



**TURUN  
YLIOPISTO**

Matemaattis-luonnontieteellinen  
tiedekunta

# **Leväsiirujen (*Idotea balthica*) ilmiasun vaihtelun ja väripolymorfian geneettinen tausta**

Joonas Jäntti

Ekologia ja evoluutiobiologia

Pro gradu -tutkielma

Laajuus: 30 op

Ohjaajat:

Sami Merilaita

Heidi Viitaniemi

27.1.2023

Turku

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

Pro gradu -tutkielma

**Pääaine:** Ekologia ja evoluutiobiologia

**Tekijä:** Joonas Jäntti

**Otsikko:** Leväsiirujen (*Idotea balthica*) ilmiasun vaihtelun ja väripolymorfian geneettinen tausta

**Ohjaajat:** Sami Merilaita, Heidi Viitaniemi

**Sivumäärä:** 38 sivua

**Päivämäärä:** 27.1.2023

---

Tutkielman tarkoituksena oli selvittää tekijöitä leväsiiran (*Idotea balthica*) väriytyksen polymorfian taustalla. Täsmensin tutkimuksen sisältämään vain yksivärisen *uniformis*- ja laikukkaan *albafusca*-morfian. Keskeiset tutkimuskysymykset tutkielmalle olivat: onko väritys yhden geenin säätelemä vai polygeenisesti säädelty, ja onko sukupuolten tai populaatioiden välillä huomattavia geneettisiä eroja? Aineistoa kerättiin useista sijainneista Saaristomerellä ja usealta vuodelta. Leväsiiroille toteutettiin kuva-analyysi, jossa leväsiirujen väriytyksestä mitattiin valkean pinta-alan osuutta ja laikkujen määrää. Lisäksi leväsiirayksilöille toteutettiin koko genomin laajuinen assosiaatiotutkimus (GWAS), jolloin yksilöiden välistä ilmiasun ominaisuuksien vaihtelua voitiin verrata geneettiseen vaihteluun. Genomin sekvensointia varten kustakin populaatiosta otettiin otos, joka koostui 20 yksilöstä, joista puolet oli koiraita ja toinen puolikas naaraita. Populaatioiden välinen vaihtelu valkoisen väriytyksen suhteellisessa alueessa ja laikkujen lukumäärässä oli tutkitussa aineistossa vähäistä. Laikullisilla yksilöillä laikkujen määrä erosi sukupuolilla, mutta tämä ero riippui myös populaatiosta. Geneettisiä menetelmiä käyttäen leväsiirayksilöiltä selvitettiin väriytyksen osiin vaikuttavat merkitsevät genomialueet ja bioinformatiikan menetelmillä selvitettiin näihin alueisiin liittyviä genejä ja niiden toimintaa. Tämän pro gradun tutkimusten mukaan näyttäisi siltä, että geneillä on merkitystä siirujen väriytykseen, etenkin valkoisen pinta-alalliseen osuuteen. Leväsiirujen laajan kirjon väriytykset, jopa populaation sisällä ja yksilöiden välillä viittaa siihen, että väritys on polygeenisesti säädelty.

---

**Avainsanat:** *Idotea balthica*, väripolymorfia, GWAS, fenotyyppi, geneettinen muuntelu

## Sisällys

1	Johdanto .....	1
2	Tutkimusmenetelmät .....	6
2.1	Menetelmien yleiskuvaus .....	6
2.2	Aineiston kerääminen .....	7
2.3	Aineiston valokuvaaminen .....	9
2.4	Kuva-analyysi .....	9
2.5	Väriytyksen ominaisuuksien analyysit .....	12
2.6	DNA:n eristys .....	12
2.7	DNA:n sekvensointi .....	13
2.8	Genomin linjaus .....	15
2.9	GWAS analyysi .....	16
2.10	Merkitsevien SNP:ien annotaatio .....	17
3	Tulokset .....	18
3.1	Värimorfien fenotyypin vaihtelu .....	18
3.2	Ilmiasun ja genomin assosiaatioanalyysi .....	21
3.3	Bedtools- ja BLAST-analyysi .....	26
4	Pohdinta .....	28
4.1	Väriytyksen ilmiasun vaihtelu .....	28
4.2	Väriytyksen ilmiasun ja genomin assosiaatio .....	31
4.3	Päätelmät .....	33
	Lähteet .....	35

# 1 Johdanto

Polymorfialla tarkoitetaan, että populaatiossa tavataan vähintään kahta tai useampaa morfologisesti selkeästi toisistaan eroavaa geneettisesti määräytynyttä yksilönmuotoa (Ford, 1945). Tällä tarkoitetaan, että populaation yksilöt voidaan jakaa jonkin tai joidenkin ominaisuuksien mukaan luokiteltuihin ryhmiin. Tällainen ryhmä on morfi. Polymorfian määritelmän mukaan morfit esiintyvät sellaisissa lukusuhteissa, että harvinaisinkaan näistä ei ole vain toistuvien mutaatioiden ylläpitämä (Ford, 1945).

Ford (1945) jakoi polymorfian kahteen kategoriaan: 1) ohimenevään (engl. transient) polymorfiaan, ja 2) tasapainotettuun (engl. balanced) polymorfiaan. Ohimenevällä polymorfialla tarkoitetaan sitä, että kahdesta genotyypistä yksi hiljalleen syrjäyttää toista, eli luonnonvalinta hävittää vähitellen toisen genotyypin populaation geenipoolista. Esimerkki tästä voisi olla tummaväritteisen morfin yleistyminen ja vaalean harvinaistuminen, kun uudet ympäristöolosuhteet suosivat tumman suojaväriin omaavia yksilöitä. Tilanne on nimensä mukaisesti ohimenevää, sillä molemmat morfit esiintyvät populaatiossa jonkin ohimenevän ajan yhdessä, ennen kuin vanhaan ympäristöön sopeutunut morfi on kokonaan syrjäytynyt. Tasapainotettu polymorfia taas tarkoittaa sitä, että vähintään kaksi morfia säilyy populaation geenipoolissa valinnan seurauksena. Yksi esimerkki tasapainotetusta polymorfiaista on geneettisesti määräytynyt sukupuoli (Ford, 1945). Frekvenssistä riippuvan valinnan seurauksena molempia sukupuolia tuotetaan suunnilleen yhtä paljon, vaikka lajista ja tilanteesta riippuen suhdeluku voi vaihdella huomattavasti.

Polymorfia ei usein ole yksinkertaisesti todennettavissa. Monissa tapauksissa polymorfia – vaikkakin usein määritelty diskreetiksi vaihteluksi ja selkeiksi eroiksi morfiin välillä – sisältää myös tosiasiaa jatkuvaa muuntelua. Eli termiä polymorfia käytetään usein, vaikka kyseessä ei olisi tarkasteltavan ominaisuuden fenotyypin epäjatkuvaa vaihtelua. Tämä voi johtua siitä, että harvan ominaisuuden fenotyypissä ei ole lainkaan muuntelua, ja myös siitä, että tutkijat ovat tutkimustaan helpottaakseen voineet keinotekoisesti luokitella ominaisuuksia, jotka todellisuudessa ovat jatkuvia (Pener, 1991). Tästä esimerkki on mustahuulitarhakotilo (*Cepaea nemoralis*). Kotiloilla on pohjaväriin vaaleanpunainen, keltainen tai ruskea. Tästä huolimatta joissain tapauksissa on vaikea erottaa mihin pohjaväriin kotilon väri kuuluu, ja värit ovat usean sävyisiä. Kotilot on silti jaettu määritettyihin morfeihin, mutta rajatapauksia ilmenee, eikä aina voi suoraan sanoa yhden yksilön edustavan puhtaasti yhtä morfia (Davison, 2019). Perinteinen käsitys morfeista on, että jokin yksi lokus

määrää yksilön fenotyypin niin vahvasti, ettei jatkuvuutta esimerkiksi värityksessä synny (Ford, 1945). Värityksen jatkuvuus näillä kotiloilla – ja esimerkiksi tutkielmani kohde-eliöllä, leväsiiroillakin (*Idotea balthica*)– herättää kysymyksen siitä, onko tämä polymorfian määritelmä toimiva ja pätevä todellisuudessa, jossa yksilöt edustavat laajaa kirjoa värityksessään ja fenotyypissään.

Eliöiden fenotyypin ominaisuudet voivat olla yhden tai useamman geenin säätelemiä. Esimerkiksi puna-apinankukilla (*Erythranthe lewisii*) ja (*E. cardinalis*) väritys on yhden lokuksen määräämä. Väritykseen vaikuttavat myös näitä pölyttävien mehiläisten ja kolibrien aiheuttama valintapaine, sillä *E. lewisii* houkuttaa vain mehiläisiä, kun taas *E. cardinalis* houkuttaa kolibreja. Lajien erilaistumista on tapahtunut kasvien evolutiivisen historian aikana muun muassa maantieteellisen eriytymisen myötä (Bradshaw & Douglas, 2003). Alun perin keltainen väritys oli kukan geenin mutaatio, mutta väritys houkutti mehiläisiä ja karkotti kolibreja synnyttäen pölyttäjäsolaation. Kumpikin värimorfi on yhtä kelpo, mutta erikoistunut eri pölyttäjien houkutteluun pölytyksen maksimoimiseksi. Värimorfit eivät kuitenkaan aina ole yksiselitteisesti yhden geenin määräämiä, sillä esimerkiksi banaanikärpäsen (*Drosophila melanogaster*) väritys on usean geenin säätelemä, missä useat rakennegeenit toimivat pigmentin syntetisoinnissa, ja sääntelygeenit pigmentin kuvion syntetisoinnissa (Protas & Nipam, 2008).

Klassinen esimerkki polymorfista on teollisuismelanismin aiheuttama väripolymorfia (Kettlewell, 1955). Tällä tarkoitetaan tilannetta, jossa saasteet tummentavat ympäristöä, jolla eliön vaaleampi kätkevä suojaväri on ennen toiminut. Tällöin suurempi osa aikaisemmin huonosti pärjääviä tumman suojaväriin omaavia yksilöitä selviää muuttuneessa ympäristössä paremmin, ja tämä väritys yleistyy. Usein esimerkkinä käytetään koivumittaria (*Biston betularia*), joka esiintyi pääasiallisesti vaaleansävyisenä ennen teollistumista. Koivumittarista esiintyy myös tummaa morfia. Normaalisti vaalea morfi on yleisemmin esiintyvä morfi, jota populaatiossa esiintyy eniten, sillä tämä suojaväri auttaa koivumittaria piiloutumaan saalistajilta. Vaalea väritys muistuttaa paljon koivurungon väritystä, mitä vasten ollessaan koivumittari ei helposti erotu saalistajille. Tumman muodon tuottava alleeli on dominoiva ja vaalean morfin tuottava alleeli on resessiivinen. Tumman morfin yksilöt sen sijaan erottuvat helpommin ja päätyvät saaliiksi useammin, jolloin kyseisen värityksen geenit vähenevät populaatiossa, eikä tätä morfia esiinny suhteessa yhtä paljon. Jos teollisuuden nokipäästöt värjäävät ympäristöä, jolloin

esimerkiksi aiemmin vaaleat koivunrungot nyt värjäytyvät tummiksi, ei koivumittarin vaalea suojaväri enää toimi yhtä tehokkaasti. Nyt tilanne on päinvastainen, jolloin tumman morfin yksilöiden suojaväri toimiikin tehokkaammin, ja tämän morfin geenit yleistyvät populaatiossa (Kettlewell, 1955; Cook et al. 2012).

Väripolymorfiaa havaitaan monilla lajeilla, hyvin erilaisissa eliöryhmissä, ja monista eri syistä. Lisäksi morfien yleisyys, eli näiden suhteellinen määrä verrattuna muihin morfeihin voi vaihdella huomattavasti saman lajin populaatioidenkin välillä. Koivumittareilla (Kettlewell, 1955), okasirkkoilla (*Tetrix subulata*) (Forsman, 2011), ja leväsiiroilla (Salemaa, 1978; Jormalainen et al. 1995; Merilaita, 2001) väripolymorfia liittyy eliöiden suojavärytykseen ja edesauttaa yksilöiden selviytymistä vaihtelevissa ympäristöissä. Leväsiiroilla esiintyy väripolymorfiaa, joka toimii suojavärinä visuaalisesti saalistavia kaloja vastaan (Merilaita, 2001). Näin eri morfit pärjäävät joissakin ympäristöissä muita morfeja paremmin, mikä saattaa edesauttaa monimuotoisuuden säilymistä. Väripolymorfia voi täten liittyä suojavärytykseen eliöiden esiintyessä vaihtelevissa ympäristöissä, edistäen näiden selviytymistä.

Suojavärytyks ei kuitenkaan ole ainoa polymorfian hyöty. Esimerkkinä mainittakoon rantaokasirkkojen lisääntynyt uusien populaatioiden perustamismenestys (engl. establishment success): on todennäköisempää, että uuden alueen asuttaminen pienellä perustajapopulaatiolla onnistuu, kun joukossa on suurempi määrä geneettisesti eroavia värimorfeja (Forsman, 2016).

Väripolymorfiaa – ja polymorfiaa yleensäkin – esiintyy useista eri syistä, kuten mustahuulitarhakotiloilla saalistuksen ylläpitämänä (*Cepaea nemoralis*) (Lamotte, 1959; Jones et al. 1980; Surmacki et al. 2013), jalokivihämäkeillä saalistusmenestyksen ylläpitämänä (*Gasteracantha fornicata*) (Kemp, 2013), tai muun muassa seksuaalivalinnan ylläpitämänä mansikkasammakoilla (*Oophaga pumilio*) (Galeano & Harms, 2016). Kaiken kaikkiaan väripolymorfia kuten myös polymorfia itsessään on jo laajalti tutkittu ilmiö, mutta käsitettä ei ole – etenkin nykytutkimusten varjolla – pystytty täsmentämään täysin kattavaksi. On oletettavaa, että polymorfia säilyy populaatioissa syystä. Aiemmin mainitusti, polymorfian määritelmän mukaan ehtona on, että se ei säily populaatiossa vain toistuvien mutaatioiden seurauksena. Jos jokin morfi säilyy populaatiossa mutaatioista

riippumatta, tarkoittaa se, että kyseinen morfi on riittävän elinkelpoinen ja kilpailukykyinen muihin morfeihin nähden (Ford, 1945).

Polymorfiasta on laajalti keskusteltu alkaen Fordin (1945) artikkelissa, ja sitä on täsmennetty yksityiskohtaisemmin, esimerkiksi lajikohtaisesti tuoreemmissa tutkimuksissa, kuten Whiten ja Kempin (2016) ja Davisonin (2019) artikkeleissa. Polymorfia ei yleisesti ole geneettiseltä taustaltaan niin yksinkertainen ilmiö kuin miltä se usein saadaan kuulostamaan, kuten klassisen koivumittari esimerkin kanssa (Kettlewell, 1955; Cook et al. 2012). Vaikkakin yhden geenin säätelemä polymorfia on mahdollista - kuten puna-apinankukilla (Bradshaw & Douglas, 2005) – on mahdollista, että morfeja ei aina ylläpidä tai säätele vain yksi tekijä. Joissain tapauksissa vaihtelun jatkuvuuden vuoksi morfien rajat voivat häilyä ja tutkijat joutuvat tekemään päätöksiä, kuinka yksilöt luokitellaan ja kuinka eri morfot rajataan. Vastaavaa huomasin henkilökohtaisesti leväsiiroja tutkiessani, missä tutkijan mukaan, siirujen värityksen luokitus saattoi erota henkilökohtaisen mielipiteen vuoksi.

Vaikkakin polymorfia on tunnettu ja tiedetty ominaisuus monilla lajeilla, on tutkimushaara tähän aiheeseen suhteellisen nuori. Ford oli ensimmäinen, joka esitti polymorfialle täsmällisen määritelmän, ja kuluneiden vuosikymmenten aikana tutkimukset polymorfian tekijöihin, sen vaikutuksiin, ja eri ilmenemistapoihin – kuten väripolymorfiaan – ovat jatkuneet ja lisääntyneet. Tutkimuksissa on edelleen vaikeuksia esimerkiksi morfien diskreetissä luokittelussa, mikä kuitenkin monilla lajeilla on tosiasiaa häilyvärajainen morfien sisäisen vaihtelun takia.

Pro graduni tarkoituksena on selvittää väripolymorfian määräytyminen leväsiiroilla ja siten lisätä myös polymorfian käsitteellistä ymmärtämistä. Leväsiirat (Kuva 1) ovat äyriäisiin (*Crustacea*) luettu laji, jota esiintyy merissä ja etenkin rannikkojen läheisyydessä kasvillisuuden suojassa. Etenkin rakkolevät (*Fucus vesiculosus*) tarjoavat yleisesti elinympäristön, suojaa sekä ravintoa leväsiiralle (Jormalainen et al. 1995). Leväsiirujen lisäksi Itämeren alueella tunnetaan kaksi muutakin *Idotea*-lajia, hoikkaleväsiira (*I. chelipes*), jota esiintyy tyypillisesti sekä Saaristomerellä että Selkämeren alueilla, ja tyrskyleväsiira (*I. granulosa*), jota esiintyy etenkin Selkämeren alueella. Aikuiset leväsiirat ovat tyypillisesti noin 10–30 mm kokoisia. Naaraat ovat koiraita tyypillisesti pienempiä (Leidenberger, et al. 2012).

Leväsiiroja pidetään väritykseltään polymorfisena. Näiden värityksen ilmiäsu vaihtelee huomattavasti populaatioiden välillä ja paikallispopulaatioidenkin sisällä (Buinheim et al. 1982; Guarino et al. 1993). Tälle mahdollinen selitys voi olla saalistuksen välttäminen, sillä leväsiiroja esiintyy alueilla, joilla esiintyy myös näiden saalistajia. Leväsiirojen väritys on kehittynyt suojaamaan näitä visuaalisesti saalistavilta kaloilta (Merilaita, 2001). Leväsiirojen väritys muistuttaa paljon niiden habitaatin väritystä – kuten esimerkiksi rakkolevää, merirokkoja (*Amphibalanus improvisus*) ja meriajokasta (*Zostera marina*) – jota leväsiirat käyttävät suojanaan ja ravintonaan. Kätkevä suojaväritys on yleinen saalistukselta suojaava ominaisuus (Bond, 2007; Forsman, et al. 2011). Väripolymorfia voi täten liittyä suojaväritykseen eliöiden esiintyessä vaihtelevissa ympäristöissä edistäen näiden selviytymistä.

Leväsiirojen pohjaväritys vaihtelee vaalean vihreästä tai ruskeasta tumman vihreään ja ruskeaan, ja naarailta jopa täysin mustaan. Koiraiden pohjaväri voi olla hyvin tumma, mutta läheisesti tarkasteltuna ei kuitenkaan täysin musta (Jormalainen, Merilaita & Tuomi, 1995). Itämerellä tyypillisesti esiintyvät leväsiirojen värimorfit on luokiteltu pohjaväriä rikkovan vaalean kuvioinnin mukaan seuraavasti Salemaan (1978) artikkelissa: *uniformis*, eli pääasiallisesti yksivärinen; *albafusca*, eli vaalea laikkuus; ja *maculata*, eli kauttaaltaan vaalean pistemäisen kuvion rikkoma pohjaväritys. Risteytyskokeilla on osoitettu, että morfit ovat geneettisesti määräytyneitä (Hamelin, 1963; Legrand-Hamelin and Legrand, 1982a,b).

*Albafuscan* ja *maculatan* ilmiäsuissa, esimerkiksi laikkujen määrässä, koossa ja muodossa on vaihtelua yksilöiden välillä merkittävässä määrin. Esimerkiksi joillakin *albafusca*-yksilöillä on useita pieniä valkeita laikkuja, kun taas toisilla valkeat laikut peittävät jopa yli puolet näiden ruumiista, ja laikut ovat pikemminkin kokonaisia valkeita alueita laikkujen sijasta. Pitkittäisraidalliset morfit *lineata* ja *bilineata* ovat meillä harvinaisempia, ja niillä ei raitojen ilmiäsuissa vaikuta olevan vaihtelua yksilöiden välillä. *Lineata* morfilla kulkee yksittäinen suora raita keskiselkää pitkin, ja *bilineata* morfilla raita kummankin kyljen mukaisesti (Kuva 1). Yksilöillä voi olla myös useiden morfiominaisuuksien yhdistelmiä, kuten *albafusca-maculata* tai *lineata-bilineata*. Minun tutkimukseni keskittyy ainoastaan *uniformis* ja *albafusca* morfeihin, eli kuviottomiin ja laikullisiin yksilöihin.





Kuva 1 – *I. balthican* Itämeren morfien. Ylä-vasen: *uniformis*, ylä-oikea: *albafusca*, ala-vasen: *lineata-bilineata*, ala-oikea: *maculata*.

Leväsiirujen monimutkaisen värityksen ja kuvioinnin vuoksi on oletettavaa, että näiden väritys ei ole yhden geenin säätelemää. Joidenkin morfien ilmiäsuissa on laajaa vaihtelua, mikä tekee epätodennäköiseksi sen, että kaikki vaihtelu on vain yhden tekijän muokkaamaa. Morfien sisällä värityksessä on suurta vaihtelua kuvioinnissa, minkä vuoksi on mahdollista, että väritykseen vaikuttavat useat geneettiset tekijät – ja mahdollisesti myös ympäristötekijät. Aikaisempien risteytyskokeiden perusteella on tehty se johtopäätös, että leväsiiran värityksen polymorfia perustuu jokaista morfia vastaavaan yhteeseen morfin ilmenemistä säätelevään lokukseen (Hamelin, 1963; Legrand-Hamelin and Legrand, 1982a,b).

Tutkimukseni päätarkoituksena oli selvittää, onko leväsiiran väripolymorfia yhden vai monen geneettisen tekijän määräämä piirre. Lisäksi selvitin sukupuolten ja populaatioiden välisiä värityksellisiä eroja. Näitä tarkoituksia varten selvitettiin geneettinen vaihtelu useassa leväsiirapopulaatiossa. Kuva-analyysi mahdollisti ilmiäsujuen mittaamisen *uniformis*- ja *albafusca*-yksilöillä ja vertaamisen sukupuolten ja populaatioiden välillä. Geneettisen tutkimuksen tuloksia voidaan yhdistää kuva-analyysiin ja siten selvittää ilmiäsujuen vaihtelun yhteyttä geneettiseen muunteluun.

## 2 Tutkimusmenetelmät

### 2.1 Menetelmien yleiskuvaus

Useista eri populaatioista kerätyiltä leväsiiroilta määritettiin sukupuoli sekä morfi. Siirat valokuvattiin ja kuva-analyysiohjelmaa käyttäen mittasin näiltä muun muassa yksilön koon, valkoisten laikkujen lukumäärän ja valkoisen pinta-alan osuuden

selkäpuolen pinta-alasta. Lisäksi siiroille toteutettiin genomilaajuinen assosiaatiotutkimus (Genome-Wide Association Study eli GWAS), jolla nimensä mukaan voidaan tutkia yksilöiltä koko genomia laajuisesti geneettisiä variantteja ja selvittää, onko variantti yhteydessä yksilön fenotyypin ominaisuuteen. Näitä yhdessä käyttäen pystyn yhdistämään siirujen värityksen ominaisuuksia geneettisen vaihteluun. Tähän tutkimukseen käytettävä leväsiira-aineisto on kerätty ensisijaisesti toista tutkimusta varten, mistä syystä aineistoa on kerätty useina vuosina, ja mikä on vaikuttanut tutkimukseen sisällytettävien populaatioiden valintaan ja otostamiseen.

## 2.2 Aineiston kerääminen

Vuonna 2021 keräsin tutkimusaineistoa Saaristomereltä sekä Selkämereltä touko-kesäkuussa. Selkämerellä keräsin aineistoa Uudenkaupungin alueelta Pyhämaasta (PYH). Saaristomerellä keräsin aineistoa kolmesta paikasta, jotka olivat Stenskär (STE), Berghamn (BER) ja Sundskär (SUN). Aineistoa tähän tutkimukseen on saatu myös aikaisemmin kerätyistä näytteistä vuosilta 2014 ja 2020, joiden keräystapa on sama kuin vuoden 2021 näytteillä. Vuoden 2014 näytteet ovat Katavaluodosta (KAT), ja 2020 näytteet Rihtniemestä (RIH), Boskäristä (BOS), Gulskäristä (GUL), ja Jurmosta (JUR). Tarkemmat tiedot on sisällytetty Taulukkoon 1 koiraiden (M) ja naaraiden (F) osin. Sisällytin aineistooni siiroja useista eri paikallispopulaatioista, jotta voin analyysissä verrata populaatioiden välisiä eroja.

Keräyspaikoilla saarten rannan läheisyydestä etsittiin rakkoleväkasvustoa, ellei tietoa kasvupaikasta ollut tiedossa etukäteen. Rannan läheisyydestä noin 0,5–1,5 metrin syvyydestä rakkoleväpuskia nostettiin haravia käyttäen veneen kyydissä oleviin saaveihin. Leväsiirat kerättiin tunnustelemalla ja ravistelemalla tai hakkaamalla rakkoleväpuskia tyhjien saavien seiniä vasten, jolloin näissä puskipissa olevat leväsiirat putoivat saavin pohjalle, josta ne kerättiin talteen merivedellä täytettyyn ämpäriin. Kerätyt näytteet säilöttiin 70-prosenttiseen etanoliliuokseen (EtOH) ja säilytettiin +4°C valokuvausta ja geneettisiä analyysejä varten.

Taulukko 1 – Valokuvatut ( $N_{total}$ ) sekä sekvensoidut näytemäärät ( $N_{sekvensoitu}$ ) per populaatio ja sukupuoli

Populaatio	Koordinaatit	Sukupuoli	$N_{total}$	$N_{sekvensoitu}$
BER	60°02,94N 21°48,96E	F	32	12
BER	60°02,94N 21°48,96E	M	36	10
BOS	60°01,97N 21°46,61E	F	23	10
BOS	60°01,97N 21°46,61E	M	40	10
GUL	60°07,16N 22°07,00E	F	18	10
GUL	60°07,16N 22°07,00E	M	24	10
JUR	59°49,32N 21°34,48E	F	19	10
JUR	59°49,32N 21°34,48E	M	19	10
KAT	60°14,02N 21°57,19E	F	36	9
KAT	60°14,02N 21°57,19E	M	51	12
PYH	60°58,45N 21°15,48E	F	21	10
PYH	60°58,45N 21°15,48E	M	49	10
RIH	61°04,06N 21°18,23E	F	11	9
RIH	61°04,06N 21°18,23E	M	35	11
STE	60°04,37N 22°02,28E	F	41	10
STE	60°04,37N 22°02,28E	M	42	10
SUN	59°57,77N 21°20,56E	F	15	12
SUN	59°57,77N 21°20,56E	M	12	10

### 2.3 Aineiston valokuvaaminen

Siirat valokuvattiin selkäpuolelta värityksen ja kuvioinnin mittaamista varten. Käytetty kamera oli PixeLink kameramikroskooppi. Kaikki kuvat otettiin kameran samoja vakioasetuksia käyttäen. Valokuvaus toteutettiin asettamalla siira 70%:sta EtOH sisältävässä petrimaljassa sinisenvärisen paperin päälle. Siiroja säilytettiin alkoholissa, etteivät näytteet kuivu lamppujen alla ja riskeeraa DNA:n hajoamista ennen sen eristystä. Sinistä paperia käytettiin, sillä siiroissa ei esiinny sinistä, jolloin taustan erottaminen siirojen kuvioinnin väreistä ja kuvioinnin analysointi oli helpompi toteuttaa. Etenkin siiran kuoren ulkoreunat ovat usein läpikuultavia. Kuvissa käytettiin harmaakorttia kameran asetusten tarkastamiseksi, ja kortti on näkyvissä joka kuvassa. Lisäksi kuviin sisällytettiin millimetripaperin pala kuva-analyysin mittakaavan suhteuttamista varten. Siirat pyrittiin asettamaan niin, että niiden selkäpuoli on suoraan ylöspäin kameraa kohti, jolloin selän väritys ja kuviointi on selkeästi näkyvillä. Valaistus asetettiin säädettävillä valoilla niin, että siiran koko selus oli mahdollisimman hyvin valaistu ilman varjoja. Siiroista otetut kuvat tallennettiin sekä .tif- että .sif-päätteisissä tiedostomuodoissa, ja tiedostot nimettiin kuvauspäivämäärän ja näiden yksilöllisten populaatiotunnisteiden mukaan. Kuva tarkennettiin keskelle siiran selkäpanssaria. Lisäksi kuvauksen yhteydessä määritettiin siiran sukupuoli tarkastamalla sukuelimet, jotka sijaitsevat siirojen kidusläppien alla.

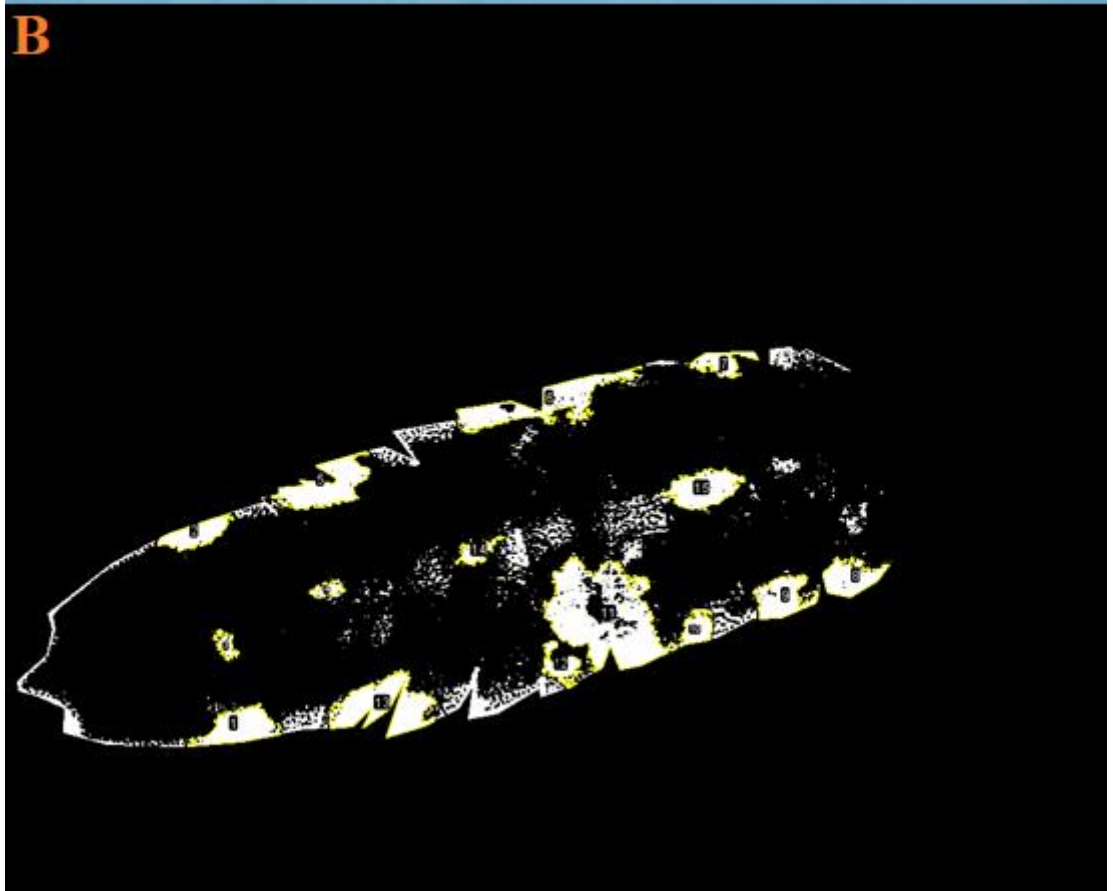
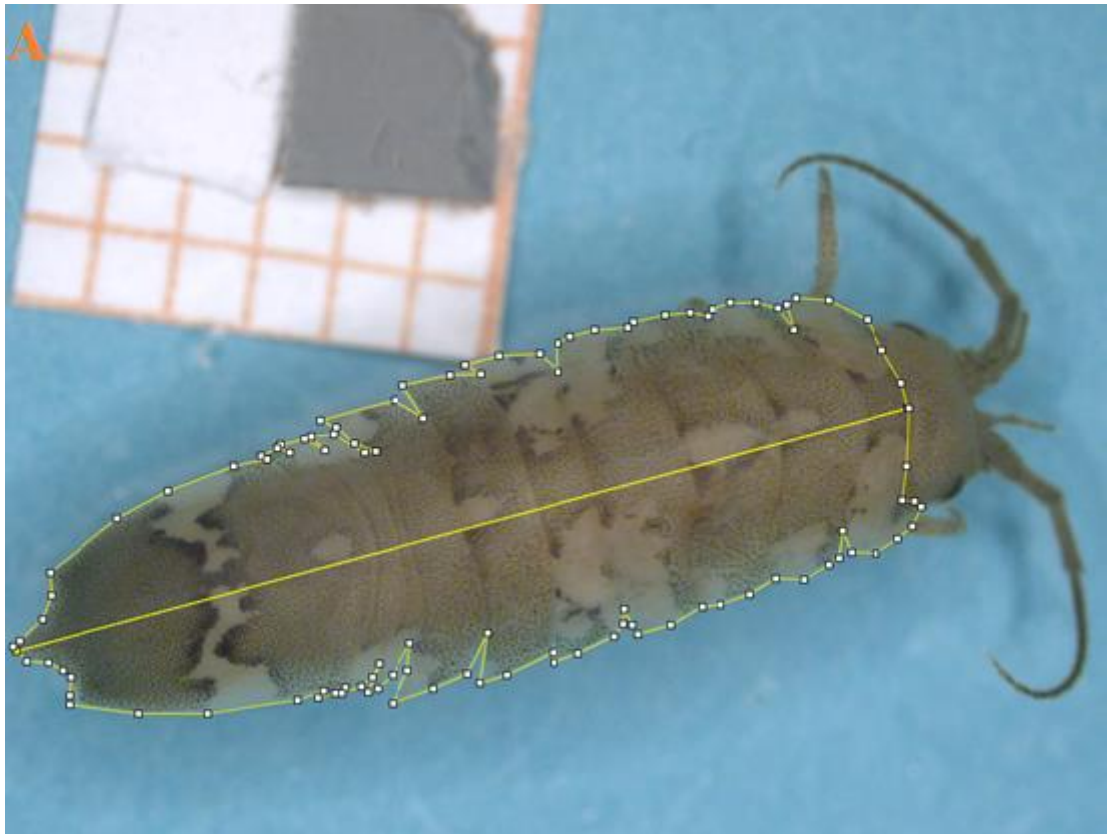
Siirojen väritys luokiteltiin Jormalaisen ja Tuomen (1989) mukaan. Pisteytys perustui siirojen tunnettuihin morfeihin ja näiden ominaisuuksien ilmentymisen vahvuuteen yksilössä. Jatkuvat ominaisuudet merkittiin välillä 0–3 – tässä tapauksessa *albafusca* ominaisuuden vahvuus – missä, 0 = ominaisuutta ei ilmene ja 1-3 = ominaisuus ilmenee, ja sen vahvuus. Rajat olivat yleisesti: 1 = ominaisuus ilmenee, mutta hyvin vähän; 2 = ominaisuus ilmenee vahvasti, mutta alle 50% siiran selkäpuolesta on vaalea; ja 3 = ominaisuus ilmenee niin, että yli 50% siiran selkäpuolesta on vaalea.

### 2.4 Kuva-analyysi

ImageJ -ohjelmaa käyttäen sekvensoiduille siiroille toteutettiin kuva-analyysi, missä siiroilta mitattiin näiden ruumiin pituus pää poisluettuna, selkäpuolen ympärysmitta ja pinta-ala, vaalean kuvioinnin pinta-alan osuus selkäpuolella, sekä yksittäisten laikkujen lukumäärä, ympärysmitta ja pinta-ala (Kuva 2).

Ensiksi ImageJ:stä valittiin haluttavat mittaustulokset, eli tutkimukseni tapauksessa: kokonaispinta-ala, eläimen ympärysmitta, ruumiin pituus, ja prosenttiosuus tumman alueen osuudesta, mikä kääntäen antoi myös vaalean alueen osuuden. Lisäksi valittiin asetus, jolla mittaukset kohdistuivat vain siiraan kuvassa.

Ensimmäinen mittaus oli siiran kehon pituus, joka tehtiin vetämällä viiva siiran pään takaosasta alkavasta jaokkeesta telsonin keskimmäisen piikin kärkeen saakka (Kuva 2A). Toinen vaihe oli rajata siiran ääri viivat. Mittaustuloksiksi saatiin siiran selkäpuolen ympärysmitta sekä pinta-ala. Kolmanneksi poistettiin kuvasta tausta ja korvattiin se mustalla värillä. Neljänneksi muutetaan monisävyinen kuva puhtaasti mustavalkokuvaksi. Samalla täytyi rajata mustan ja valkoisen välinen kynnyksarvo. Rajaus täytyi tehdä yksilökohtaisesti, sillä pienetkin muutokset valaistuksen määrässä tai siiran ruumiin asennossa kuvassa muuttivat alueiden valoisuutta/tummuutta kuvassa. Optimaalisen tuloksen saamiseksi piti silmämääräisesti arvioida, millä asetusarvolla suurin osa vaaleasta alueesta laskettiin mukaan mittaukseen, mutta samalla tummaa aluetta ei laskettu liikaa mukaan raja-arvon lähestyessä tumman värin arvoa. Mittaustulos antaa prosenttiosuuden siiran tumman värin peittämästä alueesta ja pinta-alan tälle. Vaalean värin osuus laskettiin vähentämällä tumman osuus koko pinta-alasta. Lopuksi mitattiin kukin laikku (valkoisella merkitty vaalea alue kuvassa) (Kuva 2B), ja vastaavasti näistäkin saadaan tulokseksi pinta-ala sekä ympärysmitta, kuin myös laikkujen lukumäärä. Kaikkein pienimpiä laikkuja ei ole järkevää valita, sillä jossain vaiheessa laikun koko on niin pieni, ettei mittaustulos anna merkitsevän suurta kokoa. Esimerkiksi tämän tutkimuksen kohdalla mittaukset olivat millimetreissä 3 desimaalin tarkkuudella, eli tulokset, jotka olivat 4:llä tai useammalla desimaalilla merkittynä jätettiin huomioimatta. Kaikki mittaukset suorittamalla ja kuvaa mittausten ohella muokaten tuotettiin lopullinen tuotos per mitattu siirayksilö.



Kuva 2 - Yllä (kirjaimella A merkityssä) kuvassa näkyy leväsiirayksilön ruumiin ympärysmitan sekä pinta-alan (ruumista ympäröivä keltainen alue) ja ruumiin pituuden (selän poikki kulkeva suora viiva) mittaus. Mittauksissa ei otettu huomioon yksilöiden päätä, joten se ei ole mittauksissa mukana. Alla (kirjaimella B merkityssä) kuvassa

*näkyvä leväsiiran laikkujen määrän, laikkujen pinta-alan ja laikkujen ympärysmittan mittaaminen. Kukin erillinen laikku on merkitty numerolla. Tässä kyseisessä yksilössä on laikkuja yhteensä 15.*

## 2.5 Väriytyksen ominaisuuksien analyysit

Kuva-analyysin ja numerodatan käsittelyyn käytin RStudio (RStudio Team, 2020) versiota 4.1.2. RStudio on R-ohjelmointikieltä (R Core Team, 2022) käyttävä integroitu ohjelmointiympäristö, jota käyttäen tässä tutkimuksessa kuva-analyysin tulosdataa on visualisoitu sekä käsitelty. Tämän lisäksi on laskettu valkean osuuden mediaani ja kvantiilit, sekä laikkujen keskimäärä.

Tein *albafusca*-yksilöissä valkoisen prosentuaaliselle osuudelle binomijakaumalla ja laikkujen lukumäärälle Poisson-jakaumalla yleistetyn lineaarisen sekamallin käyttäen RStudioin lm4-pakettia (Bates D. et al. 2015). Malleissa selittäjinä olivat sukupuoli, populaatio, joka oli satunnaismuuttuja, sekä näiden kahden muuttujan yhdysvaikutus. Lähtömalleja yksinkertaistettiin poistamalla selittäjä kerrallaan, jos yhdysvaikutus tai yhdysvaikutus ja sukupuoli eivät olleet merkitseviä. Satunnaismuuttujan vaikutus arvioitiin tarkastelemalla sen varianssia.

Kummaltakin sukupuolelta testattiin väriytyksen ominaisuuksien keskinäistä korrelaatioita RStudiolla Spearmanin järjestyskorrelaatiokertoimella. Analyysissä huomioitiin vain *albafusca*-morfien yksilöt. Vertailut tehtiin kummallekin sukupuolelle laikkujen lukumäärästä, valkoisen alueen prosentuaalisesta osuudesta, ja valkoisen keskimääräisestä alueesta koko yksilössä.

## 2.6 DNA:n eristys

Siirteiltä myös eristettiin ja kerättiin DNA näytteet GWAS-analyysiä varten. Eristys tehtiin ”Macherey-Nagel DNA Tissue XS” -kittiä käyttäen noudattaen valmistajan protokollaa alla mainituin muutoksin. Kultakin siirteiltä kerättiin 4–8 jalkaa sekä jalan tyven lihas suolta välttämällä eppendorf-putkeen. Kerätyt näytteet siirrettiin T1 puskuriliuokseen odottamaan muiden näytteiden valmistelua. Jokaiseen näytteeseen lisättiin proteinaasi-K ja asetettiin ravistimeen (engl. shaker) yön ylitse +60°C:een. Tämän jälkeen näytteille tehtiin RNAasi A -käsittely RNA:n hajottamiseksi ja lisättiin B3-lyysi puskuriliuos. Sentrifugilla näytteistä ajettiin kiinteät roskaosat näytteen pohjalle ja supernatantti siirrettiin uuteen putkeen ja inkuboitiin 70° C:ssa 5 minuutin ajan. Näytteisiin lisättiin 100%:sta EtOH, sekoitettiin, ja ajettiin läpi kolumnista keräysputkeen sentrifugissa 11 000 r/min yhden minuutin ajan. Keräysputkeen lisättiin kahdesti B5-pesu puskuriliuosta, ensimmäisen lisäyksen jälkeen näyte ajettiin sentrifugissa 11 000 r/min yhden minuutin ajan, ja toisen lisäyksen jälkeen kahden

minuutin ajan. Sentrifugointi toistettiin vielä kerran tyhjiltään, jotta kaikki neste varmasti poistuisi kolumnista. On teoreettinen mahdollisuus, että kaikki neste ei poistu kolumnista, koska kosteus putkessa on niin korkea. Lopuksi suoraan kolumnin membraanille lisättiin BE-puskuriliuosta kaksi kertaa (2 x 15 µl) saannon lisäämiseksi, ja kummankin lisäyksen yhteydessä näytettä ajettiin sentrifugissa samoilla asetuksilla yhden minuutin ajan.

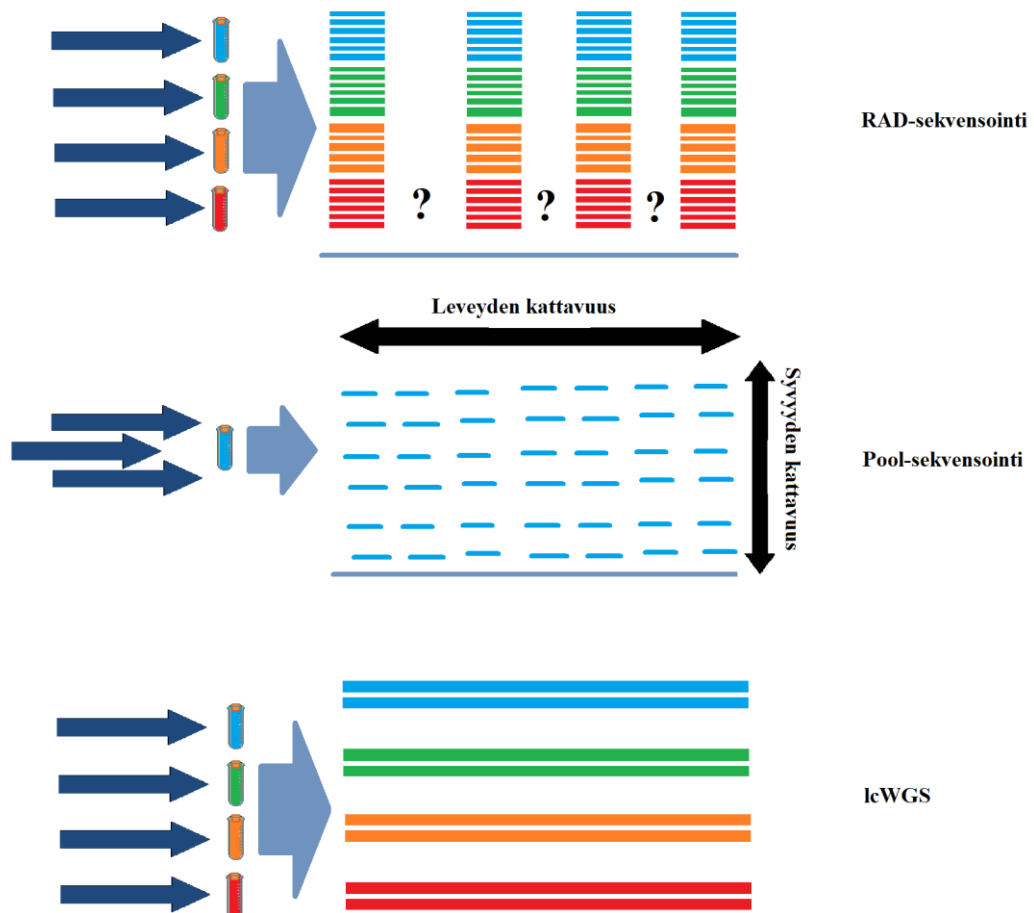
Sekvensointia varten kustakin populaatiosta otettiin otos, joka koostui 20 yksilöstä, joista puolet oli koiraita ja toinen puolikas naaraita. Yksilöitä otostettaessa otettiin huomioon populaation morfijakauma, joka perustui kaikkiin kerättyihin yksilöihin ja kuvauksen yhteydessä määritettyihin sukupuoliin sekä värityksen pisteytyksiin kustakin populaatiosta (Taulukko 1). Sekvensointikirjaston valmistus ulkoistettiin. Tätä varten lähetettiin FFGC:lle (Finnish Functional Genomics Centre) yhteensä 190 DNA näytettä, joista 5 oli duplikaatteja.

## 2.7 DNA:n sekvensointi

Genomi sekvensoitiin GWAS-analyysijä varten leväsirojen geneettisten ominaisuuksien selvittämiseen (Lou et al. 2021, Kuva 3). Genomin analysointiin on erilaisia menettelytapoja, jotka eroavat niin sanottujen leveyden (engl. breadth), eli analyysimenetelmän kattavuuden, ja syvyyden (engl. depth), eli kuinka monta kertaa genomien kukin kohta on katettu tuotetulla sekvenssillä, mukaan (Kuva 3). Kolme yleisintä menettelytapaa ovat RAD-sekvensointi, eli restriktioalue assosioitu sekvensointi (engl. RAD-seq, eli Restriction site-Associated DNA sequencing); Pool-sekvensointi, eli yhdistetty sekvensointi (engl. Pool-seq, pooled sequencing); ja lcWGS, eli matalan kattavuuden koko genomien sekvensointi (engl. low-coverage Whole Genome Sequencing). Kullakin menettelytavalla ovat vahvuutensa ja heikkoutensa. RAD-sekvensoinnilla voidaan tutkia osaa genomista syvästi, mutta genomista jää suuria osia sekvensoimatta, eli leveys ei ole kattava (Kuva 3). Pool-sekvensoinnissa kaikki yksilöt sekvensoidaan yhdistetystä näytteestä, jolloin leveys ja syvyys ovat kattavat, mutta yksilöitä on mahdotonta erottaa, jolloin tämä menetelmä toimii parhaiten koko populaation parametrien arvioimiseen. lcWGS:ää käyttäen kyetään analysoimaan koko genomi säilyttäen yksilöivä tieto, mutta syvyys ei ole kattava. lcWGS:n tehokkuus tulee metodista saada genomien tieto useasta yksilöstä. Tässä tutkielmassa sirojen genomi analysoitiin lcWGS-menetelmää käyttäen.



Aiemmin mainitusti geenikirjastojen valmistus ulkoistettiin FFGC:lle. DNA:n pitoisuus (Qubit ja Nanodrop) ja DNA eristyksen laatu ja puhtaus (Nanodrop) määritettiin Turun yliopistolla ja FFGC:llä toteutettiin kullekin DNA näytteelle Fragment Analyzer -ajo, jossa tarkastettiin kunkin näytteen DNA:n kokojakauma, jotta voitiin todentaa korkean molekyylipainon omaavan genomisen DNA:n läsnäolo. Osassa näytteistä ilmeni rappeutumista, mutta kaikki sisälsivät myös ehyttä DNA:ta. Kirjastojen valmistus kaikille näytteille aloitettiin 25 nanogramman DNA-määrällä. Kirjastot valmistettiin 4 erässä. Kirjastot sekvensoitiin Illuminan Novaseq6000 SP:tä käyttäen 150 emäsparin lukupituuteen paired-end -lukumoodissa. Tämä tarkoittaa, että kukin kirjastossa oleva kaksijuosteinen DNA-fragmentti on luettu kummastakin päästä alkaen, jolloin näiden väliin jää leväsiiran genomista peräisin oleva insertti, jonka pituus riippuu alkuperäisen fragmentin pituudesta ja sekvensoinnin lukupituudesta. Jokaiselle DNA fragmentille tuotettiin siis kaksi lukua (engl. read).



Kuva 3 - Visualisointi genomien analysoinnin menettelytavoista. Ylimpänä RAD-sekvensointi, keskellä Pool-sekvensointi, ja alimpana lcWGS. Sekvensointitavat on visualisoitu selkeyttämään eri sekvensointitapojen erot. Vaakasunnassa on leveys (breadth) ja pystysuorassa syvyys (depth). Harmahtava vaakasuoja viiva kunkin sekvensointitavan alla on referenssigenomi, johon yksilöiden genomia verrataan. Muut värit kuvastavat tutkittavien yksilöiden genomia. RAD-sekvensoinnissa jää osia genomista sekvensoimatta, mitä "?"-välit

*kuvastavat, eli leveys ei ole kattava; Pool-sekvensoinnissa yksilöt sulautuvat yhteen, mitä yksivärinen genomi kuvastaa, jolloin yksilöitä ei voida erottaa toisistaan; lcWGS:ssa jää väliä yksilöiden väliin syvyydessä, jolloin syvyyden kattavuus heikentyy.*

## 2.8 Genomin linjaus

Geenisekvenssin linjaaminen on bioinformatiikan menetelmä, joka mahdollistaa yksilöiden perimän vertailun. Genomin linjaamisen tarkoitus on löytää vastaavat kohdat referenssigenomin ja näytteen väliltä tai yksittäisten näytteiden väliltä, jos referenssiä ei ole. Useimmat nykyään käytettävät menetelmät perustuvat lyhyiden hakusekvenssien käyttöön (noin 3–7 emästä), ja genomista etsitään näiden kanssa identtisiä kohtia sekvensoiduista luvuista. Tällaisen kohdan löydyttyä verrataan, miten luvun loppuosa sekvenssistä sopii kyseiseen kohtaan referenssiä, eli kuinka luvun emäkset vastaavat verrattavaa referenssiä. Luvussa voi olla esimerkiksi yksittäisiä emäksiä välissä, jotka eivät vastaa referenssin emäksiä, tai vain luvun toinen pää vastaa referenssiä.

Sekvenssit, jotka saatiin lcWGS-menetelmällä, trimmattiin Trimmomatic-ohjelmalla (Bolger et al. 2014) poistaen sekvensointiadapteri osa sekvenssidatasta sekä huonon laatuiset emäkset luvun loppupäästä. Trimmaamisessa siis poistetaan kirjastonvalmistuksessa lisätyt sekvensoinnissa tarvittavat osat luvusta, koska ne eivät linjaudu referenssigenomiin. Trimmatun luvun voi tämän jälkeen linjata referenssigenomiin. Tässä tutkimuksessa käytetyn referenssigenomin voi löytää GenBank® verkkosivulta, joka on NIH:in (engl. National Institutes of Health, eli Yhdysvaltain terveysvirasto) geenipankki [https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA\\_023373965.1/](https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_023373965.1/). Tämä on ensimmäinen versio, joka leväsiiran genomista on julkaistu referenssinä käytettäväksi.

Tässä tutkimuksessa käytettiin bowtie2-nimistä (Langmead & Salzberg, 2012) ohjelmaa genomien linjauksessa. Bowtie2:a käyttäen pystytään tehokkaasti linjaamaan lyhyitä sekvenssejä – mukaan lukien 50:n, 100:n ja jopa tuhansien merkkien pituiset sekvenssit – referenssigenomiin. Bowtie2 luo referenssigenomista indeksitiedoston, joka sisältää kunkin lyhyen hakusekvenssin sijainnin genomissa. Tämän jälkeen bowtie2 vertaa kirjaston referenssisekvenssiä indeksitiedostoon, josta löydetään vastaavuus indeksitiedoston hakusekvenssiin. Vastaavuuden löydyttyä kirjaston yhden luvun koko sekvenssiä voidaan verrata kyseisen genomien kohdan sekvenssiin ja määrittää arvo linjauksen sopivuudelle kyseiseen genomien kohtaan ja muodostaa linjaukset. Tämä tapahtuu käyttäen apuna lyhyitä hakusekvenssejä ja näistä muodostettua indeksiä. Ohjelmalla voidaan linjata tutkimuksessa käytettyjä,

niin sanottuja paritettu loppu (engl. paired-end) -linjauksia, missä data sisältää oletukset linjattavien parien suunnasta ja etäisyydestä toisistaan alkuperäisessä DNA-molekyylissä. Tässä ohjelma tunnistaa toistuvat kohdat parien välillä ja osaa yhdistää nämä lyhyemmät pätkät yhdeksi näitä kohtia vertaamalla (Langmead & Salzberg, 2012). Tässä tutkimuksessa käytettiin paired-end -linjauksia, sillä kirjastot ovat paired-end -kirjastoja eli kukin kirjastossa oleva kaksijuosteinen DNA-fragmentti on luettu kummastakin päästä, jolloin näiden fragmenttien väliin jää insertti.

Genomin linjauksen jälkeen lcWGS-aineistolle suositellaan kirjaston valmistuksesta johtuvien artefaktien korjaamista, kuten on esitelty tarkemmin Loun ja muiden artikkelissa (Lou et al. 2021) ja myös tässä aineistossa nämä tehtiin käyttäen bamUtil (Jun et al. 2015) ja Picard Toolkit (Picard toolkit, 2019) -työkaluja. PCR ja optiset duplikaatit – kuten geeniklusterin lukeminen kahtena – vääristävät uniikkien lukujen määrää, minkä vuoksi näistä aiheutuneet duplikaatit on poistettu. Lisäksi kaikki lisäykset ja poistot (engl. Indel) aiheuttavat sen, etteivät lukujen päät linjaudu toisiinsa, mikä voidaan korjata esimerkiksi genotyyppien todennäköisyyksiä arvioivilla työkaluilla. Lopuksi paired-end -linjauksissa insertit voivat olla lyhyempiä kuin yhdistetyt luvut ja linjauksessa syntyy limittymistä, jolloin limittynyt kohta luetaan kahtena, mikä vääristää genotyyppien todennäköisyyksiä. Yleinen ratkaisu tähän on ”soft-clipping”, joka tarkoittaa, että toisen luvun pää poistetaan. Suositellut käytetyt toimenpiteet on käsitelty tarkemmin Loun ja muiden artikkelissa (Lou, et al. 2021).

## 2.9 GWAS analyysi

Genomiin linjattujen sekvenssien analysointi suoritettiin Unix-pohjaisessa ANGSD-ohjelmassa (Korneliussen et al. 2014). ANGSD-ajon päätteeksi ohjelma luo tulostiedoston, jonka voi avata esimerkiksi RStudioissa tulosten tarkastelemiseksi.

ANGSD:ia on kehitetty etenkin matalan kattavuuden sekvensointiaineiston käsittelylle ja geneettisten varianttien tunnistamiseen genomista. Lisäksi ANGSD mahdollistaa GWAS-analyysin matalan kattavuuden sekvensseille. Tässä tutkimuksessa varianteiksi valittiin yhden emäksen muutokset (single nucleotide polymorphism eli SNP). Analyysia varten määritetään ensin kriteeri varianttien tunnistamiseen populaatioista, annetaan vastemuuttajat ja kovariaatit. Variantti

hyväksyttiin analyysiin, mikäli se löytyi vähintään noin 80% näytteistä (120 yksilöltä 152 yksilöstä) sekä harvinaisemman variantin esiintyvyys oli suurempi kuin 5%.

Tämän tutkimuksen kohdalla tärkeimmiksi kovariaateiksi oletettiin populaatio ja sukupuoli, ja vastemuuttajat olivat morfi – joka oli vielä täsmennetty sisältämään vain *uniformis* ja *albafusca* yksilöt – valkean prosentuaalinen osuus sekä valkean osuus pinta-alana. Kovariaattien vaikutuksen tutkimiseksi analyysit suoritettiin molempien kovariaattien kanssa, ilman kovariaatteja, ja vain sukupuoli-kovariaatin kanssa.

Analyysi laski tilastolliset merkitsevyydet eri ominaisuuksien ja genotyyppien assosiaatiolle datassa. Useiden vertailujen vuoksi tuotetuille tiedostoille tehtiin P-arvojen korjaus väärin positiivisten tulosten poistamiseksi. Useita tilastollisia testejä samanaikaisesti tehdessä FDR (engl. False Discovery Rate), eli virheellisten tulosten suhteellinen määrä kasvaa, mikä täytyy ottaa huomioon valheellisten tulosten poistamiseksi. Tämän tein RStudioissa käyttäen Benjamini–Hochberg-menetelmää.

Lisäksi tutkituille ominaisuuksille tuotettiin ”manhattan”-kuvaaja sekä havaittuja ja teoreettisten p-arvojen kuvaaja käyttäen cowplot- (Claus, 2020) ja ggplot2-paketteja (Wickham, 2016) RStudioissa. Manhattan-kuvaajassa x-akseli kuvaa sijaintia leväsiiran genomissa, eli kromosomia ja positiota. Tässä tutkimuksessa x-akselilla on kromosomeja todella monta, sillä leväsiiran genomi on vasta ”scaffold”-tasolla, jonka suomensin telineeksi. Niin sanottuja telineitä on genomien katkonaisuuden vuoksi paljon, minkä takia kuvaajat ovat hankalasti luettavia (Kuva 7). Cowplot-kuvaajat taas kuvastavat datan teoreettisen odotetun normaalijakauman P-arvon ja tulosten välisten P-arvojen suhdetta (Kuvat 8–12). Tulosten P-arvot ovat GWAS-analyyseistä saadut korjatut P-arvot. GWAS perustuu yleistettyyn lineaariseen malliin.

## 2.10 Merkitsevien SNP:ien annotaatio

Merkitsevien SNP:ien annotaation eli toiminnallisuuden tunnistamiseksi käytetään bedtools- (Quinlan & Hall, 2010) ja BLAST-työkalua (Madden, 2002) (engl. Basic Local Alignment Search Tool). Bedtools:ia käyttäen voidaan eristää geenin sekvenssi genomitiedostosta (*getFasta* komento), sekä verrata kahden sijaintitiedoston välisiä yhtäläisyyksiä (*intersectBed* komento). Tätä varten luodaan bed-tiedosto, jossa on merkittynä merkitsevän SNP:n sijaintitiedot. Ensiksi kromosomin nimi (tässä tapauksessa telineen nimi), ja tämän jälkeen emäsparien aloitus- ja lopetusarvo välille, jossa SNP sijaitsee. Näitä voitiin kasvattaa ylä- ja alavirtaan, jos bedtools-analyysi ei löytänyt annotaatioita. Tuotettua tiedostoa verrattiin leväsiiran referenssigenomin

annotaatioon, jotta voitaisiin tunnistaa, sijaitseeko SNP geenissä, tai ylä- tai alavirtaan geenistä. Bedtools-analyysit suoritettiin laikkujen lukumäärälle, morfin vaikutukselle, valkean prosentuaaliselle osuudelle ja valkean pinta-alalliselle osuudelle.

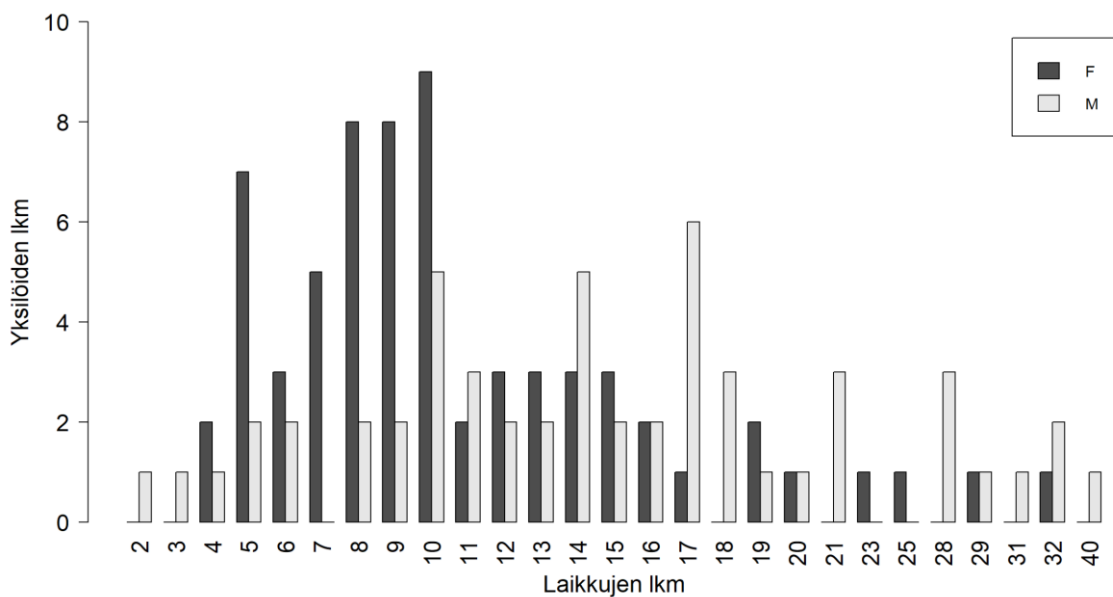
Mikäli referenssigenomin annotaatiosta ei löytynyt geenivastaavuutta SNP:lle voidaan käyttää BLAST:ia. Tällöin eristettyä sekvenssiä verrataan automaattisesti UniProt-proteiinitietokantaan käyttäen blastx hakua, joka kääntää DNA-sekvenssin aminohapposekvenssiksi ja etsii samankaltaisen proteiinisekvenssin eri lajeilta. Samankaltaisuuksien löydyttyä, voidaan pyrkiä selvittämään mahdollinen toiminta SNP:hen liittyvälle genomialueelle tutkimalla vastaavuuden tuottanutta proteiinia ja siihen liitettyjä toimintoja löydettyillä lajeilla. Proteiinin koodilla pystyy tietokantoja käyttäen löytämään tarkempaa tietoa proteiinista. Tässä tutkimuksessa käytin kahta tietokantaa: UniProt-tietokantaa <https://www.uniprot.org/>, ja GenomeNet-tietokantaa <https://www.genome.jp/>. Tietokannat listaavat proteiinin koodilla tiedot kyseisestä proteiinista ja sisältävät tarkempaa tietoa proteiinin rakenteesta, merkityksestä, ja osaluista, joita se säätelee. Verkkosivu pisteyttää annotoinnin laajuuden asteikolla 1–5, missä 1 heikoiten annotoitu ja 5 laajasti tai täysin annotoitu. Tässä tutkimuksessa jokainen tutkittu proteiini oli pisteytetty arvosanalla 1 tai 2. Täten proteiinien tarkkaa merkitystä ei voida tietää, mutta näiden yleinen merkitys on listattu, kuten esimerkiksi molekulaarinen toiminto ja tarkennetusti aminohapon sitominen. Tuloslistat ovat laajoja ja kussakin on monia osumia, liian monia, jotta jokaista voisi tai olisi järkevää käydä läpi. Sen sijaan perehdyin kustakin noin viiteen merkittävimpään proteiiniin. Lisäksi, jos mahdollista, yritin poimia proteiineja vain yhdeltä lajilta.

## 3 Tulokset

### 3.1 Värimorfien fenotyypinen vaihtelu

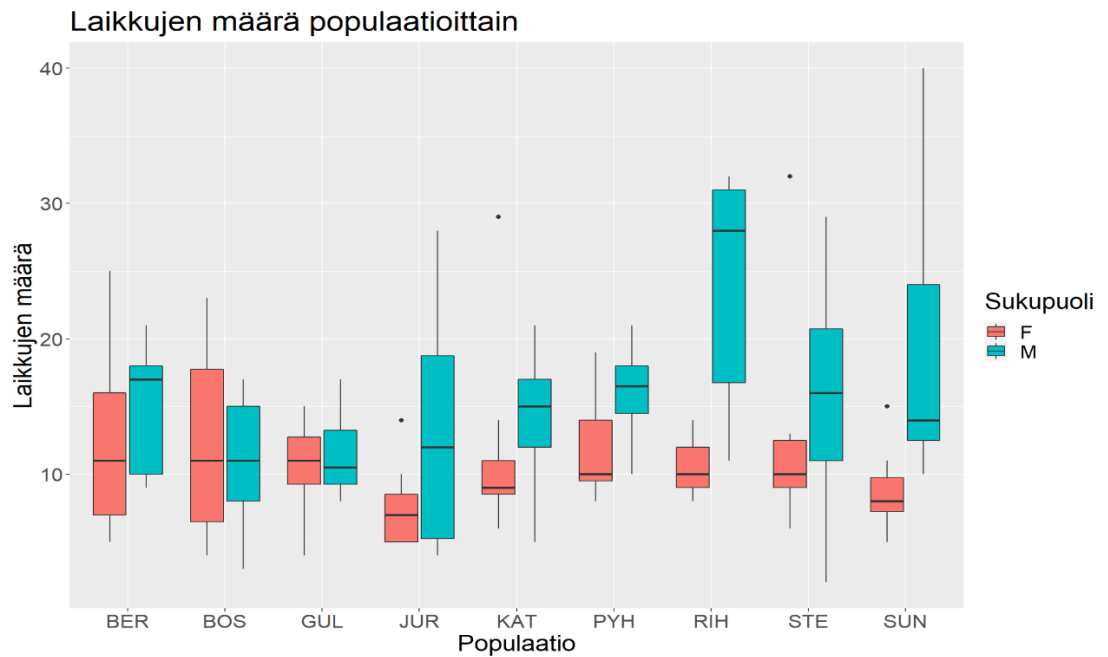
Aineistossani oli laikullisia *albafusca*-morfin yksilöitä yhteensä 120. *Uniformis*- ja *maculata*-morfien yksilöitä oli 62 yksilöä yhteensä. Aineistossani oli *uniformis*-morfin koiraita hieman enemmän kuin naaraita: 37 koirasta ja 25 naarasta ( $X^2 = 57,91$ ,  $df = 1$ ,  $P < 0,0001$ ). Laikuttomia yksilöitä on eniten Pyhämaassa (14,5% kaikista laikuttomista) ja Rihtniemessä (17,7% kaikista laikuttomista). *Uniformis*-yksilöiden frekvenssit eivät eronneet merkitsevästi populaatioiden välillä ( $X^2 = 2,51$ ,  $df = 8$ ,  $P = 0,96$ ), kuten eivät *albafusca*-yksilöidenkään ( $X^2 = 3,56$ ,  $df = 8$ ,  $P = 0,89$ ). *Albafusca*-

yksilöistä naarilla suurin määrä laikkuja oli 32 kappaletta, kun taas koirilla suurin määrä oli 40 kappaletta (Kuva 4).



Kuva 4 – Kuvaaja laikkujen lukumäärän jakaumasta albufusca-yksilöissä. Naaraita on yhteensä 66 laikullista yksilöä (tummat pylväät), ja koiraita 54 laikullista yksilöä (vaaleat pylväät). X-akselilta näkee laikkujen määrän ja y-akselin luku kertoo yksilöiden lukumäärän. Kuvaajaan ei ole sisällytetty laikkumääriä, joita yksikään yksilö ei edustanut, eli esimerkiksi yhdelläkään yksilöllä ei ollut 22 laikkua, joten tätä pylvästä ei kuvata lainkaan.

Naaraiden keskimääräinen laikkujen määrä ( $\pm$  keskihajonta) oli 10,9 ( $\pm$  5,7) laikkua, ja koirilla 15,5 ( $\pm$  8,2). Ero oli tilastollisesti merkitsevä ( $t = 3,46$ ,  $df = 91,8$ ,  $P = 0,0008$ ). Suurin määrä laikkuja on keskimääräisesti Rihtniemen koirilla (28 laikkua) (Kuva 5). Populaatiossa ero naaraisiin on myös huomattava. Rihtniemen naarilla on keskimääräisesti vain 10 laikkua. Koiraiden laikkujen määrissä on myös yleisesti enemmän hajontaa.



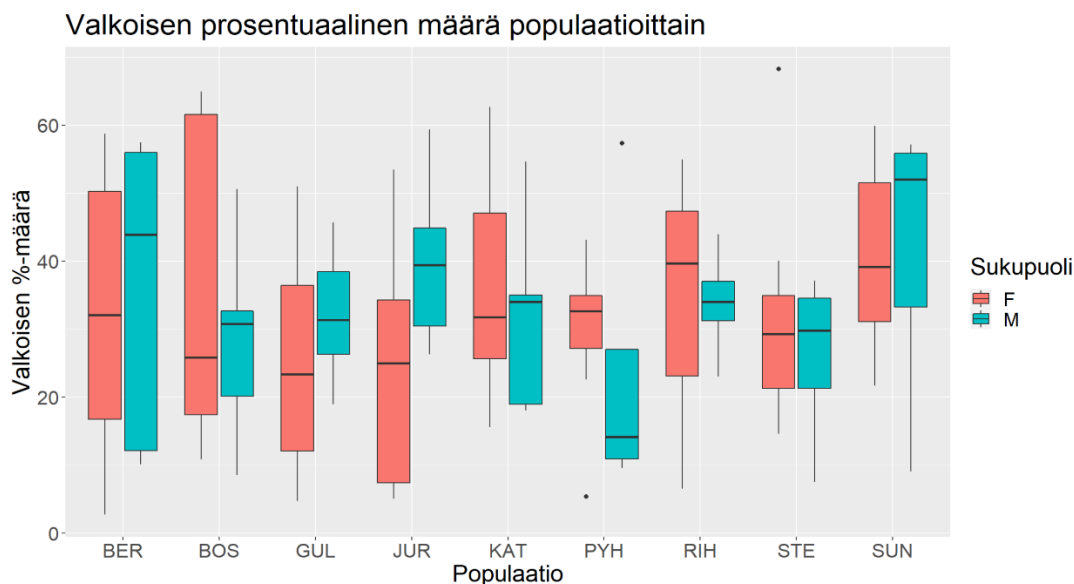
Kuva 5 – Laikkujen lukumäärän jakauma eri populaatioissa sukupuolittain. Laatikko-janakuvio kuvaa mediaanin, 25 ja 75 %, ja minimin ja maksimin kvantiilit, sekä nämä ylittävät yksittäiset arvot.

Keskimääräinen laikkujen koko naarailla on  $1,21 \text{ mm}^2 (\pm 1,9 \text{ mm}^2)$ . Koirilla koko on keskimääräisesti  $0,62 \text{ mm}^2 (\pm 0,94 \text{ mm}^2)$ . Laikkujen koossa sukupuolten välillä oli tilastollisesti merkitsevä ero ( $t = -2,23$ ,  $df = 99.201$ ,  $P = 0,028$ ).

Testasin leväsiirujen laikkujen määrän ja suhteellisen pinta-alan korrelaatioita Spearmanin järjestyskorrelaatiokertoimella *albufusca*-yksilöistä. Naarailla oli negatiivinen korrelaatio valkoisen prosentuaalisen osuuden ja laikkujen määrän välillä ( $r = -0,290$ ;  $n = 66$ ;  $P = 0,018$ ). Toisin sanoen, kun laikkujen määrä kasvoi, ne olivat samalla myös pienempiä. Koirilla tämä korrelaatio ei ollut tilastollisesti merkitsevä ( $r = 0,208$ ;  $n = 54$ ;  $P = 0,13$ ).

*Albufusca*-morfien yksilöille valkoisen prosentuaaliseen osuuteen vaikuttavien tekijöiden testauksen mukaan ei sukupuolen ja populaation yhdysvaikutus ( $X^2 = 16,88$ ,  $df = -16$ ,  $P = 0,39$ ) eikä sukupuoli ( $X^2 = 0,14$ ,  $df = -1$ ,  $P = 0,71$ ) ollut merkitseviä (Kuva 6). Lisäksi satunnaismuuttujan varianssi oli 0, eli populaatioiden väliset erot valkoisen prosentuaalisen osuuden vaihtelussa oli hyvin vähäistä. Laikkujen lukumäärään vaikuttavien tekijöiden testauksen mukaan sekä sukupuolen ja populaation yhteisvaikutus ( $X^2 = 56,1$ ,  $df = -16$ ,  $P < 0,0001$ ) ja sukupuolen päävaikutus ( $X^2 = 42,95$ ,  $df = -1$ ,  $P < 0,0001$ ) olivat merkitseviä (Kuvat 4 ja 5). Yhdysvaikutusta selittää se, että joissain populaatioissa koirilla on enemmän laikkuja kuin naarailla ja joissain populaatioissa sukupuolilla on suunnilleen yhtä

paljon laikkuja. Tässäkin testissä satunnaismuuttujan varianssi oli 0, eli populaatioiden välinen vaihtelu oli hyvin vähäistä.



Kuva 6 - Valkoisen alueen prosentuaalisen osuuden jakauma eri populaatioissa sukupuolittain. Laatikko-janakuvio kuvaa mediaanin, 25 ja 75 %, ja minimin ja maksimin kvanttiilit, sekä nämä ylittävät yksittäiset arvot.

### 3.2 Ilmiasun ja genomien assosiaatioanalyysi

GWAS-analyysillä testattiin genotyypin ja värityksen fenotyypin ominaisuuksien assosiaatiota, ja kovariaatteina analyysissä olivat sekä sukupuoli että populaatio. Analyysiin sisällytettiin 152 yksilöä. Analysoitavia SNP:tä oli yhteensä 627100743, joista suodatuksen jälkeen jäi 50546. Tiedot merkitsevistä SNP:istä p-arvojen korjausten jälkeen näkyvät taulukoista 2 ja 3.

Morfiin (*albafusca* tai *uniformis*) liittyvien SNP:ien tunnistamiseen aineistoni ei ollut tarpeeksi kattava, sillä yhtäkään merkitsevää SNP-ilmiasu -assosiaatiota ei jäänyt jäljelle P-arvojen korjausten jälkeen. Suuremmalla otoskoolla per populaatio morfien vaikutus olisi mahdollisesti tutkittavissa. Valkoisen prosentuaalisen osuuden kohdalla kaksi osumaa – eli P-arvojen korjausten jälkeen merkitsevää SNP:tä – oli samassa telineessä *Idba\_scaffold9873*, toiset kaksi telineessä *Idba\_scaffold96286*, ja toiset kolme telineessä *Idba\_scaffold98343* (Taulukko 2).

Hypoteesia, jonka mukaan valkoisen suhteellinen osuus värityksestä on geneettisesti polygeeninen ominaisuus, tukee se, että useat SNP:ien osumat ovat eri telineeseen. Tämä voi viitata fenotyypin piirteiden takana olevan useita siihen vaikuttavia genejä tai genomisia alueita.



Taulukko 2 – Merkitsevät SNP:t valkoisen pinta-alan prosentuaaliselle osuudelle selkäpuolen värityksessä. Ensimmäisessä sarakkeessa on telineen nimi, toisessa sarakkeessa vaihtelevan emäksen sijainti, mistä osuma löytyi, kolmas ja neljäs sarake nimeävät ”Major” ja ”Minor” alleelit, joista Major on referenssigenomin mukainen yleisimmän esiintyvä alleeli ja Minor harvemmin esiintyvä vaihtoehtoinen alleeli, ja viimeinen sarake on korjattu SNP:n merkitsevyyden P-arvo. Osumia oli 2 enemmän kun kovariaattina oli vain sukupuoli, tai kun kovariaatteja ei ollut yhtäkään.

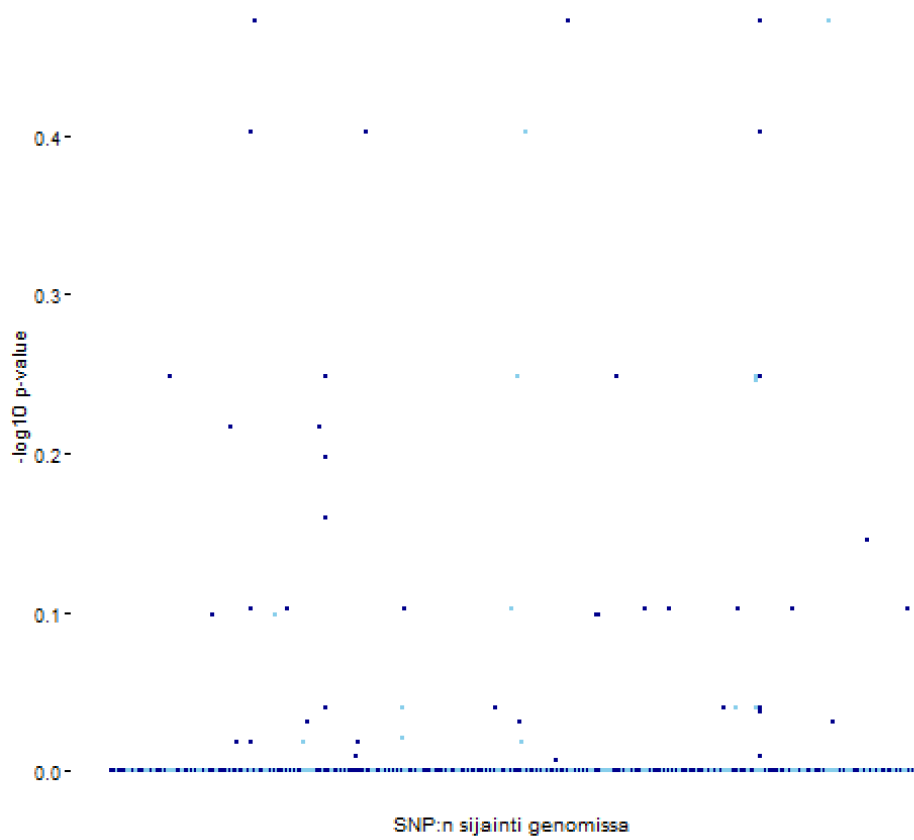
Teline	Positio	Referenssi alleeli	Vaihtoehtoinen alleeli	Korjattu P-arvo
Idba_scaffold1010	38881	C	T	1,13e-02
Idba_scaffold9873	11535	G	A	3,18e-03
Idba_scaffold9873	11572	T	A	1,21e-02
Idba_scaffold34751	8140	G	A	6,2e-05
Idba_scaffold55789	2294	C	A	4,98e-02
Idba_scaffold96286	1483	T	A	2,44e-02
Idba_scaffold96286	2074	C	T	4,51e-06
Idba_scaffold98343	2603	T	C	9,83e-07
Idba_scaffold98343	2636	T	G	8,21e-08
Idba_scaffold98343	3228	G	T	6,11e-04
Idba_scaffold136491	2853	C	T	1,27e-04

Taulukko 3 – Merkitsevät SNP:t valkoisen alueelle. Osumat olivat täysin samat kun kovariaattina oli vain sukupuoli. Osumia oli 8 enemmän kun kovariaatteja ei ollut yhtäkään.

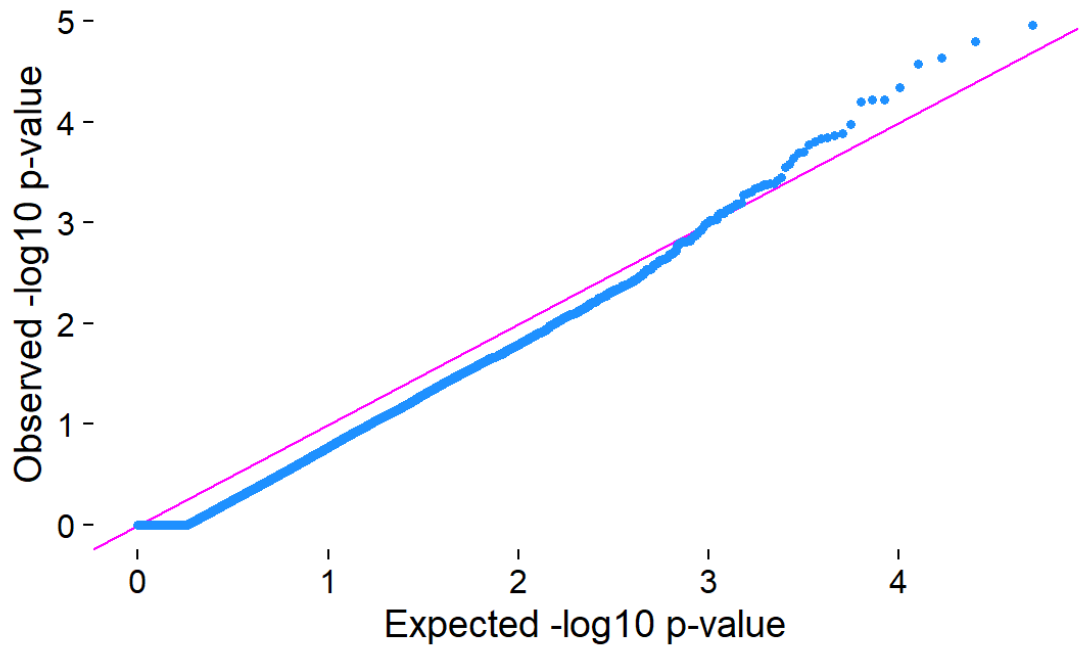
Teline	Positio	Referenssi alleeli	Vaihtoehtoinen alleeli	Korjattu P-arvo
Idba_scaffold53095	1587	G	T	7,6e-03
Idba_scaffold106507	1515	C	T	6,7e-03

Tässä tutkimuksessa käyttämäni lineaarinen malli, jossa kovariaatteina oli sukupuoli sekä populaatio, tuotti parhaimmat vastaavuudet odotetuissa ja havaituissa SNP:ien P-arvoissa morfille (Kuva 8) sekä valkoisen alueelle (Kuva 9), ja mutta myös pelkkä sukupuoli-kovariaati (Kuva 10) toimi hyvin valkoisen alueelle, verraten ilman kovariaatteja tuotettuun malliin (Kuva 11) tai valkoisen prosentuaalisen osuuden kuvaajaan (Kuva 12), missä mallit eivät tuota yhtä hyvää sopivuutta. Havaintojen ja

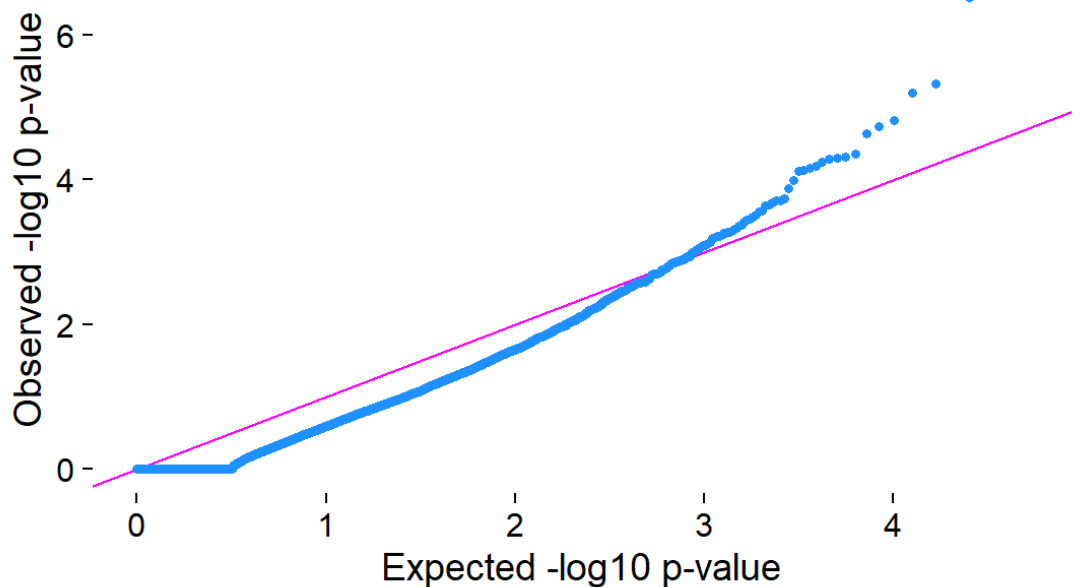
ennusteiden yhtäsuuruus tarkoittaisi, että aineiston jakauma vastaa normaalijakaumaa, jolloin pisteet erkanisivat vähemmän suoralta.



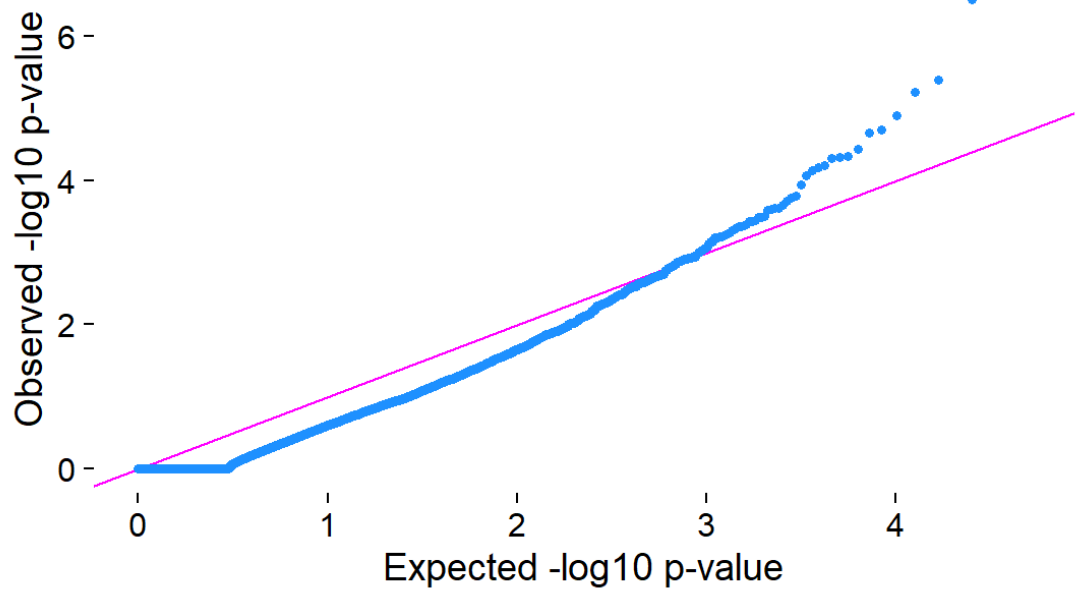
Kuva 7 - Esimerkki manhattan-kuvaajasta morfin GWAS tuloksille. Kuvaajassa kunkin datapisteen nimi kuvataan x-akselilla, ja y-akselilla on negatiivinen logaritmi p-arvojen merkitsevyydelle per SNP. Eli merkitsevimmät SNP:t, joiden p-arvo on lähimpänä nollaa, nousevat korkeimmalle y-akselilla. Suurin osa pisteistä onkin ei-merkitseviä ja ovat 0,0-tasolla.



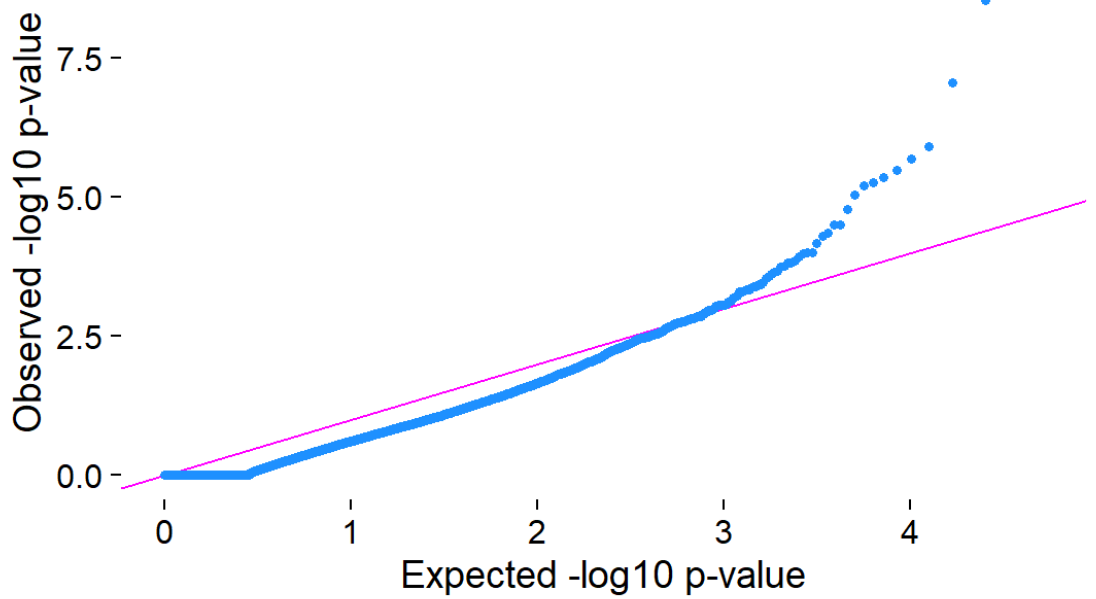
Kuva 8 - Odotettujen ja havaittujen SNP:ien ja värimorfien (*uniformis* vs. *albasusca*) assosiaatioiden merkitsevyyksien suhde aineistossa (siniset pisteet) ja täydellisen täsmäävyyden suora (punainen viiva). Käytännössä, mitä paremmin pisteet seuraavat suoraa, sitä paremmin data selittää tutkittua ominaisuutta. Se, että pisteet erkanevat hieman suoralta voi tarkoittaa, että lineaarinen malli ei täysin kuvaa todellisuutta. Pisteiden erkanemista suoralta tapahtuu, jos aineiston jakauma ei vastaa normaalia jakaumaa, joka on oletuksena tilastollisessa mallissa.



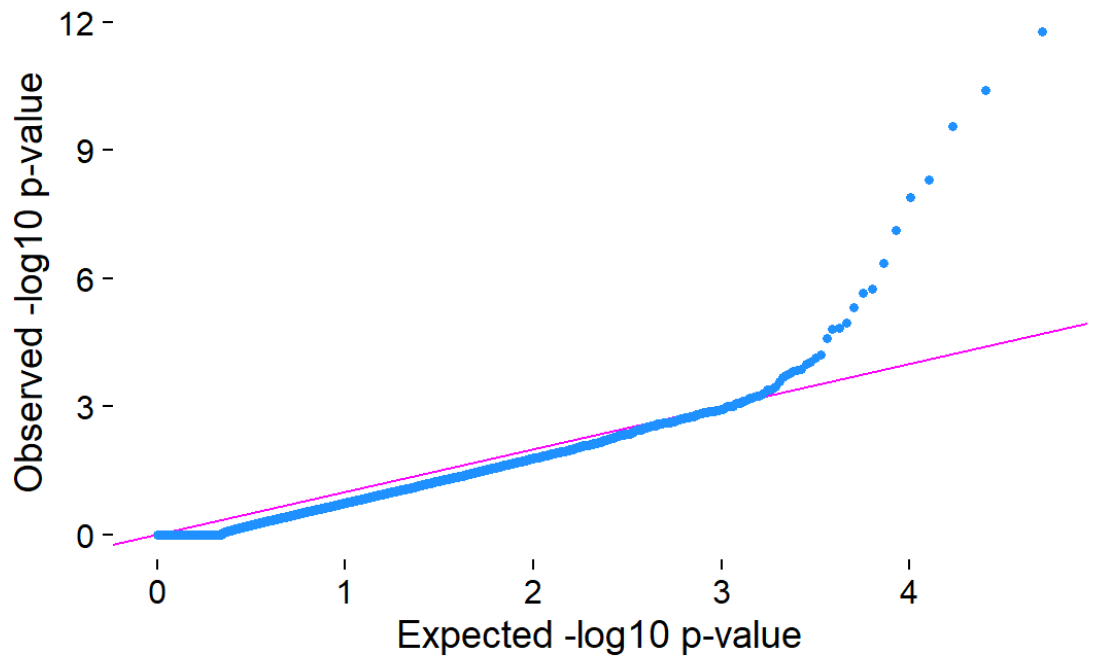
Kuva 9 - Odotettujen ja havaittujen SNP:ien suhde aineistossa valkoisen alueelle sukupuoli- ja populaatio-kovariaattien kanssa. Pisteet erkanevat enemmän suoralta kuin morfin vaikutuksen kuvaajassa, mutta havaittu jakauma on edelleen lähellä teoreettista viitaten siihen, että lineaarinen malli vastaa läheisesti todellisuutta.



Kuva 10 - Odotettujen ja havaittujen SNP:ien suhde aineistossa valkoisen alueelle pelkän sukupuoli-kovariaatin kanssa. Kuten valkoisen alue kovariaattien kanssa, havaittu jakauma on lähellä teoreettista viitaten siihen, että lineaarinen malli vastaa läheisesti todellisuutta, jopa pelkän sukupuoli-kovariaatin kanssa.



Kuva 11 - Odotettujen ja havaittujen SNP:ien suhde aineistossa valkoisen alueelle ilman kovariaatteja. Aiempiin verrattuna pisteet erkanevat enemmän suoralta.



Kuva 12 - Odotettujen ja havaittujen SNP:ien suhde aineistossa valkoisen prosentuaaliselle osuudelle sukupuoli- ja populaatio-kovariaattien kanssa. Pisteet erkanevat enemmän suoralta kuin pinta-alallisen osuuden kuvaajissa, vaikka käytössä ovat samat kovariaatit.

### 3.3 Bedtools- ja BLAST-analyysi

P-arvojen korjauksen jälkeen jäljelle jääneille merkitseville SNP:ille suoritettiin bedtools-analyysit. Valkoisen prosentuaalisen osuuden bedtools-analyysi löysi yhden geenialueen, jossa sekvenssit täsmäsivät (engl. expressed sequence match). Tämä oli tulostiedostossa ilmoitettu seuraavasti ”ID=261941:Idba\_scaffold55789:hit:239062:3.2.0.0;Name=TRINITY\_DN92960\_c6\_g1\_i1”. Tämä on ennustettu geeni, jolle ei ole vielä tarkempaa annotaatiota leväsiiran nykyisessä genomi versiossa. Sen sijaan useat muut bedtools-analyysit eivät löytäneet SNP:n läheisyydestä geeniä sisältävää aluetta, mikä johtuu todennäköisesti leväsiiran genomien pilkkoutuneisuudesta, eli siitä, että genomi koostuu lyhyistä sekvenssipätkistä: telineistä. Tilannetta hankaloittaa se, että haetut SNP:t eivät välttämättä sijaitse geenissä, vaan geenistä ylä- tai alavirtaan. Leväsiirujen genomi ei ole linkitetty kromosomeiksi, vaan aineisto koostuu lyhyemmistä telinepätkistä, joiden sijaintia suhteessa toisiinsa genomissa ei ole määritetty, mikä tekee analyyseistä hankalampia, koska referenssigenomin telinepätkät ovat lyhyempiä kuin annettu aloitus- ja lopetusemäsparin väli. Osa analyyseistä ei myöskään vain löytänyt geenialueita.

Kun bedtools-analyysit eivät tuottaneet tuloksia, niin merkittävien SNP:ien listoille suoritettiin sen sijaan BLAST-analyysit. BLAST-tuloksissa osumat on pisteytetty

korkeammaksi riippuen täsmäävien emästen määrästä. Analyysin e-arvo toimii samalla periaatteella kuin p-arvo, mitä lähempänä nollaa luku on, sitä merkitsevämpi tulos on. Henkilökohtaisesti yritin pitää tulosten kohdalla saman 0,05 merkitsevyyden raja-arvon kuin p-arvojenkin kanssa. Alla olevassa taulukossa (Taulukko 4) ovat tutkimani 5 proteiinia per analyysi.

Taulukko 4 – BLAST-tuloksista proteiinit joihin perehdyin. Osassa analyysit tuottivat tuloksia, jotka pääasiallisesti viittasivat parasiitteihin, bakteereihin tai retroviruksiin. Tämän vuoksi näistä on tutkittu alle 5 proteiinia muista poiketen.

		<b>Proteiinin koodi</b>	<b>Pistearvo</b>	<b>E-arvo</b>
Yksittäinen ennustettu geenialue	<i>Hirondellea gigas</i>	A0A6A7FXL7	175	7e-52
	<i>Hirondellea gigas</i>	A0A6A7FMS9	162	6e-47
	<i>Hirondellea gigas</i>	A0A6A7FTZ3	150	5e-43
	<i>Hirondellea gigas</i>	A0A2P2I8P3	164	7e-40
	<i>Hirondellea gigas</i>	A0A2P2I0Q3	140	2e-34
Valkoisen pinta-alallinen osuus molemmilla kovariaatteilla	<i>Daphnia magna</i>	A0A164T4Z6	38,9	0,083
Valkoisen pinta-alallinen osuus sukupuoli-kovariaatilla	<i>Daphnia magna</i>	A0A164T4Z6	38,9	0,083
Valkoisen pinta-alallinen osuus ilman kovariaatteja	<i>Stylophora pistillata</i>	A0A2B4RC22	75,1	4e-14
	<i>Stichopus japonicus</i>	A0A2G8JDJ8	74,7	7e-14
	<i>Scophthalmus maximus</i>	A0A6A4TT83	74,7	7e-14
	<i>Actinia tenebrosa</i>	A0A6P8GXQ7	74,3	1e-13
	<i>Nothobranchius kuhntae</i>	A0A1A8IG20	73,6	2e-13
Valkoisen prosentuaalinen osuus molemmilla kovariaatteilla	<i>Hirondellea gigas</i>	A0A6A7FZB1	43,1	0,002
	<i>Armadillidium vulgare</i>	A0A444SYP8	39,7	0,039
Valkoisen prosentuaalinen osuus sukupuoli-kovariaatilla	<i>Hirondellea gigas</i>	A0A6A7FZB1	43,1	0,002
	<i>Armadillidium vulgare</i>	A0A444SYP8	39,7	0,039
Valkoisen prosentuaalinen osuus ilman kovariaatteja	<i>Hirondellea gigas</i>	A0A6A7FZB1	43,1	0,002
	<i>Armadillidium vulgare</i>	A0A444SYP8	39,7	0,039
	<i>Rotaria socialis</i>	A0A821U4Q8	48,5	3e-05
	<i>Hirondellea gigas</i>	A0A2P2I5E4	40,0	0,024
	<i>Aphis craccivora</i>	A0A6G0ZQP4	39,7	0,039

## 4 Pohdinta

### 4.1 Väriytyksen ilmiasun vaihtelu

Kuva-analyysin perusteella *albafusca*-yksilöiden välillä on laajaa vaihtelua. Tämä tulee ilmi laikkujen koossa – mikä itsessään jo vaikuttaa valkean pinta-alaan sekä valkean prosentuaaliseen osuuteen – määrässä ja sijainnissa leväsiiran selkäpuolella. Morfi on yleisesti luokiteltu diskreetiksi ominaisuudeksi, eli ominaisuudeksi, joka ei

vaihtelee ja jonka tuottamat ilmiöt ovat samannäköisiä (Ford, 1945), mutta leväsiirtojen kohdalla aiemmassa tutkimuksessa on todettu, että etenkin *albafusca*-yksilöiden välillä on huomattavaa vaihtelua (Merilaita, 1998). Onkin oletettavaa, että monimutkaisen värityksen taustalla on polygeeninen piirre, fenotyypin plastisuus tai näiden yhdistelmä (White & Rabago-Smith, 2011; Aguilon et al. 2021). Tämä sallisi nopean morfologisen erilaistumisen (Van Belleghem et al. 2017), mikä puolestaan sallii lajien leviämistä laajalle alueelle (Ebel et al. 2015).

Paikallisia lukusuhteita tarkastellen vaikuttaa siltä, että keskimääräisesti koiraat edustavat yleisemmin *uniformis*-morfia. Alueet, mistä löytyi eniten *uniformis*-yksilöitä sijaitsivat Selkämerellä – tarkemmin Pyhämaan ja Rihtniemen alueilta – kun taas Saaristomeren populaatioiden välillä erot ovat huomattavasti pienempiä, ja suhde koiraiden sekä naaraiden *uniformis*-yksilöiden välillä pysyy tutkituissa populaatioissa liki samana. Tulosten perusteella sukupuolten välillä oli *albafusca*-yksilöiden kohdalla tilastollisesti merkitsevää eroa laikkujen lukumäärässä ja laikkujen koossa. Aiemmissä tutkimuksissa, joissa on tutkittu eri populaatioita ja laajempaa aluetta kuin tässä tutkimuksessa, on huomattu laajaa vaihtelua leväsiirpopulaatioiden ilmiöissä (Salemaa, 1978; Merilaita, 2001), vaikkakin selvää syytä vaihtelulle ei tiedetä. Värityksen uskotaan olevan usean abioottisen ja bioottisen tekijän summa.

Kaikista tutkituista yksilöistä eniten vaaleata aluetta omaava yksilö oli koiras, ja jopa 10:stä eniten vaaleata omaavasta yksilöstä puolet oli koiraita. Oletus oli, että naarat eroaisivat koiraista valkoisen suhteellisen pinta-alan osin, mutta tilastollisten testien mukaan merkitsevää eroa sukupuolten välillä ei tässä tutkimuksessa löytynyt. Selvästi prosentuaaliseen vaalean osuuteen vaikuttaa muukin kuin sukupuoli. Aiempien tutkimusten mukaan väritykseen vaikuttaisi etenkin saalistus, joka muokkaa suojaväritystä eri tavoin visuaalisesti erilaisissa ympäristöissä (Jormalainen & Tuomi, 1989; Jormalainen et al. 1995). Pienten populaatiokohtaisten otosten vuoksi on myös mahdollista, että tämä tilanne ei kuvasta todellisuutta ja merkitsevä ero sukupuolten välillä onkin olemassa. Tämä olisi myös varteenotettava tutkimuskohde tulevaisuuden tutkimuksille.

On mahdollista, että värityksen erot voivat tulla vahvasti ilmi paikallispopulaatioidenkin välillä, eli esimerkiksi näiden Selkämeren sijaintien välillä. Vastaavaa on havaittu esimerkiksi artikkelissa (McLean et al. 2015), joka käsittelee liskon, *Ctenophorus decresii*, polymorfisia värityseroja. Lajin jokaisesta tutkitusta



populaatiosta löydettiin samaa neljää morfia, vaikkakin morfien suhteellinen määrä vaihteli populaation mukaan. Geneettisestä populaatorakenteesta huolimatta yhteyttä morfisuhteiden ja geneettisen yhtäläisyyden, tai maantieteellisen läheisyyden ja samankaltaisuuden kanssa ei ollut. Tutkijat kumminkin huomasivat vahvan assosiaation morfisuhteiden ja ympäristötekijöiden kanssa – tarkemmin keskiarvoisen vuoden kuivuusindeksin sekä kasvipeitteen välillä. Heikko geenivirta leväsiirapopulaatioiden välillä voi selittää isommat erot populaatioiden välillä, kuten morfisuhteissa *C. decresii*n kohdalla. Toisaalta päinvastaisesti vahva geenivirta tarkoittaa, että populaatiot eroavat geneettisesti vain vähän toisistaan.

On mahdollista, että minun tutkimuksessani populaatioiden välisten erojen löytämiseen olisi tarvittu edustavampi otos, joka paranee otoskoon ja satunnaistamisen myötä. Kukin otos on määritetty vastaamaan alkuperäispopulaatiota, mutta 20 yksilön otos on niukka kuvaus todellisesta morfijakaumasta. Lisäksi joidenkin populaatioiden kohdalla täytyi esimerkiksi valita enemmän naaraita sekvensoitavaksi kuin mitä populaatiota vastaava sukupuolijakauma vastaisi, sillä muuten sekvensoitavaksi olisi tullut lähinnä vain koiraita. Tämä aineisto on kerätty ja otostettu toista koetta varten, minkä vuoksi vinoumaa ilmenee. Onkin mahdollista, että toisistaan maantieteellisesti kaukana olevat leväsiirapopulaatiot eroavat toisistaan. Rihtniemi ja Pyhämaa ovat maantieteellisesti suhteellisen lähellä toisiaan, vain vajaan 10 kilometrin etäisyydellä vesitse, kun taas Saaristomeren lähin sijainti Katavaluoto on jo noin 80 kilometrin päässä vesitse Selkämeren sijainneista. Rihtniemen ja Pyhämaan läheisyys todennäköisesti sallisi vahvan geenivirran, joka mukauttaisi populaatiot toistensa kaltaisiksi. Vastaavaa todistettiin De Witin (2020) artikkelissa, jossa todettiin, että pitkän matkan levittäytyminen on harvinaista leväsiirroilla. Leväsiirat leviävät pisimpiä matkoja passiivisesti virtausten mukana aikaisessa kehitysvaiheessa. Aikuiset leväsiirat leviävät pidempiä matkoja mahdollisesti kasvillisuuden mukana, mutta yleisimmin liikkuvat itse lyhyitä matkoja. Näin lähipopulaatiot muistuttavat lähemmin toisiaan, kun taas kaukaisemmat populaatiot eroavat suuremmin.

Toinen huomioitava asia on, että Katavaluodon näytteet on kerätty vuonna 2014 ja Rihtniemen ja Pyhämaan vuonna 2021. Jos Katavaluodosta olisi kerätty näytteitä myös vuonna 2021 olisi populaation sisäisiä eroja voitu verrata vuosien välillä, kuin myös eroa Selkämeren sijainteihin. Jos populaatioiden välillä on heikkoa geenivirtaa, on mahdollista, että geneettiset erot populaatioiden välillä olisivat suuremmat ajan

kuluessa, kun populaatioille yleiset geenit yleistyisivät kussakin populaatiossa. Nosilin artikkelissa (2009) todetaan, että osittaista lajiutumista (engl. incomplete speciation) – eli geneettisten erojen lisääntymistä, mutta laji on silti selkeästi sama eroista huolimatta – voidaan selittää useilla aikaan liittyvillä, maantieteellisillä ja geneettisillä tekijöillä. Nämä oletukset ovat täysin hypoteettisia leväsiirujen osilta, sillä tällaisen osittaisenkin lajiutumisen ilmentyminen tulisi todistaa tarkemmilla kokeilla. Tästä huolimatta on vähintään mahdollista todeta, että geneettisiä eroja voisi ilmentyä leväsiirapopulaatioiden välillä näiden maantieteellisen etäisyyden vuoksia, ja populaatioiden sisällä ajan kulun kanssa.

#### 4.2 Värikyksen ilmiasun ja genomin assosiaatio

Kuvaajista (Kuvat 8–12) pystyy näkemään, että genotyypin ohella käyttämäni kovariaatit olivat oleellisia selittämään tutkittuja värikyksen piirteitä, eli valkoisen alueen kokoon ja valkoisen prosentuaaliseen osuuteen. Kovariaateista etenkin sukupuoli vaikuttaisi olevan merkitsevä tekijä väriykselle, sillä kuvien 8 ja 9 pisteet muistuttavat läheisesti toisiaan, vaikka jälkimmäisessä kuvassa sisällytettynä kovariaattina on vain sukupuoli. Kun näitä vertaa kuvan 10 ilman kovariaatteja tuotettuun kuvaajaan, voi huomata, että ilman kovariaatteja pisteet erkanevat enemmän suoralta. Morfin kohdalla ongelmana on, että bedtools-analyysit eivät tunnistaen ainuttakaan merkitsevää SNP:ä P-arvojen korjausten jälkeen. Tämä todennäköisesti johtuu siitä, että aineistostani tuleva signaali (eli morfia selittävät SNP:t) ei ole niin vahva, että se poikkeaisi satunnaisesta. Jos näyttemäärä olisi isompi ja vastaavasti kattavuus syvempi, merkitseviä SNP-alueita saattaisi löytyä. Tämä olisi varteenotettava tutkimuskohde, mihin olisi mahdollista perehtyä syvemmin tulevaisuuden tutkimuksissa. Tätä varten voitaisiin alkujaankin ottaa populaatioittain suuremmat otokset. Myös leväsiiran genomin annotoinnin tullessa kattavammaksi, olisi mahdollista, että geneettisillä analyyseillä saataisiin parempi selvyys SNP:den vaikutuksista morfeihin – toisin kuin tämän tutkielman kohdalla.

Tutkielman ANGSD-tuloksissa erikoisinta on, että yhdenkään telineen kohdalla ei aineistosta tunnistettu yhtäkään resessiivistä homotsygoottia. Muutamaa poikkeusta lukuun ottamatta kaikkein yleisimpiä joka telineessä ovat dominoivan alleelin homotsygootit ja loput heterotsygootteja. Näissä muutamassa poikkeuksessa yleisin on heterotsygootti ja loput dominoivia homotsygootteja. Tämän tutkimuksen datan osilta tämä todennäköisesti tarkoittaa sitä, että matalan kattavuuden ja käytössä

olevan otoskoon vuoksi harvinaisten resessiivisten homotsygoottien tunnistaminen on epätodennäköisiä.

BLAST-analyysien jälkeisiä merkittäviä proteiineja tutkin tietokannoista. Osa tuloksista on todennäköisesti verrattu bakteerin perimään, sillä useassa kohtaa nähdään ”Reverse transcriptase...” tai ”Retrovirus-related Pol poly from transposon...” alkuisia proteiineja, jotka todennäköisesti ovat siiran perimään sekoittunutta bakteerin perimää. Tämä on todennäköisesti historiallista vaikutusta bakteerien toistojaksojen proteiineilta, jotka ovat jääneet osaksi siiran genomia. Näistä huolimatta useiden muiden telineiden kohdalla on merkittäviä osumia, jotka vastaavat proteiineja muilla – etenkin merellisillä – eliöillä.

BLAST tuloksista valkoisen alueelle molempien kovariaattien kanssa ja pelkän sukupuoli-kovariaatin kanssa *Daphnia magna* viittaava proteiini ”A0A164T4Z6” – vaikkei kovin korkeasti pisteytetty (38,9) tai järin merkitsevä e-arvo (0,083) – on ajon merkittävin tulos, sillä loput osumista viittaavat parasiitteihin, bakteereihin tai retroviruksiin liittyviin proteiineihin. *Daphnia magna* on äyriäisiin luettu laji kuten leväsiirakin. Proteiinia ei ole laajalti annotoitu – mikä on yhteistä monille tuloksissa esiintyville proteiineille – mutta UniProt tietokannan verkkosivu merkitsee proteiinille merkityksiksi solun rakenneosiin vaikuttaminen DNA polymeraasina kompleksina, molekylaariset toiminnot muun muassa aminohapon sitominen, ja biologisissa prosesseissa muun muassa proteolyysissä.

Leväsiiran värityksen valkoisen alueen prosentuaalisen osuuden tuloksista yhdelle telineelle merkittävin tulos on A0A6A7FZB1-proteiini, joka viittaa lajiin *Hirondellea gigas*, joka on leväsiiran tavoin äyriäisiin luettu laji. Proteiinin merkityksiksi on taas merkitty samat molekylaarisena toimintona aminohapon sitominen ja biologisissa prosesseissa DNA:n integroiminen, kuin myös solun rakenneosiin vaikuttaminen DNA polymeraasi kompleksina. Toisen telineen osumissa on laajasti vaihtelua, mutta ainakin 24 proteiinia viittaa lajiin *Cryptotermes secundus*, joka on eräs niveljalkaisiin luettu termiittilaji. Pisteytys ja e-arvo vaihtelee huomattavasti proteiinin mukaan, mutta kaikki merkitsevät osumat ovat välillä pisteytyksen osin 45,1–55,5 ja e-arvoltaan 0,002–0,0000004. Useilla näistä proteiineista merkitykseksi on mainittu biologisena prosessina DNA:n biosyntettinen prosessi, eli jokin prosessi, jonka lopputuloksena on DNA:n muodostuminen.

Eliöiden proteiinien heikon annotoinnin vuoksi on tässä vaiheessa vaikea sanoa, onko löydettyillä SNP:eillä ja niistä tunnistetuilla proteiineilla vaikutusta väritykseen tietämättä proteiinien täyttä merkitystä. Tunnetusti geneilla on monia mekanismeja vaikuttaa eliöiden väritykseen. Esimerkiksi *Clanculus margaritarius* ja *Clanculus pharaonius* ovat läheisiä merietana lajeja, joita tutkimalla huomattiin, että samankaltaisesti väritetyt ja saman muotoiset alueet määräytyivät samojen geenien mukaan kummallakin lajilla (Williams et al. 2017).

#### 4.3 Päätelmät

Tämän tutkielman tavoitteena oli vastata kysymyksiin, onko väritys yhden geenin säätelemä vai polygeenisesti säädelty, ja onko sukupuolten tai populaatioiden välillä huomattavia eroja tai näille tyypillisiä piirteitä?

Tämän pro gradun tutkimusten mukaan näyttäisi siltä, että geneilla on merkitystä siirojen väritykseen, etenkin valkoisen pinta-alalliseen osuuteen (Taulukot 2 ja 3). Läheisten siiralajien värien vertailevaa geenitutkimusta ei ilmeisesti ole toteutettu, mikä voisi lisätä tietämystä siirojen värityksen mekanismeista. Lisäksi siirojen laajan kirjon väritykset, jopa populaation sisällä ja yksilöiden välillä viittaavat siihen, että väritys on polygeeninen piirre (Aguillon et al. 2021; White & Rabago-Smith, 2011). Myös tämän tutkimuksen mukaan, esimerkiksi valkoinen alue ei täysin selittyisi pelkällä sukupuoli-kovariaatilla (Kuva 10), mikä viittaisi siihen, että tekijöitä on mahdollisesti useampia. Tutkimus täytyisi lisäksi toistaa isommalla otoskoollla. Tässä tutkielmassa populaatioista otostettiin vain 20 yksilöä, mikä on hyvin pieni määrä verraten koko populaation yksilöiden lukumäärään. Suuremmalla testijoukolla tulokset voisivat vastata läheisimmin todellisuutta, jolloin tuloksissa tämä näkyisi muun muassa cowplot-kuvaajien pisteiden seuraavan lähemmin kuvaajien suoraa.

Satunnaismuuttuja testien mukaan populaatioiden välinen vaihtelu on vähäistä. Tulevaisuudessa olisi mahdollista tuottaa tutkimus, joka keskittyisi täysin eriäviin väritykseen vaikuttaviin tekijöihin ja näiden merkitsevyyteen.

Naaraiden laikkujen määrän ja prosentuaalisen osuuden välinen negatiivinen korrelaatio tarkoittaa, että laikkujen määrän kasvaessa ne samalla pienenevät. Tämä yhdessä sukupuolten välillä tehtyjen t-testien ja väritykseen vaikuttavien tekijöiden välillä tehtyjen testien kanssa viittaa sukupuolen olevan merkitsevä piirteelle, vaikka ilmiö havaitaan vain toisessa sukupuolella. Jos tulosten perusteella pitäisi piirtää tyypillinen *albafusca*-koiras, sillä olisi paljon pieniä laikkuja, kun taas naaraalla olisi

vähän suuria laikkuja. Tämä ei suoraan välity prosentuaaliseen osuuteen, joten prosentuaalisesti koiraalla ja naaraalla voi olla 50% valkoista, mutta se miten valkoinen on selänkuvioinnissa rytmittynyt, eroaa. Eli, esimerkiksi koiraalla voi olla 25 pientä laikkua siellä täällä ja naaraalla kolme isoa vyömäistä laikkua, ja näillä olisi silti prosentuaalisesti yhtä iso osa valkoista aluetta näiden selän värityksessä. Tästäkin olisi mahdollista toteuttaa laajempi ja täsmennetty tutkimus tulevaisuudessa leväsiira sukupuolten erojen vertailemiseksi, ja värityksen ominaisuuksien korrelaatioiden vertailemiseksi.

Tämän tutkimuksen tulokset eivät siis vastaa kaikkiin tutkimuskysymyksiin toivotulla tarkkuudella. Suuntaa antavat vastauksetkin voivat toimia ohjeina tulevaisuuden tutkimuksia varten vihjaten luonnossakin esiintyviin eroihin sukupuolten tai populaatioiden välillä, tai korrelaatioihin värityksen ominaisuuksien välillä. Leväsiirapopulaatioiden välillä on kumminkin todettu olevan laajaa vaihtelua, ja populaatioita on kyetty jakamaan erottuviksi joukoiksi populaatioissa esiintyvien ilmiäsuojen suhteellisten määrien mukaan (Salemaa, 1978). Uskon, että tuloksilla on arvoa tässäkin muodossa, mutta tulevaisuuden tutkimusten tulee laajentaa ja täsmentää tutkimusaineistoa ja tutkimuskysymyksiä vastatakseen luonnossa esiintyviä populaatioita mahdollisimman tarkasti, ja tuottaakseen tarkempia tuloksia leväsiirajien ominaisuuksista.

## Lähteet

Aguillon, S. M., Walsh J., and Lovette I. J. (2021). Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proceedings of the Royal Society B* 288.1943: 20201805.

Bates D., Maechler M., Bolker B., and Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.  
doi:10.18637/jss.v067.i01.

Bolger, A.M., Lohse M., and Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30.15: 2114-2120.

Bond, A. B. (2007). The evolution of color polymorphism: crypticity, searching images, and apostatic selection. *Annual Review of Ecology, Evolution, and Systematics*. 38: 489-514.

Bradshaw Jr, H. D. and Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426.6963: 176-178.

Buinheim, H.-P., and Faya, G. (1982). Colour polymorphism and genetic variation in *Idotea baltica* populations from the Adriatic Sea and Baltic Sea. *Genetica* 59.3: 177-190.

Claus O. W. (2020). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1. <https://CRAN.R-project.org/package=cowplot>

Cook, L. M., Grant, B. S., Saccheri, I. J. and Mallet, J. (2012). Selective bird predation on the peppered moth: the last experiment of Michael Majerus. *Biology Letters* 8.4: 609-612.

Cribari-Neto F., Zeileis A. (2010). "Beta Regression in R." *Journal of Statistical Software*, 34(2), 1–24. doi:10.18637/jss.v034.i02.

Davison, A., Jackson, H. J., Murphy, E. W. and Reader, T. (2019). Discrete or indiscrete? Redefining the colour polymorphism of the land snail *Cepaea nemoralis*. *Heredity* 123.2: 162-175.

De Wit, P., Jonsson, P. R., Pereyra, R. T., Panova, M., André, C. and Johannesson, K. (2020). Spatial genetic structure in a crustacean herbivore highlights the need for local considerations in Baltic Sea biodiversity management. *Evolutionary Applications* 13.5: 974-990.

Ebel, E.R., DaCosta, J.M., Sorenson, M.D., Hill, R.I., Briscoe, A.D., Willmott, K.R. and Mullen, S.P. (2015). Rapid diversification associated with ecological specialization in Neotropical *Adelpha* butterflies. *Molecular Ecology*, 24(10), pp.2392-2405.

Ford, E. B. (1945). Polymorphism. *Biological Reviews* 20.2: 73-88.

Forsman, A., Karlsson, M., Wennersten, L., Johansson, J and Karpeström, E. (2011). Rapid evolution of fire melanism in replicated populations of pygmy grasshoppers. *Evolution*: 65.9: 2530-2540.

- Forsman, A. (2016). Is colour polymorphism advantageous to populations and species? *Molecular Ecology*: 2693-2698.
- Galeano, S. P., and Harms, K. E. (2016). Coloration in the polymorphic frog *Oophaga pumilio* associates with level of aggressiveness in intraspecific and interspecific behavioral interactions. *Behavioral Ecology and Sociobiology* 70.1: 83-97.
- Guarino, S. M., Gambardella, C., Ianniruberto, M. and de Nicola, M. (1993). Colour polymorphism in *Idotea baltica* from the Bay of Naples and its ecological significance. *Journal of the Marine Biological Association of the United Kingdom* 73.4: 785-794.
- Hamelin, E. T. (1963). Définition et analyse génétique du phénotype Pseudolineata de l'isopode valvifère *Idotea balthica* (Pallas). *Crustaceana*: 133-137.
- Jones, J. S., Selander, R. K. and Schnell, G. D. (1980). Patterns of morphological and molecular polymorphism in the land snail *Cepaea nemoralis*. *Biological Journal of the Linnean Society* 14.3-4: 359-387.
- Jormalainen, V., and Tuomi J. (1989). Sexual differences in habitat selection and activity of the colour polymorphic isopod *Idotea baltica*. *Animal Behaviour* 38.4: 576-585.
- Jormalainen, V., Merilaita, S. and Tuomi J. (1995). Differential predation on sexes affects colour polymorphism of the isopod *Idotea baltica* (Pallas). *Biological Journal of the Linnean Society* 55.1: 45-68.
- Jun, G., Wing M. K., Abecasis G. R., Kang H. M. (2015). An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome research* 25.6: 918-925.
- Kemp, D. J., Holmes, C., Congdon, B. C. and Edwards, W. (2013). Color polymorphism in spiny spiders (*Gasteracantha fornicata*): testing the adaptive significance of a geographically clinal lure. *Ethology* 119.12: 1126-1137.
- Kettlewell, H. B. D. (1955). Recognition of appropriate backgrounds by the pale and black phases of *Lepidoptera*. *Nature* 175.4465: 943-944.
- Korneliussen, T.S., Albrechtsen A., and Nielsen R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* 15.1: 1-13.
- Lamotte, M. (1959). Polymorphism of natural populations of *Cepaea nemoralis*. *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 24. Cold Spring Harbor Laboratory Press.
- Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods* 9: 357-359.
- Legrand-Hamelin, E., Legrand, J. J., Le Bote, C. and Heily, G. (1982). Déterminisme génétique de l'inversion du sexe femelle accompagnant le croisement de deux sous-espèces d'*Idotea balthica* (Pallas). II.—Résultats concernant le phénotype *bilineata-lineata* et son association avec le phénotype *albafusca*. *Reproduction Nutrition Développement* 22.6: 889-904.

- Leidenberger, S., Harding, K. and Jonsson, P. R. (2012). Ecology and distribution of the isopod genus *Idotea* in the Baltic Sea: key species in a changing environment. *Journal of Crustacean Biology* 32.3: 359-389.
- Lou, R. N., Jacobs, A., Wilder, A. P. and Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology* 30.23: 5966-5993.
- Madden T. (2002). The BLAST Sequence Analysis Tool. 2002 Oct 9 [Updated 2003 Aug 13]. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda
- McLean, C. A., Devi S.-F. and Moussalli. A. (2015). Environment, but not genetic divergence, influences geographic variation in colour morph frequencies in a lizard. *BMC Evolutionary Biology* 15.1: 1-10.
- McEntyre J., and Ostell J. (2002). The NCBI Handbook [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); Available from: <https://www.ncbi.nlm.nih.gov/books/NBK21101/>
- Merilaita, S. (1998). Crypsis through disruptive coloration in an isopod. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265.1401: 1059-1064.
- Merilaita, S. (2001). Habitat heterogeneity, predation and gene flow: colour polymorphism in the isopod, *Idotea baltica*. *Evolutionary Ecology* 15.2: 103-116.
- Nosil, P., Harmon L. J., and Seehausen O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution* 24.3: 145-156.
- Pener, M. P. (1991). Locust phase polymorphism and its endocrine relations. *Advances in Insect Physiology*. Vol. 23. Academic Press: 1-79.
- “Picard Toolkit.” (2019). Broad Institute, GitHub Repository. <https://broadinstitute.github.io/picard/>; Broad Institute
- Protas, M. E. and Patel, N. H. (2008). Evolution of coloration patterns. *Annual Review of Cell and Developmental Biology* 24: 425-446.
- Quinlan, A.R., and Hall I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26.6: 841-842.
- R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>
- Salemaa, H. (1978). Geographical variability in the colour polymorphism of *Idotea baltica* (*Isopoda*) in the northern Baltic. *Hereditas* 88.2: 165-182.



- Surmacki, A., Ożarowska-Nowicka, A. and Rosin, Z. M. (2013). Color polymorphism in a land snail *Cepaea nemoralis* (Pulmonata: Helicidae) as viewed by potential avian predators. *Naturwissenschaften* 100.6: 533-540.
- Wickham H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Van Belleghem S.M., Rastas P., Papanicolaou A., Martin S.H., Arias C.F., Supple M.A., Hanly J.J., Mallet J., Lewis J.J., Hines H.M. and Ruiz M. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. *Nature Ecology & Evolution*, 1(3), pp.1-12.
- White, D., and Rabago-Smith M. (2011). Genotype–phenotype associations and human eye color. *Journal of Human Genetics* 56.1: 5-7.
- White, T. E. and Kemp, D. J. (2016). Colour polymorphism. *Current Biology* 26.13: R517-R518.
- White, T. E. and Kemp, D. J. (2016). Color polymorphic lures target different visual channels in prey. *Evolution* 70.6: 1398-1408.
- Williams, S. T., Lockyer, A. E., Dyal, P., Nakano, T., Churchill, C. K. C. and Speiser, D. I. (2017). Colorful seashells: Identification of haem pathway genes associated with the synthesis of porphyrin shell color in marine snails. *Ecology and evolution* 7.23: 10379–10397.
- Zhao, L., Lin L. and Zheng Z. (2016). DNA barcoding reveals polymorphism in the pygmy grasshopper *Tetrix bolivari* (Orthoptera, Tetrigidae). *ZooKeys* 582: 111.