arvoon vaikuttavat lukuisat ympäristötekijät, kuten ympäröivän valuma-alueen geologia ja hydrologia. pH voi vaikuttaa sekä suorasti että epäsuorasti biofilmiyhteisöihin (Fierer ym. 2007; Wilhelm ym. 2013). pH:n onkin havaittu useissa eri tutkimuksissa korreloivan biofilmibakteereiden yhteisörakenteeseen (esim. Anderson-Glenna ym. 2008; Wilhelm ym. 2013), ja sen on havaittu olevan tärkein mikrobiyhteisöjen rakennetta selittävä muuttuja (Fierer ym. 2007). Virtavesien tilaan vaikuttaa usein yhtä aikaa useat eri tekijät, ja ne voivat kohdata yhtä aikaa esimerkiksi

rehevöitymistä ja happamoitumista. Pohjaeläimillä tehdyssä tutkimuksessa havaittiin eri taksonomisten ryhmien reagoivan hyvin eri tavoin tällaisiin kahden eri paineen aiheuttamiin muutoksiin (Mykrä ym. 2012).

4.3 Harvinaisten lajien poiston merkitys

Tämän tutkimuksen perusteella yhteisön ympäristösuhteisiin vaikuttavat erityisesti yleiset lajit. Harvinaisten lajien poistolla ei havaittu olevan suurta merkitystä, vaan merkitsevät muuttujat pysyvät lähes samoina huolimatta siitä, poistetaanko aineistosta ne lajit, jotka esiintyvät vain yhdellä paikalla vai kaikki ne lajit, jotka esiintyvät vain ≤ 40 paikalla. RDA-analyysin korjatut kokonaisselitysasteet paranivat harvinaisten OTUjen poiston myötä hieman. Kun aineistossa olivat mukana kaikki OTUt, oli RDA-analyysin kokonaisselitysaste 14,6 % ja korkeimmillaan kokonaisselitysaste oli 17,1 % kun aineistosta poistettiin kaikki ne suvut, jotka esiintyivät vain ≤ 35 paikalla (taulukko 3a). MRT-analyysien selitysaste nousi kuudesta prosentista seitsemään prosenttiin, kun aineistosta poistettiin vain ≤ 40 paikalla esiintyvät lajit.

Fierer ym. (2007) havaitsivat pH:n, liuenneen orgaanisen hiilen ja typen pystyvän selittämään eniten virtavesien bakteeriyhteisöjen vaihtelua New Hampshiressä, USA:ssa (Fierer ym. 2007). Havaintomme tukevat aiempia tutkimuksia, sillä pH valikoitui mukaan malleihin kaikilla harvinaisten OTUjen poistojen tasoilla. Tämän tutkimuksen vesikemiamuuttujavalikoimaan ei kuulunut liuennut orgaaninen hiili, mutta kokonaistyyppi valikoitui mukaan malliin yli puolessa harvinaisten OTUjen poiston eri tasoista.

Habitaatti-muuttujista malliin valikoitui pikkukivet ja mukulakivet, kun poistettavia sukuja olivat enintään ne, jotka esiintyivät vain 15 paikalla. Kun poistettavia sukuja oli tätä enemmän, valikoituivat mukaan vain mukulakivet. Uoman pohjarakenne vaikuttaa biofilmibakteeriyhteisöjen muodostumiseen (Besemer 2016), joten pikkukivien ja mukulakivien valikoitumista malliin voidaan pitää loogisena.

Harvinaisten lajien poiston merkitystä on ekologiassa tutkittu paljon erilaisilla ”perinteisemmällä” eliöryhmillä kuten kaloilla, pohjaeläimillä ja

vesikasveilla (esim. Alahuhta ym. 2014; Mouillot ym. 2013; Siqueira ym. 2012). Jatkossa mikrobien yhteisö-ympäristösuhdetta ja harvinaisten lajien poiston vaikutusta tulee tutkia vielä laajemmin erilaisia bioinformatiikan työkaluja ja tilastollisia lähestymistapoja hyödyntäen.

Tutkittaessa harvinaisten mikrobien merkitystä ympäristö-yhteisösuhteeseen on hyvä huomioida ja tunnistaa mahdolliset virhelähteet. Ne voivat liittyä esimerkiksi aineiston sekvensointiin (sequencing noise), bioinformatiikan työkalujen valintoihin (esim. käytettävät laatuarvot) tai tilastoanalyysien valintoihin ja tulosten tulkintaan liittyen (ks. esim. Lynch & Neufeld 2015).

Tulosten tulkinnassa tulee huomioida myös BLAST-kirjastoihin vielä tällä hetkellä liittyvät heikkoudet. BLAST-tietokanta kasvaa koko ajan sitä mukaa kuin uudet sekvensointiaineistot tuottavat uutta tietoa OTUjen taksonomisen luokittelun tueksi. OTUt määritellään yleensä toiminnallisiksi taksonomisiksi yksiköiksi 97 % samankaltaisuusasteen mukaan. Tämä tulee huomioida verrattaessa bakteeriaineistoja muihin biologisiin aineistoihin, joiden taksonominen luokittelu on tehty mikroskopoiden lajitasolle. BLAST-analyysi yrittää löytää osuman tarkimmalle mahdolliselle tasolle, mutta tässä tutkimuksessa havaittiin BLAST -analyysin pääsevän yleisimpien taksonien kohdalla vain harvoin sukutasolle, eikä aina saavutettu edes heimotasoa. Lahko-tai heimotason käyttäminen jatkoanalyseissä voi siis tällä hetkellä olla sukutason käyttöä perustellumpaa.

4.4 Länsisuomalaisten virtavesien biofilmibakteerit verrattuna muissa tutkimuksissa havaittuihin biofilmibakteereihin

Tässä tutkimuksessa yleisimmät havaitut pääjaksot olivat *Proteobacteria* (64,9 %), *Bacteroidetes* (21,3 %) ja *Cyanobacteria* (5,2 %). Tulokset tukevat aiempia tutkimuksia, joissa on havaittu, että juuri *Proteobacteria*, *Bacteroidetes* ja *Cyanobacteria* ovat tyypillisesti yleisimmin makean veden biofilmeillä havaittuja pääjaksoja (Besemer 2016). *Bacteroidetes* -pääjaksoon kuuluvien bakteereiden on havaittu pystyvän hyödyntämään monimutkaisia makromolekyylejä (Anderson-

Glenna ym. 2008), ja niiden uskotaan olevan merkittävässä roolissa kiintoaineen käsittelyssä (Newton ym. 2011). Näiden tekijöiden on katsottu selittävän usein hyvin yleisenä biofilmiyhteisöissä havaittua *Bacteroidetes* -pääjakson esiintyvyyttä (Besemer 2015).

Aikaisemmissa tutkimuksissa on havaittu *Acidobacteria* -pääjaksoa esiintyvän biofilmeissä erityisesti silloin kuin veden pH on alhainen (Fierer ym. 2007). Tässä tutkimuksessa *Acidobacteria* -pääjaksoa havaittiin vain 2,0 % kaikista pääjaksoista. Tutkimuspaikkojen pH oli keskimäärin 7,13 joten neutraali pH selittänee vähäistä *Acidobacteria* -pääjakson esiintyvyyttä. *Proteobacteria*-pääjaksoon kuuluvien bakteereiden on havaittu runsastuvan pH:n noustessa (Fierer ym. 2007). Tämän tutkimuksen keskimäärin neutraalit pH -olosuhteet voivat siis osaltaan selittää proteobakteerien havaittua yleisyyttä.

Yleisimmät tässä tutkimuksessa havaitut luokat olivat *Alphaproteobacteria* (28,9 %), *Betaproteobacteria* (27,7 %), *Gammaproteobacteria* (7,5 %), *Cytophagia* (7,3 %) *Saprospirae* (6,5 %) ja *Flavobacteriia* (5,1 %). Betaproteobakteerien on useissa tutkimuksissa havaittu dominoivan virtavesien ja järvien biofilmiäbakteeristoa (esim. Besemer ym. 2012; Olapade & Leff 2005). Toiset tutkimukset taas ovat osoittaneet alphaproteobakteereiden olevan biofilmiyhteisöissä joko yhtä yleisiä tai yleisempiä kuin betaproteobakteereiden (Anderson-Glenna ym. 2008; Lear ym. 2009; Wang ym. 2012). Tässä tutkimuksessa alpha- ja betaproteobakteerien havaittiin olevan lähes yhtä yleisiä (28,9 % ja 27,7 % kaikista havaituista bakteeriluokista). Alphaproteobakteereiden on havaittu pystyvän hajottamaan humusaineita ja kykenevän tuottamaan laiduntajille, eli muun muassa kivien pintojen biofilmejä ravinnokseen käyttäville pohjaeläimille, vastustuskykyisiä kantoja (Newton ym. 2011).

Yleisimmät tässä tutkimuksessa havaitut lahkot olivat puolestaan *Burkholderiales* (24,0 %), *Rhizobiales* (11,4%), *Sphingomonadales* (9,3 %), *Cytophagales* (7,3 %), *Saprospirales* (6,5%), *Xanthomonadales* (6,5 %) ja *Flavobacteriales* (5,1 %). Staley ym. (2014) ovat havainneet, että lähes 90 prosenttia heidän tutkimansa Mississippi joen yläosan virtavesiekosysteemin bakteeriyhteisöstä muodostuu vain muutama lahkoon kuuluvista bakteereista. Heidän tutkimuksessaan yleisimmät havaitut lahkot olivat: *Burkholderiales*

(54,0 %), *Actinomycetales* (10.1%), *Pseudomonadales* (8.3%), *Sphingobacteriales* (3.4%), *Methylophilales* (3.1%), *Rhodocyclales* (2.4%) ja *Rhodospirillales* (2.0%).

Myös tässä tutkimuksessa suuri osa bakteereista kuului vain muutamaankin lahkoon; 76,3 % lahkotasolle BLAST-analysoiduista bakteereista kuului yhdeksään yleisimpään lahkoon. Myös tässä tutkimuksessa *Burkholderiales* oli yleisin löydetty lahko (24,0 %), mutta esimerkiksi Staleyn ja kumppaneiden Mississippijoella kolmanneksi yleisimmäksi lahkoksi havaitsemaa *Pseudomonadales* -lahkoa ei tässä tutkimuksessa havaittu lainkaan ja heidän toiseksi yleisintä lahkoa *Actinomycetales* löydettiin tässä tutkimuksessa kaikista tutkimuspaikoista keskimäärin vain 1,1 prosenttia. On kuitenkin huomattava, että tässä tutkimuksessa tutkimme virtavesien koskipaikkojen kivien pinnoille muodostuneita biofilmibakteereita, kun taas Staley kumppaneineen eristi bakteerinäytteet suoraan vapaasti virtaavasta jokivedestä. Tämän tutkimuksen aineisto on myös kerätty huomattavasti pienemmistä luvvesistä kuin Staleyn ja kumppaneiden tutkimien jokien aineisto. Erilaiset kasvuolosuhteet ja eroavaisuudet uomien koossa on siis syytä tuloksia vertailtaessa huomioida.

Heino ym. (2014) havaitsivat, että pohjoisissa virtavesissä lehtikarikkeesta eristetyssä bakteri-DNA:ssa yleisimmät havaitut pääjaksot olivat *Proteobacteria* (54 %), *Bacteroidetes* (22,4 %) ja *Actinobacteria* (8,72 %). Havaituista proteobakteereista 45,5 % kuului luokkaan *Alphaproteobacteria*, 33,5 prosenttia kuului luokkaan *Gammaproteobacteria*, 10,3 prosenttia kuului luokkaan *Deltaproteobacteria* 10,0 prosenttia kuului luokkaan *Betaproteobacteria*.

Kahden yleisimmän pääjakson osalta tulokset ovat hyvin samansuuntaiset tässä tutkimuksessa havaittujen kanssa. Heino ym. (2014) havaitsivat kuitenkin yli 8 prosenttia tutkituista bakteereista kuuluvan *Actinobacteria* -pääjaksoon. Tässä tutkimuksessa kyseiseen pääjaksoon kuului vain 1,6 prosenttia kaikista havaituista OTUista. Paikkojen välillä *Actinobacteria* -pääjakson osuus vaihteli 0–6,1 % ollen korkein Uskelanjoki 1 -tutkimuspaikalla. Osuus vaihtelee myös samaan kolmannen jakovaiheen jokivaluma-alueeseen kuuluvien tutkimuspaikkojen sisällä, sillä Uskelanjoki 4 -tutkimuspaikalla *Actinobacteria* -pääjakson osuus on vain 0,2 %. Isojoki 4 tutkimuspaikalla pääjaksoa ei havaittu lainkaan.

Proteobakteereista *Alphaproteobacteria* -luokkaan kuului tässä tutkimuksessa 28,9 % OTUista ja Heinon ym. (2014) tutkimuksessa 45,5 %. Seuraavaksi yleisin havaittu luokka oli tässä tutkimuksessa *Betaproteobacteria* (27.7 %), ja Heinon ym. (2014) tutkimuksessa se oli *Gammaproteobacteria* 33.5 %. Heino ym. (2014) havaitsivat 10,0 prosenttia OTUista kuuluvan luokkaan *Betaproteobacteria*. Tässä tutkimuksessa havaitsimme, että keskimäärin 7,5 % OTUista kuului *Gammaproteobacteria*-luokkaan, joskin paikkojen välillä havainnot vaihtelivat Oulujoki 1 tutkimuspaikan 0,2 prosentista Perhonjoki 1 tutkimuspaikan 24 prosenttiin.

Eroavaisuuksia näiden kahden tutkimuksen välillä selittävät todennäköisesti ympäristöolosuhteiden lisäksi erilainen lähdemateriaali. Tässä tutkimuksessa bakteeri-DNA eristettiin länsisuomalaisten virtavesien koskipaikkojen kivien pinnan biofilmibakteereista, ja Heinon ym. (2014) tutkimuksessa pohjoisissa virtavesissä 30 päivää tiheissä, pohjaeläinten pussiin pääsyn estävissä kangaspusseissa olleesta harmaalepän (*Alnus incana*) lehtikarikemateriaalista. Myös sekvensointimenetelmät (454 pyrosekvensointi ja Ion Torrent -sekvensointi) ja käytetyt bakteerialukkeet (519F/926R ja 8F/518R) erosivat toisistaan, joten aineistot eivät ole suoraan vertailukelpoisia keskenään.

4.5 Soveltuuko mikrobien yhteisötutkimus biomonitoroinnin työkaluksi?

Sekvensointimenetelmien nopea kehittyminen viime vuosien aikana on mahdollistanut yhä laajempien mikrobiaineistojen kustannustehokkaan analysoinnin uuden polven sekvensointimenetelmiä hyödyntämällä. Suomalaisista vesiekosysteemeistä ei kuitenkaan tähän mennessä ole juurikaan tehty laaja-alaisia mikrobien ja erilaisten ympäristötekijöiden välisiä suhteita tarkastelevia tutkimuksia. Tämä tutkielma pyrkii omalta osaltaan vastaamaan tähän puutteeseen ja tuottamaan uutta tietoa pohjoisten virtavesien akvaattisten biofilmibakteerien yhteisöistä. Tietämykseni mukaan tutkimukseni on laajin koskaan tehty suomalaisten virtavesien mikrobiyhteisöjen analyysi ja laaja-alaisena sitä voidaan pitää myös kansainvälisellä mittapuulla.

Tulevaisuudessa tarvitaan yhä kustannustehokkaampia keinoja biomonitorointiin ja erilaisten ympäristön tilaa uhkaavien prosessien, kuten ilmastonmuutoksen vaikutusten havainnointiin ja ennustamiseen. Vesistömikrobien tutkiminen 16S -sekvensoinnilla voisi soveltua tähän tarkoitukseen hyvin, kunhan menetelmän tämänhetkiset rajoitukset ja heikkoudet tunnistetaan. Tämän tutkielman tulosten perusteella virtaavien vesien mikrobien yhteisö-ympäristösuhteet ovat heikkoja ja ympäristön tilan seurannassa tulisikin tulosten perusteella käyttää mikrobien osalta muita vastemuuttujia kuin yhteisörakennetta. Yhteisörakenteen sijaan erilaiset taksonomiset, funktionaaliset ja fylogeneettiset indeksit voivat soveltua tutkimusaineistomme kaltaisille mikrobiaineistoille (esim. Cadotte & Tucker 2017; Cadotte ym. 2009).

Mikrobinäytteen kerääminen esimerkiksi kivien pintojen biofilmimateriaaleilta on helppoa ja edullista, joten muun biologisen näytteenoton, kuten esimerkiksi pohjaeläin- ja piilevänäytteenoton yhteydessä on hyvin pienellä lisävaivalla toteutettavissa myös bakteerinäytteenotto. Myös sekvensointikustannukset ovat perinteisiin biologisten seurantamenetelmien analysointikustannuksiin verrattuna todella alhaiset (tällä hetkellä menetelmästä riippuen arviolta noin 5-15 euroa per näyte). Näytteiden analysointi bioinformatiikan työkaluilla vaatii kuitenkin erityisosaamista ja tehokkaita palvelimia. Näytteet säilyvät -20 asteessa pakastettuna useamman vuoden, joten bakteerinäytteiden keräämisen voi ajoittaa mahdollisten muiden maastonäytteiden haun kanssa samaan aikaan, vaikkei näytteiden analysointi olisikaan juuri silloin ajankohtaista. Näytteenottoajankohtaa valitessa tulee kuitenkin huomioida yhteisöjen vuodenaikaisvaihtelu, jota tulisi bakteeriyhteisöjen osalta jatkossa tutkia tarkemmin. Virtavesien biofilmiyhteisöjen muodostuminen on jatkuva prosessi, jossa virtaava vesi tuo mukanaan yläpuolisesta uomasta ja valuma-alueelta lähtöisin olevia uusia bakteeritaksoneita biofilmiin. Aikaisemmassa tutkimuksessa havaitsimme Oulun lähistöllä sijaitsevilla Muhosjoella, Tyrnävänjoella ja Juuvanjoella toteuttamassamme toukokuun puolivälistä lokakuun puoliväliin kestäneessä seurantatutkimuksessa, että Benthofluorometrillä määritetty sini-, viher-, ja piilevien määrä kasvoi virtavesien kivien pinnoilla tasaisesti kesän aikana, mutta havaitut levämäärät kääntyivät

tutkimuksen loppupuolella syys-lokakuussa jyrkkään laskuun (Aroviita ym. julkaisematon). Kivien pintojen biofilmejä ja niiden vuodenaikaisvaihteluita tulisi jatkossa myös bakteerien osalta tutkia tarkemmin. On myös huomattava, ettei biofilmin paksuus ja määrä välttämättä korreloi lajimäärän kanssa, vaan paksuakin biofilmiä voivat dominoida tietyt yleiset lajit.

Virtavesien mikrobien ja ympäristötekijöiden suhteesta tarvitaan jatkossa enemmän tietoa. Jo nykyään tiedetään esimerkiksi tiettyjen bakteerien ja vesikemiamuuttujien välisistä yhteyksistä, mutta eri taksonien tarkemmasta roolista muuttuvissa ympäristöolosuhteissa tarvitaan tarkempaa tutkimustietoa ennen kuin mikrobianalyysejä voidaan laajamittaisesti käyttää biomonitoroinnin täysimittaisena työkaluna ja ympäristön tilan kuvaajana.

4.6 Ajatuksia jatkotutkimuksille

Tässä työssä havaittiin virtaavien vesien mikrobiyhteisöjen yhteisö-ympäristösuhteiden olevan heikkoja. Ympäristön tilan seurannassa tulisikin jatkossa tulosten perusteella käyttää mikrobien osalta muita vastemuuttujia kuin BLAST-analysoituihin OTUihin perustuvaa yhteisörakennetta. Erilaiset taksonomiset, funktionaaliset ja fylogeneettiset diversiteetti-indeksit voisivat jatkossa soveltua tässä tutkielmassa kuvattuja menetelmiä paremmin virtavesiolosuhteissa elävien mikrobien yhteisöjen tutkimiseen. (esim. Cadotte & Tucker 2017; Cadotte ym. 2009).

Tulevaisuudessa bakteeriyhteisöjen toiminnallisuutta voidaan tutkia tarkemmin esimerkiksi fosmidikirjastoperusteisilla-, RNA-perusteisilla-metatraskriptomianalyysejä eli mikrobiyhteisöjen tuottamasta RNA:sta saatavaa tietoa ja metagenomianalyysejä yhdistävillä tai proteiinien määrää mittaavia metaproteomimenetelmiä ja metagenomianalyysejä yhdistävillä menetelmillä. Yhdessä tässä tutkimuksessa käytetyn bakteerien metagenomitutkimuksen kanssa nämä uudet kehittyvät menetelmät voivat tarjota tarkempia vastauksia siihen, mitkä bakteeriyhteisöjen jäsenet todella ovat toiminnallisia ja mikä niiden rooli esimerkiksi erilaisten metallien tai ravinteiden suhteen on. Myös bioinformatiikan menetelmien ja sekvenssikirjastojen kehittyminen tarjoaa jatkossa yhä

monipuolisempia ja tehokkaampia työkaluja laajojen sekvenssiaineistojen analysointiin. Tilastotieteen menetelmistä muun muassa erilaiset Bayesilaisen verkon käyttöön perustuvat menetelmät esimerkiksi fysikaalis-kemiallisten tekijöiden ja maankäytön vaikutuksesta bakteerien yhteisörakenteeseen voivat tulevaisuudessa tarjota myös uutta tietoa (ks. esim. Staley ym. 2015).

4.6.1 Taksonomiset, funktionaaliset ja fylogeneettiset diversiteetti-indeksit

Taksonomiset, funktionaaliset ja fylogeneettiset diversiteetti-indeksit voisivat jatkossa soveltua tässä tutkielmassa kuvattuja menetelmiä paremmin virtavesiolosuhteissa elävien mikrobien yhteisöjen tutkimiseen. (esim. Cadotte & Tucker 2017; Cadotte ym. 2009). Esimerkiksi Cadotte ja kumppanit (2009) havaitsivat että lajienvälisten toiminnallisten eroavaisuuksien kuvaaminen on usein haastavaa, mutta fylogeneettiset diversiteetti-indeksit soveltuvat tällaisiin tarkasteluihin monia muita lähestymistapoja paremmin (Cadotte ym. 2009).

Lajirunsauden ja esiintyvyyden suhdetta kuvaamaan pyrkivistä diversiteetti-indekseistä yleisimmät ovat Shannonin diversiteetti-indeksi (H') (Shannon 1948) ja Simpsonin diversiteetti-indeksi ($D1$) (Simpson 1949). Mikrobien taksonomista monimuotoisuutta ja diversiteettiä voidaan tutkia esimerkiksi OTU-runsauden ja sekä Shannonin että Simpsonin diversiteetti-indeksejä (Shannon 1948; Simpson 1949) yhdistävän tutkimuksen avulla (Besemer ym. 2013). Diversiteetti-indeksejä käyttävää lähestymistapaa on käytetty myös useissa mikrobien roolia hajotustoiminnassa kuvaavissa lehtikariketutkimuksissa (ks. esim. Kominoski ym. 2011). On kuitenkin huomioitava, että indeksien valinta voi vaikuttaa analyysien lopputuloksiin ja niiden valinnassa täytyy aina tarkastella huolellisesti sitä, soveltuvatko ne tarkastelun kohteena olevan aineiston analysointiin (Besemer ym. 2013). Besemer ja kumppanit (2013) testasivat useita erilaisia diversiteetti-indeksejä eri eliöryhmillä, kuten ruohovartisilla kasveilla, pensaiden juurien sienillä, maanpäällä elävillä niveljalkaisilla ja maan alla elävillä hyönteisten toukilla. He havaitsivat yleisten yksittäisten diversiteetti-indeksien toimivan yksikertaisissa analyyseissa, mutta useamman lajiryhmän välisiä moniulotteisempia yhteyksiä

tutkittaessa indeksien valinnat näyttäisivät vaikuttavan saataviin lopputuloksiin. He suosittavatkin useamman eri diversiteetti-indeksin yhtäaikaista käyttöä monimutkaisempien riippuvuussuhteiden selvittämiseksi. (Besemer ym. 2013).

Myös esimerkiksi erilaisia indikaattorilaji-analyysejä (ks esim. Dufrière & Legendre 1997) voidaan käyttää aineiston analysoinnin apuna. Yksikertaisimmillaan niissä luodaan havaittujen lajien suhteellisiin runsauksiin perustuva indikaattoriarvo (0-100). Monte Carlo -testillä tutkitaan esimerkiksi 1000 satunnaistamisella, ovatko indikaattoriarvot sattumanvaraista suurempia. Indikaattorilajien täytyy saada indikaattoriarvo > 20 ja p- arvo $\leq 0,05$ (Dufrière & Legendre 1997). Tämä lähestymistapa kuitenkin edellyttää, että tarkastelun kohteena ovat jonkinlaiset paikkaryhmät (esim. eri maantieteellisten alueiden näytteet).

4.6.2 Fosmidikirjastoperusteiset menetelmät

Bakteerien fysiologisen tilan arvioinnissa viljelymenetelmien on perinteisesti katsottu kuvaavan paremmin todellisuutta kuin ei-viljelyyn perustuvien menetelmien (Ellis ym. 2003), sillä vaikka tietty geeni olisikin havaittu, ei voida olla varmoja geenin ilmenemisestä ympäristössä. Jopa 99 prosenttia bakteerilajeista on kuitenkin sellaisia, jotka eivät suoraan kasva viljelyolosuhteissa (Amann ym. 1995). Laajojen, toiminnallisten fosmidikirjastojen tarkastelu voi olla yksi tulevaisuuden mahdollisuus mikrobiyhteisöjen toiminnallisuuden tutkimiseen (Martinez & Osburne, 2013). Fosmidikirjastot luodaan niin, että kiinnostuksen kohteena oleva DNA-insertti ensin kloonataan fosmidivektoriin ja siirretään sitten isäntäbakteeriin, joka yleensä on *E. coli*. Fosmidikirjasto (fosmid pools) muodostuu kun bakteerikloonit monistuvat. Tämän jälkeen fosmidikirjastoja voidaan edelleen koostaa ja sekvensoida pienemmissä osissa. Menetelmän heikkouksia ovat tällä hetkellä vielä suuri työmäärä ja korkeat kustannukset (Nystedt ym. 2013; Kumpula 2016) On myös huomioitava, että yllä kuvatun kaltaisessa fosmideilla toteutettavassa tutkimuksessa isännän negatiivisen selektion osuus on hyvin suuri ja tästä voi seurata se, etteivät fragmentit kloonaudu kunnolla tai ne eivät toimi heterologisessa isännässä (Auvinen, suullinen tiedonanto 4.6.2018).

Staley ym. (2015) tekivät fosmidikirjastojen avulla toiminnallisen metagenomin skreenauksen ylemmältä Mississippijoelta vuosina 2011 ja 2012 kerätyistä bakteerinäytteistä. He tutkivat bakteerien vastustuskykyä kadmiumia, kromia, kuparia, elohopeaa, mangaania ja sinkkiä kohtaan. He olivat myös aiemmin samoille näytteille tehneet 16S metagenomianalyysin ja OTUston taksonomian määrittämisen. Heidän hypoteesinsa mukaan samat tekijät, jotka vaikuttavat taksonomiseen yhteisöstruktuuriin, kuten kokonaishiili ja nitriitti- ja nitraattipitoisuudet sekä maankäyttö, vaikuttavat myös bakteerien vastustuskykyyn tiettyjä metalleja kohtaan. (Staley ym. 2015).

Näytteenotossa Staley ym. (2015) keräsivät 40 litraa jokivettä jokaisesta tutkimuspaikasta steriileihin astioihin ja suodattivat ja pellettoivat näytteet. DNA eristettiin kaupallisella kitillä ja halutut DNA –fragmentit valittiin koon perusteella elektroforeesia käyttäen. Ligatoidut fosmidit siirrettiin *E. coli*in. Jokaisen tutkimuspaikan fosmidikirjasto sisälsi vähintään 50000 kloonia. Bakteerien vastustuskykyä metalleille tutkittiin viljelemällä agar maljoille aiemmin kuvatulla tavalla tuotetut fosmidit sisältävät *E. coli*t. Maljoille lisättiin aina myös tunnettu määrä tutkittavaa metallia sisältävää liuosta ja kokeen toimivuus varmistettiin kontrollikannoilla. Vastustuskykyisiksi kannoiksi raportoitiin ne, jotka tuottivat halkaisijaltaan vähintään 1 mm kokoisen pesäkkeen. (Staley ym. 2015).

Bakteerien vastustuskyky kaikkia testattuja metalleja kohtaan jäi kokeessa alhaiseksi verrattuna aiempiin, perinteisiin viljelymenelmiin perustuneisiin kokeisiin. Eron voidaan katsoa selittyvän sekä menetelmäeroavaisuuksilla että erilaisilla tutkimusympäristöillä. Aiemmissä tutkimuksissa näytteiden suora maljalle viljely on voinut valikoida vastustuskykyisiä isolaatteja, jotka kuitenkin edustavat mikrobiyhteisön vähemmistöä. Staley ym. (2015) ovat aiemmin havainneet, että jopa 90 prosenttia nyt tutkitun ekosysteemin bakteeriyhteisöstä muodostuu vain muutamasta lajista. Nämä lajit ovat todennäköisesti ylliedustettuina fosmidikirjastoissa. Pienen bakteerivähemmistön muodostamaa metalliresistenssiä ei siis todennäköisesti fosmidikirjastomenetelmällä pystytä havaitsemaan, vaikka suuri määrä klooneja tutkittiin. Fosmidikirjastot ovat voineet myös vahingoittua pakastamisen aikana. Ympäristö-DNA ei välttämättä

ole myöskään siirtynyt fosfideihin oikein tai sen ekspressiossa on voinut olla ongelmia. (Staley ym. 2015).

Ravinnepitoisuudet ja bakteerilahkot korreloivat heikosti havaitun raskasmetallien vastustuskyvyn yleisyyteen. Poikkeuksena tähän tutkimuksessa havaittiin kuitenkin *Pseudomonadales*-bakteerien yhteys elohopea- ja sinkkiresistenssiin. Tulokset tukevat sitä, että tutkitussa virtavesiympäristössä bakteerien vastustuskyky raskasmetalleille on todennäköisesti peräisin ympäröivän maankäytön seurauksena virtavesiympäristöön päätyvistä vierasperäisistä bakteereista. (Staley ym. 2015).

4.6.3 RNA –perusteiset menetelmät

Tulevaisuudessa bakteerien eristys- ja sekvensointimenetelmien kehittyminen sekä uuden tutkimustiedon myötä kehittyvät RNA –kirjastot mahdollistavat yhä useamman bakteeriaineiston analysoinnin RNA –menetelmien avulla. Siinä missä DNA-perusteiset menetelmät eristävät kaiken ympäristönäytteessä olevan DNA:n voidaan RNA –analyysien avulla tutkia niitä bakteereja, jotka näytteenottohetkellä ovat olleet toiminnallisia bakteeriyhteisössä. Viimeaikaiset tutkimukset ovat osoittaneet, että vaikka tietyn havaitun bakteeritaksonin määrä tutkimuspaikalla olisikin korkea, ei kyseinen taksoni kuitenkaan välttämättä ole merkittävässä roolissa yhteisön toiminnallisuuden kannalta (Chen ym. 2015).

4.6.4 Metagenomi- ja metatranskriptomianalyysijä yhdistävät menetelmät

Metagenomi- ja metatranskriptomianalyysijä eli mikrobiyhteisöjen tuottamasta RNA:sta saatavaa tietoa yhdistävät menetelmät voivat jatkossa tarjota paljon uutta geenitason tietoa bakteeriyhteisöjen toiminnallisen aktiivisuuden yhä tarkempaan tutkimiseen. Myös yksittäisten bakteeriorganismien kyvystä vastata ja sopeutua muuttuneisiin ympäristöolosuhteisiin voidaan menetelmän avulla saada paljon uutta lisätietoa (Franzosa ym. 2015; Chen ym. 2015).

Chen ym. (2015) sovelsivat menetelmää eristämällä kiinalaisista kaivosympäristöistä bakteerien DNA:ta ja RNA:ta. RNA:lle tehtiin kääntö RNA:sta komplementaariseksi DNAksi (cDNA). 454 pyrosekvensoinnin jälkeen erilaisia bioinformatiikan työkaluja ja tietokantoja hyödyntäen analysoitiin mitkä tutkituista DNA ja cDNA sekvensseistä kuuluivat bakteerien 16S rRNA:han ja ne jaettiin Ribosomal Database Project Classifier:in avulla phylogeneettisiin ryhmiin.

Tämän jälkeen Chen ja kumppanit tekivät sekvenssiaineiston trimmausta CD-Hit-työkalulla ja jäljelle jääneille sekvensseille (ei-rRNA, ei-replikaatti DNA ja cDNA) haettiin NCBI:n non-redundant tietokannasta vastaavia geenejä ja tulokset käsiteltiin MEGANissa. He löysivät noin 500000 DNA ja 50000 cDNA sekvenssiä, jotka vastasivat arkeonien ja bakteerien proteiinejä koodaavia geenejä. Lopuksi Chen ym. (2015) käyttivät vielä COGs (Clusters of Orthologous Groups of proteins) ja KEGG (Kyoto Encyclopedia of Genes and Genomes) tietokantoja näiden geenien analysoimiseksi.

Yksi keskeinen tekijä metatranskriptomianalyysimenetelmien onnistumisen kannalta on ribosomaalisen RNA:n (rRNA) poistaminen totaali RNA:sta, jossa tutkijat ovat raportoineet olevan vaikeuksia (Chen ym. 2015; Stewarts ym. 2010). Myös sekvensointimenetelmien ja sitä myötä sekvensointisyvyyden kehitys mahdollistaa jatkossa yhä tarkempien metatranskriptomianalyysimenetelmien käyttämisen. Menetelmän avulla voidaan tulevaisuudessa luoda yhä tarkempia geenitason profiileja ympäristögradienttien tai aikasarjojen perusteella sekä yleisille että harvinaisille bakteeritaksoneille (Chen ym. 2015).

4.6.5 Metagenomi- ja metaproteomitutkimusta yhdistävät menetelmät

Yhdistämällä metagenomitutkimuksesta ja tuotettujen proteiinien määrää mittaavista metaproteomitutkimuksista saatua tietoa voidaan tulevaisuudessa yhä tarkemmin tutkia bakteerien ekologiaa, evoluutiota ja sopeutumista erilaisiin, joskus hyvin äärimmäisiin, olosuhteisiin (Franzosa ym. 2015; Deneff & Banfield 2012). Geenien ilmenemisestä bakteeripopulaatioissa sekä niiden paikallisesta

aineenvaihdunnallisesta aktiivisuudesta tiedetään kuitenkin vielä tänä päivänä vain vähän (Chen ym. 2015) ja tarve tarkemmille jatkotutkimuksille on ilmeinen.

5 Kiitokset

Haluan kiittää työni ohjaajia FT Jani Heinoa Suomen ympäristökeskuksesta ja Bakteerigenetiikan yliopistonlehtoria Hannu Saarilahtea Helsingin yliopistosta ohjauksesta ja tuesta graduprosessin eri vaiheissa. Janille tahdon lausua erityiskiitokset kärsivällisestä ja kannustavasta ohjauksesta sekä perehdyttämisestä tilastoanalyysien mielenkiintoiseen maailmaan. Janin Large-Scale Ecology Research Lab -tutkimusryhmässä olen saanut syventyä vuodesta 2012 mikrobien DNA-analyysien pohjattoman mielenkiintoiseen maailmaan. Ensimmäisen aineiston DNA-analyysijä aloittaessamme meillä ei ollut mitään takeita lopputuloksen onnistumisesta, ja kaikki aina alukevalinnoista ja oikeista pipetointimääristä alkaen piti kokeilla yrityksen ja erehdyksen kautta. Silti Jani uskoi työn kantavan lopulta hedelmää ja rahoitti tutkimusta. Siitä haluan lausua syvimät kiitokseni. Hannun mikrobigenetiikan ja -genomiikan kurssit ovat puolestaan opettaneet minulle paljon uutta. Hannu on myös ollut uskomattoman joustava ja mahdollistanut opintojen suorittamisen kokopäivätyön ohessa Oulusta käsin Helsinkiin.

Haluan myös kiittää koko Scale-Bio -tiimiä, johon Janin lisäksi kuuluvat FT Jenny Jyrkänkallio-Mikkola, Professori Janne Soininen, FT Kimmo Tolonen, FT Mikko Tolkkinen ja FT Virpi Pajunen, hyvästä yhteistyöstä. Projektitutkija Marko Suokas Oulun yliopiston molekyylibiologian BioSer -laboratoriosta on tehnyt tämän graduaineiston, niin kuin niin monen muunkin SYKEN aineistomme, bakteerien sekvensoinnit. Marko on opettanut minulle todella paljon viimeisten viiden vuoden aika mikrobien DNA -analyyseista. Bakteeriaineistojen analysointiin ja bioinformatiikan työkalujen käyttöön minua on perehdyttänyt FT Mikko Tolkkinen. FT Kaisa Lehosmaa puolestaan on ollut korvaamaton vertaistuki niin laboratoriossa kuin bakteeriaineistojen käsittelyssä bioinformatiikan työkaluilla. Kaisan avustuksella muun muassa tämän graduaineiston BLAST -analyysit tehtiin tehokkaalla Oulun yliopiston palvelimella. Ilman Kaisaa niin laboratoriotyöt kuin analyysitkin olisivat olleet monin verroin yksinäisempää puuhaa.

Graduaineistoni on kerätty Suomen ympäristökeskuksessa ja Helsingin yliopistossa osana Suomen Akatemian rahoittamaa tutkimusta *Biodiversiteetti eri mittakaavoilla trooppisissa ja boreaalisissa puroissa: merkitys diversiteetin kartoitukselle ja ympäristön tilan seurannalle ("Scale-Bio")*. Haluan kiittää myös työnantajaani Suomen ympäristökeskusta ja työkavereitani tuesta ja kannustuksesta graduprosessin ja maisteriopintojeni aikana. Erityiskiitokset kuuluvat työhuonetoverilleni FT Jarno Turuselle, joka on tsempannut minua läpi koko opiskelutaipaleen.

Lopuksi haluan kiittää ystäviäni ja perhettäni: vanhempiani, appivanhempiani, aviomiestäni Aria ja Vilho-koiraamme kaikesta siitä tuesta ja kannustuksesta jonka olen saanut. Suurimmat kiitokset kuuluvat syyskuun 2017 lopussa syntyneelle esikoispojallemme Onnille, joka odotusaikana ihan kirjaimellisesti potki vauhtia tämän gradun kirjoitukseen ja joka gradun hieman pitkähköksi venähtäneen viimeistelyn aikana on ehtinyt kasvaa uteliaaksi, innokkaaksi ja alati uutta oppivaksi taaperoksi.

Lähteet

- Alahuhta, J., Johnson, L. B., Olker, J., & Heino, J. (2014). Species sorting determines variation in the community composition of common and rare macrophytes at various spatial extents. *Ecological Complexity*, 20, 61-68.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215(3), 403–410.
- Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews*, 59(1), 143-169.
- Anderson-Glenna, M. J., Bakkestuen, V., & Clipson, N. J. (2008). Spatial and temporal variability in epilithic biofilm bacterial communities along an upland river gradient. *FEMS microbiology ecology*, 64(3), 407-418.
- Aroviita, J., Vuori, K. M., Hellsten, S., Jyväsjärvi, J., Järvinen, M., Karjalainen, S. M., Kauppila, P., Korpinen, S., Kuoppala, M., Mitikka, S., Mykrä, H., Olin, M., Rask, M., Riihimäki, J., Räike, A., Rääpysjärvi, J., Sutela, T., Vehanen, T., Vuorio, K. (2014). Maa- ja metsätalouden kuormittamien pintavesien ekologinen tila ja sen seuranta. Suomen ympäristökeskuksen raportteja 12/2014. Suomen ympäristökeskus, Helsinki, 96s.
- Aroviita, J., Hellsten, S., Jyväsjärvi, J., Järvenpää, L., Järvinen, M., Karjalainen, S. M., Kauppila, P., Keto, A., Kuoppala, M., Manni, K., Mannio, J., Mitikka, S., Olin, M., Perus, J., Pilke, A., Rask, M., Riihimäki, J., Ruuskanen, A., Siimes, K., Sutela, T., Vehanen, T., Vuori, K-M. (2012). Ohje pintavesien ekologisen ja kemiallisen tilan luokitteluun vuosille 2012–2013– päivitettyt arviointiperusteet ja niiden soveltaminen. Ympäristöhallinnon ohjeita 7/2012. Suomen ympäristökeskus, Helsinki, 144s.
- Austin, M. P., & Greig-Smith, P. (1968). The application of quantitative methods to vegetation survey: II. Some methodological problems of data from rain forest. *The Journal of Ecology*, 827-844.
- Auvinen, P., Hultman, J. (2010). Metagenomiikka avaa uusia ovia mikrobiologiassa. *Lääketieteellinen Aikakauskirja Duodecim*. 2010;126(11):1278-85
- Baas Becking, L.G.M. (1934) *Geobiologie of inleiding tot de milieukunde*. Haag, Alankomaat: W.P. Van Stockum & Zoon
- Batianoff, G. N., & Burgess, R. (1993). Problems in the documentation of rare plants-the Australian experience. *Biodiversity Letters*, 1(6), 168-171.

- Blanchet, F. G., Legendre, P., & Borcard, D. (2008). Forward selection of explanatory variables. *Ecology*, 89(9), 2623-2632
- Besemer, K. (2016). Microbial biodiversity in natural biofilms. Teoksessa: Romani, A. M., Guasch, H., & Balaguer, M. (toim.) *Aquatic biofilms: ecology, water quality and wastewater treatment*. Norfolk: Caister Academic Press; 2016. p. 63e87.
- Besemer, K. (2015). Biodiversity, community structure and function of biofilms in stream ecosystems. *Research in microbiology*, 166(10), 774-781.
- Besemer, K., Singer, G., Quince, C., Bertuzzo, E., Sloan, W., & Battin, T. J. (2013). Headwaters are critical reservoirs of microbial diversity for fluvial networks. *Proceedings of the Royal Society B: Biological Sciences*, 280(1771), 20131760.
- Besemer, K., Peter, H., Logue, J. B., Langenheder, S., Lindström, E. S., Tranvik, L. J., & Battin, T. J. (2012). Unraveling assembly of stream biofilm communities. *The ISME journal*, 6(8), 1459
- Bonada, N., Prat, N., Resh, V. H., & Statzner, B. (2006). Developments in aquatic insect biomonitoring: a comparative analysis of recent approaches. *Annu. Rev. Entomol.*, 51, 495-523.
- Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical ecology with R*. Springer Science & Business Media.
- Borcard, D., Legendre, P., Avois-Jacquet, C., & Tuomisto, H. (2004). Dissecting the spatial structure of ecological data at multiple scales. *Ecology*, 85(7), 1826-1832.
- Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, 153(1), 51-68.
- Borcard, D., Legendre, P., & Drapeau, P. (1992). Partialling out the spatial component of ecological variation. *Ecology*, 73(3), 1045-1055.
- Cadotte, M. W., Cavender-Bares, J., Tilman, D., & Oakley, T. H. (2009). Using phylogenetic, functional and trait diversity to understand patterns of plant community productivity. *PloS one*, 4(5), e5695
- Cadotte, M. W., & Tucker, C. M. (2017). Should environmental filtering be abandoned?. *Trends in ecology & evolution*, 32(6), 429-437.
- Cao, Y., Larsen, D. P., & Thorne, R. S. J. (2001). Rare species in multivariate analysis for bioassessment: some considerations. *Journal of the North American Benthological Society*, 20(1), 144-153.

- Caporaso J. G. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335-336.
- Chen, L. X., Hu, M., Huang, L. N., Hua, Z. S., Kuang, J. L., Li, S. J., & Shu, W. S. (2015). Comparative metagenomic and metatranscriptomic analyses of microbial communities in acid mine drainage. *The ISME journal*, 9(7), 1579.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to algorithms*. MIT press.
- Cornforth, D. M., Popat, R., McNally, L., Gurney, J., Scott-Phillips, T. C., Ivens, A., Diggle, S. P., Brown, S. P. (2014). Combinatorial quorum sensing allows bacteria to resolve their social and physical environment. *Proceedings of the National Academy of Sciences*, 111(11), 4280-4284.
- Cottenie, K. (2005). Integrating environmental and spatial processes in ecological community dynamics. *Ecology letters*, 8(11), 1175-1182.
- Davey, M. E., & O'toole, G. A. (2000). Microbial biofilms: from ecology to molecular genetics. *Microbiology and molecular biology reviews*, 64(4), 847-867.
- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *ecological modelling*, 196(3-4), 483-493.
- Dray, S., Péliissier, R., Couteron, P., Fortin, M. J., Legendre, P., Peres-Neto, P. R., Bellier, SE., Bivand, R., Blanchet, G., De Caceres, M., Dufour, A. B., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J., Wagner, H. H. (2012). Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs*, 82(3), 257-275.
- De'Ath, G. (2002). Multivariate regression trees: a new technique for modeling species–environment relationships. *Ecology*, 83(4), 1105-1117.
- Denef, V. J., & Banfield, J. F. (2012). In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. *Science*, 336(6080), 462-466
- De Wit, R., & Bouvier, T. (2006). ‘Everything is everywhere, but, the environment selects’; what did Baas Becking and Beijerinck really say?. *Environmental microbiology*, 8(4), 755-758.
- Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A-H., Soto, D., M.L.J. Stiassny, Sullivan, C. A. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews*, 81(2), 163-182.

- Dufrêne, M., & Legendre, P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological monographs*, 67(3), 345-366.
- Edgar R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460-2461.
- Ellis, R. J., Morgan, P., Weightman, A. J., & Fry, J. C. (2003). Cultivation-dependent and-independent approaches for determining bacterial diversity in heavy-metal-contaminated soil. *Applied and environmental microbiology*, 69(6), 3223-3230.
- Ezekiel, M. (1930). *Methods of correlation analysis*. John Wiley and Sons, New York, New York, USA.
- Euroopan parlamentti ja neuvosto 2000. Euroopan parlamentin ja neuvoston direktiivi 2000/60/EY, yhteisön vesienhoidon puitteista. *Euroopan yhteisön virallinen lehti* 327: 1–72.
- Fierer, N., Morse, J. L., Berthrong, S. T., Bernhardt, E. S., & Jackson, R. B. (2007). Environmental controls on the landscape-scale biogeography of stream bacterial communities. *Ecology*, 88(9), 2162-2173.
- Fox, J. (2005) The R commander: a basic statistics graphical user interface to R. *Journal of Statistical Software*, 14, 1–42.
- Franzosa, E. A., Hsu, T., Sirota-Madi, A., Shafquat, A., Abu-Ali, G., Morgan, X. C., & Huttenhower, C. (2015). Sequencing and beyond: integrating molecular ‘omics for microbial community profiling. *Nature reviews. Microbiology*, 13(6), 360.
- Friberg, N., Bonada, N., Bradley, D. C., Dunbar, M. J., Edwards, F. K., Grey, J., Hayes, R.B, Hildrew, A.G., Lamouroux, N., Trimmer, M., Woodward, G. U. Y. (2011). Biomonitoring of human impacts in freshwater ecosystems: the good, the bad and the ugly. *Advances in Ecological Research*, 44, 1-68.
- Gaston, K. J., & Kunin, W. E. (Toim.). (1997). *The biology of rarity: causes and consequences of rare-common differences*. Chapman & Hall.
- Gauch, H. G. (1982). *Multivariate analysis in community ecology* (No. 1). Cambridge University Press.
- Gerhardt, A. (2000). *Biomonitoring of polluted water*. Trans Tech Publications Ltd.
- Gilpin, M.E. & Hanski, I.A. (1991). *Metapopulation Dynamics: Empirical and Theoretical Investigations*. Academic Press, London.
- Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Molecular ecology resources*, 11(5), 759-769

- Gower, J. C., & Ross, G. J. (1969). Minimum spanning trees and single linkage cluster analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 18(1), 54-64
- Guillot, G., & Rousset, F. (2013). Dismantling the Mantel tests. *Methods in Ecology and Evolution*, 4(4), 336-344.
- Hadzic, M., Postila, H., Österholm, P., Nystrand, M., Pahkakangas, S., Karppinen, A., Arola, M., Nilivaara-Koskela, R., Häkkilä, K., Saukkoriipi, J., Kunnas, S. & Ihme, R. (2014). Sulfaattimailla syntyvän happaman kuormituksen ennakointi- ja hallintamenetelmät –SuHE-hankkeen loppuraportti. Suomen ympäristökeskuksen raportteja 17/2014. Suomen ympäristökeskus, Helsinki, 88 s.
- Hanski, I. (1991). Single-species metapopulation dynamics: concepts, models and observations. *Biological Journal of the Linnean Society*, 42(1-2), 17-38
- Hanski, I. (1998). Metapopulation dynamics. *Nature*, 396(6706), 41.
- Heino, J., Soininen, J., Alahuhta, J., Lappalainen, J., & Virtanen, R. (2017). Metacommunity ecology meets biogeography: effects of geographical region, spatial dynamics and environmental filtering on community structure in aquatic organisms. *Oecologia*, 183(1), 121-137.
- Heino, J., Nokela, T., Soininen, J., Tolkkinen, M., Virtanen, L., & Virtanen, R. (2015). Elements of metacommunity structure and community-environment relationships in stream organisms. *Freshwater Biology*, 60(5), 973-988.
- Heino, J., Tolkkinen, M., Pirttilä, A. M., Aisala, H., & Mykrä, H. (2014). Microbial diversity and community-environment relationships in boreal streams. *Journal of biogeography*, 41(12), 2234-2244.
- Heino, J., Grönroos, M., Soininen, J., Virtanen, R., & Muotka, T. (2012). Context dependency and metacommunity structuring in boreal headwater streams. *Oikos*, 121(4), 537-544.
- Jarník, V.; Über die simultanen Diophantischen Approximationen, *Math. Z.* 33, (1931), p. 505–543.
- Jyrkänkallio-Mikkola, J., Meier, S., Heino, J., Laamanen, T., Pajunen, V., Tolonen, K. T., Tolkkinen, M., Soininen, J. (2017). Disentangling multi-scale environmental effects on stream microbial communities. *Journal of Biogeography*, 44: 1512–1523.
- Kaksonen, A. H., Dopson, M., Karnachuk, O., Tuovinen, O. H., & Puhakka, J. A. (2008). Biological iron oxidation and sulfate reduction in the treatment of acid mine drainage at low temperatures. In *Psychrophiles: from Biodiversity to Biotechnology* (pp. 429-454). Springer, Berlin, Heidelberg.

- Kalia, D., Merey, G., Nakayama, S., Zheng, Y., Zhou, J., Luo, Y., Guo, M., Roembke, B. T., Sintim, H. O. (2013). Nucleotide, c-di-GMP, c-di-AMP, cGMP, cAMP,(p) ppGpp signaling in bacteria and implications in pathogenesis. *Chemical Society Reviews*, 42(1), 305-341.
- Karimi, B., Maron, P. A., Boure, N. C. P., Bernard, N., Gilbert, D., Ranjard, L. (2017). Microbial diversity and ecological networks as indicators of environmental quality. *Environmental Chemistry Letters*, 1-17.
- Karnachuk, O. V., Kadnikov, V. V., Panova, I. A., Mardanov, A. V., Beletsky, A. V., Danilova, E. V., Avakyan, M. R. & Ravin, N. V. (2017). Genome sequence of the copper resistant and acid-tolerant *Desulfosporosinus* sp. BG isolated from the tailings of a molybdenum-tungsten mine in the Transbaikal area. *Genomics data*, 11, 106-108.
- Kivinen, E. 1944. Aluna eli sulfaattimaista. *Maataloustieteen aikakauskirja* 16: 147-160.
- Kominoski, J. S., Marczak, L. B., & Richardson, J. S. (2011). Riparian forest composition affects stream litter decomposition despite similar microbial and invertebrate communities. *Ecology*, 92(1), 151-159.
- Kovaliova, A., Kadnikov, V. V., Antsiferov, D. V., Beletsky, A. V., Danilova, E. V., Avakyan, M. R., Mardanov, A. V. & Karnachuk, O. V. (2017). Genome sequence of the acid-tolerant *Desulfovibrio* sp. DV isolated from the sediments of a Pb-Zn mine tailings dam in the Chita region, Russia. *Genomics data*, 11, 125-127.
- Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical society*, 7(1), 48-50.
- Kumpula, T. (2016). Suurten kasvigenomien evoluutio ja tutkiminen. LuK-tutkielma. Oulun yliopisto. Genetiikan ja fysiologian tutkimusryhmä.
- Lane, D. J. 1991. 16S/23S rRNA sequencing, p. 115–175. Teoksessa E. Stackebrandt ja M. Goodfellow (toim.), *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, Chichester, United Kingdom.
- Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L., & Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences*, 82(20), 6955-6959.
- Langenheder, S., Wang, J., Karjalainen, S. M., Laamanen, T. M., Tolonen, K. T., Vilmi, A., & Heino, J. (2017). Bacterial metacommunity organization in a highly connected aquatic system. *FEMS microbiology ecology*, 93(4).

- Langenheder, S., & Prosser, J. I. (2008). Resource availability influences the diversity of a functional group of heterotrophic soil bacteria. *Environmental Microbiology*, 10(9), 2245-2256.
- Lear, G., Washington, V., Neale, M., Case, B., Buckley, H., & Lewis, G. (2013). The biogeography of stream bacteria. *Global Ecology and Biogeography*, 22(5), 544-554.
- Lear, G., Niyogi, D., Harding, J., Dong, Y., & Lewis, G. (2009). Biofilm bacterial community structure in streams affected by acid mine drainage. *Applied and environmental microbiology*, 75(11), 3455-3460.
- Legendre, P., & Legendre, L. F. (2012). *Numerical ecology* (Vol. 24). Elsevier.
- Legendre, P., & Gallagher, E. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129: 271-280.
- Leibold, M. A., & Chase, J. M. (2017). *Metacommunity ecology* (Vol. 59). Princeton University Press.
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecology letters*, 7(7), 601-613.
- Loukola-Ruskeeniemi, K., & Heino, T. (1996). Geochemistry and genesis of the black shale-hosted Ni-Cu-Zn deposit at Talvivaara, Finland. *Economic Geology*, 91(1), 80-110.
- Lynch, M. D., & Neufeld, J. D. (2015). Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology*, 13(4), 217.
- Manninen, M. (2014). *Ekologiset lajiryhmäjaottelut ja piilevyhteisöjen rakentuminen subarktisisissa virtavesissä. Pro gradu –tutkielma. Oulun yliopisto. 109 s.*
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1), 209-220.
- Martínez, A., & Osburne, M. S. (2013). Preparation of fosmid libraries and functional metagenomic analysis of microbial community DNA. *Methods Enzymology*, 531, 123-142.
- Martiny, J. B. H., Bohannan, B. J., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Krumins, J. A., Kuske, C.R., Morin, P. J., Naeem, S., Ovreas, L., Reusenbach, A. L., Smith, V. H., Staley, J. T. (2006). Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4(2), 102-112.
- Mayr, E. (1963). *Animal species and evolution*. Animal species and evolution.

- Mendoza, G., Kaivosoja, R., Grönroos, M., Hjort, J., Ilmonen, J., Kärnä, O. M., Paasivirta, L., Tokola, L., & Heino, J. (2018). Highly variable species distribution models in a subarctic stream metacommunity: Patterns, mechanisms and implications. *Freshwater Biology*, 63(1), 33-47.
- Mittelbach, G. G., Steiner, C. F., Scheiner, S. M., Gross, K. L., Reynolds, H. L., Waide, R. B., Willig, M. R., Dodson, S. I., Gough, L. (2001). What is the observed relationship between species richness and productivity?. *Ecology*, 82(9), 2381-2396.
- Mouillot, D., Bellwood, D. R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., Kulbicki, M., Lavergne, S., Lavorel, S., Mouquet, N. & Paine, C. T. (2013). Rare species support vulnerable functions in high-diversity ecosystems. *PLoS biology*, 11(5).
- Mykrä, H., Saarinen, T., Tolkkinen, M., McFarland, B., Hämäläinen, H., Martinmäki, K., & Kløve, B. (2012). Spatial and temporal variability of diatom and macroinvertebrate communities: How representative are ecological classifications within a river system?. *Ecological Indicators*, 18, 208-217.
- Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D., & Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews*, 75(1), 14-49.
- Nyberg, M. E., Österholm, P., & Nystrand, M. I. (2012). Impact of acid sulfate soils on the geochemistry of rivers in south-western Finland. *Environmental earth sciences*, 66(1), 157-168.
- Nystedt, B., Street, N. R., Wetterbom, A., Zuccolo, A., Lin, Y. C., Scofield, D. G., Vicedomini, R., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Käller, M., Luthman, J., Lysholm, F., Niittylä, T., Olson, Å., Rilakovic, N., Ritland, C., Rossello, J. A., Sena, J., Svensson, T., Talavera-Lopez, C., Günter, T., Tuominen, H., Vanneste, K., Wu, Z-Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Gil, R. G., Hvidsten, T. R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Thompson, S. L., Van de Peer, Y., Andersson, B., Nilsson, O., Ingvarsson, P. K., Lundeberg, J., & Jansson, S. (2013). The Norway spruce genome sequence and conifer genome evolution. *Nature*, 497(7451), 579.
- Olapade, O. A., & Leff, L. G. (2005). Seasonal response of stream biofilm communities to dissolved organic matter and nutrient enrichments. *Applied and Environmental Microbiology*, 71(5), 2278-2287.

- Pandit, S. N., Kolasa, J., & Cottenie, K. (2009). Contrasts between habitat generalists and specialists: an empirical extension to the basic metacommunity framework. *Ecology*, 90(8), 2253-2262.
- Peres-Neto, P. R., Legendre, P., Dray, S., & Borcard, D. (2006). Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*, 87(10), 2614-2625.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41(D1), D590-D596
- Rabinowitz, D. (1981). Seven forms of rarity. *Biological aspects of rare plant conservation*.
- Rao, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, 329-358.
- Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., Leamon, J. H., Johnson, K., Milgrew, M. J., Edwards, M., Hoon, J., Simons, J. F., Marran, D., Myers, J. W., Davidson, J. F., Branting, A., Nobile, J. R., Puc, B. P., Light, D., Clark, T. A., Huber, M., Branciforte, J. T., Stoner, I. B., Cawley, S. E., Lyons, M., Fu, Y., Homer, N., Sedova, M., Miao, X., Reed, B., Sabina, J., Feierstein, E., Schorn, M., Alanjary, M., Dimalanta, E., Dressman, D., Kasinskas, R., Sokolsky, T., Fidanza, J. A., Namsaraev, E., McKernan, K. J., Williams, A., Roth, G. T., & Bustillo J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348.
- Schlötterer, C., Tobler, R., Kofler, R., & Nolte V. (2014) Sequencing pools of individuals -mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* 15, 749–763.
- SFS-EN ISO 6878:en 2004: Water quality. Determination of phosphorus. Ammonium molybdate spectrometric method. Julkaisussa: SFS-käsikirja 147-3 (2011). Suomen Standardisoimisliitto SFS ry, Helsinki.
- SFS-EN ISO 11905-1 1998: Veden laatu. Typen määrittäminen. Osa 1: Peroksodisulfaattihapetus. Julkaisussa: SFS-käsikirja 147-3 (2011). Suomen Standardisoimisliitto SFS ry, Helsinki.
- SFS-EN ISO 7887:en 2011: Water quality. Examination and determination of colour. Suomen Standardisoimisliitto SFS ry, Helsinki.
- Shannon, C.E. (1948) A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423.

- Simpson, E. H. 1949. Measurement of diversity. *Nature* 163:688.
- Siqueira, T., Bini, L. M., Roque, F. O., Marques Couceiro, S. R., Trivinho-Strixino, S., & Cottenie, K. (2012). Common and rare species respond to similar niche processes in macroinvertebrate metacommunities. *Ecography*, 35(2), 183-192.
- Smith, V. H. (2007). Microbial diversity–productivity relationships in aquatic ecosystems. *FEMS Microbiology Ecology*, 62(2), 181-186.
- Song, W., Kim, M., Tripathi, B. M., Kim, H., & Adams, J. M. (2016). Predictable communities of soil bacteria in relation to nutrient concentration and successional stage in a laboratory culture experiment. *Environmental microbiology*, 18(6), 1740-175
- Staley, C., Sadowsky, M. J. (2016). Application of metagenomics to assess microbial communities in water and other environmental matrices. *Journal of the Marine Biological Association of the United Kingdom*, 96(1), 121-129. L 30.12.2004/1299.
- Staley, C., Johnson, D., Gould, T. J., Wang, P., Phillips, J., Cotner, J. B., & Sadowsky, M. J. (2015). Frequencies of heavy metal resistance are associated with land cover type in the Upper Mississippi River. *Science of the Total Environment*, 511, 461-468.
- Staley, C., Gould, T. J., Wang, P., Phillips, J., Cotner, J. B., & Sadowsky, M. J. (2014). Bacterial community structure is indicative of chemical inputs in the Upper Mississippi River. *Frontiers in microbiology*, 5.
- Stewart, F. J., Ottesen, E. A., & DeLong, E. F. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *The ISME journal*, 4(7), 896
- Sun, Z., Li, G., Wang, C., Jing, Y., Zhu, Y., Zhang, S., & Liu, Y. (2014). Community dynamics of prokaryotic and eukaryotic microbes in an estuary reservoir. *Scientific reports*, 4, 6966.
- Sutela, T., Vuori, K. M., & Louhi, P. (2012). Happamien sulfaattimaiden aiheuttamat vesistövaikutukset ja kalakuolemat Suomessa. Suomen ympäristökeskuksen julkaisuja 14/2012. Suomen ympäristökeskus, Helsinki, 61 s.
- Thornton, J.A., Rast, W., Holland, M.M., Jolankai, G., Ryding, S.-O. (Toim.), 1999. Assessment and control of non-point source pollution of aquatic ecosystems. A practical approach. *Man and the Biosphere Ser.*, V. 23 UNESCO xii, s. 1466.
- Tejesvi, M. V., Uhari, M., Tapiainen, T., Pirttilä, A. M., Suokas, M., Lantto, U., Koivunen, P., & Renko, M. (2016). Tonsillar microbiota in children with

- PFAPA (periodic fever, aphthous stomatitis, pharyngitis, and adenitis) syndrome. *European Journal of Clinical Microbiology & Infectious Diseases*, 35(6), 963-970.
- ter Braak, C. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67: 5, 1167-1179.
- Václavík, T., Kupfer, J. A., & Meentemeyer, R. K. (2012). Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). *Journal of Biogeography*, 39(1), 42-55.
- Vilmi, A., Karjalainen, S.M., Nokela, T., Tolonen, K., & Heino J. 2016. Unravelling the drivers of aquatic communities using disparate organismal groups and different taxonomic levels. *Ecological Indicators* 2016; 60: 108-118
- Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., Glidden, S., Bunn, S. E. Sullivan, C. A., Reidy Liermann, C., & Davies, P. M. (2010). Global threats to human water security and river biodiversity. *Nature*, 467(7315), 555.
- Wang, J., Soininen, J., He, J., & Shen, J. (2012). Phylogenetic clustering increases with elevation for microbes. *Environmental Microbiology Reports*, 4(2), 217-226.
- Webb, L. J., Tracey, J. G., Williams, W. T., & Lance, G. N. (1967). Studies in the numerical analysis of complex rain forest communities. II. The problem of species sampling. *Journal of Ecology* 55:525–538.
- Wentworth, C. K. (1922). A scale of grade and class terms for clastic sediments. *The Journal of Geology*, 30(5), 377-392
- Wilhelm, L., Singer, G. A., Fasching, C., Battin, T. J., & Besemer, K. (2013). Microbial biodiversity in glacier-fed streams. *The ISME journal*, 7(8), 1651.

Internet-lähteet

- Happamat sulfaattimaat -karttapalvelu. Geologian tutkimuskeskus. 2018. Viitattu 19.2.2018. Saatavilla: <http://gtkdata.gtk.fi/Hasu/index.html>
- Ion PGM and Proton Sequencers. Viitattu 31.7.2017. Saatavilla: <https://www.thermofisher.com/fi/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-publications-literature/ion-torrent-next-generation-sequencing-literature.html>
- Kindt, R. 2013. Paketti 'BiodiversityR'. Viitattu 6.8.2017. Saatavilla: <http://cran.r-project.org>, <http://www.worldagroforestry.org/resources/databases/tree-diversity-analysis>
- Laki vesienhoidon ja merenhoidon järjestämisestä. Viitattu 08.07.2017. Saatavilla: <http://www.finlex.fi/fi/laki/ajantasa/2004/20041299>
- Legendre, P., Borcard, D., Blanchet, F.G. & Dray, S. (2012) PCNM: MEM spatial eigenfunction and principal coordinate analyses. R package version 2.1-2/r106. Viitattu 30.7.2017. Saatavilla: <http://r-forge.r-project.org/projects/sedar/>.
- Millennium Ecosystem Assessment. (2005). Ecosystems and Human Well-being: Synthesis. Island Press, Washington, DC. Viitattu 08.07.2017. Saatavilla: <http://www.millenniumassessment.org/en/index.html>
- Mitikka, S. 2013. Järvien vedenlaadun vertailu. Viitattu 27.2.2018. Saatavilla: <http://www.ymparisto.fi/download/noname/%7B642BB7A7-42F9-4A67-85BA-D0FF1CC7DBCA%7D/57406>
- MultiNA Microchip Electrophoresis System for DNA/RNA Analysis MCE (2017). Viitattu 31.7.2017. Saatavilla: <http://www.shimadzu.com/an/lifescience/electrophoresis/mce/multina.html>
- Mäntykangas, M. Säävuosi 2014. Foreca. Viitattu 18.3.2018. Saatavilla: <http://blogi.foreca.fi/2014/12/saavuosi-2014-poikkeuksellista-lampoa-ja-suuria-vaihteluita/>
- Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2013) vegan: community ecology package. R package version 2.0-6. Viitattu 30.7.2017. Saatavilla: <http://CRAN.R-project.org/package=vegan>.

- Ouellette, M. H., & Legendre, P. (2013). MVPARTwrap: Additional features for package mvpart. R package, version 0.1-9.2. Saatavilla: <https://cran.r-project.org/src/contrib/Archive/MVPARTwrap/>
- Pintavesien ekologinen ja kemiallinen tila. (2016). Suomen ympäristökeskus. Viitattu 9.7.2017. Saatavilla: <http://www.ymparisto.fi/pintavesientila>
- R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Viitattu 30.7.2017. Saatavilla: <http://www.R-project.org/>.
- Roberts, D., W. (2016). labdsv: Ordination and Multivariate Analysis for Ecology. R package version 1.8-0. <https://CRAN.R-project.org/package=labdsv>
- Therneau, T. M., Atkinson, B., Ripley, B., Oksanen, J., & De'ath, G. (2004). The mvpart package. Multivariate partitioning. <https://cran.r-project.org/src/contrib/Archive/mvpart/>
- Vedenlaadun seuranta. (2017). Suomen ympäristökeskus. Viitattu 19.2.2018. Saatavilla: http://www.ymparisto.fi/fi-FI/Vesi/Vesistöjen_kunnostus/Pienvesien_kunnostus/Pienvesien_kunnostamisen_toteutuksen_ja_sen_vaikutusten_seuraaminen/Vedenlaadun_seuranta

Liite 1 Tutkimuksessa käytetyt alukkeet

Kaikille näytteille käytetty B-aluke: Ion Torrentin adapterisekvenssi trP1: CCTCTCTATGGGCAGTCGGTGAT,

926R-aluke: CCGTCAATTCCTTTTRAGTTT, koko aluke: 926trP1: CCTCTCTATGGGCAGTCGGTGATCCGTCAATTCCTTTTRAGTTT

Yksilölliset a-alukkeet:

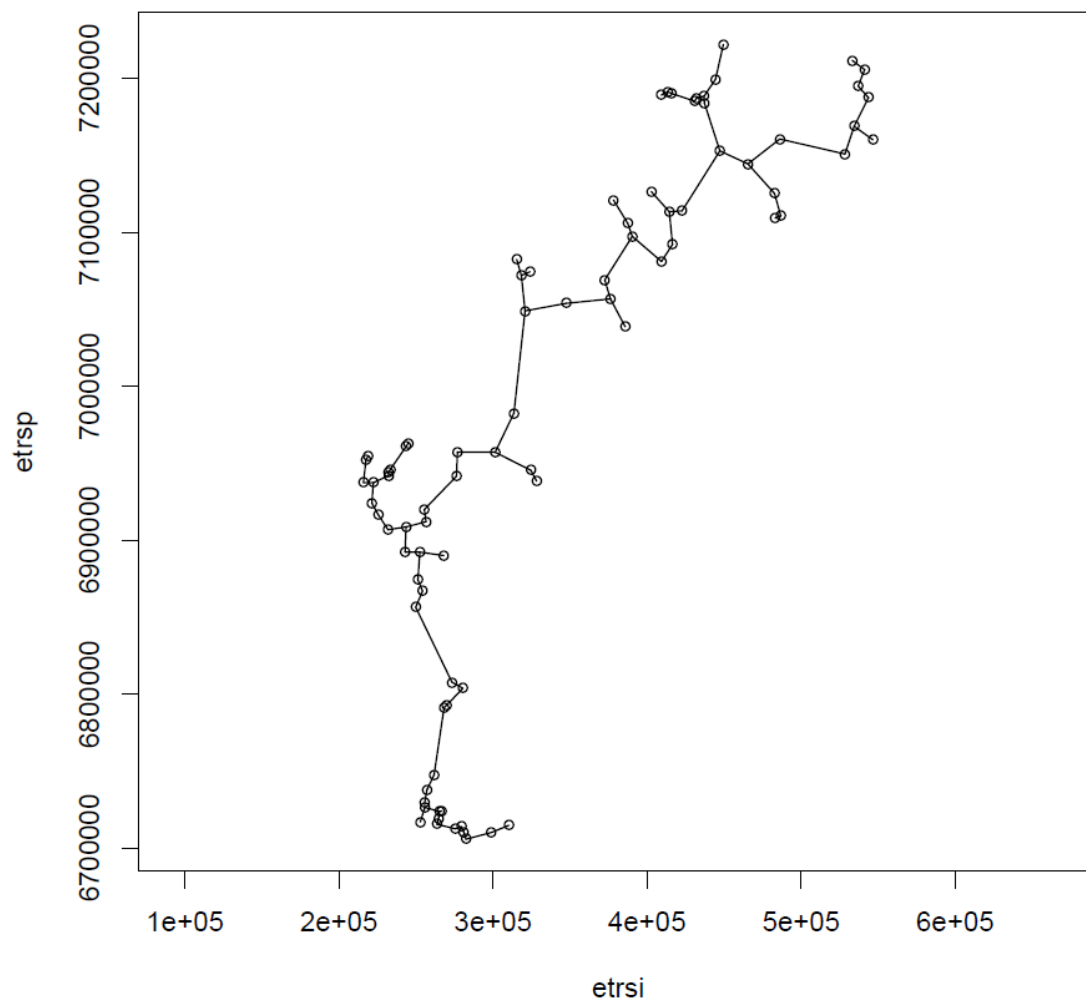
Tutkimuspaikka	Alukkeen tunnus	Adapteri	Yksilöllinen viivakoodi	Linkkeri	16S 519F	Koko aluke
Aurajoki_1	519FM84	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCGGCACTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGGCACTCACAGCMGCCGCGGTAATWC
Aurajoki_2	519FM81	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CCGTACATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCGTACATCACAGCMGCCGCGGTAATWC
Aurajoki_3	519FM110	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CATATTGCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCATATTGCTACAGCMGCCGCGGTAATWC
Aurajoki_4	519FM85	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTGACATTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGACATTCACAGCMGCCGCGGTAATWC
Aurajoki_5	519FM68	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TACTGGCTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTGGCTCACAGCMGCCGCGGTAATWC
Halikonjoki_1	519FM95	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCTAACTCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTAACTCTACAGCMGCCGCGGTAATWC
Halikonjoki_2	519FM106	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCTCTAACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCTCTAACTACAGCMGCCGCGGTAATWC
Halikonjoki_3	519FM69	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AACTGTATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACTGTATCACAGCMGCCGCGGTAATWC
Halikonjoki_4	519FM88	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GCGCGACTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGCGCGACTCACAGCMGCCGCGGTAATWC
Isojoki_1	519FM78	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GATGCACTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGATGCACTCACAGCMGCCGCGGTAATWC
Isojoki_2	519FM100	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CATGCCACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCATGCCACTCACAGCMGCCGCGGTAATWC
Isojoki_4	519FM90	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCGATACCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGATACCTCACAGCMGCCGCGGTAATWC
Isojoki_5	519FM55	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CAATCTCTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAATCTCTCACAGCMGCCGCGGTAATWC
Kalajoki_1	519FM3	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GTATAGTCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGTATAGTCTCACAGCMGCCGCGGTAATWC
Kalajoki_2	519FM6	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTAGCACTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTAGCACTCACAGCMGCCGCGGTAATWC
Kalajoki_3	519FM9	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACATGATCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGACATGATCTCACAGCMGCCGCGGTAATWC
Kalajoki_4	519FM32	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGTAGTCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGTAGTCTCACAGCMGCCGCGGTAATWC
Kalajoki_5	519FM15	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GCAGTCGTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGCAGTCGTCACAGCMGCCGCGGTAATWC

Tutkimuspaikka	Alukkeen tunnus	Adapteri	Yksilöllinen viivakoodi	Linkkeri	16S 519F	Koko alue
Karvianjoki_1	519FM94	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGTC AATCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTCAATCTACAGCMGCCGCGGTAATWC
Karvianjoki_2	519FM89	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTGTGTATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGTGTATCACAGCMGCCGCGGTAATWC
Karvianjoki_3	519FM77	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GCGATCATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGCGATCATCACAGCMGCCGCGGTAATWC
Karvianjoki_4	519FM60	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACAGTATGG	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGTATGGACAGCMGCCGCGGTAATWC
Karvianjoki_5	519FM75	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GTGACTACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGTGACTACTACAGCMGCCGCGGTAATWC
Kiiminkijoki_1	519FM41	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGTACTACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGTACTACTACAGCMGCCGCGGTAATWC
Kiiminkijoki_2	519FM36	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCGCGCATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGCGCATCACAGCMGCCGCGGTAATWC
Kiiminkijoki_3	519FM38	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGTCAGCTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTAGCTCACAGCMGCCGCGGTAATWC
Kiiminkijoki_4	519FM28	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CACGTCGTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGTCGTCACAGCMGCCGCGGTAATWC
Kiiminkijoki_5	519FM48	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGTGTGCTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTGTGCTCACAGCMGCCGCGGTAATWC
Kokemaenjoki_2	519FM76	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GTGTGATCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGTGTGATCTACAGCMGCCGCGGTAATWC
Kokemaenjoki_3	519FM80	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TACGTTATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACGTTATCACAGCMGCCGCGGTAATWC
Kokemaenjoki_4	519FM104	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AATCGCACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAATCGCACTACAGCMGCCGCGGTAATWC
Kokemaenjoki_5	519FM87	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACGCGACCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGACGCGACCTACAGCMGCCGCGGTAATWC
Kruunupyynjoki_3	519FM10	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTACGCATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACGCATCACAGCMGCCGCGGTAATWC
Kyronjoki_1	519FM57	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGAAGAGCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAAGAGCTACAGCMGCCGCGGTAATWC
Kyronjoki_2	519FM102	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TATCGACCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTATCGACCTACAGCMGCCGCGGTAATWC
Kyronjoki_3	519FM96	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGTCAGCCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGTCAGCCTACAGCMGCCGCGGTAATWC
Kyronjoki_4	519FM67	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACCAGAGCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGACCAGAGCTACAGCMGCCGCGGTAATWC
Lapuanjoki_1	519FM103	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ATTAGAGTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGATTAGAGTCACAGCMGCCGCGGTAATWC
Lapuanjoki_2	519FM83	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTGCATACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTGCATACTACAGCMGCCGCGGTAATWC
Lapuanjoki_3	519FM79	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GATGCGTCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGATGCGTCTACAGCMGCCGCGGTAATWC
Lapuanjoki_5	519FM91	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTGGTATCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGTATCTACAGCMGCCGCGGTAATWC
Lestijoki_1	519FM12	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCATGTACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCATGTACTACAGCMGCCGCGGTAATWC
Lestijoki_2	519FM4	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCACATATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCACATATCACAGCMGCCGCGGTAATWC
Lestijoki_3	519FM21	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCGCGTCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCGCGTCTACAGCMGCCGCGGTAATWC
Lumijoki_1	519FM8	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ATACGACTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGATACGACTCACAGCMGCCGCGGTAATWC
Lumijoki_2	519FM16	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCAGTGCTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGTGCTCACAGCMGCCGCGGTAATWC

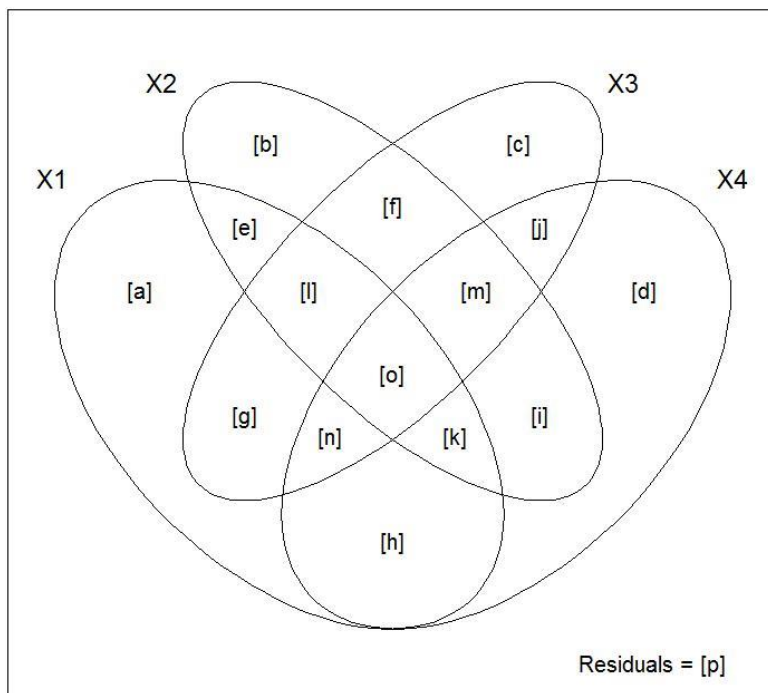
Tutkimuspaikka	Alukkeen tunnus	Adapteri	Yksilöllinen viivakoodi	Linkkeri	16S 519F	Koko aluke
Narpionjoki_1	519FM86	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGGTCTCTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGGTCTCTCACAGCMGCCGCGGTAATWC
Narpionjoki_2	519FM82	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ATGCATAGG	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGCATAGGACAGCMGCCGCGGTAATWC
Narpionjoki_3	519FM108	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGTGTGCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGTGTGCTACAGCMGCCGCGGTAATWC
Narpionjoki_4	519FM70	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TACTGTAGG	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTGTAGGACAGCMGCCGCGGTAATWC
Narpionjoki_5	519FM107	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AATATACTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAATATACTCACAGCMGCCGCGGTAATWC
Oulujoki_1	519FM200	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGCATCATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCATCATCACAGCMGCCGCGGTAATWC
Oulujoki_2	519FM34	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCGTATGTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGTATGTCACAGCMGCCGCGGTAATWC
Oulujoki_3	519FM39	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGTCATATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGTCATATCACAGCMGCCGCGGTAATWC
Oulujoki_4	519FM45	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TATATCACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTATACTACTACAGCMGCCGCGGTAATWC
Oulujoki_5	519FM51	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GCACAGCTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGCACAGCTCACAGCMGCCGCGGTAATWC
Paimionjoki_1	519FM61	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ACAGTTATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGACAGTTATCACAGCMGCCGCGGTAATWC
Paimionjoki_2	519FM58	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTACGCACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTACGCACTCACAGCMGCCGCGGTAATWC
Paimionjoki_3	519FM71	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GACTGTACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGACTGTACTACAGCMGCCGCGGTAATWC
Paimionjoki_5	519FM50	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TTATACATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTTATACATCACAGCMGCCGCGGTAATWC
Perhonjoki_1	519FM33	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTGCATATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCTGCATATCACAGCMGCCGCGGTAATWC
Perhonjoki_3	519FM7	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ATAGCTGTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGATAGCTGTACAGCMGCCGCGGTAATWC
Perhonjoki_4	519FM5	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTATATGTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTATATGTACAGCMGCCGCGGTAATWC
Perhonjoki_5	519FM1	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CTATACACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTATACACTACAGCMGCCGCGGTAATWC
Pyhajoki_2	519FM20	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCGCACTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCGCACTCACAGCMGCCGCGGTAATWC
Pyhajoki_4	519FM11	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCATGCGTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCATGCGTCACAGCMGCCGCGGTAATWC
Pyhajoki_5	519FM17	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGTACACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGACTACACTACAGCMGCCGCGGTAATWC
Siikajoki_1	519FM42	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TATGCTGCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTATGCTGCTACAGCMGCCGCGGTAATWC
Siikajoki_2	519FM46	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGTGTCGTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTGTCGTACAGCMGCCGCGGTAATWC
Siikajoki_3	519FM19	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AGCTATGTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAGCTATGTACAGCMGCCGCGGTAATWC
Siikajoki_4	519FM40	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGTACTATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTACTATCACAGCMGCCGCGGTAATWC
Siikajoki_5	519FM35	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ATGACTATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGATGACTATCACAGCMGCCGCGGTAATWC
Temmesjoki_1	519FM44	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CATATCATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCATATCATCACAGCMGCCGCGGTAATWC

Tutkimuspaikka	Alukkeen tunnus	Adapteri	Yksilöllinen viivakoodi	Linkkeri	16S 519F	Koko aluke
Temmesjoki_3	519FM26	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TACTGATCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTACTGATCTACAGCMGCCGCGGTAATWC
Temmesjoki_4	519FM27	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGCATCACT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGCATCACTACAGCMGCCGCGGTAATWC
Teuvanjoki_1	519FM92	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCGATGTTT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCGATGTTTACAGCMGCCGCGGTAATWC
Teuvanjoki_2	519FM73	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GACGTATCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGACGTATCTACAGCMGCCGCGGTAATWC
Teuvanjoki_3	519FM63	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TGCTAGTTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCTAGTTCACAGCMGCCGCGGTAATWC
Teuvanjoki_4	519FM65	CCATCTCATCCCTGCGTGTCTCCGACTCAG	AACACATCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGAACACATCTACAGCMGCCGCGGTAATWC
Teuvanjoki_5	519FM66	CCATCTCATCCCTGCGTGTCTCCGACTCAG	ATCTCCTCT	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGATCTCCTCTACAGCMGCCGCGGTAATWC
Uskelanjoki_1	519FM74	CCATCTCATCCCTGCGTGTCTCCGACTCAG	GACGTGCTC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGGACGTGCTCACAGCMGCCGCGGTAATWC
Uskelanjoki_4	519FM59	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CCATGTATC	A	CAGCMGCCGCGGTAATWC	CCATCTCATCCCTGCGTGTCTCCGACTCAGCCATGTATCACAGCMGCCGCGGTAATWC

Liite 2. Tutkimuspaikkojen sijaintia kuvaava pienin puu



Liite 3 Mallikuva vaihtelun ositusten tulosten tulkintaan



Tässä tutkimuksessa:

- [aeghklno] = X1, Vesikemia-muuttujat = kem
- [befiklmo] = X2, habitaatti-muuttujat = hab
- [cfgjlmno] = X3, valuma-alueuuttujat = va
- [dhijkmno] = X4, spatiaalimuuttujat = spa
- [abcdefghijklmno] = Kokonaisselitysaste
- [abefghiklmno] = kem+fys
- [acefghjklmno] = kem+va
- [adeghijklmno] = kem+spa
- [bcefgijklmno] = fys+va
- [bdefhijklmno] = fys+spa
- [cdfghijklmno] = va+spa
- [abcefgijklmno] = kem+fys+va
- [abdefghijklmno] = kem+fys+spa
- [acdefghijklmno] = kem+va+spa
- [bcdefghijklmno] = fys+va+spa
- [p] = Selittämätön vaihtelu

Liite 4 Tutkimuspaikkojen vesikemiatulokset

Tutkimuspaikka	Pohjois- koordinaatti ETRS-TM35FIN	Itä- koordinaatti ETRS- TM35FIN	Kokonais -typpi µg/l	Kokonais -fosfori µg/l	Väri mgPt/l	Sähkön- johtavuus µS/cm	pH
Aurajoki_1	6716879	252740	1250.0	74.0	50	189.1	7.73
Aurajoki_2	6747439	261700	1212.5	129.3	350	75	7.14
Aurajoki_3	6737890	257102	1300.0	107.9	375	88.4	7.34
Aurajoki_4	6729190	255535	1281.3	91.1	250	80.6	6.98
Aurajoki_5	6726348	255715	1500.0	139.1	110	116.4	7.77
Halikonjoki_1	6712493	275382	1262.5	84.0	120	75.5	7.61
Halikonjoki_2	6713799	279456	2175.0	356.9	20	128.4	7.7
Halikonjoki_3	6710512	280920	612.5	65.2	110	128.2	7.77
Halikonjoki_4	6705676	282416	593.8	58.9	80	137.9	7.81
Isojoki_1	6892198	242858	275.0	13.6	60	42.4	7.6
Isojoki_2	6908522	243546	312.5	12.7	60	27.3	7.6
Isojoki_4	6906738	231810	637.5	22.1	100	32.6	7.1
Isojoki_5	6916487	225462	762.5	55.5	120	103.6	7.61
Kalajoki_1	7097266	390360	737.5	91.1	150	60.8	6.76
Kalajoki_2	7120905	378172	1125.0	99.2	275	121.1	6.93
Kalajoki_3	7081001	409327	425.0	121.4	75	127.4	6.73
Kalajoki_4	7092013	416479	762.5	43.7	120	97.8	7.61
Kalajoki_5	7105706	387439	975.0	89.4	300	167.3	7.04
Karvianjoki_1	6856364	249796	462.5	39.0	160	31.4	7.16
Karvianjoki_2	6866894	254220	350.0	33.6	20	43.7	7.7
Karvianjoki_3	6874212	251198	837.5	88.3	200	65.9	7.8
Karvianjoki_4	6892178	252453	293.8	12.3	30	56.4	7.5
Karvianjoki_5	6889941	268054	875.0	63.1	160	49.7	7.16
Kiiminkijoki_1	7211551	533108	312.5	11.9	60	28.9	7.6
Kiiminkijoki_2	7194890	536983	162.5	0.6	30	48.4	7.5
Kiiminkijoki_3	7187349	543702	218.8	3.3	60	15.5	7.6
Kiiminkijoki_4	7205612	541226	150.0	0.1	15	33.6	7.05
Kiiminkijoki_5	7221918	449467	437.5	7.0	100	60.7	7.1
Kokemaenjoki_2	6791261	268218	2012.5	146.7	200	143.4	7.8
Kokemaenjoki_3	6792660	269791	1087.5	46.0	75	65.2	6.73
Kokemaenjoki_4	6807662	273349	481.3	15.2	30	119.5	7.5
Kokemaenjoki_5	6803708	280376	518.8	18.3	30	64.2	7.5
Kruunupyynjoki_3	7048881	320765	587.5	77.4	150	42.2	6.76
Kyronjoki_1	6911944	256508	437.5	28.0	100	25.4	7.1
Kyronjoki_2	6919874	255387	437.5	38.8	100	48.3	7.1
Kyronjoki_3	6957219	276861	2350.0	241.8	200	156.4	7.8
Kyronjoki_4	6941524	276276	1087.5	78.0	75	74.5	6.73
Lapuanjoki_1	6957050	301316	712.5	26.8	140	64.2	6.41
Lapuanjoki_2	6945360	324501	887.5	99.6	200	59.8	7.8
Lapuanjoki_3	6938335	328257	725.0	55.7	125	51.7	7.2
Lapuanjoki_5	6982459	313545	750.0	61.5	200	49.7	7.8
Lestijoki_1	7069017	372505	675.0	17.1	200	31.4	7.8
Lestijoki_2	7038529	385791	606.3	35.0	180	18.7	5.99
Lestijoki_3	7056658	376186	675.0	19.2	200	18.8	7.8
Lumijoki_1	7191150	413341	612.5	54.7	110	82	7.77
Lumijoki_2	7188998	409152	535.0	78.6	105	71	6.88
Narpionjoki_1	6937339	215852	850.0	67.8	100	169.6	7.1
Narpionjoki_2	6952099	217512	687.5	56.2	120	116.3	7.61
Narpionjoki_3	6954450	218864	587.5	76.6	150	103	6.76
Narpionjoki_4	6960732	243509	987.5	89.4	125	50.1	7.2

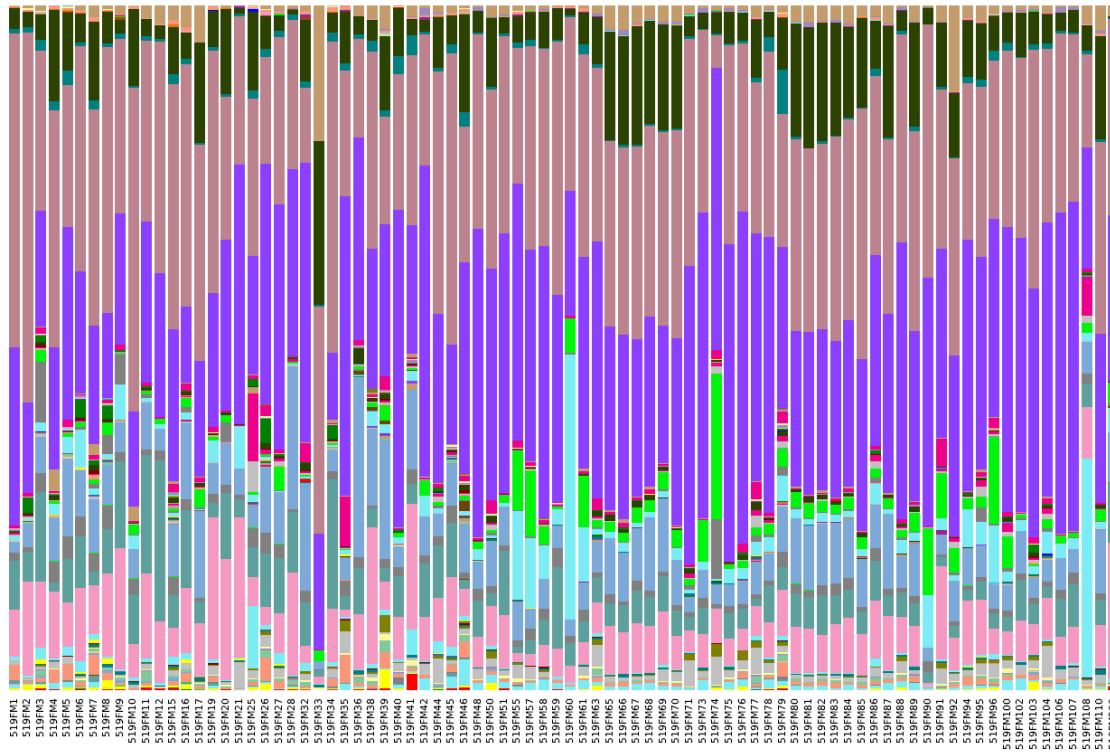
Tutkimuspaikka	Pohjois- koordinaatti ETRS- TM35FIN	Itä- koordinaatti ETRS-TM35FIN	Kokonais -typpi µg/l	Kokonais -fosfori µg/l	Väri mgPt/l	Sähkön- johtavuus µS/cm	pH
Narpionjoki_5	6962864	244875	575.0	41.5	140	118.7	6.41
Oulujoki_1	7160029	546570	212.5	2.7	50	27.9	7.73
Oulujoki_2	7160519	486118	275.0	182.5	60	20.6	7.6
Oulujoki_3	7168747	534432	312.5	8.4	60	21.3	7.6
Oulujoki_4	7199319	444397	1181.3	20.7	180	38.6	5.99
Oulujoki_5	7150766	528405	200.0	6.1	50	62.6	7.73
Paimionjoki_1	6723930	266588	1950.0	146.0	80	202.1	7.81
Paimionjoki_2	6723724	265116	1925.0	132.6	30	191.5	7.5
Paimionjoki_3	6719303	264673	1337.5	72.6	70	127.9	6.99
Paimionjoki_5	6715395	263564	1187.5	69.1	40	128.9	8.13
Perhonjoki_1	7053767	347452	2162.5	111.9	200	48.4	7.8
Perhonjoki_3	7082271	315322	1012.5	70.2	150	37.2	6.76
Perhonjoki_4	7071916	318371	1337.5	105.0	70	100.2	6.99
Perhonjoki_5	7074283	323980	1725.0	120.6	90	74.4	6.21
Pyhajoki_2	7126365	402721	637.5	37.6	100	50.6	7.1
Pyhajoki_4	7113835	422678	812.5	50.0	62.5	82.9	7.02
Pyhajoki_5	7113267	414409	662.5	34.0	80	28.3	7.81
Siikajoki_1	7125093	482726	875.0	82.1	160	19.7	7.16
Siikajoki_2	7152795	447224	700.0	10.9	100	13.9	7.1
Siikajoki_3	7109164	483041	512.5	33.1	110	13.2	7.77
Siikajoki_4	7144333	465452	487.5	82.8	105	22.1	6.88
Siikajoki_5	7110529	486774	687.5	119.2	120	16.9	7.61
Temmesjoki_1	7184065	437139	600.0	84.3	150	207.8	6.76
Temmesjoki_3	7187224	432088	2762.5	39.5	275	619	6.93
Temmesjoki_4	7185037	430947	2075.0	167.4	200	402.7	7.8
Teuvanjoki_1	6944187	232047	1412.5	127.8	100	88.6	7.1
Teuvanjoki_2	6946074	233423	1237.5	197.6	200	86.9	7.8
Teuvanjoki_3	6941973	232301	1087.5	79.4	75	96.6	6.73
Teuvanjoki_4	6937229	222240	662.5	150.1	80	94.2	7.81
Teuvanjoki_5	6923900	221332	675.0	38.6	200	58.4	7.8
Uskelanjoki_1	6714695	310248	1462.5	99.4	40	130.3	8.13
Uskelanjoki_4	6710038	298786	912.5	207.7	40	190.2	8.13

Liite 5 Tutkimuspaikkojen yksilölliset a-alukset

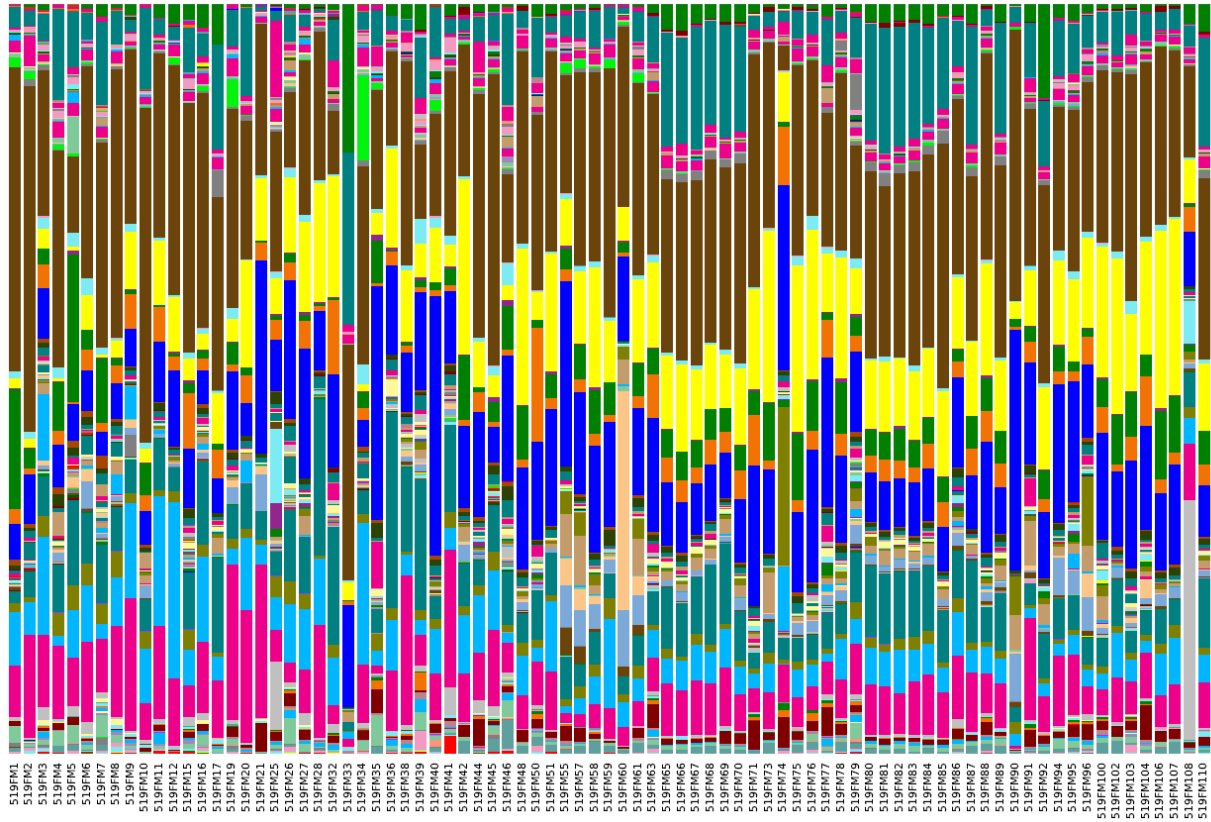
Tutkimuspaikka	Alukkeen tunnus	Tutkimuspaikka	Alukkeen tunnus	Tutkimuspaikka	Alukkeen tunnus
Perhonjoki_5	519FM1	Kiiminkijoki_1	519FM41	Lapuanjoki_3	519FM79
Kalajoki_1	519FM3	Siikajoki_1	519FM42	Kokemaenjoki_3	519FM80
Lestijoki_2	519FM4	Temmesjoki_1	519FM44	Aurajoki_2	519FM81
Perhonjoki_4	519FM5	Oulujoki_4	519FM45	Narpionjoki_2	519FM82
Kalajoki_2	519FM6	Siikajoki_2	519FM46	Lapuanjoki_2	519FM83
Perhonjoki_3	519FM7	Kiiminkijoki_5	519FM48	Aurajoki_1	519FM84
Lumijoki_1	519FM8	Paimionjoki_5	519FM50	Aurajoki_4	519FM85
Kalajoki_3	519FM9	Oulujoki_5	519FM51	Narpionjoki_1	519FM86
Kruunupyynjoki_3	519FM10	Isojoki_5	519FM55	Kokemaenjoki_5	519FM87
Pyhajoki_4	519FM11	Kyronjoki_1	519FM57	Halikonjoki_4	519FM88
Lestijoki_1	519FM12	Paimionjoki_2	519FM58	Karvianjoki_2	519FM89
Kalajoki_5	519FM15	Uskelanjoki_4	519FM59	Isojoki_4	519FM90
Lumijoki_2	519FM16	Karvianjoki_4	519FM60	Lapuanjoki_5	519FM91
Pyhajoki_5	519FM17	Paimionjoki_1	519FM61	Teuvanjoki_1	519FM92
Siikajoki_3	519FM19	Teuvanjoki_3	519FM63	Karvianjoki_1	519FM94
Pyhajoki_2	519FM20	Teuvanjoki_4	519FM65	Halikonjoki_1	519FM95
Lestijoki_3	519FM21	Teuvanjoki_5	519FM66	Kyronjoki_3	519FM96
Temmesjoki_3	519FM26	Kyronjoki_4	519FM67	Isojoki_2	519FM100
Temmesjoki_4	519FM27	Aurajoki_5	519FM68	Kyronjoki_2	519FM102
Kiiminkijoki_4	519FM28	Halikonjoki_3	519FM69	Lapuanjoki_1	519FM103
Kalajoki_4	519FM32	Narpionjoki_4	519FM70	Kokemaenjoki_4	519FM104
Perhonjoki_1	519FM33	Paimionjoki_3	519FM71	Halikonjoki_2	519FM106
Oulujoki_2	519FM34	Teuvanjoki_2	519FM73	Narpionjoki_5	519FM107
Siikajoki_5	519FM35	Uskelanjoki_1	519FM74	Narpionjoki_3	519FM108
Kiiminkijoki_2	519FM36	Karvianjoki_5	519FM75	Aurajoki_3	519FM110
Kiiminkijoki_3	519FM38	Kokemaenjoki_2	519FM76	Oulujoki_1	519FM200
Oulujoki_3	519FM39	Karvianjoki_3	519FM77		
Siikajoki_4	519FM40	Isojoki_1	519FM78		

Liite 6. Luokkatasolle BLAST-analysoidut bakteerit

Yleisimmät havaitut luokat ovat *Alphaproteobacteria* 28.9% (kuvassa violetilla), *Betaproteobacteria* 27.7% (kuvassa vaaleanruskealla), *Gammaproteobacteria* 7.5% (kuvassa tummanvihreällä), *Cytophagia* 7.3% (kuvassa vaaleanpunaisella), *Saprosirae* 6.5% (kuvassa siniharmaalla) ja *Flavobacteriia* 5.1% (kuvassa mintunvihreällä).



Liite 7. Lahkotasolle BLAST-analysoidut bakteerit.



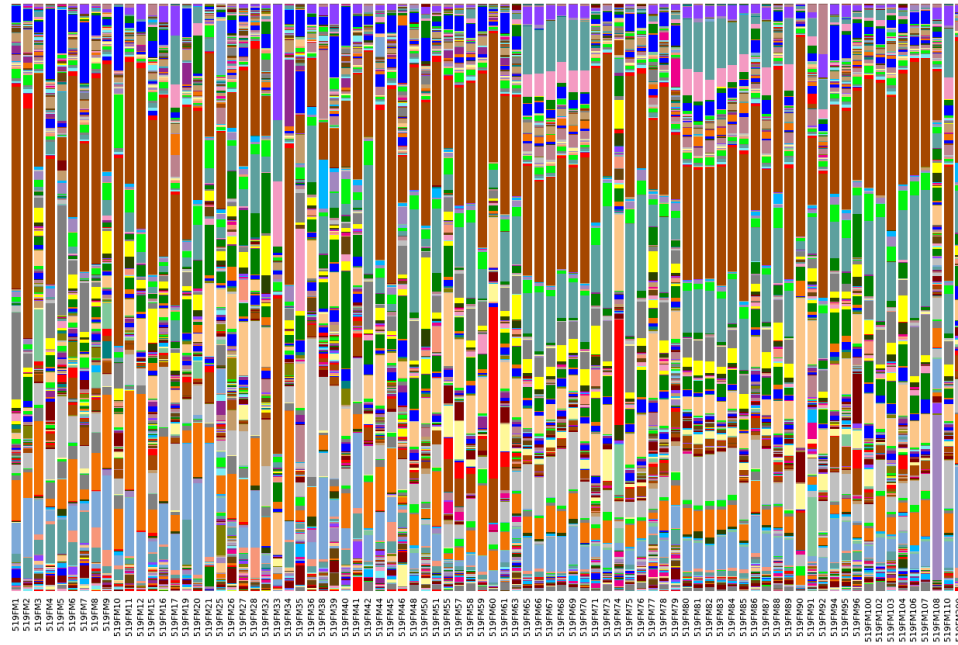
Yleisimmät havaitut lahkot ovat Burkholderiales 24.0% (kuvasa ruskealla), Rhizobiales 11.4% (kuvasa sinisellä) Sphingomonadales 9.3% (kuvasa keltaisella), Cytophagales 7.3% (kuvasa pinkillä), Saprospirales 6.5% (kuvasa sammaleenvihreällä), Xanthomonadales 6.5% (kuvasa sammaleenvihreällä, mutta ylempänä kuin Saprospirales) ja Flavobacteriales 5.1% (kuvasa vaaleansinisellä)

Liite 8. Heimotasolle BLAST-analysoidut bakteerit.



Yleisimmät havaitut heimot ovat Comamonadaceae 21.6% (kuvassa sammaleenvihreällä), Sphingomonadaceae 7.9% (kuvassa vaaleankeltaisella), Cytophagaceae 7.2% (kuvassa siniharmaalla, Chitinophagaceae 5.8% (kuvassa myös sammaleenvihreällä, mutta alempana kuin Comamonadaceae), Flavobacteriaceae 4.8% (kuvassa vaaleankeltaisella, alempana kuin Sphingomonadaceae), lahkotasolla Rhizobiales: yksi määrittynyt, mutta heimotasolla ilman tarkennusta jääneet 4.7% (kuvassa lappapuuronpunaisella) ja Hyphomicrobiaceae 4.5% (kuvassa kirkaan sinisellä).

Liite 9. Sukutasolle BLAST-analysoidut bakteerit.



Yleisimmät havaitut suvut ovat heimotasolla Comamonadaceae:ksi määrittynyt, mutta sukutasolla ilman tarkempaa määrettä jääneet 18.7% (kuvassa ruskealla), Flectobacillus 4.6% (kuvassa siniharmaalla), heimotasolla Sphingomonadaceae:ksi määrittynyt, mutta sukutasolla ilman tarkempaa määrettä jääneet 6.1% (kuvassa turkoosilla), heimotasolla Chitinophagaceae:ksi määrittynyt, mutta sukutasolla ilman tarkempaa määrettä jääneet 5.2% (kuvassa harmaalla), Flavobacterium 4.8% (kuvassa oranssilla), lahkotasolla Rhizobiales:iksi määrittynyt mutta heimo- ja sukutasolla ilman tarkennusta jääneet 4.7% (kuvassa aprikoosin värisellä).