



**TURUN
YLIOPISTO**

Matemaattis-luonnontieteellinen
tiedekunta

**DNA-metaviivakoodausmenetelmän soveltaminen liejutaskuravun
(*Rhithropanopeus harrisii*) ravinnonkäyttöanalyysiin**

Biologia

LuK-tutkielma

Nellie Paavola

Laajuus: 8 op

9.3.2026

Turun yliopisto

Turun yliopiston laatu järjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu
Turnitin OriginalityCheck -järjestelmällä.

Pääaine: Biologia

Tekijä: Nellie Paavola

Otsikko: DNA-metaviivakoodausmenetelmän soveltaminen liejutaskuravun (*Rhithropanopeus harrissii*) ravinnonkäyttöanalyysiin

Ohjaaja: Heidi Viitaniemi

Sivumäärä: 24 sivua

Päivämäärä: 9.3.2026

Uudessa elinympäristössä menestyvät haitalliset vieraslajit ovat uhka alueen alkuperäiselle ekosysteemille. Vieraslajien vaikutuksia uusiin elinympäristöihin voidaan tutkia tarkastelemalla alueen ravintoverkkoja analysoimalla yksittäisten eliöiden ravinnonkäyttöä. DNA-metaviivakoodaus on yksi yleistynyt menetelmä, jolla voidaan tuottaa eliöiden ravinnonkäyttöanalyseja. Menetelmä mahdollistaa useiden lajien samanaikaisen tunnistamisen yhdestä DNA-näytteestä perustuen jokaisen eliön uniikkiin emäsvaihteluun. DNA-metaviivakoodauksen periaatteena on eläimen ruoansulatuskanavan sisällöstä eristetystä DNA:sta valikoitujen geenialueiden monistaminen. Näistä yleisimmin käytettyjä ovat mitokondriaalinen *COI*-geenialue sekä nukleaariset *18srRNA*- (*18S*) ja *ITS*-geenialueet. Tutkielmassani sovellan ensimmäistä kertaa DNA-metaviivakoodausmenetelmää vuonna 2009 Suomen Saaristomereen saapuneen liejutaskuravun (*Rhithropanopeus harrissii*) ravinnonkäytön analysointiin. Tavoitteenani oli kolmen eri puolilta Saaristomerta kerätyn rapuyksilön avulla arvioida kolmen edellä mainitun geenialueen tuottaman informaation määrää ja laatua ruoansulatuskanavan sisällöstä eristetystä DNA:sta. Jatkotutkimuksia varten arvioin eri geenialueiden avulla saatujen tulosten mahdollisia päällekkäisyyksiä ja eroavaisuuksia sekä mahdollisuutta vähentää käytettävien geenialueiden määrää. Jokaisesta kolmesta liejutaskurapuyksilöstä eristettiin ruoansulatuskanavan sisällön DNA, josta monistettiin polymeerasiketjureaktion avulla *COI*-, *18S*- sekä *ITS*-geenialueet. Tämän jälkeen yhdeksän DNA-näytettä sekvensoitiin. Sekvensointitulokset analysoitiin siihen soveltuvan DADA2-prosessin avulla ja lopulliset päätelmät tehtiin DNA-näytteistä tunnistettujen heimotason sekvenssien perusteella. Sekvensoiduista DNA-näytteistä yksi *ITS*-geenialueen näyte oli huonolaatuinen, mutta kaikki muut näytteet olivat onnistuneita. Tunnistettujen heimojen lukumäärä oli geenialueesta riippuvaista; *COI*:n avulla tunnistettiin yhteensä 20, *18S*:n avulla 10 ja *ITS*:n avulla 7 heimoa. Sekä *COI* että *18S* tunnistivat paljon akvaattisissa ympäristöissä tavattavia heimoja, jotka kuuluvat liejutaskuravun normaaleihin ravintokohteisiin. *ITS*-geenialueen tunnistukset kohdistuvat pääasiassa terrestristen alueiden heimoihin. Tulosteni perusteella *ITS*-geenialue on *COI*- sekä *18S*-geenialueita tehottomampi liejutaskuravun ravinnonkäytön selvittämisessä DNA-metaviivakoodauksen avulla, ja näin ollen jatkotutkimuksissa on perusteltua hyödyntää vain kahta muuta geenialuetta.

Avainsanat: DNA-metaviivakoodaus, ravinnonkäyttöanalyysi, *COI*, *18S*, *ITS*

Sisällysluettelo

1	Johdanto	4
2	Aineisto ja menetelmät	6
2.1	Liejutaskurapujen kerääminen ja preparointi.....	6
2.2	DNA:n eristäminen.....	6
2.3	Eristetyn DNA:n laadun varmistaminen	7
2.4	Polymeraasiketjureaktio (PCR).....	8
2.5	Sekvensointi.....	11
2.6	Datan käsittely	12
3	Tulokset	13
4	Pohdinta	18
4.1	Sekvenssien laatu ja määrä	18
4.2	<i>COI</i> -, <i>ITS</i> - ja <i>18S</i> -geenialueiden käyttö DNA-metaviivakoodauksessa	19
4.3	Heimotason sekvenssitunnistukset	20
4.4	Geenialueiden verrannollisuus	21
	Kiitokset	22
	Lähdeluettelo	23

1 Johdanto

Ilmastonmuutoksen aiheuttamaa paikallista biodiversiteetin muuttumista on havaittu joka puolella maapalloa. Kuten muutkin meret, myös Itämeri on ilmastonmuutoksen takia jatkuvasti muuttuva elinympäristö, jossa on havaittavissa lämpötilan nousua, happamoitumista, saastumista sekä happikatoa (Reusch ym. 2018). Ekosysteemeissä tapahtuvia muutoksia voidaan selvittää kartoittamalla eliöiden ravinnonkäyttöä. Sen avulla voidaan tarkastella lajien välisiä vuorovaikutussuhteita ja tunnistaa ravintoverkkoja, jotka auttavat ymmärtämään ekosysteemin pysyvyyttä ja lajien keskinäisiä riippuvuuksia.

Muita Itämerta muuttavien tekijöiden lisäksi sinne saapuneet vieraslajit uhkaavat ja muuttavat alkuperäislajien elinympäristöä. Vieraslajit ovat ihmisten vaikutuksesta uuteen ympäristöön saapuneita lajeja, jotka eivät itsenäisesti pystyisi leviämään kyseisille alueille (Suomen luonnonsuojeluliitto, 2025). Vieraslajeista suurin osa katoaa nopeasti lajille epäsuotuisten olosuhteiden vuoksi. Lajit, jotka menestyvät uudessa elinympäristössään hyvän kilpailukykyä ansiosta ja aiheuttavat merkittäviä haittoja alkuperäislajeille, luokitellaan haitallisiksi vieraslajeiksi. Vieraslajit muuttavat uuden elinympäristönsä habitaattia muun muassa muokkaamalla ravintoverkkoja. Vieraslajien sekä haitallisten vieraslajien aiheuttamia vaikutuksia voidaan arvioida tekemällä kyseisten lajien ravinnonkäyttöön liittyviä tutkimuksia.

Saaristomeri on Suomen lounaisosassa sijaitseva osa Itämerta, ja sinne vuonna 2009 saapunut liejutaskurapu (*Rhithropanopeus harrisi*, Gould) on todettu haitalliseksi vieraslajiksi. Liejutaskurapu on kotoisin Pohjois-Amerikan itärannikolta, jossa sitä esiintyy eteläiseltä Saint Lawrencen lahdelta Meksikonlahdelle asti. Laji on levinnyt Suomen lounaisrannikolle laivojen painolastivesien mukana ensimmäisenä Naantaliin, josta sen levinneisyysalue on laajentunut ympäröiville alueille Saaristomerellä (Fowler ym. 2013). Liejutaskurapujen on havaittu olevan omnivoreja sekä opportunisteja elinympäristönsä suhteen ja niitä on havaittu monenlaisissa habitaateissa myös Saaristomerellä (Hegele-Drywa and Normant 2009, Aarnio ym. 2015). Näistä yksi on rakkohaurun (*Fucus vesiculosus*) antamassa suojassa (Fowler ym. 2013). Rakkohauru on yksi Itämeren avainlajeista ja se luo habitaatin monille muille eliöille, muun muassa leväkotiloille (*Theodoxus fluviatilis*) ja leväsiirroille (*Idothea baltica*), joiden määrät

ovat laskeneet huomattavasti liejutaskuravun levittäytyessä ympäri Saaristomerta (Jormalainen ym. 2016).

Eliöiden ravinnonkäyttöä voidaan tutkia useilla eri menetelmillä. Tutkielmassani keskityn DNA-metaviivakoodausmenetelmään (engl. DNA metabarcoding), jonka avulla voidaan selvittää tutkittavan lajin ravinnonkäytön kohteet lajitason tarkkuudella. DNA-metaviivakoodaus mahdollistaa useiden lajien samanaikaisen tunnistamisen yhdestä näytteestä. Menetelmässä hyödynnetään eläimen ruoansulatuskanavasta eristettyä DNA:ta ja tietoa jokaisen eliön ainutlaatuisesta emäsjärjestyksen vaihtelusta (Pompanon ym. 2012).

DNA-metaviivakoodauksessa eukaryootteja ravintonaan käyttävien eläinten ruoansulatuskanavasta eristetyistä DNA:sta monistetaan geenialueita, joiden avulla voidaan selvittää eliön ravinnonkäyttökohteet. Geenialueiden muuntelu vaihtelee eri eliöryhmien välillä, minkä vuoksi metaviivakoodauksessa käytettävät geenialueet valitaan tutkittavien eliöryhmien mukaan. Näitä geenialueita ovat usein eläinkunnan lajiston tunnistamiseen käytetty mitokondriaalinen *sytokromi C oksidaasi I (COI)*, eläin- sekä kasvikunnille käytetty nukleaarinen RNA-alue *18srRNA (18S)* sekä kasvi- ja sienikuntien lajiston tunnistamiseen käytetty nukleaarinen *sisäisesti transkriptoitu välike* (engl. *internal transcribed spacer*) -geeni (*ITS*). Geenien monistamiseen käytetään polymeerasiketjureaktiota. Emäsjärjestyksen selvittämiseen eli sekvensointiin on olemassa useita erilaisia menetelmiä. Tutkielmassani menetelmäksi on valittu uuden sukupolven sekvensointi (engl. next generation sequencing).

Tutkielmassani selvitän *COI*-, *18S*- sekä *ITS*-geenialueiden käyttökelpoisuutta liejutaskuravun ravinnonkäytön tutkimisessa DNA-metaviivakoodausmenetelmää hyödyntäen. Kandidaattitutkielmani on pilottitutkimus suuremmalla mittakaavalla toteutettavalle tutkimukselle, jonka tarkoituksena on hyödyntää kyseistä menetelmää tunnistamaan liejutaskurapujen käyttämää ravintoa lajitason, ja lajin eri populaatioiden habitateista johtuvia mahdollisia eroja ravinnon suhteen. Tutkielmani selvittää minkälaista informaatiota eristetyistä DNA:sta saadaan käyttämällä kaikkia kolmea edellä mainittua geeniä, niiden antamien tulosten mahdollisia eroavaisuuksia sekä sitä, onko jatkotutkimuksessa mahdollista hyödyntää vähempää määrää metaviivakoodaukseen käytettäviä geenialueita.

2 Aineisto ja menetelmät

2.1 Liejutaskurapujen kerääminen ja preparointi

Liejutaskuravut kerättiin kesän 2024 aikana kolmesta paikasta Saaristomereltä: kolme rapuyksilöä ulkosaaristosta ja yksi välisaaristosta. Rapujen pyynti tapahtui syöttikalojen pyyntiin tarkoitetuilla katiskoilla, joiden syöttinä käytettiin kalkkunanakkia. Pyydytyt ravut säilöttiin 75 % etanoliin Falcon -putkiin ja niiden keräyspaikka merkittiin jatkokäyttöä varten. Rapuyksilöt eli tutkimukseni näytteet säilöttiin -20 °C:en odottamaan jatkokäsittelyä. Näytteitä käsitellessäni maaliskuussa 2025 valitsin satunnaisesti yhden ravun kustakin ohjaajani valitsemasta keräyspaikasta ja määritin rapujen sukupuolen sekä selkäkilven (engl. carapax) leveyden (taulukko 1). Näitä tietoja voidaan mahdollisesti hyödyntää myöhemmässä tutkimuksessa yksilöiden väliseen vertailuun ravinnonkäytön ja elinympäristön suhteen.

Taulukko 1. Näytteiden keräyspaikat, selkäkilven leveys sekä sukupuoli.

Näyte	Sijainti	Selkäkilven leveys (mm)	Sukupuoli
1	Stenskär, Berghamn	19,6	Koiras
2	Paraisten Portti	17,0	Koiras
3	Stenskär, Birskär	14,1	Naaras
4	Stenskär, Birsskär	16,7	Koiras

Preparoin liejutaskurapuyksilöiltä 1–3 ruoansulatuselimistön sisällön mikroskoopin alla hyödyntäen useita erilaisia pinsettejä. Keräsin yksilöistä saadut näytteet Eppendorf -putkiin ja nimesin ne keräyspaikkojen sekä preparointijärjestyksen (1–3) mukaan. Preparoidut näytteet säilytettiin -20 °C:ssa niiden parhaimman säilyvyyden takaamiseksi. Preparoitujen näytteiden lisäksi sain tutkielmaani mukaan yhden valmiiksi preparoidun ja suolasaostamalla käsitellyn näytteen (taulukko 1, näyte 4).

2.2 DNA:n eristäminen

DNA:n eristämiseen käytettävä protokolla on jokaisessa tilanteessa lähes samanlainen. Halutun DNA:n sisältämä näyte kerätään ja näytteen solurakenne hajotetaan lymfaattin muodostamiseksi.

DNA erotetaan seuraavaksi lymfaattisista soluista ja muusta materiaalista, jota ei voida hyödyntää, jonka jälkeen DNA kiinnitetään puhdistusalustalle. DNA pestään proteiineista ja muista epäpuhtauksista huuhtelemalla se tarkoituksenmukaisilla liuoksilla. Viimeisenä vaiheena eristetty DNA eluoidaan eli irrotetaan puhdistusalustalta.

Käytin liejutaskurapunäytteiden 1–3 DNA:n eristämiseen siihen tarkoitettua valmista DNeasy PowerSoil Pro Kit -pakkausta (Qiagen (Saksa), versio: toukokuu 2019). Pakkaus sisälsi kaikki DNA:n eristämiseen käytettävät liuokset sekä ohjeet pakkauksen käyttöön, joita seurasin tarkasti DNA:n eristämisen aikana. Näytteen 4 DNA oli valmiiksi eristetty suolasaostamalla mukaillen Aljanabi & Martinez alkuperäistä protokollaa (Aljanabi 1997). Suolasaostus DNA:n eristämisessä perustuu DNA:n negatiivisen varauksen ja ionien väliseen vuorovaikutukseen. Suolasaostuksessa lisätty suola neutraloi DNA:n negatiiviset varaukset, jolloin DNA-molekyylit voivat kerääntyä yhteen ja saostua liuoksesta alkoholin (esimerkiksi etanolin) avulla.

2.3 Eristetyn DNA:n laadun varmistaminen

Ennen puhdistettujen DNA-näytteiden jatkokäsittelyä tarkistin niiden laadun kolmella eri tavalla: elektroforeesilla sekä NanoDrop 2000- ja Qubit v2-laitteilla. Elektroforeesin hyödyntäminen DNA:n laaduntarkistukseen liittyy näytteen aminohappoketjujen pituuteen. Lyhyemmät DNA-fragmentit liikkuvat agarosigeelillä nopeammin ja näin ollen pidemmälle elektroforeesijon aikana kuin pidemmät DNA-fragmentit. Jos DNA on epäpuhdasta, geelille syntyy halutun DNA:n muodostaman alueen lisäksi yksi tai useampia muita alueita. NanoDrop mittaa DNA:n konsentraatiota ultraviolettivalon absorbanssin aallonpituudella 260 nm ja laskemalla siitä näytteen DNA-pitoisuuden tunnetun absorbanssikeroimen perusteella. Qubit mittaa kaksijuosteisen DNA:n konsentraatiota fluoresenssin avulla.

Elektroforeesissa käyttämäni agarosigeeli oli agarosipitoisuudeltaan 1,5 %. Puskurina agarosigeeliin käytin 1x TBE:tä, johon sekoitin valmistajan ohjeiden mukaan sopivan määrän agarosia. Värjäsin tilavuudeltaan 9 µl olevat DNA-näytteet 1µl:lla 10x loading dye -väriainetta. Molekyylipainomarkkerina käytin Generuler 1 kb DNA-markkeria. Agarosigeeli

ajettiin BioRad -elektroforeesilaitteella 90 V teholla. NanoDrop- ja Qubit-mittaukset tein noudattamalla laitteiden valmistajien ohjeita.

2.4 Polymeraasiketjureaktio (PCR)

Ennen varsinaisen polymeraasiketjureaktiomenetelmän aloitusta tein suolasaostetulle näytteelle laimennossarjan. Laimennossarjan tarkoituksena on vähentää DNA:ssa mahdollisesti esiintyvien inhibiittoreiden konsentraatiota. Inhibiittorit estävät PCR:ssä käytettävän polymeraasin toimintaa, joka vähentää ja hidastaa DNA:n monistumista. Tein laimennossarjaan, johon valmistin 5 näytettä: 1:1, 1:5, 1:10, 1:25 ja 1:100. Laimensin näytteet käyttämällä alkuperäistä suolasaostettua näytettä ja ultrapuhdasta nukleaasivapaata vettä (Invitrogen 10977-035).

PCR-näytteiden valmistamiseen käytin jokaista kolmea geenä varten yleispäteviä etu- ja taka-alukkeita (engl. forward primer ja reverse primer), jotka rajaavat halutun geenin osuuden DNA:sta. Käyttämieni etu- ja taka-alukkeiden nimet ovat kuvattuna taulukossa 2. Jokaiselle näytteelle käytettiin alun perin samaa polymeraasia (valmistaja: New England Biolabs (NEB), valmistusnumero: M0494), mutta *COI*-näytteille polymeraasi vaihdettiin myöhemmin uudessa PCR-ajossa toiseen polymeraasiin (valmistaja: Qiagen, valmistusnumero: 206143). Yhden näytteen PCR-reaktioon käytin 12,5 µl 2x puskuriliuosta, 10 µl etualuketta, 10 µl taka-aluketta, 5 µl ultrapuhdasta nukleaasivapaata vettä sekä 5 µl DNA-näytettä.

Taulukko 2. PCR-reaktiossa käytettyjen etu- ja taka-alukkeiden nimet, alukkeiden avulla saatavan monistetun DNA-fragmentin pituus emäspareina (bp) sekä käytettyjen polymeraasien nimet geenikohtaisesti.

Geeni	Etualuke	Taka-aluke	Lopputuotteen pituus	Puskuriliuos
<i>COI</i>	mICOIintF-XT	jgHCO2198	303 bp	Q5 HotStart HiFidelity (NEB), QMP 2X (Qiagen)
<i>18S</i>	V4_If	TAReukREV3	420 bp	Q5 HotStart HiFidelity (NEB)
<i>ITS</i>	ITS5-1737F	ITS2-2043R	450 bp	Q5 HotStart HiFidelity (NEB)

Ajoin PCR-reaktiot jokaisen geenin kohdalla erikseen, joten ajoja oli yhteensä kolme. Kaikissa kolmessa reaktiossa mukana oli yhteensä seitsemän näytettä sekä negatiivinen kontrolli ultrapuhdasta nukleasivapaata vettä. Lisäksi jokaisessa PCR-reaktiossa oli mukana neljä ylimääräistä näytettä, joilla ei ollut tutkielmani kannalta merkitystä. Näytteet PCR-reaktioissa olivat näytteistä 1–3 sekä näytteestä 4 laimennossarjan neljä laimeinta laimennosta.

Toteutin PCR-reaktiot valmiiden ohjeiden mukaan. Reaktiossa on useita vaiheita ja syklejä, jotka alkavat alkudenaturoinnista, jonka aikana kaksijuosteisen DNA:n juosteet irtoavat toisistaan. Alkudenaturointia seuraa monistettavasta geenistä ja tarvittavasta lopputuotteen määrästä riippuvainen määrä syklejä, joiden aikana denaturoinnissa kaksijuosteisen DNA:n juosteet irtoavat toisistaan. Syklin liittymisvaiheessa näytteissä olevat alukkeet kiinnittyvät näytteen yksijuosteisiin DNA-fragmenteihin, ja syklin lopussa DNA-polymeraasi rakentaa uusia DNA-juosteita käyttäen alukkeita sekä DNA-mallijuostetta. Sykliä jälkeen PCR-reaktiossa on viimeinen laajentumisvaihe, jonka aikana varmistetaan kaikkien DNA-ketjujen rakentuminen loppuun. Varsinaisen PCR-reaktion loputtua PCR-laitteisto jäädytetään 4 °C:en siihen asti, kunnes näytteet siirretään jatkokäyttöön tai jääkaappiin. Toteutin *COI*-geenin näytteiden PCR-reaktion taulukon 3 mukaisesti, *18S*-geenin reaktion taulukon 4 mukaisesti ja *ITS*-geenin PCR-reaktion taulukon 5 mukaisesti.

Taulukko 3. PCR-reaktion lämpösykliä asetukset *COI*-geenille. Taulukossa on esitetty DNA:n monistamisen vaihe, vaiheeseen käytettävä lämpötila ja aika sekä syklien määrä.

Vaihe	Lämpötila (°C)	Aika	Syklien määrä
Alkudenaturointi	98	10 min	1
Denaturointi	94	1 min	35
Liittymisvaihe	45	1 min	
Laajentumisvaihe	72	1 min	
Lopullinen laajentuminen	72	5 min	1

Taulukko 4. PCR-reaktion lämpösykliä asetukset *18S*-geenille. Taulukossa on esitetty DNA:n monistamisen vaihe, vaiheeseen käytettävä lämpötila ja aika sekä syklien määrä.

Vaihe	Lämpötila (°C)	Aika	Syklien määrä
Alku-denaturointi	98	2 min	1
Denaturointi	98	30 s	15
Liittymisvaihe	53	30 s	
Laajentumisvaihe	72	45 s	
Denaturointi	98	30 s	20
Liittymisvaihe	48	30 s	
Laajentumisvaihe	72	45 s	
Lopullinen laajentuminen	72	10 min	1

Taulukko 5. PCR-reaktion lämpösykliä asetukset *ITS*-geenille. Taulukossa on esitetty DNA:n monistamisen vaihe, vaiheeseen käytettävä lämpötila ja aika sekä syklien määrä.

Vaihe	Lämpötila (°C)	Aika	Syklien määrä
Alku-denaturointi	98	2 min	1
Denaturointi	94	30 s	32
Liittymisvaihe	55	30 s	
Laajentumisvaihe	72	45 s	
Lopullinen laajentuminen	72	10 min	1

Tarkistin PCR-reaktiolla monistettujen näytteiden laadun samoilla menetelmillä, kuin DNA:n eristämisen jälkeen. Geenin *COI* monistaminen ei laaduntarkastuksen mukaan onnistunut, joten tein sille kaksi PCR-reaktiota lisää. Nämä reaktiot tehtiin QIAGEN Multiplex PCR (valmispakkaus, versio: maaliskuu 2016) sekä DreamTaq DNA Polymerase (valmis pakkaus, versio: huhtikuu 2022, Thermo Scientific EP0701) protokollien mukaan. Käytin molemmissa reaktioissa samoja etu- ja taka-alukkeita, kuin aiemmassa *COI*:n PCR-reaktiossa, mutta puskuriliuos vaihdettiin QMP 2X -puskuriliuokseksi. Ajossa käytettyjen reagenssien määrät vastasivat aiempia. Noudatin reaktioiden tekemisessä tarkasti pakkauksien valmistajien ohjeita, mutta muutin PCR-reaktion lämpötiloja taulukoiden 6 ja 7 mukaisiksi. Varmistin vielä molempien reaktioiden laadun, ja totesin QIAGEN Multiplex PCR-reaktion käyttökelpoisemmaksi.

Taulukko 6. QIAGEN Multiplex PCR -reaktion muokatut lämpösykliä asetukset. Taulukossa on esitetty DNA:n monistamisen vaihe, vaiheeseen käytettävä lämpötila ja aika sekä syklien määrä.

Vaihe	Lämpötila (°C)	Aika	Syklien määrä
Alku-denaturointi	95	15 min	1
Denaturointi	95	1 min	35
Liittymisvaihe	57–50	90 s	
Laajentumisvaihe	72	90 s	
Lopullinen laajentuminen	72	5 min	1

Taulukko 7. Thermo Scientific DreamTaq DNA Polymerase - reaktion muokatut lämpösykliä asetukset. Taulukossa on esitetty DNA:n monistamisen vaihe, vaiheeseen käytettävä lämpötila ja aika sekä syklien määrä.

Vaihe	Lämpötila (°C)	Aika	Syklien määrä
Alku-denaturointi	95	10 min	1
Denaturointi	95	1 min	35
Liittymisvaihe	57–50	1 min	
Laajentumisvaihe	72	1 min	
Lopullinen laajentuminen	72	10 min	1

PCR-reaktioiden jälkeen puhdistin DNA-näytteet. *18S*-, *ITS*- ja paremmaksi todetun *COI*-geenin PCR-näytteiden puhdistamisessa noudatin valmista QIAquick PCR Purification -pakkausta (Qiagen, versio: heinäkuu 2018). Tämän jälkeen mittasin puhdistettujen näytteiden kaksijuosteisen DNA:n konsentraation Qubit -menetelmällä.

2.5 Sekvensointi

Sekvensoitaviksi näytteiksi lähetettiin kaikki *ITS*- ja *18S*-geenialueiden näytteet, parhaimmiksi todetut *COI*-geenin PCR-näytteet sekä suolasaostamalla eristettyjen laimennettujen näytteiden 1:5 ja 1:100 sekoitus vaaditun näytemäärän saavuttamiseksi. Näytteet lähetettiin sekvensointiin Genewiz -yritykselle Saksaan. Sekvensointitapana käytettiin uuden sukupolven DNA-sekvensointimenetelmää (engl. next-generation sequencing, NGS), joka perustuu DNA:n emäsjärjestyksen selvittämiseen fluoresoivien merkkien avulla. Sekvensoinnissa käytettiin lukupituutta 250bp paired-end-lukumuodossa (Illumina).

Sekvensointitulokset saatiin Genewiz yritykseltä sähköisinä fastq-tiedostoina, jotka analysoitiin jatkokäyttöä ja tuloksia varten. Yritys ilmoitti, että näytteen 2 sekvensointitulokset eivät olleet standardien mukaisia, eikä niiden tuloksia saatu lainkaan. Tutkielmaa jatkettiin näytteillä 1, 3 ja 4.

2.6 Datan käsittely

Käsittelin sekvensoinnin raakadatan RStudiolla (versio: r-env/451) Puhti.csc -ympäristössä. Raakadata muokattiin sen läpikäyntiä varten oikeanlaiseen muotoon; rajasin sekvensoinnista saadut lukemat (engl. read) laadun perusteella vähintään sadan emäksen pituisiksi, jonka avulla minimoin niiden sisältämiä mahdollisia informaatiovirheitä. Tämän jälkeen käsittelin rajatut sekvensointitulokset R-paketilla DADA2 V1.25 ([DADA2](#), 2025). DADA2 suorittaa datan laadunhallinnan, mahdollisten virheiden korjauksen sekä sekvenssien tunnistamisen vertailemalla niitä DADA2:lle annettuihin tietokantoihin. NGS sekvensoinnissa kaksijuosteiset DNA-fragmentit sidotaan sekvensoinnissa käytettävälle alustalle ja luetaan ensin yhteen suuntaan, jonka jälkeen ne luetaan toisen kerran vastakkaiseen suuntaan. Näitä syntyneitä lukemia kutsutaan eteen- ja taaksepäinlukemiksi. DADA2 yhdistää saman DNA-juosteen eteen- ja taaksepäinlukemat yhdeksi lukemaksi, jota voidaan verrata menetelmälle annettuun tietokantaan.

Analysoin jokaisen geenialueen (*COI*, *18S* ja *ITS*) sekvenssidatan erikseen DADA2-menetelmällä. Vertailin *COI*-geenialueen sekvenssejä BOLD Systems'in Baltic Sea (2024) tietokantaan, *ITS*-sekvenssejä UNITE community:n luomaan tietokantaan (Abarenkov ym. 2025) ja *18S* sekvenssejä valmiiseen SILVA v132 tietokantaan (2025). *COI*:n sekä *ITS*:n tietokantojen fastaformaattit muokattiin DADA2:n oletusten mukaiseen muotoon. *18S*:n tietokanta oli valmiiksi oikeassa muodossa.

Tietokantoihin verratuista sekvensointituloksista analysoin geenien välisiä eroavaisuuksia sekä samankaltaisuuksia. Päätin pitäytyä tutkielmani tuloksissa heimotason taksonomisessa

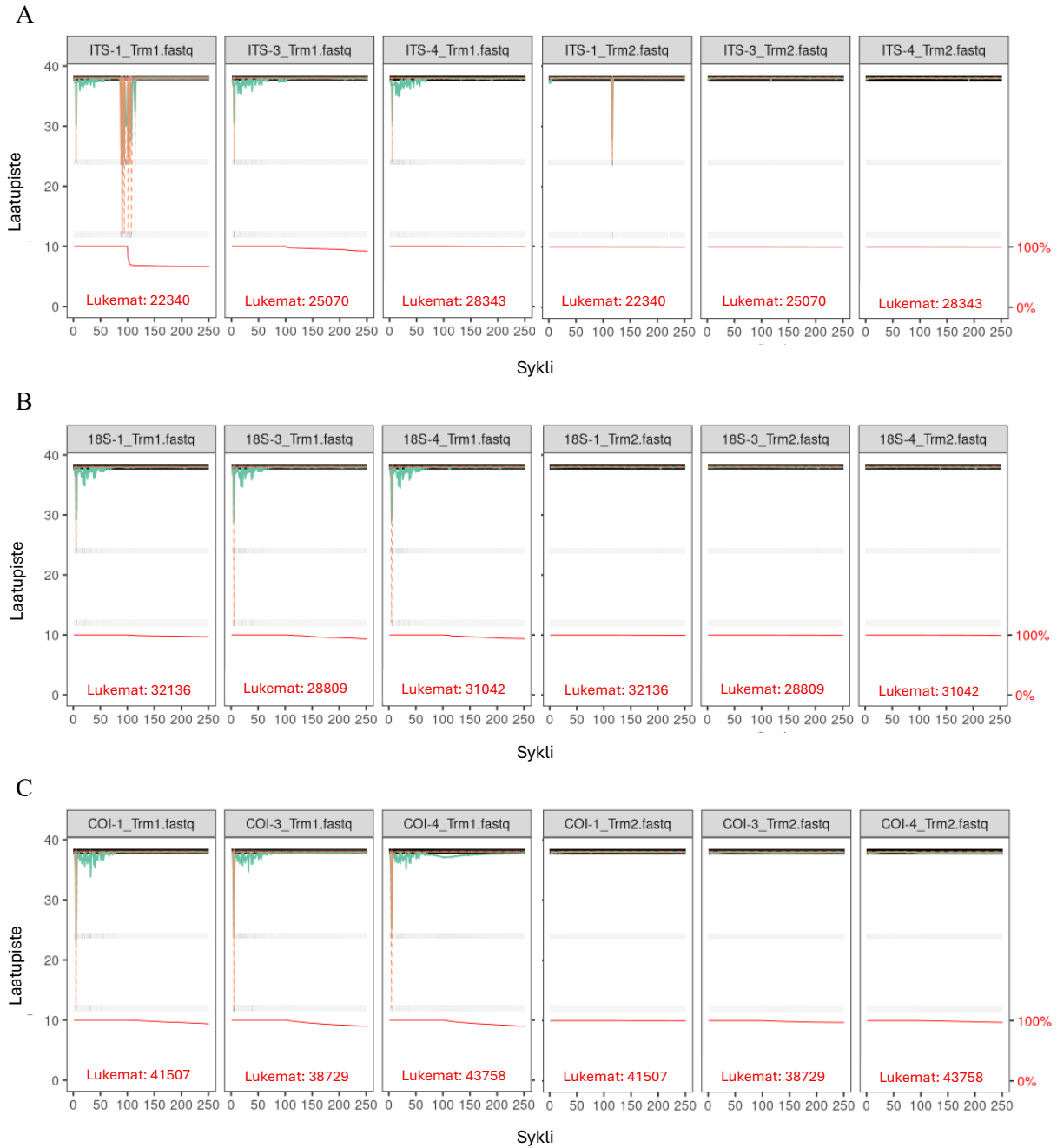
tunnistuksessa. Vertailin näitä heimotason tunnistuksia myös kesän 2024 aikana Turun Saaristomerellä tehtyihin kasvillisuus- ja eläinkartoituksiin.

3 Tulokset

Näytteiden 1 ja 3 preparoinnissa ei tehty normaalista poikkeavia havaintoja. Liejutaskurapunäytteen 2 preparoinnin aikana havaittiin kyseisen yksilön olevan normaalista poikkeavan tumma ulkoisen tukirankansa sisäpuolelta. Lisäksi yksilön suoli oli tyhjä ja sen vatsassa oli kellertävää nestemäistä massaa normaalin tummanruskean ja kiinteämmän sisällön sijasta. Kyseisestä näytteestä ei saatu sekvensointituloksia, joten sitä ei käsitelty pidemmälle tutkielmassa.

Sekvenssidatasta DADA2 käsittelyn avulla tuotettujen laatukuvaajien avulla tarkasteltiin jokaisen näytteen sekvensoinnin onnistumista. Laatukuvaajat näyttävät eteen- ja taaksepäinlukemien laadun jokaisesta näytteestä. Sekvensoinnin alussa sekvensointilaite kalibroiduu ensimmäisten 12 nukleotidin aikana, mikä havainnoidaan eteenpäinlukemien kuvaajissa vihreän ja oranssin viivan epätasaisuutena (kuvat 1 A–C). Tämän jälkeen sekvensoinnin laatu paranee. Taaksepäinlukemien sekvensoinnin aikana sekvensointilaite ei kalibroidu, joten kuvaajat ovat usein hyvälaatuisia onnistuneilla näytteillä.

Geenien *COI* ja *18S* näytteet 1, 3 ja 4 olivat hyvälaatuisia (kuvat 1 B ja C). *ITS*-geenin 1. näyte oli kuvaajan mukaan huonolaatuinen. Tämä nähdään kuvasta 1 A ensimmäisen näytteen eteenpäin- ja taaksepäinlukemien kuvaajien oranssin viivan laskemisesta sekvensoinnin keskivaiheilla. Näytteen 1 eteenpäinlukemat olivat kuvan 1 A mukaan huonolaatuisempia, kuin taaksepäinlukemat. Näytteet 3 ja 4 olivat puolestaan *ITS*-geenillä hyvälaatuisia.



Kuva 1. DADA 2 käsittelyllä tehty visualisaatio (A) *ITS*-, (B) *COI*- ja (C) *18S*-geenien sekvensointien laadusta. Kuvaajissa kolme ensimmäistä paneelia vasemmalta ovat näytteiden 1, 3 ja 4 eteenpäinlukemien (Trm1.fastq) laatukuvaajat ja kolme ensimmäistä paneelia oikealta puolelta luettuna taaksepäinlukemien (Trm2.fastq) laatukuvaajat. Kuvaajat näyttävät x-akselilla nukleotidin sijainnin sekvenssissä ja vasemmalla y-akselilla nukleotidin laatupisteen, eli todennäköisyyden sekvensoinnissa oikein luetusta nukleotidista. Vihreä viiva esittää nukleotidien keskimääräistä laatupistettä ja oranssi viiva laatupisteiden hajontaa kvartiileina. Punainen viiva kuvaa lukemien osuutta, jotka kattavat kyseisen emäksen sijainnin (y-akseli oikealla). Mitä alemmas vihreä tai oranssi viiva laskee, sitä epätodennäköisempää yksittäisen nukleotidin sekvensointi oikein on ollut. Lukemat kertovat sekvensoitujen lukemien määrän. Kuva A kuvaa *ITS*-geenialueen, B *COI*:n ja C *18S*:n sekvensointilaatua.

Analysoiduista sekvensointituloksien lukemista (taulukko 8) *COI*-geenillä oli korkeimmat eteen- ja taaksepäinlukemien määrät (34689–40520 kpl). Vastaavien *ITS*:n lukemien määrät olivat kolmesta geenistä vähäisimmät (22080–28122). *I8S*:n eteen- ja taaksepäinlukemien määrät jäivät edellisten väliin (27823–31895). Yhdistettyjen lukemien määrä verrattuna niitä vastaaviin eteen- ja taaksepäinlukemiin oli matalampi (taulukko 8). Tietokantoihin verrattujen lukemien (taulukko 8, lukemat jälkeen) määrät olivat seuraavat; *ITS*-näytteiden arvot (10210–24672) olivat suurempia kuin *COI*-geenin (9228–10915) mutta pienempiä kuin *I8S*-geenin (18379–20668) vastaavat arvot. Jokaisen tutkitun geenialueen osalta suolasaostuksella käsiteltyjen näytteiden (näytteet 4) tietokantoihin verrattut lukemamäärät olivat suurimpia.

Taulukko 8. Käsiteltyjen sekvensointitulosten lukemien määrät. Taulukon ensimmäinen sarake kertoo analysoidun geenin ja -näytteen numeron, toinen sarake eteenpäinlukemien määrän ja kolmas sarake taaksepäinlukemien määrän. Neljännessä sarakkeessa on laatutarkastettujen- ja sen jälkeen yhdistettyjen lukemien lukumäärät. Viides sarake kertoo onnistuneesti geenien tietokantoihin verrattujen lukemien määrät.

Näyte	Eteenpäinlukemat	Taaksepäinlukemat	Yhdistetyt lukemat	Lukemat jälkeen
ITS 1	22080	22144	10712	10210
ITS 3	24183	24451	19530	17975
ITS 4	27673	28122	26850	24672
COI 1	39718	40520	35695	9228
COI 3	34689	36300	28986	9597
COI 4	38745	40206	26371	10915
I8S 1	31597	31895	30180	20411
I8S 3	27823	28409	26153	18379
I8S 4	30464	30878	29132	20668

Sekvenssien laatuvaajien paljastama 1. näytteen *ITS*-geenin huonolaatuisuus (kuva 1 A) johtaa todennäköisimmin sen matalaan lukemamäärään (taulukko 8, yhdistetyt lukemat ja lukemat jälkeen). Näytteen huonolaatuisuus on johtanut siihen, että eteen- ja taaksepäinlukemien yhdistäminen on onnistunut vain alle puolissa yhdistämättömistä lukemista. Tämän vuoksi näytettä ITS1 ei ole verrattu yhtä moneen *ITS*-geenin tietokannan

sekvenssiin, kuin näytteitä ITS3 tai ITS4. Näytteen ITS1 yhdistettyjen lukemien lukumäärä oli myös matalampi verrattuna eteen- ja taaksepäinlukemien määrään kuin näytteessä ITS4, jossa sekvenssien lukumäärät pysyivät suhteellisen tasaisina jokaisessa lukemien käsittelyvaiheessa.

COI-geenin avulla saatujen lukemien määrä oli keskimääräisesti korkein eteen- ja taaksepäinsekä yhdistettyjen lukemien osalta kaikkia kolmea geeniä vertailemalla (taulukko 8). Tietokantaan verrattujen lukemien määrä oli kuitenkin kolmesta käytetystä geneistä keskimääräisesti alhaisin. *18S*-geenialueen näytteet osoittautuivat toimiviksi ja hyvää dataa tuottaviksi. Kyseisten näytteiden lukemien määrän voimakasta vähenemistä niiden yhdistämisen tai tietokantaan vertailun aikana ei havaittu, toisin kuin geenien *ITS* ja *COI* kohdalla (taulukko 8). Taulukosta 8 voidaan havaita, kuinka jokaisen kolmen geenin kohdalla näytteillä 4 oli suurimmat lukemamäärät tietokantoihin verratuissa lukemissa (taulukko 8). Näytteet 4 poikkesivat muista näytteistä DNA:n eristysmenetelmän osalta; kyseisten näytteiden DNA eristettiin suolasaostamalla, kun muiden näytteiden DNA:n eristysmenetelmänä käytettiin valmispakkausta.

Tietokantoihin verratuissa sekvensseissä oli jokaisen kolmen geenin välillä määrällisiä eroja (taulukko 8). *18S*-geenillä tunnistettiin suurin määrä heimoja (20), *COI*-geenillä seuraavaksi eniten (10) ja *ITS*-geenillä vähiten (7) (taulukko 9). *18S*:n tunnistamat heimot ovat pääasiassa erilaisia leville ja mereneläville kuuluvia heimoja. *COI*:lla tunnistettiin muun muassa planktoneille-, simpukoille, kaloille sekä äyriäisille kuuluvia heimoja. *ITS*-geenin avulla tunnistettiin heimotasolla muun muassa sammalia, alkueläimiä sekä ristikukkaiskasveja (taulukko 9).

Taulukko 9. DADA2 käsittelyn avulla tunnistetut heimotason sekvenssit: lukumäärät ja tieteelliset nimet. Ylimmällä rivillä on geenin nimi, toisella rivillä tunnistettujen sekvenssien lukumäärät ja kolmannella rivillä tunnistettujen heimojen tieteelliset nimet. *18S*-geenialue tunnisti yhteensä 20 heimoa, *COI* 10 heimoa ja *ITS* 7 heimoa.

Geeni	<i>18S</i>	<i>COI</i>	<i>ITS</i>
Lukumäärä	20	10	7
Tunnistetut heimot	Brassicales Chilopoda Chlorophyceae Chytridiaceae Dorylaimia Ectocarpales Eugregarinorida Eumalacostraca Eustigmatales Gromochytriaceae Mamiellales Mamiellophyceae Neopterygii Solanales Sphaeropleales Spumellaria Suessiaceae Tetrapoda Thoracosphaeraceae Ulvophyceae	Acartiidae Arcticidae Centropagidae Diastylidae Loliginidae Osmeridae Palaemonidae Polymorphidae Stylochidae Tellinidae	Brassicaceae Bryoxiphiaceae Grimmiaceae Notocotylidae Pottiaceae Schizoporaceae Trypanosomatidae

Geenien heimotason tunnistuksia verrattiin Saaristomerellä kesällä 2024 tehtyihin eläin- ja kasvillisuuskartoituksiin (Nousiainen 2025; Ukkonen, julkaisematon aineisto). Vertailun tuloksena tunnistettiin kaksi heimoa, jotka löytyivät sekä metaviivakoodausmenetelmällä tehdyistä tunnistuksista että kartoitusten aineistoista. *18S* tunnisti viherlevien pääjaksoon kuuluvan *Ulvophyceae* -heimon ja *COI* *Palaemonidae* -heimon, johon kuuluvat katkaravut.

Kahden heimon lisäksi *COI*-geeni tunnisti kaksi lajia, jotka olivat listattu elän- ja kasvillisuuskartoituksiin: kolmipiikki, *Gasterosteus aculeatus* ja leväkatkarapu *Palaemon adspersus*.

4 Pohdinta

4.1 Sekvenssien laatu ja määrä

Erityisesti ITS1 näytteen huonolaatuisuus on voinut johtaa näytteen eteen- ja taaksepäinlukemien yhdistämisen epäonnistumiseen. Lukemien yhdistämisen onnistumisella on havaittavissa verrannollisuus tietokantoihin verrattuiden lukemien määrään, mutta vaikutuksen suuruus vaihtelee näytteittäin. Näytteen ITS3 sekvenssien lukumäärän vähäisyys yhdistettyjen- sekä tietokantaan verrattujen lukemien määrässä voi osittain johtua samoista seikoista, kuin näytteellä ITS1. Tämä ei kuitenkaan paljastunut näytteen laatuvaajasta, joten on mahdollista, että lukemien vähenemiseen vaikuttavat muut havaitsemattomat tekijät, joita ei ole mahdollista määrittää tutkielmani aineistolla.

COI-geenialueen PCR-reaktion haastava toimivuus on luultavasti vaikuttanut kyseisen geenin näytteiden lopulliseen alhaiseen lukemien määrään. *COI*-näytteille tehtiin kolme eri PCR-reaktiota ja vastaavasti kahdelle muulle geenille riitti vain yksi. Tämä ei kuitenkaan selitä sitä, miksi näytteiden COI1, COI3 ja COI4 lukemat olivat ennen tietokantaan vertaamista kaikista näytteistä korkeimmat.

Näytteet 4 tuottivat jokaisen geenialueen kohdalla suurimmat tietokantoihin verratut lukemamäärät. Kyseiset näytteet poikkesivat muista näytteistä ainoastaan siten, että niiden DNA oli eristetty suolasaostamalla – menetelmällä, joka on usein tehokas ja valmiita eristämispaketteja edullisempi vaihtoehto genomisen DNA:n eristämiseen (Aljanabi 1997). Tutkielmani näyteköön ollessa pieni, saamani tulokset eivät kuitenkaan selitä sitä, miksi suolasaostamalla käsitellyt näytteet tuottivat korkeimmat tietokantoihin verrattujen sekvenssien lukumäärän. Näytteiden 4 sekvenssien lukumäärät eivät kuitenkaan olleet suurimpia kaikkia näytteitä verrattaessa, vaan korostuvat vain tietokantoihin verrattuiden lukemien osalta

yksittäisten geenialueita tarkasteltaessa. Suolasaostettujen näytteiden hyvä suorituskky voi olla tuloksissani sattumaa.

DADA2-menetelmä luokittelee tunnistamansa eliöryhmät hierarkkisesti taksonomisiin ryhmiin, mutta tunnistuksien tuloksia tarkastelemalla voidaan todeta niiden olevan hieman epätarkkoja. Esimerkiksi *18S*:n tunnistama Brassicales on kasvikunnan lahko ja *ITS*:n tunnistama Brassicaceae kyseisen lahkon sisälle taksonomisesti sijoittuva heimo. DADA2:n epätarkkuuden lisäksi käytetyn tietokannan tietojen tarkkuudessa voi olla virheitä, jotka voivat selittää taksonomisen hierarkian epätarkkuutta. Tutkielmassani käytän epäselvyyksistä huolimatta DADA2:n tekemää taksonomista luokittelua.

4.2 *COI*-, *ITS*- ja *18S*-geenialueiden käyttö DNA-metaviivakoodauksessa

COI-geenin käyttö DNA-metaviivakoodauksessa on hyvin yleistä, sillä sille on saatavilla eniten vertailusekvenssejä useista eri taksoneista verrattuna muihin metaviivakoodauksessa usein käytettyihin merkkigeeneihin. *COI*:n käyttö DNA-metaviivakoodauksessa on todettu kuitenkin hieman kiistanalaiseksi, sillä kyseinen geenialue ei ole eri eliöryhmien välillä erityisen konservoitunut. DNA:n monistamisen näkökulmasta tämä tarkoittaa sitä, että *COI*-geenin alukkeiden sitoutumiskohdat vaihtelevat eri lajien välillä. Näin ollen joidenkin lajien *COI*-geenialueet monistuvat tehokkaammin kuin toisten, jolloin yksittäisten taksonien monistumatta jääminen voi peittyä onnistuneesti monistuneiden lajien alle (Deagle ym. 2014).

Nukleaarinen *sisäisesti transkriptoitu välike*, *ITS*, on suunniteltu sekä paljon käytetty kasvi- ja sienikuntien lajien tunnistamiseen. Se voidaan jakaa kahteen osaan, *ITS-1*- ja *ITS-2*-alueisiin, joista käyttämäni etu- ja taka-alukkeet sijaitsevat *ITS-1*-alueella. *ITS-2*-aluetta voidaan käyttää tunnistamaan kasvilajeja sekä täydentämään *COI*-geenialueen tunnistamia eläinlajeja DNA-metaviivakoodauksessa (Yao ym. 2010). *ITS-1*-alueen on kuitenkin sanottu olevan parempi eukaryoottien tunnistamisessa (Wang ym. 2015). Kyseisessä tutkimuksessa *ITS-1*-geenialueen avulla oli tunnistettu muun muassa enemmän kalalajeja, kuin *ITS-2*-alueella. Tutkielmani tuloksissa ei kuitenkaan havaittu lainkaan eläinkunnan heimoja *ITS*-geenin avulla.

Nukleaarista *18S*-geenialuetta käytetään tunnistamaan eukaryoottisia alkueläimiä, ja sieniin sekä leviin kuuluvia lajeja. Sitä ollaan käytetty akvaattisissa ympäristöissä muun muassa kalojen ravintoverkkotutkimuksiin (Albaina ym. 2016). Lisäksi kyseisen geenialueen on todettu olevan hyvä planktoneiden tunnistamisessa (Zhao ym. 2021), mitä tutkielmani tulokset tukevat.

DNA-metaviivakoodauksen tavoitteena on tunnistaa useita eliölajeja yhdestä DNA-näytteestä. Kuten tutkielmassani, jokaisesta näytteestä tunnistettavaa lajia ei aina tiedetä etukäteen, jolloin käytettävien geenialukkeiden tulee olla suunniteltu useita lajeja sekä niiden genomeja ajatellen. Universaalialukkeet ovat suunniteltu tunnistamaan tietty geenialue monista eri lajeista. Kyseiset alukkeet sisältävät DNA:ssa olevien emästen vaihtelukohtia, joiden avulla alukkeet voivat sitoutua lajien välillä hieman konservoituneisiin geeneihin. DNA-metaviivakoodauksen tehokkuus riippuu voimakkaasti käytettyjen universaalialukkeiden toimivuudesta (Chen ym. 2025).

4.3 Heimotason sekvenssitunnistukset

Eniten heimotason tunnistuksia tehtiin *18S*-geenialueella (20), toiseksi eniten *COI*-geenialueella (10) ja vähiten *ITS*:llä (7). *18S*-geenialueen heimotason tunnistukset painoutuivat eniten selkärangattomiin- ja selkärankaisiin eläimiin, leviin, sieniin sekä kasveihin. Kyseisellä geenialueella tehtiin paljon tunnistuksia, joita voidaan olettaa esiintyvän Saaristomerellä esimerkiksi planktonissa ja joita *18S*-geenialueen tulisi tunnistaa. *COI*-geenialueen tunnistuksiin sisältyivät nilviäiset, äyriäiset, planktoneliöt, yksi kalaheimo sekä matoihin ja loisiin kuuluvia heimoja. *COI* tunnisti laajasti erilaisia eliöitä, joita voidaan *18S*:n tunnistuksiin verraten löytää meriympäristöistä. *ITS*-geenialueen avulla tunnistettiin heimotasolla muun muassa sammalia, kukkakasveja, sieni sekä yksisolainen alkueläin. Kyseiset *ITS*:n tunnistukset eivät viittaa yhtä vahvasti meriympäristöön kuin kahden muun geenialueen heimotason tunnistukset.

Jokaisen geenialueen tunnistuksissa oli huomiota herättäviä heimoja, joista lähes kaikki voidaan selittää esimerkiksi kontaminaatiolla tai erilaisten lajien kulkeutumisella meriympäristöön sattumalta. Muun muassa *ITS*:n tunnistamat sammalheimot (Bryoxiphiaceae ja Pottiaceae) esiintyvät kosteissa ympäristöissä ja täten osia kyseisistä kasveista on voinut

kulkeutua liejutaskurapujen elinympäristöön ja tutkielmani DNA-näytteisiin. *COI*-geenialueen tunnistamat Tetrapoda (nelijalkaiset) ja Brassicales (vrt. Brassicales, ristikukkaikasveja) johtuvat luultavasti samasta syystä kuin *ITS*:n tunnistamat sammalkasvit. Vaihtoehtoisesti kyseiset tunnistetut heimot voivat olla peräisin laboratorioissa DNA:n eristyksen, PCR-reaktion tai sekvensoinnin yhteydessä tapahtuneesta kontaminaatiosta. Eliöiden ravinnonkäyttöä koskevia analyysejä tehtäessä ei voida luottaa ainoastaan DNA-viivakoodauksesta saatuihin tuloksiin. Tuloksia tarkasteltaessa on olennaista ottaa huomioon tutkimuskohteen ja sen lähisukulaisten normaali ravinnonkäyttö, jonka perusteella voidaan osittain arvioida näytteiden ja tulosten luotettavuutta

Taulukoita 8 ja 9 vertailemalla havaitaan yhteys tietokantoihin verrattujen lukemien sekä tunnistettujen heimojen lukumäärien välillä. *18S*-geenialueen tietokantaan verrattujen lukemien määrän sekä heimotason tunnistusten lukumäärä oli kaikista kolmesta geenialueesta korkein. *COI*:n ja *ITS*:n tunnistamien heimojen lukumäärät ja tietokantoihin verrattujen lukemien määrät olivat runsaasti matalampia, pois lukien *ITS4* -näytteen tietokantaan verratut lukemat, joita oli enemmän kuin minkään muun näytteen lukemia tietokantoihin vertaamisen jälkeen.

4.4 Geenialueiden verrannollisuus

Tutkielmassani käytetyt kolme geenialuetta tunnistivat liejutaskurapujen ruoansulatuskanavan DNA-näytteistä eri heimoja, eikä geenialueiden tuottamissa sekvensointituloksissa ollut päällekkäisyyksiä. Tulosten perusteella voidaan olettaa, että *COI*-, *ITS*- ja *18S*-geenialueet antavat informaatiota eri eliöryhmistä ja näin ollen kaikki kolme geenialuetta ovat hyödyllisiä DNA-viivakoodaukseen perustuvissa ravinnonkäyttöanalyseissa. Geenialueiden toimivuutta on kuitenkin suositeltavaa testata tapauskohtaisesti eri tutkimuksissa pilottitutkimuksen tapaisesti. Jatkotutkimuksissa käytettävät geenialueet voidaan valita niiden antaman informatiivisuuden perusteella vertailemalla tulosten laatua aiempiin tutkimuksiin, jotka ovat perustuvat joko tutkimuslajiin tai sen lähisukulaisiin.

Kolmesta geenistä *COI*-geenialue oli informatiivisin heimotason tunnistuksessa sekä heimomäärän että tunnistettujen heimojen perusteella. Kyseinen geeni tunnisti eniten akvaattisissa ekosysteemeissä tavattavia heimoja. Aiemman tutkimuksen mukaan kyseisten eliöiden voidaan olettaa olevan osa liejutaskurapujen käyttämiä ravintokohteita (Aarnio ym. 2015). *18S*-geenialueen tunnistuksia voidaan verrata heimojen perusteella *COI*:n tunnistamiin heimoihin. Ne ovat niin ikään akvaattisia eliöitä ja niiden voidaan olettaa olevan osa liejutaskuravun ravintoa.

ITS-geenialueen avulla tunnistetut heimot eivät viittaa yhtä vahvasti meriympäristöissä esiintyviin eliöihin kuin kahdella muulla geenialueella tunnistetut heimot. *ITS*:n tunnistamista heimoista ei vastaavasti ole mainintoja aiemmissa liejutaskuravun ravinnonkäyttöä koskevissa tutkimuksissa. Saamieni tulosten perusteella voin todeta *ITS*-geenialueen olevan *COI*:ta sekä *18S*:ää tehottomampi liejutaskuravun ravinnonkäytön selvittämisessä DNA-metaviivakoodauksen avulla.

Aiemmissa DNA-metaviivakoodauksen avulla tehtyjen akvaattisten eläinlajien ravinnonkäyttötutkimuksissa käytetyt geenialueet ovat lähes kaikissa tapauksissa olleet *COI*-geenialueita. *COI*:n lisäksi useassa tutkimuksessa ollaan käytetty, tai ollaan jatkoa ajatellen ehdotettu käytettäväksi, myös *18S*-geenialuetta (Barbato ym. 2019, Batta-Lona ym. 2024). Tutkielmani tulosten sekä aiempien tutkimusten pohjalta *ITS*-geenialueen käyttö ei ole yhtä perusteltua kuin näiden kahden muun geenin käyttö. Näin ollen jatkossa tehtävissä liejutaskurapujen ravinnonkäyttöanalyysissä olisi hyvin perusteltua hyödyntää metaviivakoodauksessa vain *COI*- ja *18S*-geenialueita, käyttämättä lainkaan *ITS*-geenialuetta.

Kiitokset

Kiitokset Heidi Viitaniemelle LuK-tutkielmani ohjaamisesta, avusta näytteiden preparointiin, laboratoriotyöhön ja näytteiden lähettämisestä sekvensointiin sekä opetuksesta tilastojen käsittelyyn liittyvissä asioissa. Kiitos projektin rahoituksesta Suomen Kulttuurirahaston Varsinais-Suomen maakuntarahastolle. Kiitokset Veijo Jormalaiselle ja hänen tutkimusryhmälleen ideoista ja tuesta tutkielmaani koskien.

Lähdeluettelo

- Aarnio, K., Törnroos, A., Björklund, C. and Bonsdorff, E. 2015. Food web positioning of a recent coloniser: the North American Harris mud crab *Rhithropanopeus harrisi* (Gould, 1841) in the northern Baltic Sea. - *Aquatic Invasions* 10: 399–413.
- Abarenkov, K., Zirk, A., Piirmann, T., Pöhönen, R., Ivanov, F., Nilsson, R. H. and Kõljalg, U. 2025. UNITE general FASTA release for eukaryotes.
- Albaina, A., Aguirre, M., Abad, D., Santos, M. and Estonba, A. 2016. 18S rRNA V9 metabarcoding for diet characterization: a critical evaluation with two sympatric zooplanktivorous fish species. - *Ecology and Evolution* 6: 1809–1824.
- Aljanabi, S. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. - *Nucleic Acids Research* 25: 4692–4693.
- Barbato, M., Kovacs, T., Coleman, M. A., Broadhurst, M. K. and De Bruyn, M. 2019. Metabarcoding for stomach-content analyses of Pygmy devil ray (*Mobula kuhlii cf. eregoodootenkee*): Comparing tissue and ethanol preservative-derived DNA. - *Ecology and Evolution* 9: 2678–2687.
- Batta-Lona, P. G., Wojcicki, M. L., Zavell, M. D., Bucklin, A. and Baumann, H. 2024. Using DNA metabarcoding to reveal prey diversity in diets of juvenile black sea bass (*Centropristis striata*) in Long Island Sound in the Northwest Atlantic Ocean. - *Fisheries Bulletin* 123: 22–33.
- BOLD systems, Baltic Sea, COI. 2024.
[https://portal.boldsystems.org/result?query=%22Baltic%20Sea%22\[geo\]](https://portal.boldsystems.org/result?query=%22Baltic%20Sea%22[geo]) [Luettu 14.10.2025]
- Chen, X., Yan, Z., Li, S. and Yao, M. 2025. Advancing aquatic biodiversity assessments of invertebrates using EDNA metabarcoding: A systematic evaluation of primers for marine and freshwater communities. - *Methods in Ecology and Evolution* 16: 2408–2430.
- Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F. and Taberlet, P. 2014. DNA metabarcoding and the cytochrome *c* oxidase subunit I marker: not a perfect match. - *Biology Letters* 10: 20140562.
- Fowler, A., Forsström, T., Von Numers, M. and Vesakoski, O. 2013. The North American mud crab *Rhithropanopeus harrisi* (Gould, 1841) in newly colonized Northern Baltic Sea: distribution and ecology. - *Aquatic Invasions* 8: 89–96.
- Hegele-Drywa, J. and Normant, M. 2009. Feeding ecology of the American crab *Rhithropanopeus harrisi* (Crustacea, Decapoda) in the coastal waters of the Baltic Sea. - *Oceanologia* 51: 361–375.
- Jormalainen, V., Gagnon, K., Sjöroos, J. and Rothäusler, E. 2016. The invasive mud crab enforces a major shift in a rocky littoral invertebrate community of the Baltic Sea. - *Biological Invasions* 18: 1409–1419.

- Nousiainen, E. 2025. The effects of a novel invasive predator (*Rhithropanopeus harrisi*) on rocky littoral invertebrate community in the Northern Baltic
- Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N. and Taberlet, P. 2012. Who is eating what: diet assessment using next generation sequencing. - *Molecular Ecology* 21: 1931–1950.
- Reusch, T. B. H., Dierking, J., Andersson, H. C., Bonsdorff, E., Carstensen, J., Casini, M., Czajkowski, M., Hasler, B., Hinsby, K., Hyytiäinen, K., Johannesson, K., Jomaa, S., Jormalainen, V., Kuosa, H., Kurland, S., Laikre, L., MacKenzie, B. R., Margonski, P., Melzner, F., Oesterwind, D., Ojaveer, H., Refsgaard, J. C., Sandström, A., Schwarz, G., Tonderski, K., Winder, M. and Zandersen, M. 2018. The Baltic Sea as a time machine for the future coastal ocean. - *Science Advances* 4: eaar8195.
- Silva v132 eukaryotic 18S tietokanta DADA2-pipelinelle. 2025.
<https://zenodo.org/records/10444891> [Luettu 10.10.2025]
- Suomen luonnonsuojeluliitto, vieraslajit. 2025. <https://www.sll.fi/opi-lisaa/vieraslajit/> [Luettu 4.4.2025]
- Wang, X., Liu, C., Huang, L., Bengtsson-Palme, J., Chen, H., Zhang, J., Cai, D. and Li, J. 2015. ITS 1: a DNA barcode better than ITS 2 in eukaryotes? - *Molecular Ecology Resources* 15: 573–586.
- Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., Pang, X., Xu, H., Zhu, Y., Xiao, P. and Chen, S. 2010. Use of ITS2 Region as the Universal DNA Barcode for Plants and Animals (B Hansson, Ed.). - *PLoS ONE* 5: e13102.
- Zhao, L., Zhang, X., Xu, M., Mao, Y. and Huang, Y. 2021. DNA metabarcoding of zooplankton communities: species diversity and seasonal variation revealed by 18S rRNA and COI. - *PeerJ* 9: e11057.