

DNA-KOETTIMIEN SUUNNITTELU JÄRVISEDIMENTIN MIKROBIEN TUNNISTAMISEKSI LDR-SIRULLA

Vipuvoimaa
EU:lta
2007-2013



Euroopan unioni
Euroopan aluekehitysrahasto

LAHDEN AMMATTIKORKEAKOULU
Tekniikan ala
Ympäristötekniikan koulutusohjelma
Ympäristöbiotekniikka
Opinnäytetyö
Syksy 2010
Atte Halttunen

HALTTUNEN, ATTE: DNA-koettimien suunnittelu järvisedimentin mikrobien tunnistamiseksi LDR-sirulla

Ympäristöbiotekniikan opinnäytetyö 49 sivua, 9 liitesivua

Syksy 2010

TIIVISTELMÄ

Tämä opinnäyte oli osa YMLI-hanketta, jonka tavoitteena on mm. DNA-sirutekniikan kehittäminen ympäristön mikrobien tunnistamiseksi. Opinnäytetyön tarkoituksena oli suunnitella lajitasoa edustavia koettimia sedimentin mikrobien seurantaan Ligation Detection Reaction (LDR) -sirulla. Opinnäytteessä käytetty sedimentti oli peräisin projektista, jossa tutkittiin rehevöityneiden järvien kunnostamista sedimenttiin levitettävällä happea vapauttavalla kemikaalilla. Koetinsuunnittelun tavoitteena oli ensisijaisesti suunnitella käsittelyä indikoivia koettimia ja toiseksi koettimia monimuotoisuuden seurantaan.

Koettimien suunnittelemiseksi sedimentistä eristettiin DNA, josta monistettiin polymeerasiketjureaktiolla (PCR) osaa 16S rDNA:sta. Monistettujen 16S rDNA kappaleiden emäsjärjestys selvitettiin pyrosekvensoinnilla. Sekvensointiaineisto määritettiin OTU:iksi 97 %:n samankaltaisuusasteella, joka edustaa lajitasoa. Sedimentin mikrobisyhteisön selvittämiseksi sekvensointiaineistoa tarkasteltiin Ribosomal Data Base Projectin (RDP) Classifier- ja MEGAN-ohjelmalla. Näillä muodostettujen taksonomisesti hierarkkisten puiden avulla koetinsuunnitteluun valittiin 138 DNA-sekvenssiä, jotka edustivat OTU:ja.

Koettimia suunniteltiin yhteensä 87, joista 73 oli bakteereille ja 14 arkeille. Koettimista mahdollisesti 18 indikoi happikäsittelyn vaikutusta ja lopuilla seurattiin pelkästään monimuotoisuutta. Mikrobit olivat yleisesti ottaen tunnistettavissa heikosti, mihin vaikutti pyrosekvensoinnista saatujen DNA-sekvenssien lyhyys ja tutkittava ympäristö, sillä sedimentin mikrobistoa ei vielä tunneta kokonaisuudessaan kovinkaan hyvin. Tämän vuoksi pyrosekvensoinnilla tuotettua DNA-sekvenssiaineistoa ei pystytty hyödyntämään kokonaisuudessaan koetinsuunnittelussa.

Avainsanat: LDR-siru, koetinsuunnittelu, sedimentti, bakteeri, arkki

Lahti University of Applied Sciences
Degree Programme in Environmental Technology

HALTTUNEN, ATTE: Designing DNA probes for surveying
microbes in lake sediment with an LDR
chip

Bachelor's Thesis in Environmental Biotechnology 49 pages, 9 appendices

Autumn 2010

ABSTRACT

This thesis was part of YMLI project, which for example, aimed at developing DNA microarray technology for surveying microbes at an environment. The aim of the thesis was to design species level probes for surveying microbes in sediment with a Ligation Detection Reaction (LDR) chip. The sediment used in the study was from a project, which studied the renovation of a eutrophic lake by distributing an oxygen releasing chemical in the sediment. The probes were, first of all, designed to indicate the effect of the oxygen releasing chemical, and, secondly to observe diversity.

To design the probes, DNA was extracted from the sediment, and a part of the 16S rDNA was amplified by a Polymerase Chain Reaction (PCR). The base order of the 16S rDNA sequences were defined by pyrosequencing. The reads derived from the pyrosequencing were assigned as OTUs at 97 % identity that represent the species level. Taxonomical assignments were conducted with two programs, the Classifier of the RDP and MEGAN. With these programs, 138 reads, which represented the OTUs, were chosen to probe designing.

Altogether 87 probes were designed, of which 73 were for bacteria, and 14 for archaea. Possibly 18 probes indicated the effect of the oxygen releasing chemical, and the rest indicated only diversity. The microbes in the sediment were poorly recognizable. This was because the reads obtained from pyrosequencing were quite short, and the microbes of sediment have not been studied very much. Therefore, the data produced were not completely utilizable because of its poor recognizability.

Key words: LDR chip, probe designing, sediment, bacterium, archaea

SISÄLLYS

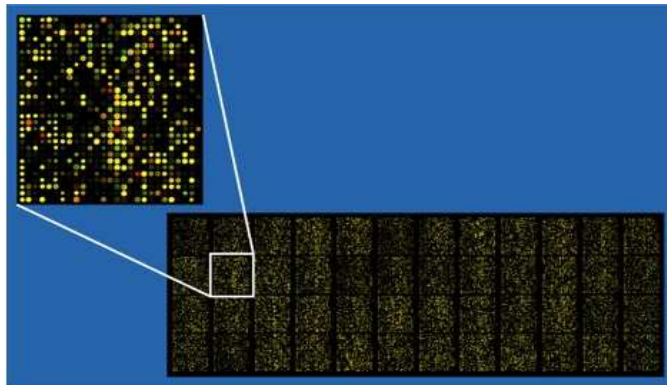
1	JOHDANTO	1
2	DNA TIEDON TUOTTAMINEN LDR-SIRUA VARTEN	3
2.1	Ligation Detection Reaction (LDR) -siru	3
2.2	Polymeraasiketjureaktio	5
2.3	Pyrosekvensointi	6
2.4	Bioinformatiikka	8
3	MATERIAALIT JA MENETELMÄT	11
3.1	DNA:n eristyksestä DNA-kirjastoon	11
3.1.1	PCR lyhyillä alukkeilla	13
3.1.2	PCR-tuotteen puhdistus	14
3.1.3	PCR pitkillä alukkeilla	15
3.1.4	PCR-tuotteen puhdistus	17
3.1.5	PCR-tuotteen koon ja konsentraation määrittäminen	17
3.2	Sekvensointidatasta koettiin	18
3.2.1	Laadukkaiden fasta-tiedostojen luonti pyrosekvensointidatasta	19
3.2.2	Näytteiden sekvensien luokittelu OUT:ihin 97 %:n samankaltaisuusasteella	19
3.2.3	Sekvensoinnin ja sekvensointiaineiston tarkastelu	20
3.2.4	OTU:jen valinta koetinsuunnitteluun	23
3.2.5	Koetinsuunnittelu	23
4	TULOKSET JA TULOSTEN TARKASTELU	26
4.1	Laboratoriotyöt ennen sekvensointia	26
4.2	Sekvensoinnin tulokset	28
4.3	Sekvensointiaineiston sisältö	31
4.4	Koettimet	37
4.5	Koettimet ja sedimentti	39
5	JOHTOPÄÄTÖKSET	43
	LÄHTEET	45
	LIITTEET	50

LYHENTEET

LYHENNE	ENGLANNIKSI	SUOMEKSI
16S rRNA	A part of the prokaryotic ribosome	Osa prokaryoottien ribosomia
ATP	Adenosine-triphosphate	adenosiinitrifosfaatti
BLAST	Basic local alignment search tool	-
Bp	Base pair	Emäs pari
DNA	Deoxyribonucleic acid	Deoksiribonukleiinihappo
DDBJ	DNA Data Bank of Japan	-
EMBL	The European Molecular Biology Laboratory	-
Em-PCR	Emulsion PCR	Emulsio-PCR
GG	GreenGenes	-
LCA	Lowest Common Ancestor	Viimeinen yhteinen sukulainen
LDR	Ligation Detection Reaction	Ligaation havaitsemisreaktio
MEGAN	MEtaGenome ANalyzer	-
NCBI	National Center for Biotechnology Information	-
OTU	Operational Taxonomic Unit	Operatiivinen taksonominen yksikkö
PCR	Polymerase Chain Reaction	Polymeraasiketjureaktio
RDP	Ribosomal Data Base Project	-
RNA	Ribonucleic acid	Ribonukleiinihappo
YMLI	-	Uudet ympäristömittausmenetelmät liiketoimintamahdollisuutena
DGGE	Denaturing gradient gel electrophoresis	Denaturoivan gradientin geleelektroforeesi

1 JOHDANTO

DNA-sirujen rakenne ja toiminta ovat yksinkertaisia. Ne koostuvat alustasta (esim. lasia), johon on kiinnitetty koettimia. Koettimet ovat synteettisestä DNA:sta koostuvia lyhyitä sekvenssejä, jotka ovat komplementaarisia (täydentäviä) kohde-DNA:n tai RNA:n kanssa. Tutkittava näyte voidaan värjätä esim. fluoresoivalla värillä, jonka perusteella saadaan selville, mihin koettiin näytettä on kiinnittynyt (hybridisoitunut) (kuvio 1). Sirulta saatu informaatio käsitellään tilastollisin menetelmin.



KUVIO 1. DNA-siru, jolle on hybridisoitu näyte. Värilliset pisteet ilmaisevat koettimia, joihin on hybridisoitunut näytettä, mustat kohdat ovat koettimia, joihin ei ole hybridisoitunut näytettä (Wikipedia 2010)

DNA-sirujen sovellusala on laaja. Niitä voidaan käyttää mm. molekyylibiologian tutkimuksessa (Hoheisel 2006, 200), ympäristönäytteiden tutkimisessa (Shiu & Borevitz 2006, 141) ja lääketieteellisten ongelmien ratkaisemisessa (Loy & Bodrossy 2005, 106). Ensimmäiset nykyisen kaltaiset DNA-sirut kehitettiin 90-luvun puolivälissä, jonka jälkeen kehitys on ollut huimaa (Shiu & Borevitz 2006, 141). Nykyään DNA-sirulla voidaan tarkastella tuhansista jopa yli miljoonaan genomien ominaisuutta samanaikaisesti (Shiu & Borevitz 2006, 141).

DNA-siruja on käytetty jo jonkin aikaa ympäristönäytteiden tutkimiseen ja tulevaisuudessa sirujen käyttö lisääntyy, kun sekvensointiaineisto tietokannoissa kasvaa (Shiu & Borevitz 2006, 141). DNA-siruilla on tutkittu mm. geenien määrällistä ilmentymistä ympäristönäytteistä (Cho & Tiedje 2001), kompostissa esiintyviä

mikrobeja (Hultman, Ritari, Romantschuk, Paulin & Auvinen 2008), syanobakteerien diversiteettiä järvisedestä (Castiglioni ym. 2004) ja jätevedessä esiintyviä patogeenejä (Lee, Lauder, Cruwys, Falletta & Beaudette 2007).

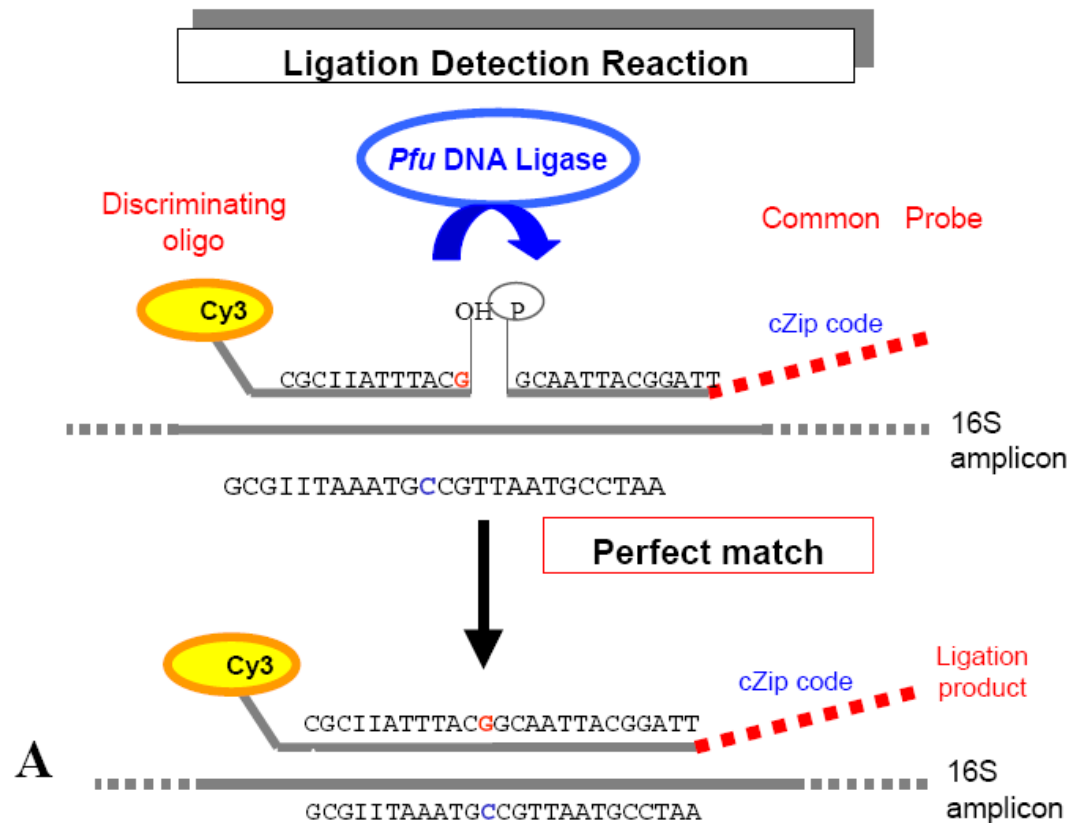
Tämä opinnäyte oli osa Euroopan aluekehitysrahaston rahoittamaa YMLI-hanketta (Uudet ympäristömittausmenetelmät liiketoimintamahdollisuutena). Hankkeen tavoitteina on kehittää ja kaupallistaa nopeita, edullisia ja tarkkoja mitausmenetelmiä mm. DNA-tasolle. Hankkeen yhtenä tavoitteena on DNA-sirutekniikan kehittäminen ympäristön mikrobien tarkkailuun. Hankkeen päätoimittaja on Helsingin yliopiston Ympäristötieteiden laitos ja muita osanottajia ovat mm. Lahden ammattikorkeakoulu ja Biotekniikan instituutti. (Lahden ammattikorkeakoulu 2010.)

Tässä opinnäytteessä oli tavoitteena suunnitella n. 100 lajitasoa edustavaa koetinta järvisedimentin mikrobien seurantaan LDR-sirulla. Sedimenttinäytteet DNA-koettimien suunnitteluun saatiin Maj ja Tor Nessling -säätiön rahoittamasta projektista, jossa tutkitaan rehevöityneiden järvien kunnostamista sedimenttiin levitetävän happea vapauttavan kemikaalin avulla. Ensisijaisena tavoitteena oli suunnitella koettimia, jotka indikoivat happikäsittelyn vaikutusta, ja toiseksi koettimia monimuotoisuuden seurantaan. Koettimien suunnittelemiseksi sedimenttinäytteistä eristettiin DNA, josta monistettiin polymeerasiketjureaktiolla (PCR) osaa 16S rRNA -geenistä. Monistetun DNA:n emäsjärjestys selvitettiin pyrosekvensoinnilla, josta saatua DNA-tietoa käsiteltiin bioinformatiikan ohjelmilla. Lähtökohtaisesti koettimet suunniteltiin tunnistamaan bakteereja ja arkkeja laji- tai sukutasolle.

2 DNA TIEDON TUOTTAMINEN LDR-SIRUA VARTEN

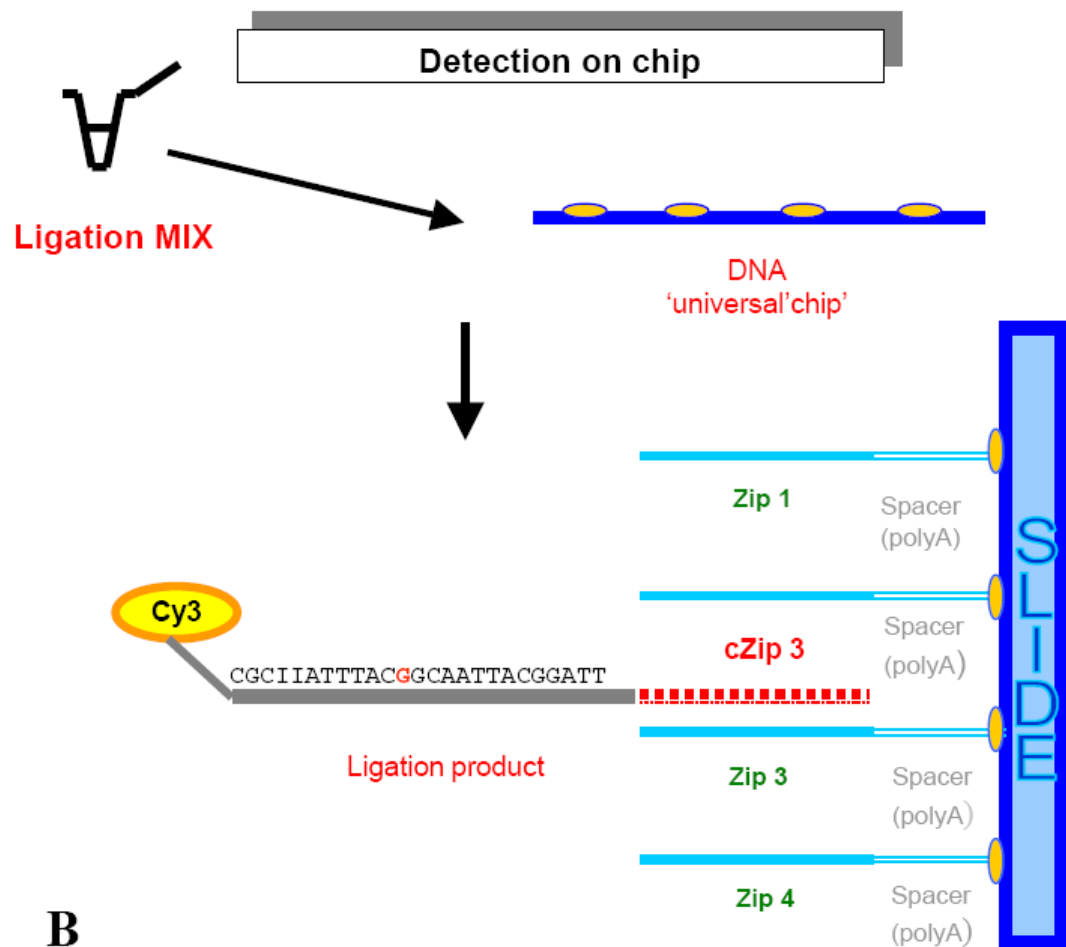
2.1 Ligation Detection Reaction (LDR) -siru

LDR-sirua käytetään yhden nukleotidin mutaatioiden tunnistamisessa (Gerry, Witowski, Day, Hammer, Barany & Barany 1999) tai tutkimuksissa, joissa hyödynnetään genomien konservoituneita alueita, kuten 16S rDNA:ta (Busti, Bordoni, Castiglioni, Monciardini & Sosio 2002). LDR-sirut poikkeavat hiukan suunnittelultaan ja toimintaperiaatteeltaan useimmiten käytetyistä DNA-siruista. LDR-sirulla on koettimien sijasta zip-koodeja, joihin koettimet hybridisoidaan (kuvio 3). Koetin koostuu erottelevasta koettimesta, d-koetin (discriminating), ja yleisestä koettimesta, c-koetin (common). D-koettimen 5' päässä on fluoresoiva leima ja 3' päässä erotteleva nukleotidi (kuvio 2). C-koettimen 3' päässä on Zip-koodia vastaava sekvenssi cZip (complementary). Kohde DNA:na voidaan käyttää suoraan näytteestä eristettyä DNA:ta tai PCR:llä monistettua sekvenssiä. Jos koettimet vastaavat täysin kohdesekvenssiä, niin ne kiinnittyvät peräkkäin kohteeseen ja liitetään ligaatio-reaktion avulla toisiinsa (kuvio 2). Jos kohdesekvenssissä on poikkeavuus d-koettimen 3' pään nukleotidin kohdalla, niin koettimet eivät yhdisty. Koettimissa, jotka ovat yhdistyneet, on 5' päässä fluoresoiva leima ja 3' päässä cZip-koodi (kuvio 2). (Gerry ym. 1999, 252.) Yhdistyneiden koettimien määrää voidaan kasvattaa lineaarisesti termisellä kierrätyksellä tai eksponentiaalisesti PCR:llä (Hultman ym. 2008, 2).



KUVIO 2. Ligoitumis reaktio sekä d- ja c-koettimien rakenne (Busti ym. 2002)

Ligaation jälkeen koettimet hybridisoidaan sirulle cZip-koodeja vastaaviin kohtiin (kuvio 3). Sirulle hybridisoiduista koettimista vain ne, jotka ovat ligoituneet emittoivat säteilyä ja voidaan tunnistaa. Zip-koodien laatua ja sirulle hybridisoituneiden koettimien määrää voidaan arvioida Zip-koodeihin lisätyn kontrollikoettimen avulla. Kontrollikoettimessa on fluoresoiva leima, joka emittoi eri aallonpituutta kuin d-koetin ja voidaan näin erottaa toisistaan. Ligoituneiden- ja kontrollikoettimien emittoimaa säteilyä mitataan skannerilla ja säteilymäärien perusteella tulokset normalisoidaan ja analysoidaan. (Hultman ym. 2008, 5.)



B
KUVIO 3. Koettimen hybridisoiminen sirulle (Busti ym. 2002)

2.2 Polymeraasiketjureaktio

PCR (Polymerase Chain Reaction), eli polymeraasiketjureaktio, on tekniikka, jolla voidaan monistaa palasta DNA:ta satoja miljoonia kopioita. Reaktioon tarvitaan malli-DNA:ta, alukkeita, DNA-polymeraasientsyymiä, yksittäisiä nukleotideja, puskuriliuosta ja kationeita. Malli-DNA on kohde, jota kopioidaan. Alukkeet ovat yksijuosteisia 5 - 50 emästä pitkiä synteettisesti valmistettuja DNA-sekvenssejä. DNA kostuu kahdesta juosteesta, johtavasta ja vastakkaisesta, joiden kopioimiseen tarvitaan omat alukkeensa. Alukkeet tarttuvat juosteiden 3' päihin emäsjärjestyksensä vastaaviin kohtiin rajaten samalla kopioitavan alueen. Polymeraasientsyymi käyttää aluketta aloituskohtana ja rakentaa nukleotideista uuden juosteen, joka on yhteensopiva malli-DNA:n kanssa. Useimmiten käytetään Taq-polymeraasientsyymiä, mutta muitakin vaihtoehtoja on (Lahr & Katz 2009, 857). Nukleotidit ovat DNA:n perusrakenneosia, ne koostuvat fosfaattiosasta, sokerimo-

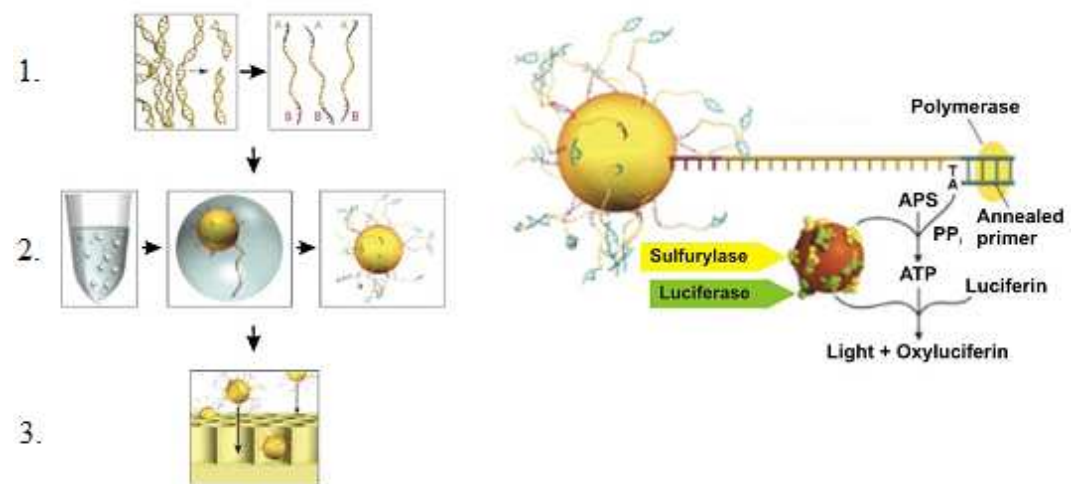
lekyylistä ja emäksestä, joka on adeniini (A), tyymiini (T), guaniini (G) tai sytosiini (C). Puskuriliuoksen tehtävänä on vakauttaa entsyymin toimintaa ja tehdä olosuhteet optimaalisiksi. Kationit (Mg^{2+} tai Mn^{2+}) parantavat entsyymin tarkkuutta. (Cha & Thilly 1993.)

PCR perustuu lämpöolosuhteiden täsmälliseen toistuvaan muutokseen eli termiseen kierrätykseen. PCR:n vaiheet ovat denaturaatio (94 - 98 °C), alukkeiden kiinnittymien (45 - 65 °C) ja pidennys (72 °C). PCR aloitetaan yleensä normaalia pidemmällä, n. 2 minuutin, denaturaatiolla, jolla varmistetaan, että juosteet ovat varmasti irronneet toisistaan. (1) Normaali denaturoimis aika termisen kierrätyksen aikana on 20 - 30 sekuntia ja lämpötila 94 - 98 °C. (2) Denaturoinnin jälkeen lämpötila lasketaan alukkeille sopivaan kiinnittymislämpötilaan, joka vaihtelee 45 ja 65 °C:n välillä. Kiinnittymislämpötila riippuu alukkeiden pituudesta ja emäskoostumuksesta. Alukkeet kiinnittyvät yksijuosteeseen DNA:han emäsjärjestyksen mukaiseen paikkaan ja määräävät kopioitavan alueen. (3) Pidennysvaiheessa lämpötila nostetaan DNA-polymeraasientsyymille sopivaksi. Entsyymille optimilämpötila on 75 - 80 °C, mutta useimmiten käytetään 72 °C. Viimeisen syklin jälkeen tehdään vielä yksi pidennys 70 - 74 °C:ssa 5 - 15 minuutin ajan. Viimeisen pidennyksen tarkoituksena on varmistaa, että kaikki juosteet on rakennettu loppuun asti. Tämän jälkeen lämpötila lasketaan 4 - 15 °C:seen, jossa PCR-tuotetta voidaan säilyttää väliaikaisesti lyhyt aika. Vaiheita 1 - 3 toistetaan yleensä 25 - 45 kierrosta, ja joka kierroksella kopioitavan DNA:n määrä kaksinkertaistuu. Jos kierroksia on 30, niin yhdestä DNA-molekyylistä saadaan 2^{28} eli n. 270 miljoonaa kopiota. (Cha & Thilly 1993.)

2.3 Pyrosekvensointi

Pyrosekvensointi on tekniikka, jonka avulla voidaan luopua aikaa vievistä kloonauksesta ja pesäkkeiden valinnasta. Pyrosekvensoinnilla saadaan sekvensoitua yhdellä ajolla tavallisesti n. 100 miljoonaa hyvälaatuista emästä. Pyrosekvensoinnissa on neljä vaihetta: DNA-kirjaston valmistelu, em-PCR (emulsio-PCR), sekvensointi (kuvio 4) ja sekvensointi-informaation analysointi. (Roche Diagnostics Oy 2010.)

DNA-kirjasto voidaan luoda kahdella tavalla: pilkkomalla ympäristönäytteen genomin DNA sopivan pituisiksi kappaleiksi tai PCR:n avulla monistamalla kiinnostuksen kohteena olevaa genomien aluetta. DNA-kirjaston luonnin yhteydessä lyhyisiin kappaleisiin lisätään A- ja B-adapterit sekä avainsekvenssit näytteen tunnistamista varten. B-adaptereissa on magneettiset biotiinimolekyylit, joiden avulla erotetaan ne kappaleet, joissa on sekä A- ja B-adapterit ja näin saadaan yksijuosteinen DNA-kirjasto. (Roche Diagnostics Oy 2010.)



KUVIO 4. Pyrosekvensoinnin vaiheet 1 – 3 ja sekvensointivaiheen valoa tuottava reaktio (Ansoerge 2009)

Em-PCR:ssä DNA-kirjasto monistetaan sekvensoinnissa käytettävän helmen pintaan. Em-PCR:ään tarvitaan PCR-tuotetta, jossa on adapterit ja avainsekvenssi, pieni helmi, nukleotideja, entsyymejä, ja muita yleisesti PCR:ssä tarvittavia reagensseja. PCR-reagenssit sekoitetaan vesi-öljyseokseen, jota ravistellaan, niin että muodostuu pieniä kuplia, mikroreaktoreita. PCR-tuotteen ja helmien määrää säädetään niin, että yhteen kuplaan päätyy korkeintaan yksi helmi ja yksi juoste kirjastosta sekä sopiva määrä muita reagensseja. Helmen pinnalla on A-adapteria vastaavia koettimia, joihin PCR-tuote tarttuu ja kopioituu helmeen kiinni. Kun helmen pintaan on monistettu riittävän monta kopiota, mikroreaktorit rikotaan, monistettuja vastaavat juosteet huuhdellaan pois ja helmet ja niissä kiinni oleva tuote on valmis sekvensointia varten. (Roche Diagnostics Oy 2010.)

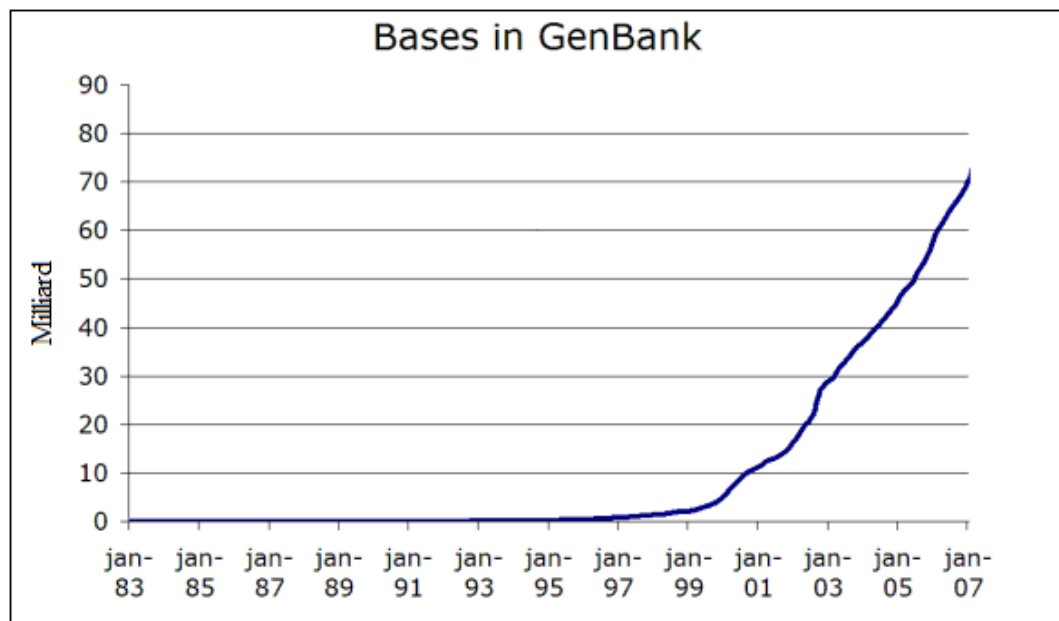
Pyrosekvensointi perustuu valoreaktioon, joka syntyy sen seurauksena, että sekvensoitavaan juosteeseen liittyy nulkeotidi. Pyrosekvensoinnista saadaan jopa 400 tuhatta n. 250 bp:n sekvenssiä yhdellä ajolla. Sekvensointiin tarvitaan em-PCR:stä helmiä, joista jokaisen pintaan on kopioitu yhtä juostetta. Helmet laitetaan pikotiitterilevyille, jossa on pieniä kaivoja ja joihin jokaiseen mahtuu vain yksi helmi. Lisäksi kaikkiin kaivoihin tulee pakkaushelmiä ja entsyymihelmiä, jotka sisältävät sulfurylaasia ja lusiferaasia. Helmissä kiinni oleviin sekvensseihin hybridisoidaan B-adaptoria vastaavat alukkeet, jotka toimivat sekvensoinnin aloituskohtana. Sekvensoinnissa pikotiitterilevyille lisätään vuorotellen yhtä emästä (A, T, C, tai G), joka pestään pois ennen seuraavan lisäystä. Jos emäs kiinnittyy sekvensoitavaan juosteeseen, syntyy valoreaktio, joka tallennetaan CCD-kameralla. Nukleotidin kiinnittyessä juosteeseen vapautuu pyrofosfaatti, joka reagoi sulfurylaasin kanssa tuottaen ATP-molekyylin. Reaktiosarja jatkuu ATP:n ja lusiferaasin reaktiolla, jonka seurauksena syntyy valoa ja lusiferaasi hapettuu (kuvio 4). Mitä useampi nukleotidi liittyy kerralla sekvensoitavaan juosteeseen, sitä voimakkaampi valoreaktio syntyy. Tallentuneiden valoreaktioiden perusteella saadaan selville DNA-fragmenttien emäsjärjestys. (Roche Diagnostics Oy 2010.)

2.4 Bioinformatiikka

Bioinformatiikan käsite syntyi, kun biologiset sekvenssiaineistot (DNA- ja aminohapposekvenssit) kasvoivat sellaisiin mittoihin, että niiden tehokkaaseen käsittelyyn tarvittiin tietoteknisiä ratkaisuja. Tieteenalana bioinformatiikka on varsin uusi, eikä sitä ole vielä täsmällisesti määritelty. (Tuimala 2005, 12.) Yksinkertaisesti bioinformatiikan voidaan sanoa olevan työkalu biologian ymmärtämiseen analysoimalla biologisia sekvenssiaineistoja tietotekniikan avulla (Gibas & Jambek 2001, 11). Bioinformatiikka voidaan määritellä tarkemmin kolmen eri näkökulman kautta. Ensimmäiseksi bioinformatiikan tarkoituksena on biologisen tiedon säilyttäminen ja ylläpitäminen tietokannoissa. Oleellista on, että tieto on helposti tutkijoiden saatavilla ja että tutkijat voivat tallentaa tietokantoihin omia aineistoja tiedon kartuttamiseksi. Toiseksi bioinformatiikkaan kuuluu olennaisena osana tietoteknisten ratkaisujen, kuten ohjelmien ja algoritmien, kehittäminen. Näiden avulla aineistoja voidaan tutkia ja käsitellä. Kolmannesta näkökulmasta

bioinformatiikka on aineistojen ja tietoteknisten ratkaisujen hyödyntämistä, niin että saadaan järkevästi tulkittavaa biologista tietoa. (Tuimala 2005, 12.)

Tietokannat ovat keskeinen osa bioinformatiikkaa. Tietokantoja, joihin on tallennettu ajan myötä valtavat määrät mm. DNA- ja aminohapposekvenssejä sekä muuta biologista tietoa, kutsutaan tarkemmin biotietokannoiksi. Tietokannat ovat avoimia kaikille, jolloin tutkijat pääsevät hyödyntämään niitä täysipainoisesti ja tallentamaan moniin niistä uutta tietoa omista tutkimuksistaan. Tietokannat ovat kasvaneet eksponentiaalisesti viime vuosina ja näyttää, että kasvu tulee myös jatkumaan lähitulevaisuudessa (kuvio 5). Tietokantoja on monia erilaisia, kuten nukleotidisekvenssi-, aminohappo-, yhdistelmä- ja julkaisutietokantoja. Esimerkiksi nukleotidisekvenssitietokantoja ovat EMBL (Euroopassa), GenBank (Pohjois-Amerikassa) ja DDBJ (Japanissa). (Tuimala 2005, 25 - 26, 31 - 33, 37.) Jokaiseen niistä oli tallennettuna yli 117 miljardia emästä elokuuhun 2010 mennessä (NCBI 2010; Tuimala 2005, 26).



KUVIO 5. GenBank:iin tallennettujen emästen määrä vuodesta 1983 vuoteen 2007 (CLC Bio 2008; NCBI 2010)

Biologisen aineiston käsittelyyn on monia eri menetelmiä ja työkaluja, joista tärkeimpänä voidaan pitää kahden sekvenssin rinnastamista eli vertaamista toisiinsa (Gibas & Jambeck 2001, 12). Rinnastuksen tarkoituksena on saada selville, kuinka hyvin kaksi sekvenssiä vastaa toisiaan. Rinnastaminen voidaan tehdä kahdella tavalla: 1) koko matkalta, jolloin puhutaan kokonaisrinnastuksesta, tai 2) etsimällä sekvensseistä parhaiten toisiaan vastaavia alueita, jolloin kyseessä on paikallisrinnastus. Rinnastettavat sekvenssit ovat harvoin identtisiä, minkä vuoksi rinnastuksessa voi olla yhteen sopimattomia merkkejä tai lisättyjä aukkoja, jos se on tarpeellista sekvenssien yhteen sovittamiseksi (kuvio 6). (Tuimala 2005, 17, 20, 48.) Rinnastamista käytetään mm. tietokantahauissa, evoluutiopuiden muodostamisessa ja genomien vertailussa (Gibas & Jambeck 2001, 142).

Query	1	TGGGGAATTTTGGACAATGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGAAGAAGG	60
Sbjct	298	TGGGGAATTTTGGACAACGGGGGCAACCCTGATCCAGCCATGCCGCGTGAGTGAAGAAGG	357
Query	61	CCTTCGG-TTGIAAAGCTCTTTGTCCGGGAAGAAAAGGC-TCACCTAATACGGTGG-G	117
Sbjct	358	CCTTCGGGTGIAAAGCTCTTTGTCCGGGAAGAAAAGGCATCGC-CTAATA-GGAGATG	415

KUVIO 6. Esimerkki BLAST-haun rinnastuksesta: Query on hakusekvenssi, Sbjct (subject) on tietokannasta löydetty sekvenssi, pystyviiva (|) tarkoittaa, että merkit vastaavat toisiaan, tyhjä kohta () tarkoittaa, että merkit eivät vastaa toisiaan ja vaakaviiva (-) on lisätty, että sekvenssit vastaisivat paremmin toisiaan

Tämän opinnäytetyön kannalta mainitsemisen arvoinen menetelmä on myös sekvenssihaut, joka on luultavasti eniten käytetty työkalu bioinformatiikassa. Sekvenssihauilla etsitään tietokannoista hakusekvenssille vastaavuuksia. Jos tuloksena saadut sekvenssit vastaavat tarpeeksi hyvin hakusekvenssiä, voidaan tutkimus tarkoituksesta ja sekvenssistä riippuen päätellä esim. mistä geenistä on kyse tai mikä on proteiinin funktio. Nykyisin eniten käytetty sekvenssihakuohjelma (algoritmi) on BLAST (Basic local alignment search tool). (Tuimala 2005, 58 - 60.)

3 MATERIAALIT JA MENETELMÄT

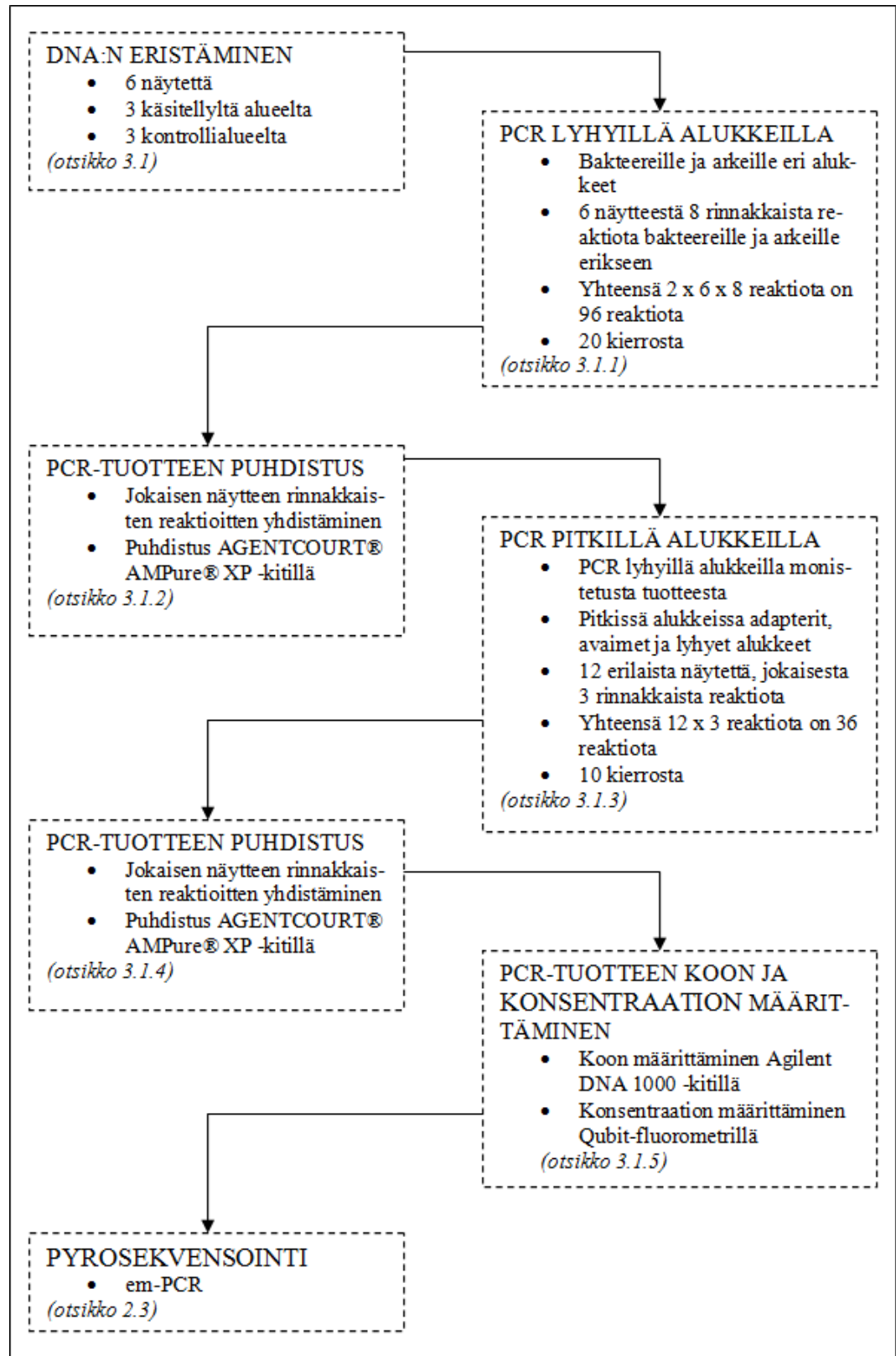
3.1 DNA:n eristyksestä DNA-kirjastoon

Sedimenttinäytteet DNA-kirjastoa varten otettiin Hollolassa sijaitsevasta Työtjärvestä kahdelta eri alueelta, joista ensimmäinen oli happea hitaasti vapauttavalla kemikaalilla käsitelty syväne ja toinen kontrollialue. Kummaltakin alueelta otettiin näyte kolme kertaa (taulukko 1). Näytteitä säilytettiin -80 C:ssa näytteenoton jälkeen.

TAULUKKO 1. Näytekoodit ja näytteenottopäivät

Näytteenottopäivä	BAKTEERIEN NÄYTTEET		ARKKIEN NÄYTTEET	
	Käsitelty alue	Kontrollialue	Käsitelty alue	Kontrollialue
8.10.2009	B1	B4	A1	A4
2.12.2009	B2	B5	A2	A5
12.1.2010	B3	-	A3	-
13.1.2010	-	B6	-	A6

DNA eristettiin MO Bio Power Soil -kitillä (MO BIO Laboratories, Inc.) valmistajan ohjeen mukaan pienin muutoksin. Pakastetuista näytteistä louhittiin jäätyntä sedimenttiä 60 ml:n falconputkiin, joissa sedimentin annettiin sulaa. Sulanutta sedimenttiä sentrifugoitiin Hareaus Multifuge 1 S - R:llä (Kendro Laboratory Products, Inc) 3 minuuttia 4000 rpm:ssä, minkä jälkeen erottunut vesi kaadettiin pois ja sedimenttiä punnittiin n. 0,5 g eppendorffputkiin (kaksi kertaa enemmän kuin ohjeessa). Tästä eteenpäin noudatettiin kitin ohjetta. Eristyksen onnistuminen tarkistettiin elektrofooresilla agarosigeelillä. Vaiheet DNA:n eristämisestä DNA-kirjaston luomiseen on esitetty lyhyesti kuviossa 7.



KUVIO 7. Työn vaiheet DNA:n eristämisestä sekvensointiin

3.1.1 PCR lyhyillä alukkeilla

Eristetystä DNA:sta monistettiin 16S rRNA -geenin aluetta V3 (kuvio 8). Usein yksittäisissä PCR-reaktioissa monistuu vain osa näytteessä esiintyvistä lajeista, mikä voi johtaa tutkimuksen vääristymiseen. Tämän vuoksi näytteiden koko monimuotoisuuden monistamiseksi on suositeltavaa tehdä rinnakkaisia reaktioita, jotka yhdistetään PCR:n lopuksi. (Lahr & Katz 2009, 857.) Lyhyillä alukkeilla tehdyssä PCR:ssä jokaisesta näytteestä (taulukko 1) tehtiin 8 rinnakkaista reaktiota, jotka yhdistettiin ennen puhdistusta (kuvio 7). PCR tehtiin DNA Engine Dyad™ Peltier Thermal Cycler -laitteella (MJ Research Genetic). Lyhyillä alukkeilla tehdyn PCR:n ohjelmaa toistettiin 20 kierrosta (taulukko 2).



KUVIO 8. 16S rRNA -geenin konservoituneet ja vaihtelevat alueet E.Colin mukaan (C (conseved) = konservoitunut ja V (variable) = vaihteleva) (Petrosino, Highlander, Luna, Gibbs & Versalovic 2009)

Lyhyillä alukkeilla tehdyssä PCR:ssä käytettiin bakteereille ja arkeille universaaleja alukkeita, jotka sitoutuivat 16S rDNA -alueelle. Bakteerialukkeet olivat seuraavanlaiset: 341f 5'CCTACGGGAGGCAGCAG3' (Muyzer, De Waal & Uittierlinden 1993, 696) ja D' 5'GTATTACCGCGGCTGCTG3' (Edwards, Rogall, Blocker, Emde & Bottger 1989, 7844). Arkeille käytettiin seuraavanlaisia alukkeita: 344f 5'ACGGGGCGCAGCAGGCGCGA3' (Bano, Ruffin, Ransom & Hollibaugh 2004, 782), CREN512 5'CGGCGGCTGACACCAG3' (Jürgens ym. 2000) ja 519 5'ATTACCGCGGCGGCTG3', jota on muunnettu alkuperäisestä (Muyzer ym. 1993, 696).

PCR-reaktiot toteutettiin 50 µl:n tilavuudessa. Bakteerien PCR-reaktioihin tuli 10 µl 5x HF-puskuria (Finnzymes), 1 µl 10 mM dNTP (Finnzymes), 0,5 µl Phusion-entsyymiä (Finnzymes), 1 µl alukkeita (Oligomer Oy), 1 µl malli-DNA:ta ja loput vettä. Arkkien PCR-reaktioihin tuli 5 µl 5x-puskuria (Biotools), 2 µl MgCl₂ (Bio-

tools), 1 µl 10 mM dNTP (Finnzymes), 0,25 µl DNA-polymeraasientsyymiä (Bio-tools), pipetin kärjellä vähän PFU-entsyymiä (Fermentas), 1 µl alukkeita (Oligomer Oy), 1 µl malli DNA:ta ja loput vettä. PFU-entsyymi lisättiin, niin että pipetin kärkeä kastettiin entsyymissä, minkä jälkeen pipetin kärkeen tarttunut entsyymi sekoitettiin reaktioseokseen

TAULUKKO 2. PCR-ohjelmat arkeille ja bakteereille

ARKIT			BAKTEERIT		
Kohta	Lämpötila	Inkubointi	Kohta	Lämpötila	Inkubointi
1	94 °C	5 min	1	98 °C	30 sec
2	94°C	30 sec	2	98°C	10 sec
3	60 °C	30 sec	3	65 °C	30 sec
4	72 °C	2 min	4	72 °C	10 sec
5	paluu kohtaan 2, x kertaa		5	paluu kohtaan 2, x kertaa	
6	72 °C	5 min	6	72 °C	5 min
7	4 °C	5 min	7	4 °C	5 min
8	4 °C	ikuisesti	8	4 °C	ikuisesti

3.1.2 PCR-tuotteen puhdistus

PCR-tuote puhdistettiin AGENCOURT® AMPure® XP -kitillä. Puhdistustuloksen optimoimiseksi Agencourt AMPure XP Magnetic Particle -liuoksen toimivuutta testattiin eri suhteilla Ladder-laimennoksessa (taulukko 3). Ladder-laimennokseen tuli 240 µl ladderia ja 360 µl dH₂O:ta. Tuloksien pohjalta bakteereille käytettiin 0,9:1 suhdetta ja arkeille 1:1 suhdetta. Lyhyillä alukkeilla tehdyn PCR:n tuote puhdistettiin kuoppalevyillä ohjeen mukaan. Eluointiin käytettiin 40 µl 1x TE-puskuria (10 mM Trisbase, 1mM EDTA).

TAULUKKO 3. Taulukko eri seossuhteista Agencourt AMPure XP –liuoksen puhdistustuloksen optimoimiseksi

	AMPure Xpliuos:Ladderi	AMPure XP-liuos	Ladderi- laimennos
1.	0,5:1	25	50
2.	0,6:1	30	50
3.	0,7:1	35	50
4.	0,8:1	40	50
5.	0,9:1	45	50
6.	1:1	50	50
7.	1,1:1	55	50
8.	1,2:1	60	50
9.	1,3:1	65	50
10.	1,4:1	70	50
11.	1,5:1	75	50

3.1.3 PCR pitkillä alukkeilla

Lyhyillä alukkeilla tehdystä PCR:stä oli tuloksena yhteensä 12 erilaista reaktiota: kuusi bakteereille ja kuusi arkeille. Jokaisesta 12 reaktiosta tehtiin 3 rinnakkaista reaktiota pitkillä alukkeilla (kuvio 7). PCR-ohjelmaa toistettiin 10 kierroksen verran (taulukko 2).

Pitkillä alukkeilla tehdyssä PCR:ssä käytettiin jokaiselle näytteelle hiukan erilaista aluketta. Ne koostuivat adapterista, avaimesta ja lyhyestä alukkeesta (taulukko 4). Adaptereita oli kaksi erilaista A: 5'GCCTCCCTCGCGCCATCAG3' ja B: 5'GCCTTGCCAGCCCGCTCAG3'. A tuli johtavaa juostetta monistaviin alukkeisiin ja B vastakkaista juostetta monistaviin alukkeisiin. A adapterin tehtävänä oli toimia kohtana, joka kiinnittyy emPCR:ssä käytettävien helmien pinnassa olevaan koettimeen kiinni. B adapteri puolestaan on hybridisoitumiskohta alukkeelle, josta alkaa sekvensointi. Avaimen tehtävänä oli toimia sekvenssinä, jonka avulla tunnistettiin näytteet toisistaan ja muiden näytteistä pyrosekvenssoinnin jälkeen. Avaimia oli yhteensä 12 erilaista, joista 6 oli bakteereille ja 6 arkeille (taulukko 4).

PCR-reaktiot toteutettiin 50 µl:n tilavuudessa. Bakteerien PCR-reaktioihin tuli 10 µl 5x HF-puskuria (Finnzymes), 1 µl 10 mM dNTP (Finnzymes), 0,5 µl Phusion-

entsyymiä (Finnzymes), 2 µl jokaista näytettä vastaavaa aluketta (Oligomer Oy) (taulukko 4), 2 µl malli-DNA:ta ja loput vettä. Arkkien PCR-raktioihin tuli 5 µl 5x-puskuria (Biotools), 2 µl MgCl₂ (Biotools), 1 µl 10 mM dNTP (Finnzymes), 0,25 µl DNA-polymeraasientsyymiä (Biotools), 0,1 µl PFU-entsyymiä (Fermetas), 2 µl jokaista näytettä vastaavaa aluketta (Oligomer Oy) (taulukko 4), 4 µl malli-DNA:ta ja loput vettä. Malli-DNA:na PCR:ssä käytettiin puhdistettua PCR-tuotetta, joka oli monistettu lyhyillä alukkeilla.

TAULUKKO 4. Pitkien alukkeiden rakenne: näyte, jonka monistamiseen aluketta on käytetty; adapteri; avain näytteen tunnistamiseksi; hybridisoitumiskohtaa vastaavan alukkeen nimi

BAKTEERIT				ARKIT			
KÄSITELTY ALUE				KÄSITELTY ALUE			
Näyte	Adapteri	Avain	Aluke	Näyte	Adapteri	Avain	Aluke
B1	A	ACTCAG	341f	A1	A	ACTATA	344f
B1	B	ACTCAG	D'	A1	B	ACTATA	CREN512
				A1	B	ACTATA	519
B2	A	ATAGTA	341f	A2	A	ACTCTC	344f
B2	B	ATAGTA	D'	A2	B	ACTCTC	CREN512
				A2	B	ACTCTC	519
B3	A	ACACTG	341f	A3	A	ATGAGC	344f
B3	B	ACACTG	D'	A3	B	ATGAGC	CREN512
				A3	B	ATGAGC	519
KONTROLLIALUE				KONTROLLIALUE			
B4	A	CGCTAC	341f	A4	A	CATAGC	344f
B4	B	CGCTAC	D'	A4	B	CATAGC	CREN512
				A4	B	CATAGC	519
B5	A	CTGCTA	341f	A5	A	TACATA	344f
B5	B	CTGCTA	D'	A5	B	TACATA	CREN512
				A5	B	TACATA	519
B6	A	TAGTAT	341f	A6	A	TACGCG	344f
B6	B	TAGTAT	D'	A6	B	TACGCG	CREN512
				A6	B	TACGCG	519

3.1.4 PCR-tuotteen puhdistus

PCR-tuote puhdistettiin AGENCOURT® AMPure® XP -kitillä samoin kuin aiemmin, mutta puhdistus tehtiin PCR-putkissa, jolloin inkubointiajat venyivät pidemmiksi kuin ohjeessa. Eluointiin käytettiin toisen PCR:n jälkeen 30 µl 1x TE-puskuria (10 mM Trisbase, 1mM EDTA) 40 µl:n sijaan.

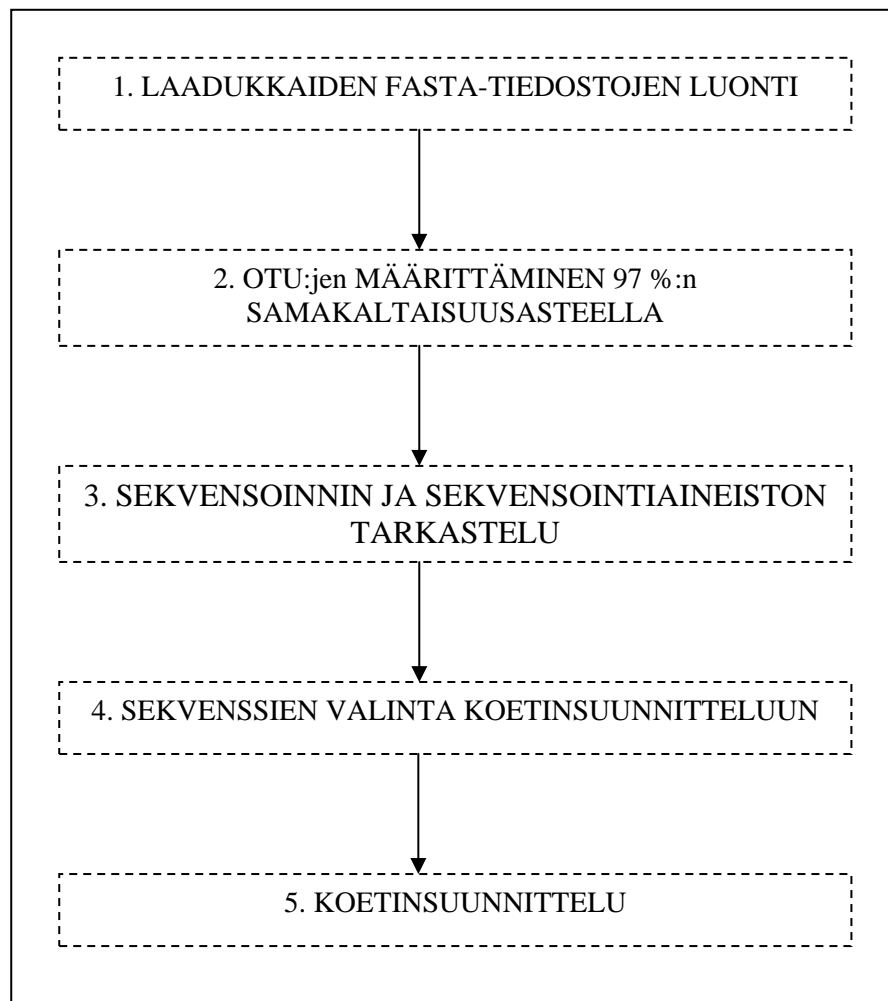
3.1.5 PCR-tuotteen koon ja konsentraation määrittäminen

PCR-tuotteen koko tarkistettiin Agilent DNA 1000 -kitillä. Kitissä oli mukana DNA-siru, johon pipetoitiin reagenssit ja PCR-tuote analyysiä varten. Sirua ravisteltiin Vortex Mixerissä (IKA -model MS2-S8/ MS2-S9) 2400 rpm:ssä 60 sekuntia, minkä jälkeen siru asetettiin Bioanalyzer asemaan ja näytteet analysoitiin 2100 Expert ohjelmistolla.

PCR-tuotteen konsentraatio määritettiin Qubit® -fluorometrillä. Työliuos valmistettiin sekoittamalla Quant-iT™ -reagenssia Quant-iT™ -puskuriin suhteessa 1:200. Kalibrointiliuoksiin tuli standardeja 10 µl ja työliuosta 190 µl. Näytteiden analysointiin tuli 1 µl näytettä ja 199 µl työliuosta. Näytteet ja standardit vortexsoitiin, minkä jälkeen niitä inkuboitiin 2 minuuttia ennen analysointia.

3.2 Sekvensointidatasta koettiin

Laboratoriotöiden jälkeen PCR-tuotteen emäsjärjestys selvitettiin pyrosekvensoinnilla. Siitä saatua aineistoa muokattiin niin, että sitä voitiin tarkastella bioinformatiikan ohjelmilla ja hyödyntää koetinsuunnittelussa. Työn vaiheet sekvensoinnin raakadatan muokkaamisesta koetinsuunnitteluun on kuviossa 9.



KUVIO 9. Työn vaiheet sekvensointidatasta valmiisiin koettiin

3.2.1 Laadukkaiden fasta-tiedostojen luonti pyrosekvensointidatasta

Pyrosekvensointi tehtiin Biotekniikan instituutissa Helsingissä, josta data luovutettiin sff-muodossa. Raakadata muunnettiin aluksi fasta-muotoon, jonka jälkeen Ribosomal Data Base Projectin (RDP) Pyrosequencing Pipeline Initial Process -ohjelmalla (Cole ym. 2008; Cole ym. 2007) poistettiin huonolaatuiset DNA-sekvenssit, avainsekvenssit, adapterit ja sekvenssit, jotka olivat alle 100 bp:n pituisia.

Seuraavaksi näytteet linjattiin ja linjauksia tarkasteltiin Bioedit-ohjelmalla. Oudosti linjautuneet sekvenssit tarkistettiin NCBI:n tietokannasta, ja jos sekvenssi oli tuntematon, väärään yläkuntaan kuuluva (bakteerit, arkit, eukaryootit) tai muuten kehno, niin se poistettiin. Tämän jälkeen poistettiin alukkeet Bioedit -ohjelmassa ja linjaukset tekstinkäsittelyohjelmalla, jolloin tuloksena oli laadukkaat näytekohtaiset fasta-tiedostot.

3.2.2 Näytteiden sekvenssien luokittelu OUT:ihin 97 %:n samankaltaisuusasteella

Sekvensointiaineiston DNA-sekvenssit voidaan jakaa samankaltaisuusasteen perusteella ryhmiin, joita kutsutaan myös OTU:iksi (operational taxonomic unit). Samankaltaisuusasteen perusteella voidaan määrittää, millä taksonomisella tasolla OTU:ja tarkastellaan. Jos esim. halutaan tarkastella sekvensointiaineistoa lajitasolla, ryhmitellään DNA-sekvenssit OTU:ihin 97 %:n samankaltaisuusasteella, joka tarkoittaa, että yhteen OTU:uun kuuluvat DNA-sekvenssit poikkeavat toisistaan korkeintaan 3 %. Yleisesti 97 %:n samankaltaisuusastetta pidetään lajirajana (Schloss & Handelsman 2005, 1501).

Sekvensointiaineisto määriteltiin OTU:iksi, jotta näytteiden lajikirjoja olisi helpompi tarkastella. Aluksi bakteeri- ja arkinäytteiden sekvensseistä muodostettiin molemmille tiedosto, joissa oli kaikkien bakteeri- tai arkinäytteiden sekvenssit koottuna yhteen. Seuraavaksi tiedostot linjattiin RDP:n Aligner-ohjelmalla (Cole

ym. 2008; Cole ym. 2007) ja linjatuista sekvensseistä muodostettiin etäisyysmatriisi RDP:n Column Distance Matrix -ohjelmalla (Cole ym. 2008; Cole ym. 2007). Etäisyyden laskemiseen käytettiin Jukes-Cantor -mallia, Distance Cutoff -arvoa 0,2 ja Minimum Sequence Overlap -arvoa 25. Etäisyys matriisien avulla ryhmiteltiin sekvenssit OTU:iksi 97 % samankaltaisuudella Mothur -ohjelmassa (Schloss ym. 2009) ja näistä ryhmistä valittiin yksi sekvenssi edustamaan omaa OTU:aan. OTU-tiedostoille ja näytetiedostoille tehtiin BLAST-haut EMBL:n prokaryoottinukleotidisekvenssitietokantaa vastaan, josta tuloksena oli 25 parasta osumaa ja linjausta jokaista sekvenssiä kohti. BLAST-haun tuloksia käytettiin tarkasteltaessa näytteiden taksonomista sisältöä MEGANilla (Huson, Auch, Qi, & Schuster 2007).

3.2.3 Sekvensoinnin ja sekvensointiaineiston tarkastelu

Mothurilla arvioitiin OTU:jen todellista määrää näytteissä ja bakteerien sekä arkien yhdistetyissä tiedostoissa. OTU-runsauden arviointiin käytettiin vääristymän huomioivaa (1) Chao1 -bc (bias corrected) -menetelmää (Chao 1984), joka hyödyntää arvioinnissa OTU:ja, joihin kuuluu 1 tai 2 DNA-sekvenssiä.

$$(1) \quad S_{chao1} = S_{obs} + \frac{n_1(n_1 - 1)}{2(n_2 + 1)}$$

S_{chao1} = arvioitu OTU:jen runsaus

S_{obs} = havaittujen OTU:jen määrä

n_1 = OTU:t , joihin kuuluu vain 1 DNA-sekvenssi

n_2 = OTU:t , joihin kuuluu vain 2 DNA-sekvenssiä

Lisäksi Mothurilla määritettiin kuvaajat, joista ilmeni OTU:jen kertyminen verrattuna sekvensoituihin DNA-sekvensseihin (rarefaction-kuvaajat). Kuvaajista voidaan päätellä, kuinka suuri osuus OTU:ista on saatu sekvensoitua käytetyllä sekvensointimäärällä ja kuinka suuri työ on vielä tehtävä, että havaittaisiin uusi OTU. Kun kuvaajat alkavat tasaantua ja lähestyä niin sanottua asymptoottia, lähes kaikki OTU:t on sekvensoitu. (Lai, Cao, Tan, Fang, Huang, & Zhou 2007, 758.)

RDP:n Classifieria (Wang, Garrity, Tiedje & Cole 2007) käytettiin erityisesti muodostamaan yleiskuva mikrobistosta, mutta myös joidenkin koetinsuunniteluun valittujen sekvenssien nimeämiseen, jos Classifierin määritelmä oli tarkempi kuin MEGANin. Classifier määrittää jokaiselle sekvenssille, mihin taksoneihin se kuuluu kaaresta sukuun ja laskee jokaiselle taksonille luotettavuusarvon, joka voi vaihdella 0 % ja 100 %:n välillä. Mitä korkeampi luotettavuusarvo on, sitä todennäköisemmin taksoni on määritetty oikein. Classifier muodostaa tuloksista hierarkkisen puun, jossa esitetään kaikki taksonit, jotka ylittävät tarkasteluun valitun luotettavuusrajan. Esimerkiksi, jos tarkasteluun valitaan rajaksi 80 %, niin kaikki taksonit, joiden luotettavuusarvo on alle 80, tiputetaan pois puusta. Tässä työssä käytettiin luotettavuusrajana 50 %:a, joka on lähes yhtä luotettava kuin 80 % luotettavuusraja alle 250 bp:n sekvensseille, ja lisäksi pystyy määrittämään taksonin huomattavasti suuremmalle osuudelle DNA-sekvenssejä (Claesson ym. 2009, 4).

MEGANia käytettiin lajitason tarkasteluun sekä sekvenssien valintaan koetinsuunnittelua varten. MEGAN esittää aineiston graafisena puuna, johon sekvenssit on sijoitettu LCA-algoritmin (Lowes Common Ancestor) perusteella. LCA-algoritmi sijoittaa sekvenssit viimeiseen yhteiseen taksoniin, eli jos sekvenssi tunnistetaan kahdeksi lajiksi yhdestä suvusta, niin se sijoitetaan sukuun. LCA-algoritmissa on neljä erilaista arvoa, Min Support, Min Score, Top Percent ja Win Score, joita muuttamalla voidaan vaikuttaa MEGANin muodostamaan puuhun ja sen täsmällisyyteen (taulukko 5). Tiedostojen tarkastelussa käytetyt LCA:n arvot ovat taulukossa 6. Arvot valittiin niin, että suurin osa huonosti linjatuista sekvensseistä jäi tarkastelun ulkopuolelle.

TAULUKKO 5. LCA-algoritmin arvojen vaikutus taksonomiseen puuhun

LCA-algoritmin arvojen selitteet	
Min Support	Tämä kohta määrittelee kuinka monta DNA-sekvenssiä on vähintään oltava määritettynä lajille tai puun viimeiselle lehdelle, että se esitetään puussa. Jos esim. min support -arvoksi on asetettu 5, niin sellaisia taksoneita ei esitetä, joihin DNA-sekvenssejä on määrätty 1, 2, 3, tai 4, näitten taksoneitten DNA-sekvenssit sijoitetaan ei määritettyjen DNA-sekvenssien joukkoon (not assigned). Pientämällä tätä arvoa, puussa esitetään myös sellaiset lajit, joita näytteissä on ollut vähän. Toisaalta yksittäiset DNA-sekvenssit voivat olla sekvenssoinnista tai PCR:stä johtuvia virheitä.
Min Score	Määrittää kuinka paljon huomioitavista linjauksista on vähintään saatava pisteitä. Niitä linjauksia, joiden pisteet jäävät raja-arvon alle, ei huomioida. Jos kaikki linjaukset ovat alle raja-arvon, DNA-sekvenssi sijoitetaan ei-määritettyjen DNA-sekvenssien joukkoon (Not Assigned). Oletusarvona on 35, jota nostamalla saadaan lyhyet ja huonot linjaukset tiputettua pois puusta.
Top Percent:	Määrittää mitkä DNA-sekvenssin linjauksista otetaan mukaan, kun määritetään DNA-sekvenssin taksonia. Huomioitavien linjausten pisteiden on oltava tietyn prosentiosuuden sisällä parhaan linjauksen pisteistä. Jos Top Percen -arvo on 10 % ja paras linjaus on saanut 250 pistettä, niin niitä linjauksia ei huomioida, jotka saavat alle 225 pistettä. Pientämällä tätä arvoa DNA-sekvenssit määritetään lähemmäs lajitasoa, mutta samalla määrityksen virheellisyys on todennäköisempää.
Win Score	Ne DNA-sekvenssien linjaukset, joiden pisteet ylittävät win score -arvon, määräävät taksonin. Jos Win Score -arvo on 150, niin vain ne linjaukset huomioidaan, joiden arvo on 150 tai yli.

TAULUKKO 6. MEGANissa käytettyjen LCA:n arvot bakteereille ja arkeille

LCA:n ARVOT					
BAKTEERIT			ARKIT		
Tiedosto	B1 - B6	OTU	Tiedosto	A1 - A6	OTU
Min Support	2	1	Min Support	2	1
Min Score	180	180	Min Score	80	80
Top Percent	7	7	Top Percent	7	7
Win Score	0	0	Win Score	0	0

3.2.4 OTU:jen valinta koetinsuunnitteluun

DNA-sekvenssit valittiin koetinsuunnitteluun MEGANin tulosten perusteella. Ensisijaisesti pyrittiin valitsemaan OTU:ja sellaisista lajeista ja suvuista, jotka indikoivat sedimentin käsittelyä happea vapauttavalla kemikaalilla. Käsittelyn vaikutusta taksoneihin tarkasteltiin taksonien suhteellisen koon muutoksena käsitellyllä ja kontrollialueella ajan funktiona. Käsittelyä indikoivina taksonina pidettiin sellaisia taksonia, joiden suhteellisen koon muutos oli erilainen käsitellyllä ja kontrollialueella ajan kuluessa. Esimerkiksi, jos lajin suhteellinen koko kasvoi käsitellyllä alueella ja pieneni kontrollialueella ajan kuluessa, niin kyseessä oli käsittelyä indikoiva laji. Käsittelyn vaikutuksen seurantaan käytettiin käsittelyn alueen näytteitä B1 ja A1 (otettu 8.10.2009 ennen happea vapauttavan kemikaalin levitystä), B3 ja A3 (12.1.2010) sekä kontrollialueen näytteitä B4 ja A4 (8.10.2009) sekä B6 ja A6 (13.1.2010). Jos näistä ei saatu riittävästi koettimia, niin seuraavaksi pyrittiin valitsemaan monimuotoisuutta ilmentäviä OTU:ja, jotka olivat tunnistettavissa laji- ja sukutasolle.

3.2.5 Koetinsuunnittelu

Koettimia oli tarkoitus suunnitella n. 50 - 100 kpl, joista n. 80 % bakteereille ja n. 20 % arkeille. Koettimet suunniteltiin niin, että ensin etsittiin sekvenssistä paikka, jossa on poikkeava emäs eli erotteleva kohta. Tämän jälkeen suunniteltiin alustava versio d-koetimesta, jota varten valittiin pitkä alue sekvenssiä niin, että viimeisenä oli 3' päässä poikkeava emäs. Alustava c-koetin alkoi välittömästi erottelevan kohdan oikealta puolelta, josta valittiin koetinta varten pitkä alue sekvenssiä. Alustavia koettimia varten otettiin pidemmät alueet sekvenssiä kuin lopullisessa koettimessa tulisi olemaan koetinosaa sen vuoksi, että myöhemmin pystyttiin suunnittelemaan sulamislämpö oikeanlaiseksi.

Ennen koetinsuunnittelun aloittamista haettiin täysipitkät (n. 1500 bp) 16S rRNA-geenit NCBI:n tietokannasta niille koetinsuunnitteluun valituille sekvensseille, jotka oli tunnistettu lajiksi ja vastasivat sitä vähintään 94 prosenttisesti. 16S rRNA

-geenejä käytettiin koetinsuunnittelussa, koska ne olivat sekvensoinnista saatuja DNA-sekvenssejä pidempinä ja helpottivat näin suunnittelua. Muuten koetinsuunnittelussa hyödynnettiin pyrosekvensoinnista saatuja DNA-sekvenssejä. Lisäksi koetinsuunnitteluun valitut lyhyet DNA-sekvenssit linjattiin GreenGenesin NAST Alignment -ohjelmalla (DeSantis ym. 2006a) ja OTU:jen edustajat (bakteerit ja arkit, yhteensä 3403 DNA-sekvenssiä) linjattiin Biotekniikan Instituutin linjausohjelmalla. Linjatuista tiedostoista ja GreenGenesin (DeSantis ym. 2006b) sivustolta peräisin olevasta tietokannasta, jossa oli vähintään 1250 bp:n pituisia prokaryoottien 16S rRNA -geenejä, muodostettiin kaksi erilaista referenssitiedostoa. Ensimmäisessä referenssitiedostossa oli GreenGenesin tietokanta ja kokopitkät linjatut 16S-geenit, toisessa referenssitiedostossa oli linjatut OTU:ja edustavat sekvenssit. Koetinsuunnittelussa käytettiin R-ohjelmistoa ja NCBI:n sivustolla olevia PrimerBlast- sekä BLAST-ohjelmia.

Koetinsuunnittelu aloitettiin R-ohjelmistolla, jolla pyrittiin löytämään kaikille suunnitteluun valituille sekvensseille koetin. Koetinsuunnitteluun valittuja sekvenssejä verrattiin R-ohjelmistolla referenssitiedostoihin. Suunnittelun kohteena olleista sekvenssistä etsittiin yhden emäksen poikkeamia verrattuna referenssitiedostossa olleisiin sekvensseihin. Kun ensimmäinen poikkeava kohta löytyi, siihen suunniteltiin koetin. Jos yhden emäksen eroja ei löytynyt, niin koetin suunniteltiin sellaiseen kohtaan, jossa sama emäs oli mahdollisimman harvalla sekvenssillä. Joillekin koetinsuunnitteluun valituille sekvensseille ei löytynyt tällä tavalla koetinta tai koetin oli yhteinen liian monen sekvenssin kanssa, jolloin koetinta oli etsittävä muilla keinoin.

NCBI:n PrimerBlast-ohjelmaa käytettiin koetinsuunnittelussa, jos R:llä ei löytynyt sopivaa koetinta. PrimerBlast-ohjelmaan syötettiin sekvenssi, jolle suunniteltiin koetinta; sulamislämpötilaksi asetettiin min 50, max 60 ja max difference 10; PCR-tuotteen kooksi asetettiin vaihtelevia arvoja 80 - 1000; alukkeen täsmällisyyden arvot asetettiin niin, että neljästä viimeisestä emäksestä kahden tuli olla huteja. Ohjelman antamia alukkeita tarkasteltiin BLAST-ohjelmalla, ja jos BLAST-tulos antoi vähän täsmällisiä osumia varsinkin viimeisen nukleotidin osalta, aluke hyväksyttiin koettimeksi.

Viimeinen vaihtoehto koetinsuunnittelussa oli käyttää NCBI:n BLAST-ohjelmaa. Suunnitteluun valitulle sekvenssille tehtiin BLAST-haku ja tuloksen linjauksista etsittiin kohtia, joissa oli poikkeava nukleotidi tai aukko. Näistä kohdista otettiin lyhyt alue linjausta, joka tarkistettiin BLAST-ohjelmalla. Jos tuloksena oli linjauksia, joissa suurimmassa osassa viimeinen nukleotidi oli poikkeava, niin sekvenssi hyväksyttiin koettimeksi.

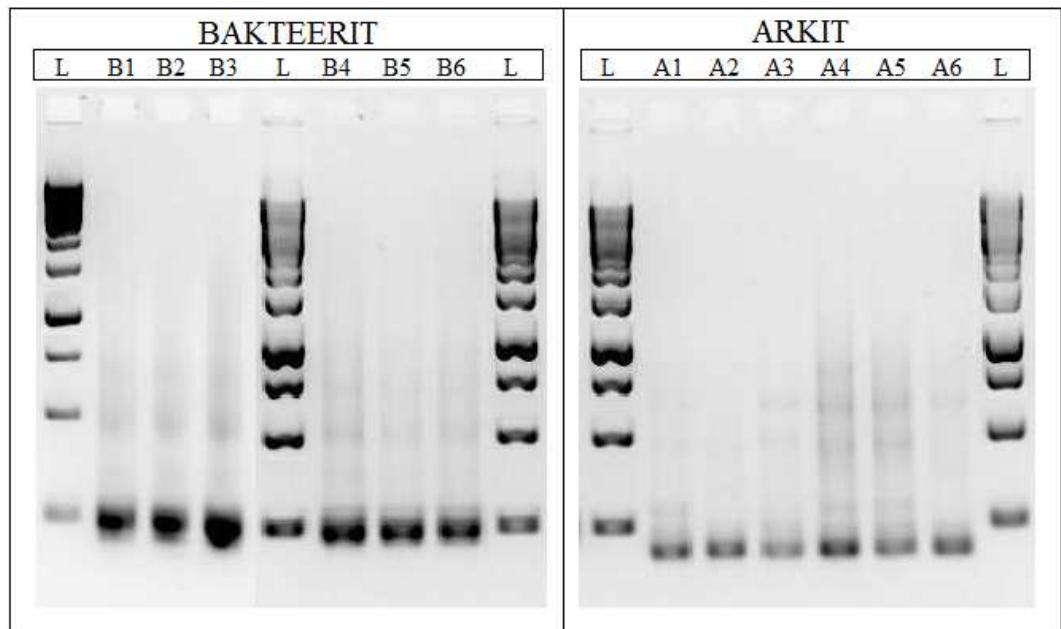
R-ohjelmistolla, PrimerBLAST:illa ja BLAST-ohjelmalla löydettyjen d-koettimien täsmällisyys tarkastettiin R:n avulla sedimentin mikrobistoa kuvaavaa referenssitietokantaa vastaan. Tiedostossa oli linjaamattomina GreenGenisin prokaryoottitietokanta, kaikki koetinsuunnitteluun valitut kokopitkät 16S-geenit ja arkkien sekä bakteerien OTU:jen edustajat, yhteensä 8379 sekvenssiä. Tarkastuksen tuloksista otettiin mukaan ne koettimet, joille tuli 1 - 4 väärää osumaa, riippuen siitä mitä osumat olivat. Lisäksi löydettyjen koettimen toimivuutta arvioitiin BLAST-haulla vertaamalla koettimien 20 viimeisen emäksen pituista aluetta NCBI:n nr/nt-tietokantaa vastaan. Jos vertailun tuloksena koetin sai erittäin paljon osumia sedimentissä esiintyviin mikrobeihin tai koettimen sukulaislajeihin, koetin hylättiin.

Viimeiseksi Biotekniikan Instituutissa valittuja koettimia muokattiin niin, että niiden sulamislämpötilaksi tuli d-koettimille 60 °C ja c-koettimille 64 °C. Seuraavaksi suunniteltiin kaksiosainen koetin, jonka 1-osa koostui c-koettimesta, zipkoodista ja vastaavan alukkeen puolikkaasta (VA1), 2-osa koostui toisesta vastaavan alukkeen puolikkaasta (VA2), universaalista alukkeesta (UA) ja erottelevasta koettimesta. Näistä saadaan yhtenäinen sirkulaarinen koetin, kun osat yhdistetään ligaatiolla vastaavaa aluketta vastaavan sekvenssin avulla.

4 TULOKSET JA TULOSTEN TARKASTELO

4.1 Laboratoriotyöt ennen sekvensointia

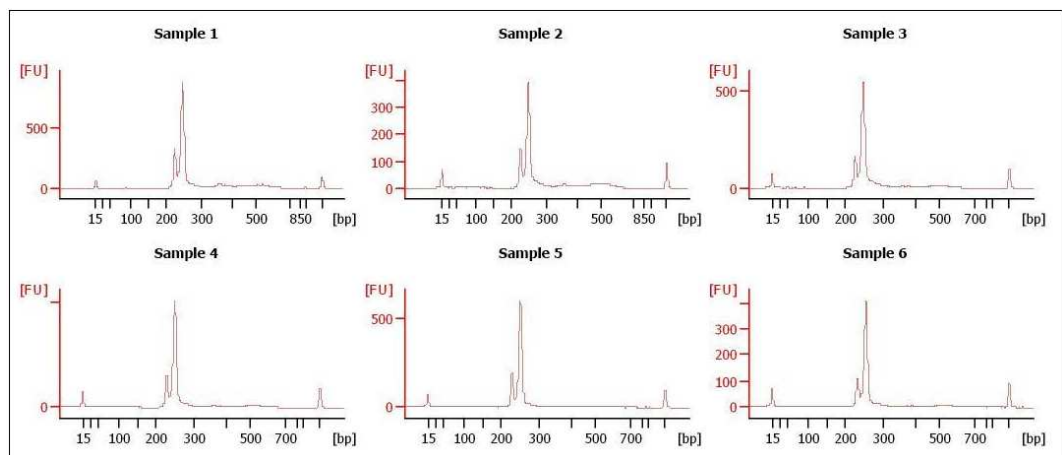
Ennen sekvensointia sedimentinäytteistä eristettiin DNA, jota monistettiin PCR:llä kahdessa vaiheessa. PCR-tuote puhdistettiin lyhyistä sivutuotteista kummankin PCR:n jälkeen. DNA:n eristyksen, PCR-monistuksen ja puhdistuksen onnistumien tarkistettiin elektroforeesigeelillä. Bakteerien DNA:n monistus pitkillä alukkeilla ja puhdistus onnistuivat hyvin elektroforeesikuvan perusteella (kuvio 10). Elektroforeesikuvasta nähdään, että PCR-tuote erottuu selvästi, eikä kuvassa ole havaittavissa jälkiä alukkeista tai PCR-tuotteen ”häntimisestä”. Arkkien DNA:n monistus pitkillä alukkeilla ja puhdistus ei onnistunut aivan yhtä hyvin kuin bakteereilla (kuvio 10). Elektroforeesikuvassa PCR-tuote erottu melko hyvin ja puhdistus on onnistunut alukkeiden osalta, mutta varsinkin näytteissä A4 ja A5 on havaittavissa ”häntimistä”, mikä olisi voinut häiritä pyrosekvensointia.



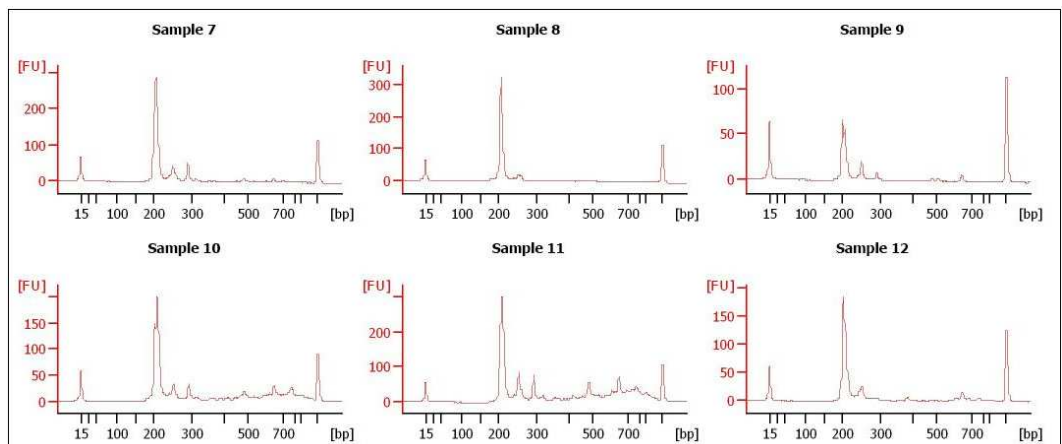
KUVIO 10. Elektroforeesikuvat bakteerein ja arkkien PCR:istä pitkillä alukkeilla ja puhdistuksen tulokset

PCR-tuotteen koko ja sen määrä tarkistettiin vielä Agilentin Bioanalyzer -laitteella. Bakteerien PCR onnistui hyvin, sillä 250 bp:n pituisen tuotteen kohdalla

oli suuri piikki eikä sivutuotteita ollut havaittavissa (kuvio 11). Arkkien PCR-tuote oli havaittavissa hiukan yli 200 bp:n kohdalla, mutta sen lisäksi joihinkin näytteisiin oli muodostunut myös pitkä sivutuotetta (kuvio 12). Eniten pitkiä sivutuotteita oli näytteissä A4 ja A5, mutta näytteet kelpasivat kuitenkin sekvensointiin (kuvio 12). PCR-tuotteen määrän riittävyys tarkistettiin mittaamalla DNA:n konsentraatit Qubit-fluorometrillä. Kaikkien näytteiden DNA-konsentraatit ylittivät tavoitearvon 10 µg/ml (taulukko 7). DNA:n konsentraatit oli oleellista tietää, jotta näytteet voitiin laimentaa oikein pyrosekvensointia varten.



KUVIO 11. Agilentin kuvat bakteerien PCR-tuotteesta



KUVIO 12. Agilentin kuvat arkkien PCR-tuotteesta

TAULUKKO 7. Bakteeri- ja arkinäytteiden DNA-pitoisuudet

NÄYTTEIDEN DNA-PITOISUUDET			
BAKTEERIT		ARKIT	
Näyte	DNA-pitoisuus (µg/ ml)	Näyte	DNA-pitoisuus (µg/ ml)
B1	29,2	A1	10,8
B2	17,3	A2	10,7
B3	21,8	A3	15,2
B4	29,6	A4	20
B5	12,9	A5	13,4
B6	14,2	A6	15,3

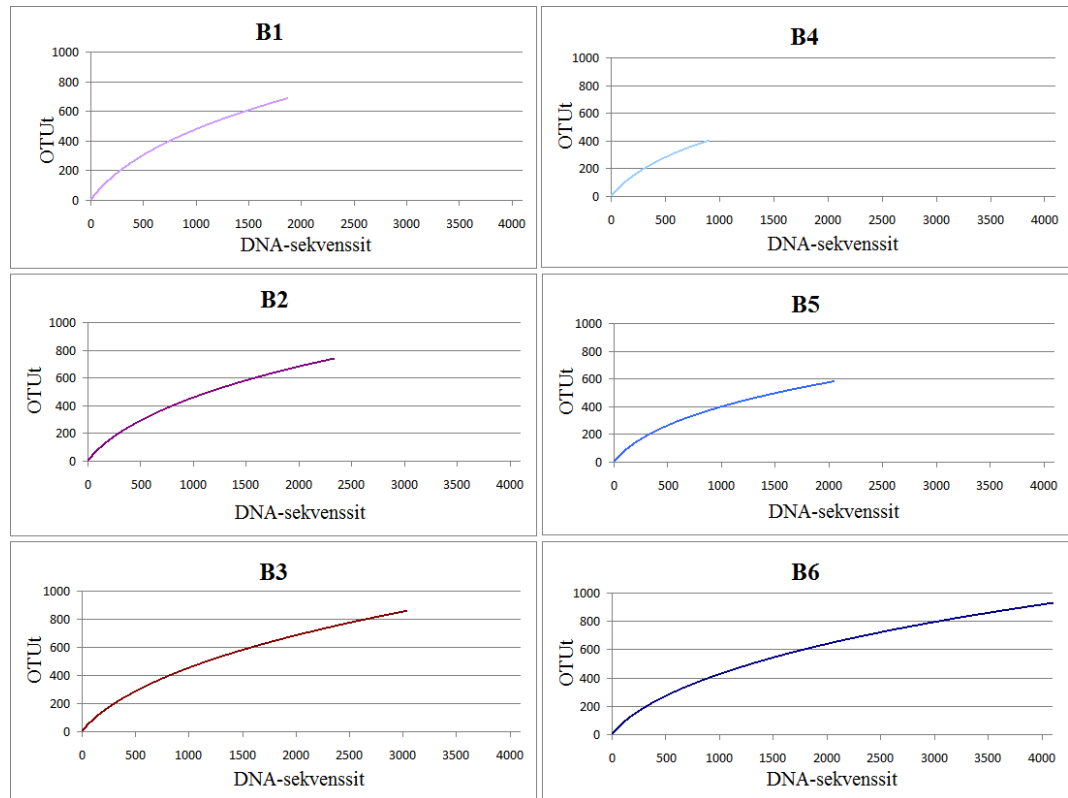
4.2 Sekvensoinnin tulokset

Huonolaatuisten ja lyhyiden DNA-sekvenssien poistamisen jälkeen pyrosekvensoinnista oli tuloksena yhteensä 14 292 keskimäärin 153 bp:n pituista (alukkeet poistettu) DNA-sekvenssiä kaikista bakteerinäytteistä ja vastaavasti 39 999 keskimäärin 109 bp:n pituista (alukkeet poistettu) DNA-sekvenssiä kaikista arkinäytteistä (taulukko 8). Bakteerinäytteiden DNA-sekvenssien määrät vaihtelivat 905:n (näyte B4) ja 4 103:n (näyte B6) välillä, arkeilla vastaavat luvut olivat 5 796 (näyte A4) ja 8 378 (näyte A1) (taulukko 8).

TAULUKKO 8. Pyrosekvensoinnista saatuja tuloksia näytteittäin: sekvensoinnista saatu DNA-sekvenssien määrä, OTU:jen määrä ja Chao1, joka arvioi OTU:jen todellista määrää sedimentissä (OTU:t yhteensä tarkoittaa erilaisten OTU:jen määrää)

Näyte	DNA-sekvenssit	OTU:t	Chao1
B1	1 871	687	1 310
B2	2 333	738	1 281
B3	3 035	857	1 526
B4	905	403	700
B5	2 045	580	1 055
B6	4 103	928	1 544
Bakteerit yhteensä	14 292	2 075	3 279
A1	8 378	692	1 005
A2	5 950	520	739
A3	6 237	562	907
A4	5 796	626	932
A5	7 055	648	947
A6	6 583	336	478
Arkit yhteensä	39 999	1 328	1 836

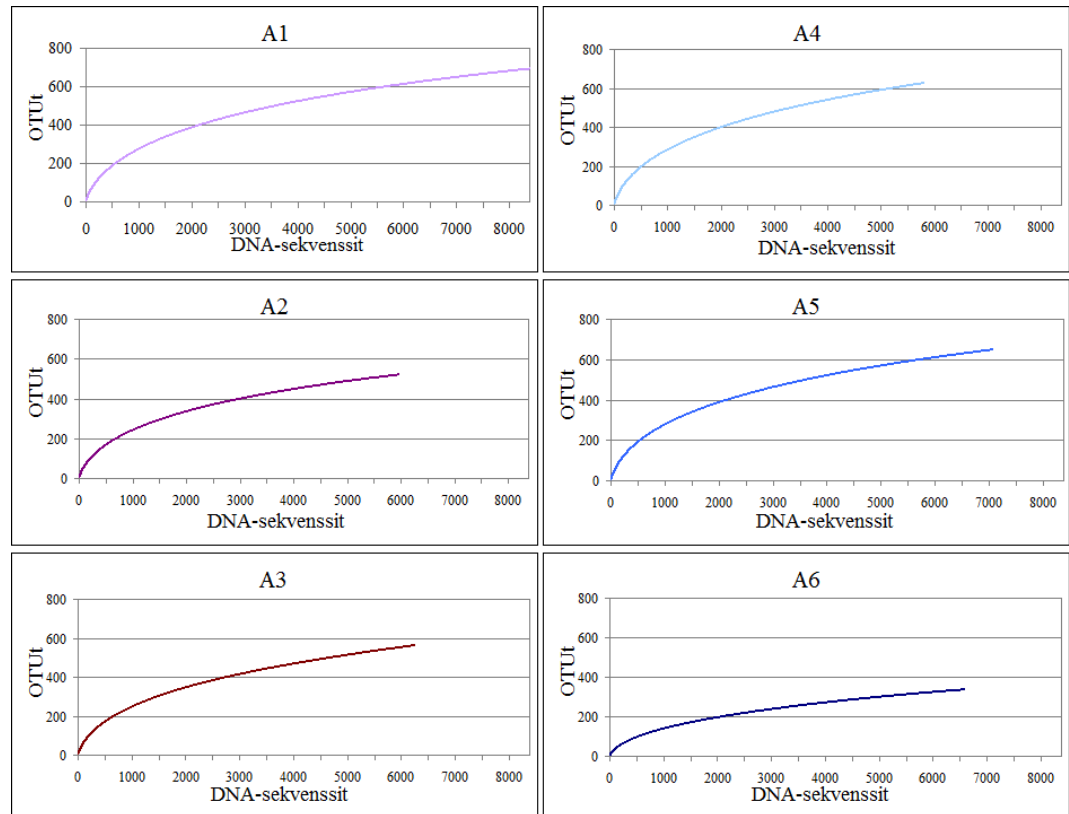
DNA-sekvenssit määritettiin OTU:iksi 97 %:n samankaltaisuusasteella, jota yleisesti pidetään lajirajana 16S rRNA -geenien yhteydessä. OTU:t määritettiin jokaiselle bakteeri- ja arkkinäytteelle erikseen. Lisäksi kaikkien bakteeri- ja arkkinäytteiden DNA-sekvenssit koottiin omiin tiedostoihin, joiden perusteella määritettiin, kuinka monta erilaista OTU:a näytteissä on yhteensä. OTU:jen todellista runsautta näytteissä arvioitiin Chao1-menetelmällä. Taulukossa 8 on OTU:jen määrät näytteittäin ja lisäksi yhdistettyjen bakteeri- ja arkkinäytteiden OTU:jen määrät. Yhteensä havaittuja bakteerien OTU:ja oli 2075 ja arvioitu OTU-runsaus Chao1:n mukaan oli 3279, vastaavat arvot arkeille olivat 1328 havaittua OTU:a ja 1836 OTU:a Chao1:n mukaan.



KUVIO 13. Kuvaajat, joiden perusteella voidaan päätellä kuinka suuri osuus bakteerinäytteiden OTU:ista on löydetty, ja kuinka suuri työ pitää vielä tehdä, jos halutaan löytää lisää OTU:ja

Pyrosekvensoinnista onnistumista voidaan vertailla monella tavalla. Esim. sekvensoinnista saatujen bakteeri- ja arkkinäytteiden DNA-sekvenssien välinen ero oli melko suuri. Jos verrataan kaikkien DNA-sekvenssien määrää, niin voidaan sanoa arkkien sekvensoinnin onnistuneen lähes kolme kertaa paremmin kuin bak-

teerien. Kuvioissa 13 ja 14 olevien kuvaajien avulla voidaan arvioida, kuinka suuri osuus OTU:ista on onnistuttu sekvensoimaan ja lisäksi sekvensoinnin onnistumista. Kuvaajissa on y-akselilla sekvensoinnista saatujen eri OTU:jen määrä ja x-akselilla sekvensoitujen DNA-sekvenssien määrä. Kuvaajan muodon perusteella voidaan päätellä, kuinka suuri osuus näytteen OTU:ista on onnistuttu sekvensoimaan. Jos kuvaaja nousee vielä jyrkästi, niin suuri osuus OTU:ista on vielä löytämättä. Jos kuvaaja puolestaan on alkanut tasaantua, niin suurin osa OTU:ista on löydetty. Kun kuvaaja saavuttaa vaakatason ja lähestyy niin sanottua asymptoottia, niin näytteestä on löydetty kaikki OTU:t. Kuvaajan perusteella voidaan arvioida, kuinka suuri työ OTU:jen sekvensoimiseksi on tehty, ja myös kuinka suuri työ pitää tehdä, jos halutaan löytää lisää OTU:ja. Kuviossa 13 on bakteerinäytteiden kuvaajat, joista mikään ei ole vielä alkanut tasaantua, mutta kaikki ovat kuitenkin ohittaneet jyrkimmän nousun. Kuvaajien perusteella voidaan päätellä, että suurin osa kaikkien paitsi B4 näytteen OTU:ista on onnistuttu sekvensoimaan. Kuviossa 14 on arkinäytteiden kuvaajat, joista kaikki ovat ohittaneet jyrkimmän nousun ja A6 näyte on alkanut selkeästi tasaantua. Tarkasteltaessa kaikkia näytteitä havaitaan, että suurin osa OTU:ista on sekvensoitu.



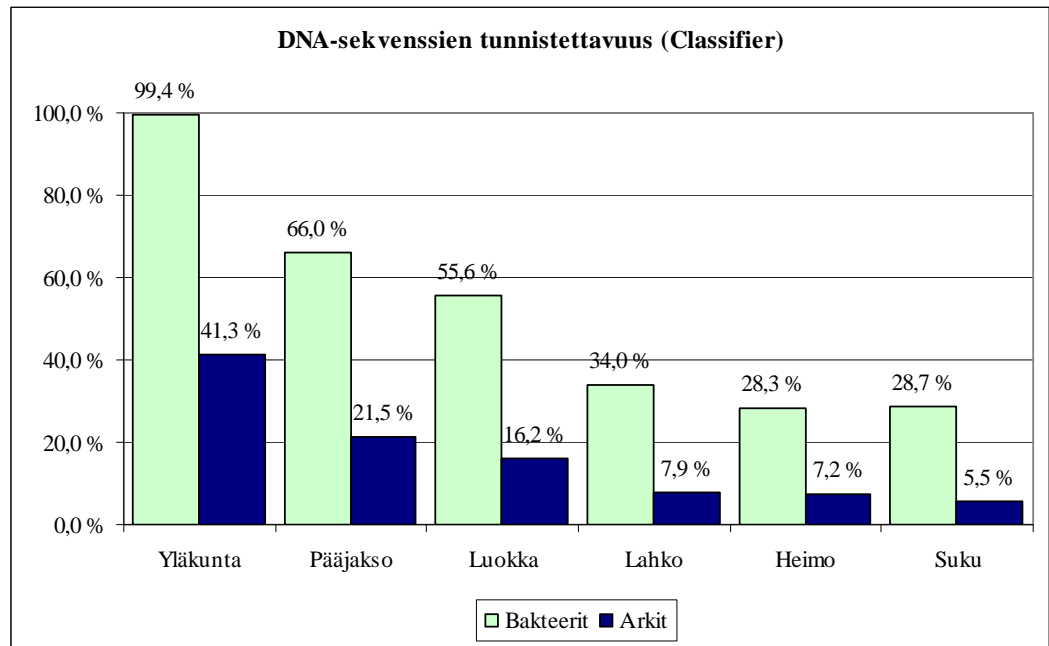
KUVIO 14. Kuvaajat, joiden perusteella voidaan päätellä kuinka suuri osuus arkinäytteiden OTU:ista on löydetty, ja kuinka suuri työ pitää vielä tehdä, jos halutaan löytää lisää OTU:ja

4.3 Sekvensointiaineiston sisältö

Sedimentin mikrobiston tunnistamiseen käytettiin Classifier- ja MEGAN-ohjelmia. Classifierilla voidaan tunnistaa mikrobit sukutasolle ja MEGANilla lajitasolle asti. Classifierissa käytettiin luotettavuusrajana 50 %:a, joka antaa lähes yhtä luotettavan tuloksen kuin 80 %:n luotettavuusraja ja lisäksi pystyy määrittämään taksonin huomattavasti suuremmalle osuudelle DNA-sekvenssejä. MEGANin algoritmin arvot (taulukko 6) määritettiin, niin että koetinsuunnittelun kannalta vain merkittävät lajit otettiin huomioon, minkä vuoksi suuri osa huonosti linjatuista tuloksista jäi tarkastelun ulkopuolelle. Classifierilla oli tarkoituksena muodostaa yleiskuva näytteiden mikrobistosta ja MEGANia käyttää lajitason tarkasteluun sekä taksoneitten valintaan koetinsuunnittelua varten. Erilaisten raja-arvojen valinnan vuoksi Classifierin ja MEGANin tuloksia ei voida verrata suoraan toisiinsa.

Sekvensoinnista saatujen DNA-sekvenssien tunnistettavuus vaihteli paljon bakteerien ja arkkien sekä eri taksonomisten tasojen välillä. Classifierilla bakteerinäytteiden DNA-sekvensseistä oli tunnistettavissa bakteereiksi 99,4 % ja arkinäytteiden DNA-sekvensseistä arkeiksi 41,3 % (kuvio 15). MEGANilla vastaavat arvot olivat bakteereille 32,6 % ja arkeille 28,6 %. Kun tarkasteltiin DNA-sekvenssien tunnistettavuutta eri taksonomisilla tasoilla, niiden tunnistettavuus laski samankaltaisesti sekä Classifierilla että MEGANilla siirryttäessä kohti pienempiä tasoja. DNA-sekvenssien tunnistettavuus heikkeni siirryttäessä kohti sukutasoa niin, että bakteerien DNA-sekvensseistä oli tunnistettavissa sukutasolle enää 28,7 % ja arkkien DNA-sekvensseistä vain 5,5 % (kuvio 15). MEGANilla bakteerien DNA-sekvensseistä oli tunnistettavissa lajitasolle 6 % ja arkkien DNA-sekvenssistä 17 %. Arkkien parempi tunnistettavuus lajitasolla selittyy sillä, että arkeilla oli bakteereihin verrattuna muutamia huomattavan suuria OTU:ja (sisälsi paljon DNA-sekvenssejä), joista osan MEGAN kykeni tunnistamaan.

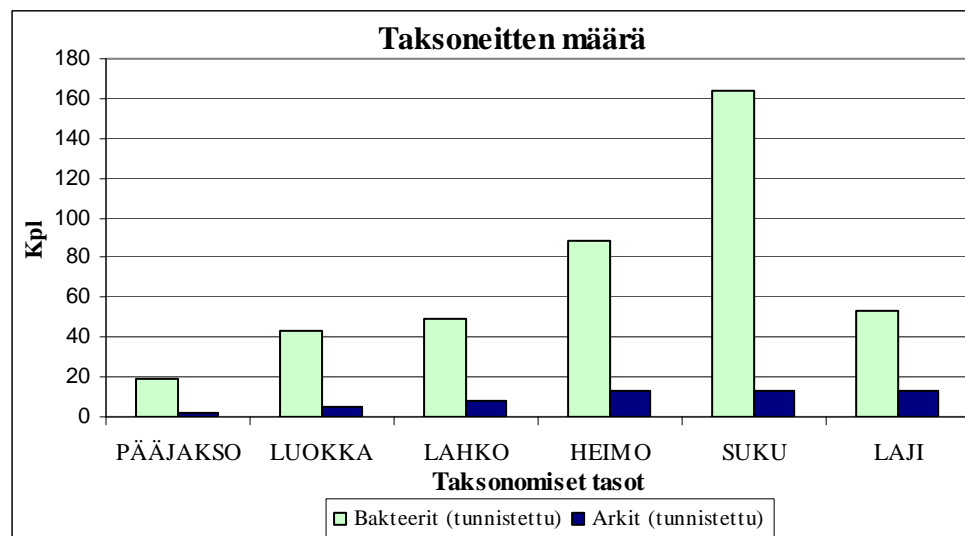
Tulosten perusteella oli havaittavissa, että bakteerit olivat selvästi paremmin tunnistettavissa kuin arkit, mikä ei ole yllättävää, kun otetaan huomioon, että bakteereja on tutkittu huomattavasti enemmän kuin arkkeja. Tutkimuksen keskittyminen bakteereihin voidaan havaita mm. vertaamalla luonnossa esiintyvien mikrobien suhteellisia osuuksia tietokantojen sisältöön. On arvioitu, että n. 25 % kaikista mikrobeista on arkkeja (Ettema 2005, 2), mutta esim. RDP:n tietokannassa olevista 16S rRNA -sekvensseistä 95 % (643 916) on bakteerien sekvenssejä ja vain 5 % (33 082) arkkien sekvenssejä (Cole ym. 2008, 141). Tuloksista havaitaan lisäksi, että DNA-sekvenssien tunnistettavuus suku- ja lajitasolla oli melko huono. Tunnistettavuuteen vaikuttaa jo pelkästään tarkastelun kohteena oleva ympäristö. Liu ym. (2008) ovat tutkineet kolmesta erilaisesta ympäristöstä peräisin olevien 16S rRNA -geenien V1 - V9 alueiden 100 - 400 bp:n pituisten DNA-sekvenssien tunnistettavuutta sukutasolla mm. RDP:n Classifierilla (50 %:n luotettavuusrajalalla) (Liu, DeSantis, Andersen, & Knight 2008). V3-alueen 250 bp:n pituisten DNA-sekvenssien tunnistettavuuden tulokset vaihtelivat lähes 90 %:n tunnistettavuudesta reiluun 30 %:iin, niin että parhaiten oli tunnistettavissa ihmisen ruuansulatusjärjestelmän bakteerit, toiseksi parhaiten hypersuolaisen mikrobimaton bakteerit ja heikoiten hiiren ruuansulatusjärjestelmän bakteerit (Liu ym. 2008, 120).



KUVIO 15. Bakteerien ja arkkien DNA-sekvenssien tunnistettavuus Classifierilla eri taksonomisilla tasoilla

Bakteerien ja arkkien erilaisten tunnistettavien taksonoiden määrät sekvensointiaineostossa määritettiin erikseen jokaiselle taksonomiselle tasolle pääjaksosta lajiin. Taksonoiden määrät tasoilla pääjaksosta sukuun määritettiin Classifierilla ja lajien määrät MEGANilla. Bakteerien taksonoiden määrät kasvoivat eksponentiaalisesti siirryttäessä pääjaksosta sukuun, niin että erilaisia sukuja oli tunnistettavissa 164 (kuvio 16). Siirryttäessä lajitasolle tunnistettavuus romahti ja eri lajeja oli havaittavissa bakteereille vain 53 kpl (kuvio 16). Arkkien taksonoitteita oli tunnistettavissa kaiken kaikkiaan erittäin vähän, pääjaksotasolle kaksi, sukuja 13 ja saman verran lajeja (kuvio 16). Kun verrataan kuvioita 16 ja 15, niin voidaan päätellä, että suurinta osaa taksonoista ei ole kyetty tunnistamaan. Vaikka taksonoiden määrä kasvaa siirryttäessä kohti pienempiä tasoja, niin samaan aikaan DNA-sekvenssien tunnistettavuus heikkenee. Sama ilmiö on havaittavissa, kun verrataan OTU:jen määrää havaittuihin lajeihin ja myös muiden tasojen taksonoiden määriin. OTU:ja oli bakteereille 2075 ja arkeille 1328. Kun näitä lukuja verrataan tunnistettujen lajien määrään, niin lajeja pystyttiin tunnistamaan vain murto-osa siitä, mitä OTU:jen määrän perusteella sekvensointiaineistossa oli.

Tarkasteltaessa taksoneiden määriä havaitaan, että bakteerien ja arkkien välillä on huomattava ero, joka selittyy osittain bakteerien paremmalla tunnistettavuudella, muttei kokonaan. Todellista eroa voidaan arvioida vertaamalla Chao1-menetelmällä saatuja arviota OTU:jen todellisesta määrästä sedimentissä. Chao1:n mukaan bakteeri-OTU:ja oli 3279 ja arkki-OTU:ja 1836, minkä perusteella bakteerien lajeja olisi sedimentissä 78,6 % enemmän (tutkimuksen aikana 8.10.2009 - 13.1.2010).

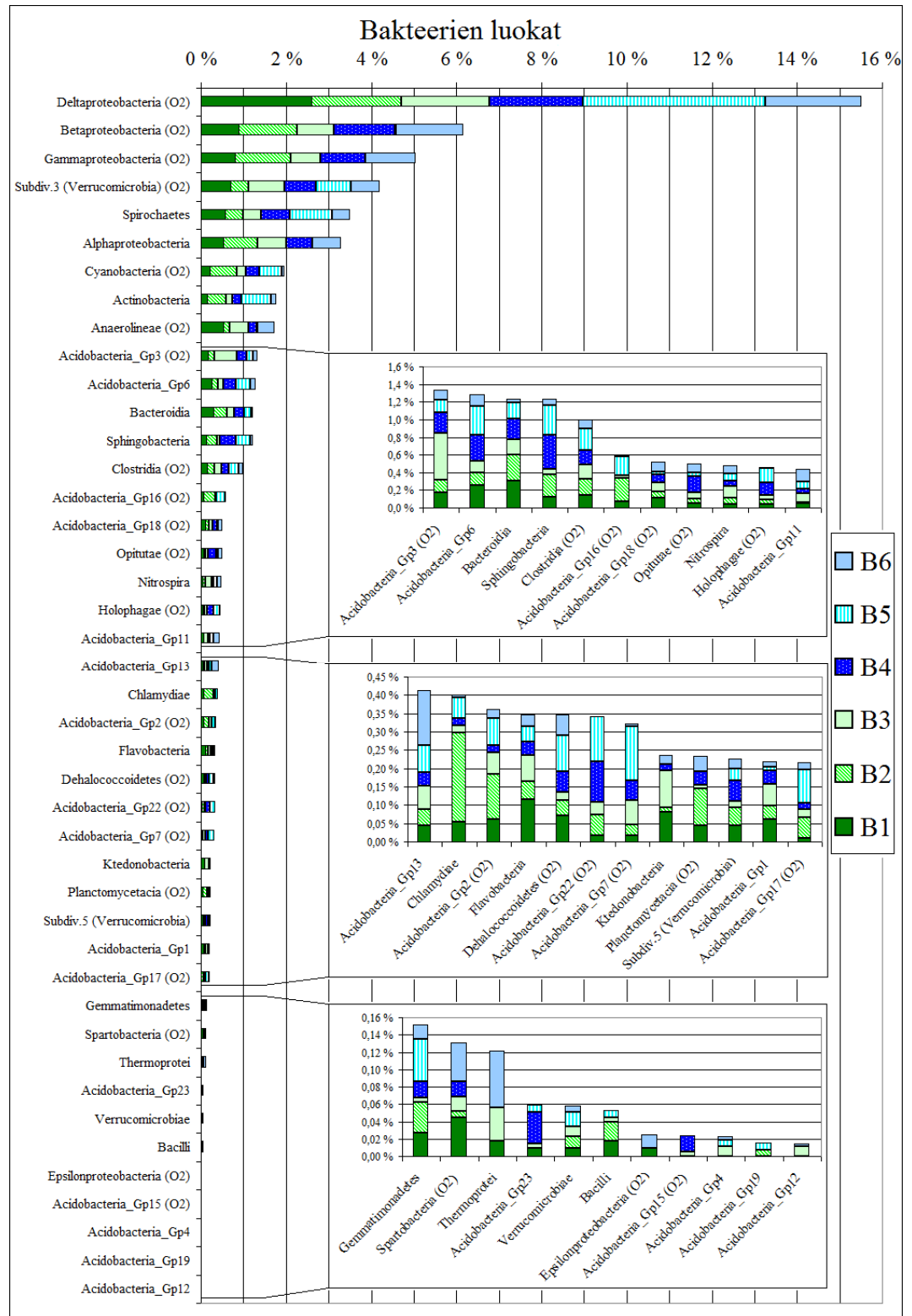


KUVIO 16. Bakteerien ja arkkien erilaisten taksoneiden määrät pääjaksotasolta lajitasolle

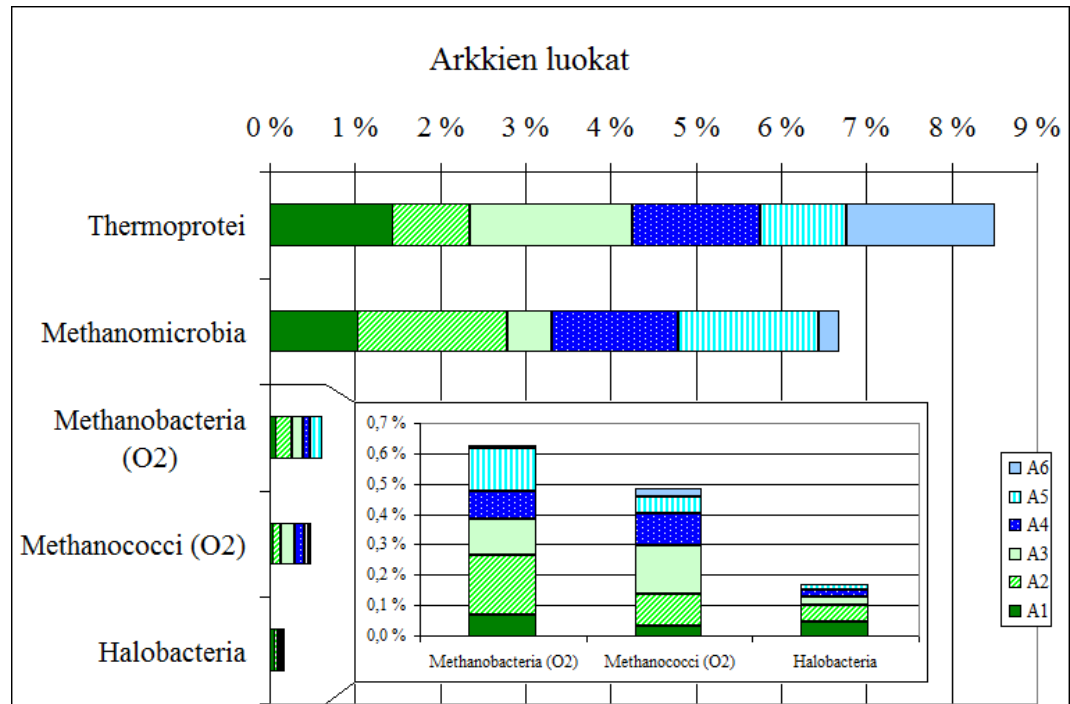
Tarkasteltaessa DNA-sekvenssejä luokkatasolla bakteerien sekvensseistä oli tunnistettavissa 56 %. Kuviossa 17 on esitetty Classifierilla tunnistettujen bakteeriluokkien suhteelliset osuudet kaikista DNA-sekvenssistä ja näytteiden osuudet luokasta. Näytteenottojen välisenä aikana deltaproteobacter oli selvästi bakteerien suurin luokka 15,5 %:n osuudella. Muita merkittäviä luokkia suuruusjärjestyksessä olivat betaproteobacteria (6,2 %), gammaproteobacteria (5,0 %), subdivision3, joka kuului verrucomicrobiakaareen (4,2 %), spirochaetes (3,5 %) ja alphaproteobacteria (3,3 %). Lisäksi acidobacteriakaaren luokat ovat huomionarvoisia, sillä niitä oli yhteensä 17 ja pääjaksotasolla acidobacteria oli kolmanneksi suurin ryhmä.

Tuloksista huomataan, että luokkien koko pienenee nopeasti ja kymmeneen ensimmäiseen luokkaan sisältyy jo lähes 79 % tunnistetuista sekvensseistä. Vastavasti 43 luokasta 26 pienintä sisältää vain n. 10 % kaikista tunnistetuista sekvensseistä. Sedimentin käsittelemisellä happea vapauttavalla kemikaalilla näytti olevan vaikutusta selvästi 18 luokkaan. Suhteellisesti suurin vaikutus käsittelyllä oli luokkiin acidobacteria Gp3, anaerolineae ja cyanobacteria. Suurista luokista käsittelyllä oli vaikutusta muutoksen suuruuden mukaan seuraaviin: subdivision3, deltaproteobacteria, gammaproteobacteria ja betaproteobacteria.

Arkkien DNA-sekvensseistä luokkatasolle oli tunnistettavissa 16 %. Kuviossa 18 on kuviota 17 vastaavat tiedot arkeista. Arkeille oli tunnistettavissa vain 5 luokkaa, jotka suuruusjärjestyksessä olivat thermoprotei (8,5 %), methanomicrobia (6,7 %), methanobacteria (0,6 %), methanococci (0,5 %) ja halobacteria (0,2 %) (Kuvio 18). Happikäsittelyn vaikutus on nähtävissä selvimmin luokissa methanobacteria ja methanococci. Arkkiyhteisön tarkastelu on vaikeaa heikon tunnistavuuden vuoksi jopa luokkatasolla, eikä esim. hallitsevia luokkia ole mahdollista määritellä luotettavasti.



KUVIO 17. Bakteeriluokkien suhteellinen osuus DNA-kirjastoissa ja näytteiden osuudet luokasta (luokat, joihin sedimentin käsittelyllä on ollut vaikutusta, on merkitty O2:lla)



KUVIO 18. Arkkiluokkien suhteellinen osuus sekä näytteiden osuudet luokasta

4.4 Koettimet

Koetinsuunnitteluun valittiin 138 DNA-sekvenssiä, joista 116 oli bakteereille ja 22 arkeille. Koetinsuunnitteluun valittiin käytännössä kaikki mahdolliset lajeja edustavat DNA-sekvenssit, jotka olivat tunnistettavissa riittävän tarkasti. Koetinsuunnittelun pohjana käytetyt sekvenssit oli määritetty OTU:iksi 97 %:n samankaltaisuudella, mikä voi käsittää lajiksi, eli käytännössä koettimet edustavat lajeja, jotka ovat tunnistettavissa eri taksonomisille tasoille. Koettimien tunnistettavuus vaihteli laji- ja lahkotason välillä. Bakteereille suunniteltiin 73 koetinta, joista 32 oli tunnistettavissa lajitasolle, 36 sukutasolle, 4 heimotasolle ja yksi lahkotasolle (taulukko 9). Bakteerien koettimista 15:llä oli mahdollista seurata käsittelyn vaikutusta sedimentin mikrobistoon. Käsittelyä indikoivista koettimista 14 oli tunnistettavaksi lajiksi ja 1 suvuksi (taulukko 9). Arkkikoettimia suunniteltiin yhteensä 14, joista 11 oli tunnistettavissa lajitasolle ja 3 sukutasolle (taulukko 9). Arkkien koettimista mahdollisesti yksi, lajitasolle tunnistettava koetin, indikoi käsittelyn vaikutusta. Loput arkkikoettimet ilmensivät pelkästään monimuotoisuutta (taulukko 9). Koettimien taksonomiset määritelmät ovat liitteissä 1 - 5 ja koettimien d-osan, c-osan ja zip-koodin sekvenssit ovat liitteessä 6.

TAULUKKO 9. Bakteerien ja arkkien koettimien määrät eri kategorioissa: kaikki koettimet, koettimien tunnistettavuus, käsittelyä indikoivien koettimien määrät ja monimuotoisuutta indikoivien koettimien määrät

	Kaikki	Tunnistettavuus				Muutosta indikoivat		Monimuotoisuus	
		Laji	Suku	Heimo	Lahko	Laji	Suku	Laji	Suku
Bakteerit	73	32	36	4	1	14	3	18	33
Arkit	14	11	3	0	0	1	0	10	3
Yht	87	43	39	4	1	15	3	28	36

Koettimien laatua arvioitiin tutkimalla kuinka hyvin koettimet tunnistavat oikean kohteen. Kohteen tunnistamista tutkittiin vertaamalla koettimien d-osan sekvenssien 20 viimeistä nukleotidia BLAST-haulla NCBI:n nr/ nt-tietokantaan ja R:llä sedimentin mikrobistoa kuvaavaan referenssitietokantaan (tietokannan sisältö luvussa 4.2.5). NCBI:n tietokantaan on valtavan paljon suurempi kuin sedimentin mikrobistoa kuvaava referenssitietokanta oli. NCBI:n tietokantaan sisältyy miljoonia eripituisia sekvenssejä (mm. kaikki kokonaiset ja osittaiset 16S rRNA - geenit), kun taas sedimentin mikrobistoa kuvaavaan tietokantaan kuului vain 8 379 sekvenssiä. Verrattaessa koettimia NCBI:n tietokantaan tulokset vaihtelivat paljon ja olivat vain suuntaa antavia. Tärkeämpiä olivat tulokset koettimien vastaavuudesta sedimentin mikrobistoa kuvaavaan tietokantaan. Koettimien laatua arvioitiin ensisijaisesti sen perusteella, kuinka montaa sekvenssiä koettimet vastasivat ja minkälaisia vastaavuudet olivat. Erinomaiset koettimet vastasivat vain yhtä sekvenssiä sedimentin mikrobistoa kuvaavassa tietokannassa. Hyvät vastasivat korkeintaan kahta sekvenssiä, joista toisen piti kuulua GG:n 16S rRNA - tietokantaan ja lisäksi koettimille löytyi NCBI:n tietokannasta yli 10 osumaa. Kohtalaisille koettimille löytyi sedimentin mikrobistoa kuvaavasta tietokannasta oman sekvenssin lisäksi 2 - 3 osumaa, jotka olivat pieniä OTU:ja, tunnistettavissa samaksi lajiksi tai osumat kuuluivat GG:n 16S rRNA -tietokantaan. Huonot koettimet vastasivat oman sekvenssin lisäksi vähintään yhtä suurta OTU:a, minkä vuoksi koettimet saattavat antaa todennäköisesti väärän signaalin. Erinomaisia koettimia oli 52, hyviä 25, kohtalaisia 8 ja huonoja 2 (taulukko 10).

TAULUKKO 10. Koettimien laadun arviointi asteikolla erinomainen – huono. Taulukossa on koettimien määrät laadun mukaan ja kuinka monta vastaavuutta koettimille on tietokannoista löytynyt laatuluokassaan

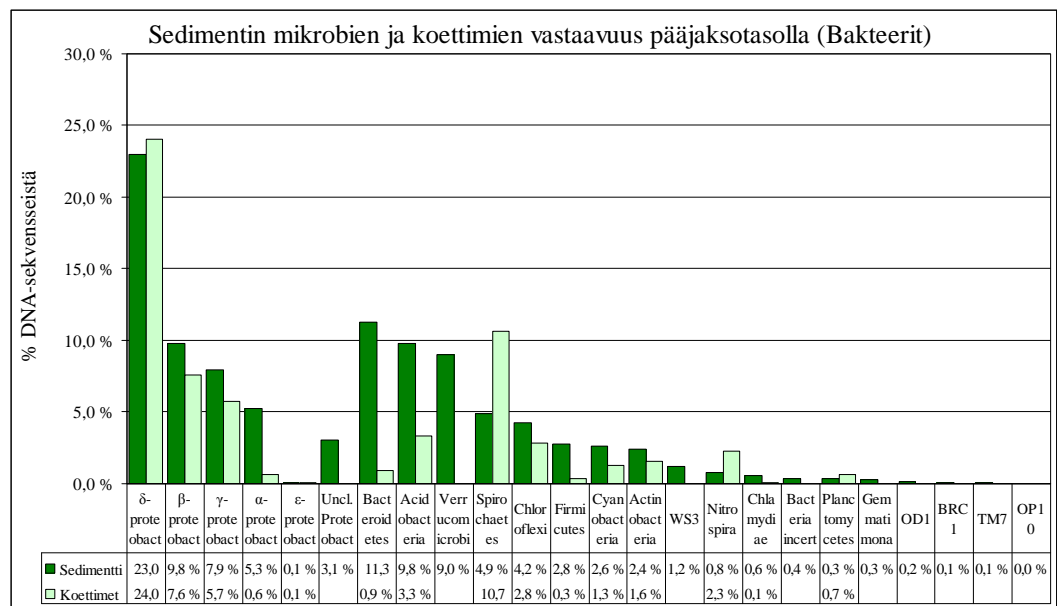
Laatu	Koettimien määrä	Sedimentti	NCBI
Erinomaisia	52	1	1 - 10
Hyviä	25	1 - 2	11 - 100
Kohtalaisia	8	2 - 4	0 - 100
Huonoja	2	2 - 4	4 - 9

4.5 Koettimet ja sedimentti

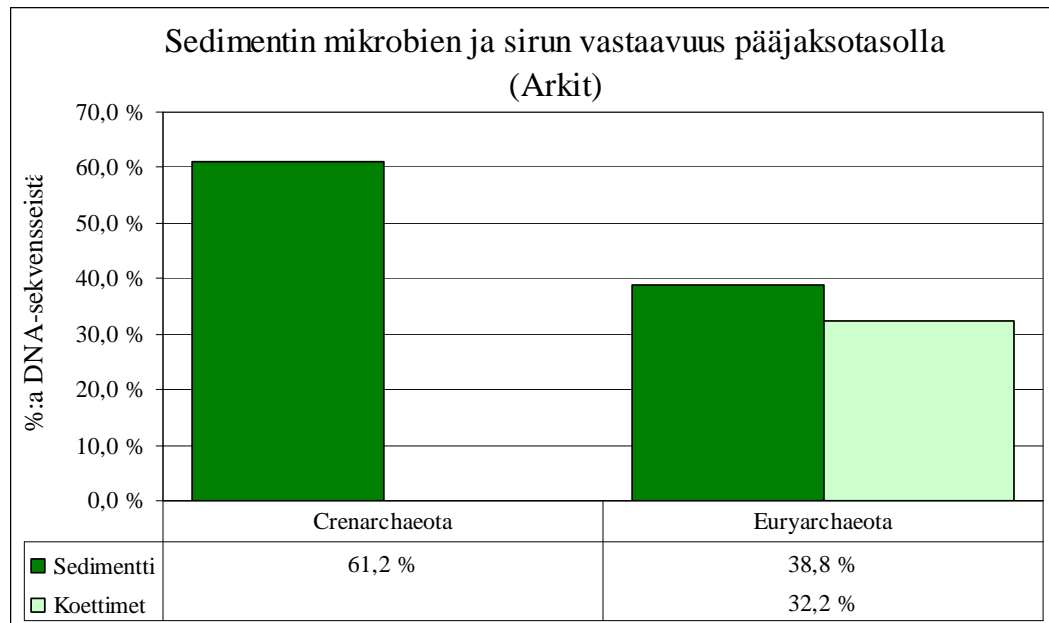
Koettimien ja sedimentin mikrobiston vastaavuuden tarkastelemiseksi määritettiin pääjaksojen suhteelliset osuudet tunnistetuista DNA-sekvensseistä sekä sedimentin mikrobeille että koettimille, minkä jälkeen niiden pääjaksoja verrattiin keskenään (kuviot 19 ja 20). Sedimentin mikrobiston pääjaksojen suhteelliset osuudet määritettiin vertaamalla DNA-sekvenssien määriä kaikkiin pääjaksotason tunnistettuihin DNA-sekvensseihin, joita oli bakteereilla 9430 ja arkeilla 8601. Koettimien pääjaksojen suhteellisten osuuksien määrittäminen tehtiin samalla periaatteella. Koettimien pääjaksoihin kuuluneiden DNA-sekvenssien määrät määritettiin koettimia vastaavien OTU:jen kokojen perusteella. Eli aluksi selvitettiin, mihin pääjaksoihin koettimet kuuluivat, minkä jälkeen OTU:jen kokojen perusteella määritettiin pääjaksoon kuuluvien DNA-sekvenssiä määrää. Tämän jälkeen koettimien pääjaksojen kokoa verrattiin kaikkiin koettimia edustaviin DNA-sekvensseihin, joita oli bakteereilla 3205 ja arkeilla 3193. Tarkasteltaessa sedimentin mikrobiston DNA-sekvenssien jakautumista pääjaksoihin on huomioitava, että luvut edustavat vain keskimääräisiä arvoja ja pääjaksojen suhteelliset osuudet vaihtelevat näytteittäin ajankohdan mukaan.

Vertailtaessa bakteereja selvästi suurin pääjakso oli proteobacteria 49 %:n osuudella sedimentissä ja 38 %:n osuudella koettimista. Suuren koon vuoksi se jaettiin luokkiin vertailun helpottamiseksi. Sedimentin mikrobiston ja koettimien vertailu osoitti, että muutama taksoni sedimentissä ja koettimien välillä oli melko samankokoisia (δ -, β - ja γ -proteobacteria), mutta enimmäkseen pääjaksot eivät vastan-

neet toisiaan. Lisäksi 8 pääjaksoa (tyhjät kohdat koettimien rivissä kuviossa 19) ei ollut edustettuna ollenkaan koettimissa (kuvio 19), joten bakteerien koettimien voidaan sanoa edustavan melko huonosti sedimentin mikrobeita pääjaksotasolla. Arkeille oli tunnistettavissa kaikkiaan kaksi pääjaksoa, joista vain toinen edusti koettimia (kuvio 20). Arkkikoettimien sedimentin mikrobiston vastaavuus oli samankaltainen bakteerien kanssa eli huono. Arkkien tulokset ovat parhaimmillaan vain suuntaa antavia, sillä arkkien DNA-sekvensseistä vain 22 % oli tunnistettavissa pääjaksotasolle (kuvio 15).

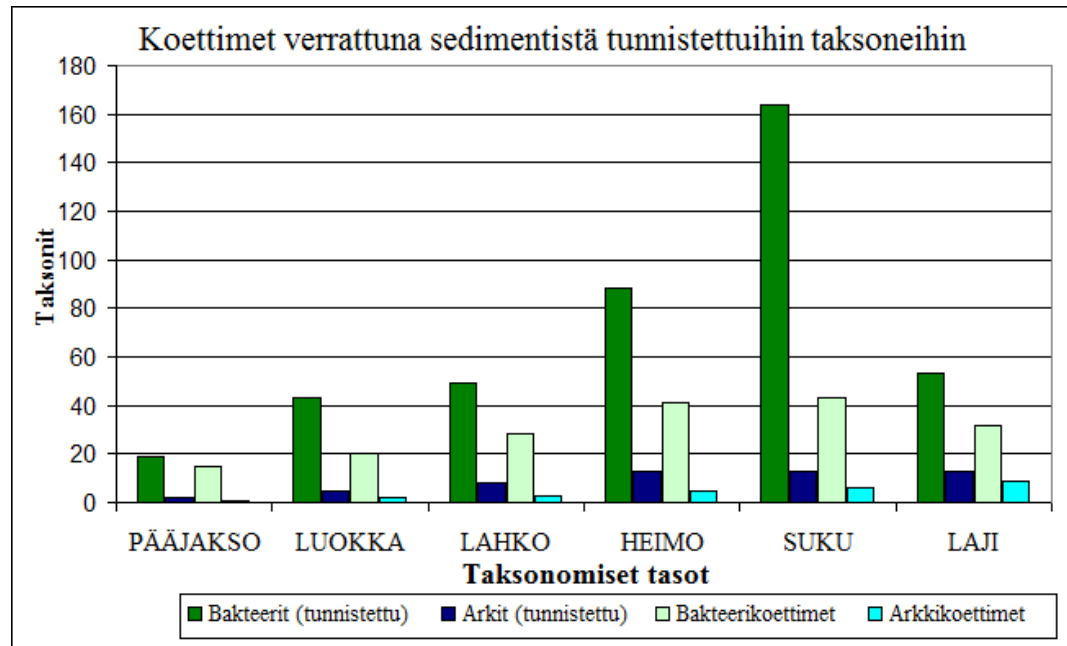


KUVIO 19. Bakterikoettimien ja sedimentin mikrobiston vastaavuuden tarkastelu pääjaksotasolla (alussa proteobacteria pääjakso on jaettu luokkiin)



KUVIO 20. Arkkikoettimien ja sedimentin mikrobiston vastaavuuden tarkastelu pääjaksotasolla

Koettimien kattavuutta sedimentin mikrobistossa tarkasteltiin vertailemalla koettimien ja sedimentistä tunnistettujen taksoneiden määrää keskenään (kuvio 21). Vertailun perusteella havaitaan, että koettimien ja sedimentistä tunnistettujen taksoneiden ero kasvaa nopeasti bakteerien kohdalla, ja arkeilla ero pysyy melko pienenä (kuvio 21). Kun lisäksi verrataan koettimien määrää (87) havaittujen OTU:jen määrään (2075 bakteerit ja 1328 arkit), huomataan, että koettimet kattavat sedimentin mikrobistosta murto-osan. Koska koettimet edustavat vain pientä osuutta kaikista lajeista, niin tutkittaessa jotain näytettä, voi iso osuus näytteen lajeista jäädä otannan ulkopuolelle, mikä saattaa vaikuttaa tulokseen merkittävästi verrattaessa koettimien ja sedimentin mikrobien välistä vastaavuutta. Monimuotoisuuden seuraamisen kannalta, ideaalissa tilanteessa, päällekkäisyyksiä on mahdollisemman vähän. Suunnitelluista koettimista sukutasolla 14 oli tunnistettavissa samaksi suvuksi. Suunniteltujen koettimien osalta päällekkäisyyksien välttäminen oli kuitenkin vaikeaa, sillä koetinsuunnitteluun valituissa DNA-sekvensseissä ei ollut paljoa valinnanvaraa, ja lisäksi koettimia ei saatu suunniteltua tavoitemäärää.



KUVIO 21. Koettimien vertaus sedimentistä tunnistettuihin taksoneihin

5 JOHTOPÄÄTÖKSET

Tavoitteena oli suunnitella 100 koetinta, jotka olivat tunnistettavissa laji- tai sukutasolle. Koettimilla oli tarkoitus seurata happikäsittelyn vaikutusta sedimentin mikrobeihin ja mikrobiston monimuotoisuutta. Koettimia saatiin suunniteltua 87, joista 43 oli tunnistettavissa lajitasolle ja 18:lla oli mahdollista seurata käsittelyn vaikutusta.

Lajitasolle tunnistettavien koettimien suunnittelu osoittautui haasteelliseksi. Usein DNA-sekvenssit oli tunnistettavissa vähintään kahdeksi eri lajiksi, samalla tai lähes samalla luotettavuusarvolla, minkä vuoksi DNA-sekvenssien tarkka määrittäminen ei onnistunut. Tähän vaikutti mm. DNA-sekvenssien pituus, joka oli bakteereilla keskimäärin 153 bp ja arkeilla 109 bp, kun koko 16S rRNA -geeni on n. 1500 bp:n pituinen. Lyhyet DNA-sekvenssit eivät yksinkertaisesti sisältäneet riittävästi DNA-tietoa, joka olisi riittänyt lajitason tunnistettavuuteen. Sekvensointitekniikoiden kehittyessä tästä ongelmasta on kuitenkin mahdollista päästä eroon, ainakin osittain. Esim. jo tällä hetkellä on uusi pyrosekvensointimenetelmä käytössä, jolla voidaan sekvensoida keskimäärin 400 bp:n DNA-sekvenssejä (454 Life Sciences, a Roche company 2010).

DNA-sekvenssien tunnistettavuuteen vaikuttaa myös tietokantoihin tallennetun tiedon määrä. Vaikka tietokantoihin on tallennettu DNA-sekvenssejä valtavia määriä, niin siitä huolimatta vasta pientä osaa mikrobeista on tutkittu, ja harvan lajin koko 16S rRNA -geeni on tallennettuna tietokantoihin. Mikrobien tunnistettavuus tulee kuitenkin parantumaan, kun DNA-tiedon määrä kasvaa tietokannoissa.

Kokonaisuudessaan tavoitteesta jäätiin 13 koettimen päähän. Tosin tätä ei voi pitää kovin merkittävän lukuna, koska ennen koetinsuunnittelua ei tiedetty, kuinka monta koetinta on mahdollista suunnitella. 100 koetinta oli vain ennen suunnittelua asetettu summittainen tavoite. Suunnitelluilla koettimilla käsittelyn vaikutuksen seuraaminen pitäisi onnistua kohtuullisesti. Koettimien ja sedimentin mikrobien vastaavuus ei ollut paras mahdollinen. Kun otetaan huomioon, että koettimen

suunnittelu oli melko haasteellista, niin lopputulos on kuitenkin hyvä, vaikkei aivan ideaali olekaan.

Sirutekniikalle on myös vaihtoehtoja, joilla voidaan tunnistaa mikrobeja ympäristönäytteistä. Yleisimmin käytetty menetelmä on ollut 16S rDNA:n monistaminen, kloonaukset, DGGE (Denaturing gradient gel electrophoresis) ja Sanger-sekvensointi. Menetelmällä saadaan sekvensoitua koko 16S rRNA -geenin alue. Verrattaessa menetelmää pyrosekvensointiin se on huomattavasti hitaampi, kymmenen kertaa kalliimpi per emäs (Liu, Lozupone, Hamady, Bushman & Knight 2007, 7) ja menetelmällä on työlästä sekvensoida kattavasti tutkimuksen kohteena olevaa näytettä. Vaikka pyrosekvensoinnista saadut DNA-sekvenssit ovat lyhyitä, niin vastapainona on huomattavasti kattavampi sekvensointi, minkä myötä mikrobipopulaatiosta saadaan tarkempi kuva. Laajamittaiseen mikrobien seurantaan edullisin vaihtoehto on pyrosekvensoinnin, tai jonkin muun vastaavan menetelmän, ja sirutekniikan yhdistäminen. Lisäksi sirulla näytteiden analysointi on nopeaa ja hyvin suunniteltujen koettimien avulla tarkkaa. Vaikka sirutekniikassa on edelleen ongelmia, niin tutkimusmenetelmien (esim. pyrosekvensointi ja LDR-sirut) kehittämisen myötä niistä on mahdollista päästä eroon.

LÄHTEET

Painetut lähteet

Ansorge, W. J. 2009. Next-generation DNA sequencing techniques. *New Biotechnology*, Volume 25, Number 4, April 2009.

Bano, N., Ruffin, S., Ransom, B., & Hollibaugh J. T. 2004. Phylogenetic composition of arctic ocean archaeal assemblages and comparison with Antarctic assemblages. *Appl. Environ. Microbiol.* 70: 781–789.

Busti, E., Bordoni, R., Castiglioni, B., Monciardini, B. & Sosio, M. 2002. Bacterial discrimination by means of a universal array approach mediated by LDR (ligase detection reaction). *BMC Microbiology*, 2:27, pages 1–13.

Castiglioni, B., Rizzi, E., Frosini, A., Sivonen, K., Rajaniemi, P., Rantala, A., Mugnai, M., Ventura, S., Wilmotte, A., Boutte, C., Grubisic, S., Balthasart, P., Consolandi, C., Bordoni, R., Mezzelani, A., Battaglia, C., & Bellis, G. 2004. Development of a Universal Microarray Based on the Ligation Detection Reaction and 16S rRNA Gene Polymorphism To Target Diversity of Cyanobacteria. *Applied and Environmental Microbiology*, Dec. 2004, p. 7161–7172.

Cha, R. S. & Thilly, W. G. 1993. Specificity, efficiency, and fidelity of PCR. *Genome Res.* 1993 3: S18–S29.

Chao, A. 1984. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* 11:265–270.

Cho, J. & Tiedje, J. 2001. Quantitative Detection of Microbial Genes by Using DNA Microarrays. *Applied and Environmental Microbiology*, Mar. 2002, p. 1425–1430.

Claesson, M. J., O'Sullivan, O., Wang, Q., Nikkilä, J., Marchesi, J. R., Smidt, H., de Vos, W. M., Ross, R. P. & O'Toole, P. W. 2009. Comparative Analysis of Pyrosequencing and a Phylogenetic Microarray for Exploring Microbial Community Structures in the Human Distal Intestine. *PLoS ONE*. August 2009, Volume 4, Issue 8.

Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Marsh, T., Garrity, G. M. & Tiedje, J. M. 2008. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, 2009, Vol. 37, Database issue D141–D145.

Cole, J. R., Chai, B., Farris, R. J., Wang, Q., Kulam-Syed-Mohideen, A. S., McGarrell, D. M., Bandela, A. M., Cardenas, E., Garrity, G. M., & Tiedje, J. M. 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Res.* 35 (Database issue): D169–D172.

DeSantis, T. Z., Hugenholtz, P., Keller, K., Brodie, E. L., Larsen, N., Piceno, Y. M., Phan, R., & Andersen, G. L. 2006a. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res* 34:W394–9.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., & Andersen, G. L. 2006b. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied Environmental Microbiology*, 72:5069–72.

Edwards, U., Rogall, T., Blocker, H., Emde, M. D. & Bottger, E. C. 1989. Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res* 17, 7843–7853.

Ettema, T. J. G. 2005. Functional and comparative genomics of the Archaea. The Netherlands, Wageningen, Ph.D Thesis Wageningen University.

- Gerry, N., Witowski, N., Day, J., Hammer, R., Barany, G. & Barany, F. 1999. Universal DNA Microarray Method for Multiplex Detection of Low Abundance Point Mutations. *J. Mol. Biol.* (1999) 292, 251–262.
- Gibas, V. & Jambeck, P. 2001. *Developing Bioinformatics Computer Skills*. Published by O'Reilly & Associates, Inc. Sebastopol, CA.
- Hoheisel, J. 2006. Microarray technology: beyond transcript profiling and genotype analysis. *Nature*, Vol 7, p. 21–210.
- Hultman, J., Ritari, J., Romantschuk, M., Paulin, L & Auvinen, P. 2008. Universal ligation-detection-reaction microarray applied for compost microbes. *BMC Microbiology* 2008, 8:237.
- Huson, D. E., Auch, A. F., Qi, J. & Schuster, S. C. 2007. MEGAN analysis of metagenomic data. *Genome Res.* 2007 17: 377–386.
- Jürgens, G., Glöckner, F.-O., Amann, R., Saano, A., Montonen, L., Likolammi, M. & Munster, U. 2000. Identification of novel archaea in bacterioplankton of a boreal forest lake by phylogenetic analysis and fluorescent in situ hybridization. *FEMS Microbiol. Ecol.* 34:45–56.
- Lahr, D. J. G. & Katz, L. A. 2009. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *BioTechniques* Vol. 47, No. 4, 2009.
- Lai, X., Cao, L., Tan, H., Fang, S., Huang, Y. & Zhou, S. 2007. Fungal communities from methane hydrate-bearing deep-sea marine sediments in South China Sea. *The ISME Journal* (2007) 1, 756–762.
- Lee, D-Y., Lauder, H., Cruwys, H., Falletta, P. & Beaudette, L. A. 2007. Development and application of an oligonucleotide microarray and real-time quantita-

tive PCR for detection of wastewater bacterial pathogens. *Science of the total environment* 398 (2008), p. 203–211.

Liu, Z., DeSantis, T. Z., Andersen, G. L. & Knight, R. 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research*, 2008, Vol. 36, No. 18.

Liu, Z., Lozupone, C., Hamady, M., Bushman, F. D. & Knight, R. 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Research*, 2007, Vol. 35, No. 18.

Loy, A. & Bodrossy, L. 2005. Highly parallel microbial diagnostics using oligonucleotide microarrays. *Clinica Chimica Acta* 363 (2006) s. 106–119.

Muyzer, G., De Waal, E. C. & Uitierlinden, A. G. 1993. Profiling of Complex Microbial Populations by Denaturing Gradient Gel Electrophoresis Analysis of Polymerase Chain Reaction-Amplified Genes Coding for 16S rRNA. *Applied and Environmental Microbiology*, Mar. 1993, P. 695–70, Vol. 59, No. 3.

Petrosino, J. F., Highlander, S., Luna, R. A., Gibbs, R. A. & Versalovic, J. 2009. Metagenomic Pyrosequencing and Microbial Identification. *Clinical Chemistry* 55:5, 856–866, (2009).

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, Dec. 2009, Vol. 75, No. 23.

Schloss, P. D. & Handelsman, J. 2005. Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. *Applied and Environmental Microbiology*, Mar. 2005, Vol. 71, No. 3.

Shiu, S. & Borevitz, J. 2006. The next generation of microarray research: applications in evolutionary and ecological genomics. *Heredity* (2008) 100, 141–149.

Tuimala, J. 2005. Bioinformatiikan perusteet. CSC – Tieteellinen laskenta Oy. 1. painos. Painopaikka: Picaset Oy, Helsinki 2005.

Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Applied Environmental Microbiology*, 73(16):5261-7.

Sähköiset lähteet

454 Life Sciences, a Roche company. 2010. GS FLX Titanium Series [viitattu 19.10.2010]. Saatavissa: <http://454.com/products-solutions/system-features.asp>

CLC Bio. 2008. Bioinformatics explained: Biological databases [viitattu 14.9.2010]. Saatavissa: <http://www.clcbio.com/index.php?id=1238>.

Lahden ammattikorkeakoulu. 2010. Uudet ympäristömittausmenetelmät liiketoimintamahdollisuutena (YMLI) [viitattu 3.8.2010]. Saatavissa: <http://www.lamk.fi/tekniikka/tutkimus/hankkeet/yml.html>

NCBI. 2010. NCBI-GenBank Flat File Release 179.0 [viitattu 14.9.2010]. Saatavissa: <ftp://ftp.ncbi.nih.gov/genbank/gbrel.txt>

Roche Diagnostics Oy. 2010. Genome Sequencer FLX [viitattu 4.8.2010]. Saatavissa: http://www.roche-applied-science.com/publications/multimedia/genome_sequencer/flx_multimedia/wbt.htm

Wikipedia. 2010. DNA microarray [viitattu 19.10.2010]. Saatavissa: http://en.wikipedia.org/wiki/DNA_microarray

LIITTEET

LIITE 1: Bakteerien taksonomiset määritelmät lajitasolla

LIITE 2: Bakteerien taksonomiset määritelmät sukutasolla

LIITE 3: Bakteerien taksonomiset määritelmät heimo- ja lahkotasolla

LIITE 4: Arkkien taksonomiset määritelmät lajitasolla

LIITE 5: Arkkien taksonomiset määritelmät sukutasolla

LIITE 6: Koettimien rakenne ja sekvenssit

LIITE 1. Bakteerien taksonomiset määritelmät lajitasolla

JÄRJESTYS NUMERO	SEKVENSIIN TUNNUS	LAHKO	LAJI	OHJEL- MA
B1	JNQ01B34BY	Acidimicrobiales	<i>Iamia majanohamensis</i>	M
B2	JNQ03G46ID	Acidobacteriales	<i>Acidobacteria bacterium</i> KBS 96	M
B3	JNQ01BYIQV	Anaerolineales	<i>Leptolinea tardivitalis</i>	M
B4	JNQ03G2RY6	Anaerolineales	<i>Levilinea saccharolytica</i>	M
B6	JNQ03GKOLY	Desulfuromonadales	<i>Geobacter pelophilus</i>	M
B7	JNQ03F8PFM	Desulfuromonadales	<i>Geobacter sp. Ply1</i>	M
B9	JNQ03FI6YO	Ei lahkoo (Bacteroidetes)	<i>Bacteroidetes bacterium</i> zo30	M
B10	JNQ01A58U0	Ei lahkoo (Lentisphaerae)	<i>Cytophaga sp.</i> Dex80-64	M
B11	JNQ03HHH8M	Ei lahkoo (Firmicutes)	<i>Firmicutes bacterium</i> WSF2-15	M
B12	JNQ03HDX1H	Ei lahkoo (Proteobacteria)	<i>Syntrophorhabdus aromaticivorans</i>	M
B13	JNQ03G01FR	Ei lahkoo (Proteobacteria)	unidentified sulphate reducing bacterium DSB-DSb99-3	M
B14	JNQ03FVT1N	Holophagales	<i>Geothrix fermentans</i>	M
B15	JNQ01BLC62	Myxococcales	<i>Myxobacterium</i> KC	M
B16	JNQ03F6PN9	Nitrospirales	<i>Nitrospira sp.</i>	M
B17	JNQ03HAMVQ	Oscillatoriales	<i>Oscillatoria sp.</i> INT-06	M
B18	JNQ03GMZR1	Rhizobiales	Filamentous symbiotic bacterium of <i>Methylobacterium sp.</i> , strain: TFF	M
B19	JNQ03HEQWW	Rhodocyclales	<i>Rhodocyclus sp.</i> , mixed culture isolate R6	M
B20	JNQ03F9YZE	Rhodospirillales	<i>Rhodospirillaceae bacterium</i> L34	M
B21	JNQ03HG8YW	Spirochaetales	<i>Spirochaeta sp. isolate</i> TM3	M
B22	JNQ03F342U	Syntrophobacteriales	<i>Syntrophus sp.</i> , Clone B2 tai Clone B3	M
B30	JNQ03GR8QS	Burkholderiales	Arctic sea ice bacterium ARK10172	M
B31	JNQ03GY7KL	Burkholderiales	<i>Derrxia gummosa</i> (IAM14990)	M
B32	JNQ03F1Q9M	Chlamydiales	<i>Chlamydiales bacterium</i> CRIB 32	M
B39	JNQ03F4UHB	Desulfobacteriales	<i>Desulfobacterium anilini</i> , strain AK1	M
B47	JNQ03FTCSE	Ei lahkoo (Bacteroidetes)	<i>Bacteroidetes bacterium</i> PPF50E2	M
B49	JNQ03FLFNX	Ei lahkoo (Chlorobi)	<i>Chlorobi bacterium</i> Mat9-16	M
B51	JNQ03GM20Q	Ei lahkoo (Chloroflexi)	<i>Dehalococcoides sp.</i> BHI80-15 tai BHI80-52	M
B52	JNQ03F9O10	Ei lahkoo (Lentisphaerae)	<i>Lentisphaerae bacterium</i> 7509 tai 7548	M
B58	JNQ03GYHSQ	<i>Myxococcales</i>	<i>Byssophaga cruenta</i> , strain DSM 14553T	M
B68	JNQ01ALVYP	Planctomycetales	<i>Isosphaera-like str.</i> OJF2 tai <i>Planctomycete</i> OJF8	M
B72	JNQ01CGKS2	Sphingobacteriales	<i>Pedobacter terricola</i> strain DS-45	M
B73	JNQ03G4WB4	Spirochaetales	<i>Spirochaetes bacterium</i> SA-8 tai SA-10	M

LIITE 2. Bakteerien taksonomiset määritelmät sukutasolla

JÄRJESTYS- NUMERO	SEKVENSIIN TUNNUS	LAHKO	SUKU	OHJEL- MA
B8	JNQ03GTSA7	Campylobacterales	Sulfurospirillum	C
B23	JNQ03G5YIJ	Rhizobiales	Hyphomicrobium	M
B24	JNQ01AI4MC	"Acidobacteria"	Gp3	C
B25	JNQ03GT9HC	Actinomycetales	Mycobacterium	M
B26	JNQ03GAE8K	Actinomycetales	Rhodococcus	M
B27	JNQ03GX9XJ	Anaerolineales	<i>Levilinea</i>	C
B28	JNQ01A9G82	Bdellovibrionales	Bdellovibrio	M
B29	JNQ03F6I5N	Burkholderiales	Schlegelella	M
B33	JNQ03G99HI	Chloroflexales	Roseiflexus	M
B34	JNQ03HDRU4	Chromatiales	Thiocapsa	M
B35	JNQ03FTZ9M	Chroococcales	Cyanobacterium	M
B36	JNQ03GOIHY	Clostridiales	Clostridium	M
B37	JNQ03HFK5S	Clostridiales	Clostridium	M
B38	JNQ03G8ALT	Clostridiales	Acetivibrio	C
B40	JNQ03G686H	Desulfuromonadales	Geobacter	C
B42	JNQ03GLJIX	Desulfuromonadales	Geobacter	M
B43	JNQ03F05IS	Desulfuromonadales	Geobacter	M
B44	JNQ01BL7PX	"Acidobacteria"	Gp4	C
B45	JNQ03F2T3S	Holophagales	Holophaga	C
B46	JNQ03GXGEM	"Acidobacteria"	Gp2	C
B55	JNQ03FP3AZ	Methylococcales	Methylobacter	M
B56	JNQ03G2IMC	Myxococcales	Anaeromyxobacter	M
B57	JNQ01AYY5T	<i>Myxococcales</i>	<i>Chondomyces</i>	M
B59	JNQ01B40UY	Myxococcales	Cystobacter	M
B61	JNQ03G3UF7	Myxococcales	<i>Sorangium</i>	C
B62	JNQ03G5O40	Nitrosomonadales	Nitrosospira	M
B63	JNQ03FWWCN	Nitrosomonadales	(Nitrosospira)	M
B64	JNQ03GOM4G	Nitrosomonadales	(Nitrosospira)	M
B65	JNQ03GMGMJ	Nitrospirales	Nitrosospira	M
B67	JNQ03FUAEX	Planctomycetales	Singulisphaera	C
B70	JNQ03F0B9J	Solirubrobacterales	Solirubrobacter	M
B74	JNQ03FZANX	Spirochaetales	Spirochaeta	M
B78	JNQ03FRDT3	Spirochaetales	Spirochaeta	M
B79	JNQ03G4L3J	Syntrophobacterales	Desulfoglaeba	C
B80	JNQ03FQW33	Rhodocyclales	Azoarcus	M
B81	JNQ03GPE57	Thiotrichales	Achromatium	M

LIITE 3. Bakteerien taksonomiset määritelmät heimo- ja lahkotasolla

JÄRJESTYS- NUMERO	SEKVENSIIN TUNNUS	LAHKO	HEIMO	OHJEL- MA
B60	JNQ03HBCII	Myxococcales	Myxococcaceae	M
B69	JNQ03FMOOL	Rhodocyclales	Rhodocyclaceae	M
B71	JNQ03G4TNV	Sphingobacteriales	Flexibacteraceae	M
B41	JNQ03GCOXI	Desulfuromonadales	- (Geobacteraceae)	M
B66	JNQ01CHCWP	Oscillatoriales	-	M

LIITE 4. Arkkien taksonomiset määritelmät lajitasolla

JÄRJESTYS- NUMERO	SEKVENSIIN TUNNUS	LAHKO	LAJI	OHJEL- MA
A1	JNQ03GL3V2	Mathanosarcinales	Methanosarcina sp.	M
A2	JNQ03GUYF1	Mathanosarcinales	Methanosarcina sp.	M
A4	JNQ03GIF4T	Methanobacteriales	Methanobacterium aarhusense	M
A7	JNQ01BKQMT	Methanomicrobiales	Methanolinea tarda	M
A8	JNQ03HECDK	Methanomicrobiales	Candidatus Methanoregula boonei strain SN19 tai strain SN21	M
A9	JNQ03GCNRK	Methanomicrobiales	Methanomicrobiales archaeon SMSP	M
A10	JNQ03GZWGD	Methanomicrobiales	methanogenic endosymbiont of Caenomorpha-like sp. 1	M
A11	JNQ03GC453	unclassified Archaea	anaerobic methanogenic archaeon ET1-8	M
A12	JNQ03G8RTU	unclassified Archaea	anaerobic methanogenic archaeon ET1-8	M
A13	JNQ01BZ7ZS	unclassified Archaea	anaerobic methanogenic archaeon ET1-10	M
A14	JNQ03FRDXU	unclassified Archaea	anaerobic methanogenic archaeon E15-7	M

LIITE 5. Arkkien taksonomiset määritelmät sukutasolla

JÄRJESTYS- NUMERO	SEKVENSIIN TUNNUS	LAHKO	SUKU	OHJEL- MA
A5	JNQ03HEGB2	Methanomicrobiales	Methanoculleus	M
A6	JNQ03FMTND	Methanomicrobiales	Methanospirillum	M
A3	JNQ03F4X9I	Mathanosarcinales	Methanosaeta	C

LIITE 6/ 1. Bakteeri- ja arkki koettimien koko koettimen rakenne ja c-osan, zip-
koodin sekä d-osan emäsjärjestykset

KOODI	C-OSA	ZIP	VA 1/2	VA 2/2	UA	D-OSA
B1	TGCCGGGGACT CGTAGGAGACT GC	GATATAGGA ATGGCGCAT AC	VA1	VA2	UA	TGTCCTATGTT GCCAGCGGGT AG
B2	GGAAACGCTGG CTTTCCCGCAA GG	CTCATCGGA AGGGCTCGT AA	VA1	VA2	UA	AACGGCTTCG GACGGCTGGC
B3	AGGTAACGGCC CACCAAGGCGA	ACAGATGGA AAGGCAGTT CT	VA1	VA2	UA	CTGCGTACCAT CAGCTAGTTG GTG
B4	GGTCGCGAAAC CGCGAGGTGGA	TTTGGTAGCT GAGTGCCCT A	VA1	VA2	UA	CACACGCTAC AATGGCAAGT ACAATA
B6	CATCGGATGAG TCCGCGTCCCA TC	TAACTGTTT GACGCCACG C	VA1	VA2	UA	GGGACCTTCG GGCCTCCCGA
B7	CCAAGGTAGCC GTAAGGGGA GATGG	TAATTGAGC TGACGGCGC AC	VA1	VA2	UA	TCGGTGAAGTC GGGTGAAGTC GTC
B8	GGGGTAAAATC CGTAGATATCA CTAGGAATACC G	TTGTTGCTAC TCTGGCCCG A	VA1	VA2	UA	AGGGGGAGAT GGAATTAGTG GTGTAA
B9	CCGCTGGGGA GTACGGCCGC	TTCCGTGCAT AGTATAGGG A	VA1	VA2	UA	GTCGCAGCTA ACGCGATAAG TTCG
B10	CTGCCCGTGTC AAGCGGGAGG A	TTATGCGAC TTATCTCGG GA	VA1	VA2	UA	ATGCTGGGGA CTCTGGCGAG G
B11	GAGTTCCTAGC GGCGAACGGGT G	TGTATAGGA TTATGTCCGC G	VA1	VA2	UA	GCTCATCTCCC GCAAGGGAGG C
B12	CGGGGTAGAA GGTGGCTTGTA TTCGTA	CTGCTAGGA ATATGAGCT AC	VA1	VA2	UA	CGGATAAGAC CACACGACAC ATGTCG
B13	CGAGGGTGAGC CAATCCAGAA AGCA	CTTCTGTCAA TATGGGTAC G	VA1	VA2	UA	GCTGTACAAA GGGTTGCAAG CCCG
B14	GCTGGGATGCT GGAGGTGAAA CCTGG	TATTCGAG ATATGAGGC GC	VA1	VA2	UA	AGTGGTAATA CCGCATGAGA CCTTCG
B15	GGCAGGTCCT AAGGAGTGGCT GGT	TTGATCGTA GATTCGTGA GC	VA1	VA2	UA	GCCGTGCCAA CCCACAAGGA
B16	GGTAAGCCCTC CGTGAATCTC CGGG	CGAGATTAC AATTCACGA GC	VA1	VA2	UA	ACAGGGGGCG TAGGCGGTTA
B17	ACCGCGTGAGG GATGAAGGCTT GC	TGGTGTCTA GCTTCAGC CT	VA1	VA2	UA	GGGCGAAAGC CTGACAGAGC AAC
B18	GGACCTGCGTC CCATCAGCTAG TTGG	TGAGGTCAC GGTTCATGC TA	VA1	VA2	UA	ACTAAAGGAG TAATCCGCTTT GGGAA
B19	CAGAGGGTTGC CAACCCGCGAG	TGGTACTG GTATATGCC GC	VA1	VA2	UA	CTCACACGTC ATACAATGGT CGGTC

LIITE 6/ 2

B20	GGAAACGGGG TCCTTCAGTTC GGCT	CCGAGTGCA GAATAAAC CG	VA1	VA2	UA	TTGACATGGG ACGTATGGGC CTC
B21	CTTCAGGGTGA AAGGTGCAATT GCATT	GCGGTCTCA ATACAAACT CA	VA1	VA2	UA	CGCATGAGGC CCTGAGGGTA GAG
B22	CCCGGAGACAG GTGCTGCATGG C	GAAGTACC ATACGCGAG CA	VA1	VA2	UA	AACTAAGAAG TGCCTTCGGG GAG
B23	GATAATGACGG TACCCGGAGAA TAAGTCCC	GATTCACAC TGACCCATG TA	VA1	VA2	UA	GGTTGTAAAG CTCTTTGCCG GGAC
B24	TAAAGCTCTTT CGGCAGGAAC GAATTCT	TAAATAGAT TGGAGACGC GC	VA1	VA2	UA	GAGCGATGAA GCCCTTCGGGT G
B25	TGCAGAAGAA GCACCGCCAA CTC	GCATTAGAA GGTCTGGAC TA	VA1	VA2	UA	GACGAAGCGG AAGTGACGGT AGA
B26	CAAGTGACGGT AGGTACAGAA GAAGCACG	ATTGGCATA ACGTATTGC GC	VA1	VA2	UA	CCTCTTTCAGT ACCGACAAG CGT
B27	GAACCAGCAAC GCCGCGTGTGC	CAGGACTGA AGATCGAGT AC	VA1	VA2	UA	GGAATATTGG TAATCTGCGG AAGCGG
B28	GGTACTCTGCA AGAAAGGATC GGCTAACG	TAGAGTCAG TCATAGCTC GA	VA1	VA2	UA	CTGTTGCAGG GGAATAACAC AATGAA
B29	TTCGGGAAGAA GGAGTCGGGGC T	TTTATCGTAG CTGGCTGCC C	VA1	VA2	UA	GGCCTTCGGTT GTAAACCGCT TTTA
B30	AATACGGTGGG TTCTTGACGGT ACTGG	AGGATTAGA ACCTACGCA CC	VA1	VA2	UA	GTCCGGGAAG AAAAGGCTCA CTCT
B31	CGCTTGCTAAT ACCAGGCGCTT ATGACA	GCCGTGAGA CCACTGTAC TA	VA1	VA2	UA	GCACTTTGGC GGGACGAAAC GG
B32	AAAGAAGCAC CGGCTAACTCC GTGC	GACGCTGAA TCCTATTGAC A	VA1	VA2	UA	ACAATATTCTG TTGATTTGATG GTACTTGGT
B33	GATGACGGTAC CCCTGGAATAA GCCCC	CGCCTAAGG ATCGTGAAG TA	VA1	VA2	UA	ACCCTTTTTTG GGGGGACGAT
B34	TGAGTTGAACA GACGCAGGGG GTGA	CGACGACGA AGCTGCATG AA	VA1	VA2	UA	CCTTTTGGCGG GGAAGAAAGC CC
B35	GTAAACCTCTT TTCTCAAGGAA GAAGTTCTGAC	ACTCGAATA ACAGCATCT CG	VA1	VA2	UA	GAGGGATGAA TGCCCGTGGG TA
B36	TGGGACGATAA TGACGGTACCA GAGGA	CCCCTAAGC ATGGCACAG AT	VA1	VA2	UA	ACCTTCGGGTT GTAAAGTTCT GTCTTC
B37	GGGACGATAAT GACGTACCTGA GGAGA	CAATACAAG ATTACGGCC TC	VA1	VA2	UA	AGTCTTCGGAT TGTAATAACTCT GTCTTCA
B38	ATCAGGGACGA AAATGACGGTA CCTACG	GATCAGAAT CTATGGTAC GC	VA1	VA2	UA	GAAGGTTTTTC GGATTGTAAA CTCCTTTG

LIITE 6/ 3

B39	GTGAATAGCCA GGAAGACTGAT GGTACCA	TCTGTGTACT GCTCGCCAA T	VA1	VA2	UA	CTGTCAAGTG GGAAAAAAGC TTCTG
B40	GGTGCTAATAC CACTTTTACTT GACGGTACCC	ATATTTGGA ACGCAGCTC AC	VA1	VA2	UA	AGCTCTGTCA GAGGGAAAAGA AATGTAGA
B41	AATACCTTCCA TCCTTGACGGT ACCATCA	TGCAGTATC GCAGCGGTC TA	VA1	VA2	UA	CAGATGGGAA GAAATGGTTG GAAGCT
B42	GTTACTTGACGG TACCCTCAAGG AAGCA	GGGCAATGT TTATCCACA GA	VA1	VA2	UA	GTGTACGGTG GCTAATACCC GCT
B43	GTCGTAAAGCC CTGTCAGGTGG GAA	CTGACCGAA TCCAGCAGA GA	VA1	VA2	UA	CGCGTGGGTG ATGAAGGCCT TGG
B44	TGTAAGCCCCG GCTAACTCCGT GC	GATCGTGAA TCCGCGCAC TA	VA1	VA2	UA	TGGGAAGAAC AAAATGACGG TACCATA
B45	TTGACGGTACC CGGAGAGGAA GCC	GAGCCGTAA TCCGAGCGA TA	VA1	VA2	UA	CGGTGGGGAG GTTAATAGCCT CTTTAC
B46	AAGAGGAGTG ACGGTACGGGT ATAAGAA	TACTCTGA CGACTTAGG CA	VA1	VA2	UA	CGACCCGGAC GAATGCCTCG T
B47	CGGTACCCCCG GAATAAGGATC GGCA	TGCTGTCACT CGGCGTCTA T	VA1	VA2	UA	TCCCGATGCGT CGGGAGATGC
B49	GAAAAATCTCG ATCAAATCGAG AACTGACTGT	GTACTAGCA TATCATCGA CG	VA1	VA2	UA	AAGGTTGTAA ACTCTGTAG AGGGAA
B51	AAGGACGGTAC CTCAGGAATAA GTCTCGC	TATCGCATA GATCAGTGA GC	VA1	VA2	UA	TCTAAACCTCT TTTCTGGGGAC GAGC
B52	TGACGGTACCG GGAGAGGAAG CCA	TACGGGCAG CCAGGTACT TT	VA1	VA2	UA	CTGTCACCCG GGACGAAGAC AG
B55	TGATCCAGCAA TACCGCGTGTG TGA	GTTCATCAC GAGTGCGTA GA	VA1	VA2	UA	GGAATATTGG ACAATGGGCG AAGCC
B56	AACATCCACGG GCCTTGACGGT A	CATGTATCA AGATGGCTG AC	VA1	VA2	UA	GAGGGAAGAA CAAGCCTGTG GCC
B57	ATGACGGTATC TCCTTAGCAAG CACCG	GGTTCGCGC ATTCCAGCA TA	VA1	VA2	UA	GGGGTCGAAT ACCCCCCGCC G
B58	ATGACGGTACC TCCTTAGCAAG CACC	GCACATATC TAGCGACAT CT	VA1	VA2	UA	GGAATATGGG TTGGCTAATAT CCAACCTTG
B59	CTGACGGTACC GAGAGAAGAA GCACCA	ACGCGGCTA AAGGTAGAT AC	VA1	VA2	UA	ACCTCTGGCTA ACATCCAGGG GC
B60	TTGACTGTACC GGGAGAAGAA GCACCA	CACTGCCCA CAAGATGTA GA	VA1	VA2	UA	ACCCATTGGC CAACATCCAA TGGC

LIITE 6/ 4

B61	ATGACGGTACC CTTTTAGCAAG CACCA	GGATTTACA TGGCCTAGC AA	VA1	VA2	UA	CAGGGTTGGC TAACACCCAA CTCG
B62	TGGTGAATATC CGTGACTGATG ACGGTA	CATGACACA GAATCGACC GT	VA1	VA2	UA	TTCGGACGGG ACGAAATGGT CA
B63	ACGGCTAATAC CCGTGACTAAT GACGGT	AGAGGCATA AATGAGTCT CC	VA1	VA2	UA	GCTCTTTCAGA CGGAAAGAAA CGGTT
B64	ACTCTTTCGGA CGGAACGAAAT GGTTCT	TGAGTAGTA CGTTACGCC TG	VA1	VA2	UA	AGTGAAGAAG GCCTTCGGTTG TAAG
B65	TTCGGATTGTA AACCCCTTTCG GGAGG	CGATAGCGA AGGAGTCCA CA	VA1	VA2	UA	CGCCGCGTGG GGATGAAGGT C
B66	AAAGCCTGACG GAGCAATACCG CG	ACACTCTGA AAGACGCGA CG	VA1	VA2	UA	GGGAAGTTTT ACCGTCAATG GGCGG
B67	CGCAAGGTTGA CTGATCCCTGG AGGA	GTCTTAATGT TGGGCAAAC G	VA1	VA2	UA	ACTCCTGTCSA GGGGAAAGAA GC
B68	CCCTGGTAGGA AGCACGGGCTA AGT	GTTATCGAC TACGCTGTA CT	VA1	VA2	UA	GGAAGGGGCA ACCTTGACCG AT
B69	GAAATCGCGTA CGTGAATAACG TGCGT	TCGTGAGAC CGTCGTCAG TA	VA1	VA2	UA	TTGTAAAGCTC TTTCGGCCGG AAC
B70	GTGACGGTACC TTCAGAAGAAG CCCCG	GACAGCGCA GTACAGGTA AT	VA1	VA2	UA	GCTTCCAGGG TTAATAGCCC ATGGG
B71	GAAGTACGGT ACCGTAGGAAT AAGCAC	CGTACAGTA AGTATGATG CC	VA1	VA2	UA	AGAAACGCTC CCTGCGGGGA
B72	CGTGCACCGGA GCTGATGGTAC CG	GTCACAGAT TAGTAGGCA CG	VA1	VA2	UA	TCTTTTATACG GGAAAAAACCC TCCGGT
B73	TGCCCGCGTGA TGACATGAACC GA	TCGTACCTGT ATTCAGCGC G	VA1	VA2	UA	GGGGAAGAAT AAGCGTGGGA GGTAA
B74	GAAGAATAGA GGTGGGAGTGG AAAGCCC	TTAATCCGCT GTAGCCCAA A	VA1	VA2	UA	CCGTAAGGTT GTAAAATCCTT TTGTTGGC
B78	GGAGTGGA GCCCTGGTGAT GACGT	TTAATTGACT TCGCTCCAG C	VA1	VA2	UA	TCCTTTTGTCTG GTGAAGAATA AGCCA
B79	GGATTTGACGG TACCACCAGAG GAAGCC	TAAACTAGC ATTGAGCCC AC	VA1	VA2	UA	AAGGTTCTGC CAAGAAACGG GTG
B80	GGAAAGAAAT CGCCCTCCCA ATACGG	AAATCAGCA AACGGGCTC CG	VA1	VA2	UA	CCGGGTTGTA AACCGCTTTCG GAC
B81	GGGGTGCTAAT ACCACCTTGAC TTGACG	GAATTGATA ATCGCAGCC AC	VA1	VA2	UA	AGCACTTTCG GGAGGGAAGA AAAGC
A1	TCAATGCGTAC GCTTTTCAGAT GCCTA	ACGGGATAA CAACGCAGC CT	VA1	VA2	UA	CGCGAAAAGCG CGATAAGGGA CC
A2	GCCTAAAAAGC ATGTGGAATAA GAGCCGGG	AGAAGATCA ACAGCTCGT CC	VA1	VA2	UA	TGCGCAATGC GTACGCTTTTC ATAC

LIITE 6/ 5

A3	GCATACAATGC TGGCTGTCCAG ATGCC	ATAAGATCA AGACCTGTG CC	VA1	VA2	UA	CGATAAGGGG ACCTCGAGTG CCG
A4	GGATGCCGGGC ACACTAAGGGG A	ATTAGATTA AGACCAGCG CC	VA1	VA2	UA	CCCTTAGTTAC CAGCGGATCC TTTCA
A5	GTGTATAAAAAG GCACCGGGAG AAAGGGA	ATATAATCA AGACTGGCG CG	VA1	VA2	UA	GCCCGTAAAT TCGGGCTGTTC CG
A6	AAAACCTGGTGA AGAAAGGGCC GGGC	AGCATATAA CCACTGATC CG	VA1	VA2	UA	AGGCTGGCTG TCCACCAGTTT AC
A7	ATCTGTGTAAA AAGCAGGTGG AGAAAGGG	ACACTATTA AAGCTGCTC CG	VA1	VA2	UA	CTGCCCGTAA ATTCCGGGCTGT TA
A8	AAAAAACCCAGT GAAGAAAGGG CCGGGT	CAATGTATA AGACTCTCG CC	VA1	VA2	UA	GAAAATCGGG CTGTCCATTGG TGA
A9	CTGGGCTACAC GCGGGCTACAA G	CACTAATTC AGACGAAGC CG	VA1	VA2	UA	TAGGTCAGCA TGCCCCGAATT AA
A10	AAGAAAGGGC CGGGCAAGACC GG	GACCCTATC AGACAGATG CA	VA1	VA2	UA	AGGCTGTCCA TCTGTTTAAAT ACAGGTG
A11	TGACGGGCCAT CCCAGTGCCA	CACGCATCA AGACAGTAT CG	VA1	VA2	UA	CGGAAACTTC GCAATAAGCG AAGCT
A12	GGCGTCGCCCA AGGATGAGGCC	CAGCTCCTA AGACTTGGA CA	VA1	VA2	UA	AAGACGCTTG CGTTATGCTCG TG
A13	GAATAAGGAG AGGGCAAGTCT GTCGTTG	GGTATCATA GGACATTCG CA	VA1	VA2	UA	TTTACCCAGTC TAGAACGCTG GGA
A14	AAGGGCCGGG CAAGACCGGTG	GGTTACATG GATATAGCA CC	VA1	VA2	UA	CTGTCGTTGTG TCTAAAAACA CAATATAGC