

# Koiran ja suden genomiikka

Santtu Urpilainen

LuK-tutkielma

Biologian tutkinto-ohjelma, biotiede

Oulun yliopisto

Toukokuu 2018

Avainsanat: *Canis lupus familiaris*, *Canis lupus*, genomiikka, introgressio, domestikaatio

## Sisällysluettelo

1. Johdanto .....	3
2. Koiran genomiikka .....	4
2.1. Koiran alkuperä .....	4
2.2. Genominen data .....	6
2.2.1. Bokseri .....	6
2.2.2. Koreanjindonkoira .....	8
3. Suden genomiikka .....	9
3.1. Suden genomi .....	9
4. Koiran ja suden genomien vertailu .....	10
4.1. Suden ja koiran introgressio .....	11
4.1.1. Tiibetin mastiffi ja susi .....	11
4.1.2. Italian susi .....	11
4.2. Suden ja koiran populaatiogenetiikka .....	12
5. Pohdinta ja johtopäätökset .....	14
Lähteet .....	15

Lyhenteet ja termistö:

SNP: single nucleotide polymorphism

Jatkumo: contig

Kehys: scaffold / supercontig

N50: on pienimmän jatkumon pituus, jota tarvitaan peittämään 50 % koko genomien koosta. Sitä käytetään kuvaamaan genomisekvenssin kokonaisuutta.

LD: linkage disequilibrium / kytkentäepätasapaino

# 1. Johdanto

Freedmanin ja Waynen 2016 mukaan koira (*Canis lupus familiaris*) on ensimmäinen ja ainoa ihmisen kesyttämä suurpeto. Sen alkuperäisen domestikaatioprosessin jälkeen noin viimeisen 200 vuoden aikana koirilla on nopeasti tapahtunut fenotyypin radiaatiota. Ihminen on vahvasti ohjannut valintaa niin, että uusia koirarotuja on muodostunut ja niitä on ylläpidetty. Koirilla fenotyypin diversiteetti on laajempaa kuin suurella osalla selkärangkaisista. Tätä omalaatuista jalostamisen historiaa voidaan hyödyntää tautien, morfologisen variaation ja käyttäytymisen tutkimisessa. Koiran domestikaation historian kautta koiralle on kehittynyt suurta morfologista diversiteettiä, jonka seurauksena nykyisillä koiraroduilla useat taudit ovat yleistyneet. Ihmisellä esiintyy samoja tauteja, minkä vuoksi koiran genomia voidaan käyttää hyväksi ihmisen tautien ja genomien tutkimuksessa (Lindblad-Toh ym., 2005). Serres-Armero ym. 2017 mukaan koiran ja suden (*Canis lupus*) välistä geneettistä variaatiota on tutkittu käyttäen enimmäkseen yhden nukleotidin polymorfioita (Single Nucleotide Polymorfism) ja mikrosatelliitteja. Näiden tutkimusten pohjalta on saatu selville, että koiran nukleotididiversiteetti on matalampi kuin sudella, mikä on seurausta domestikaatiosta. Gopalakrishnan ym. 2017 mukaan susien ja koirien läheisestä sukulaisuudesta johtuen niiden harvinaiset ja yksityiset variaatiot antavat informaatiota lajien eroista, minkä vuoksi referenssigenomin valinta on näiden varianttien kannalta tärkeää. Kuitenkin domestikaation geneettiset seuraukset, populaatioiden historiallinen isolaatio ja demografinen vaihtelu ovat johtaneet siihen, että suden ja koiran populaatiot eroavat toisistaan suuresti mikrosatelliittien alleelifrekvenssien osalta. Useiden lajille spesifien alleelien läsnäolo viittaa siihen, että suden ja koiran populaatioiden geenienvaihto on ollut vähäistä (Randi & Lucchini, 2002).

Genomitutkimuksella on mahdollista tunnistaa geeneihin liittyvät fenotyypiset ominaisuudet ja taudit hyödyntämällä keskimääräistä määrää geneettisesti muuntelevia paikkoja. Kuitenkin koirilla on samankaltaisia ominaisuuksia fiksoituneina useisiin eri rotuihin ja siten fenotyypin geneettistä taustaa voidaan tutkia toiston kautta (Wayne & vonHoldt, 2012). Referenssigenomin valinta vaikuttaa populaatiorakenteen havaitsemiseen ja demograafisten tapahtumien ajoittamiseen, kun tutkitaan useita toisilleen sukua olevia lajeja. Ongelman odotetaan muuttuvan haastavammaksi mitä läheisempiä lajit ovat keskenään (Gopalakrishnan ym., 2017). Kytkeäpäätasapainon

laajuutta luonnollisissa selkärangaispopulaatioissa on aikaisemmin ollut vaikeaa arvioida suurten genomitutkimusten ollessa mahdollisia vain mallilajeilla. Nykyään tehokkaamman genotyyppityksen suoritustehon ja genomien sekvensointiprojektien avulla on mahdollista hahmottaa helpommin luonnollisia populaatioita (Grey ym., 2009).

Vilá ym. 2003 mukaan hybridisaatio tapahtuu yleensä vain toisille läheisten taksonien välillä, minkä seurauksena hybridisoivien populaatioiden erottaminen toisistaan tapahtuu alleelifrekvenssien eroavaisuuksien avulla lajikohtaisten alleelien yleisyyden sijasta. Lajikohtaisten geneettisten markkerien puuttuminen tekee hybridien tunnistamisen haastavaksi, mutta viimeaikaisten tunnistamismenetelmien kehityksen ansiosta hybridit on mahdollista tunnistaa isäntäpopulaation alleelien perusteella. Autosomaalisten ja parentaalisten markkerien yhteysvaikutuksella on mahdollista tehdä päätelmiä hybridisaatiotapahtumien suunnasta. Kuitenkaan niin tarkkaa tietoa ei ole ollut mahdollista kerätä, mikä on seurausta Y-kromosominmarkkerien yleisestä puutteesta.

Tämän kirjallisuuskatsauksen tarkoituksena on tuoda esille, millaista tutkimusta on olemassa suden ja koiran genomiikasta, sekä havainnoida tutkimusten tuloksista kyseisten lajien sukulaissuhdetta, hybridisaatiota ja introgressiota. Tutkimuskysymykseni ovat:

Miten sudet ja koirat eroavat toisistaan genomisen datan pohjalta?

Miten referenssigenomin valinta vaikuttaa suden ja koiran sukulaisuuden tutkimisen tuloksiin?

## 2. Koiran genomiikka

### 2.1. Koiran alkuperä

Freedmanin ja Waynen 2016 mukaan riippumatta siitä, että koiran fenotyyppisten piirteiden geneettinen tausta on nykyään enemmän tunnettu ja niiden genomisen datan sekvensointi on tullut helpommaksi, ei koiran domestikaation ajoituksesta, maantieteellisestä alkuperästä eikä ekologisesta kontekstista olla yksimielisiä. Pääasiallisena syynä voidaan pitää sitä, että nykysuden ja -koiran eroaminen on

tapahtunut lyhyessä ajassa, minkä seurauksena tämän haarautumisen tapahtumat ovat selvittämättä. Analyysiä ovat vaikeuttaneet epätäydelliset sukulinjojen lajittelut (incomplete lineage sorting) ja divergenssin jälkeinen geenivirtaus. Harmaasutta luultavammin koiran esi-isä on ollut myöhäisen pleistoseenikauden aikainen susi. Fylogeneettistä rekonstruointia on vaikeuttanut se, että sudet ja koirat ovat domestikaation aikana eläneet samoissa ympäristöissä. Fylogeneettisen substituutiodatan vähyyden vuoksi muinaisten koirapopulaatioiden ajoittaminen ja niiden havaitseminen on haastavaa (Freedman & Wayne, 2016).

Freedmanin ja Waynen 2016 mukaan varhaisimmat koiran jäänteet on löydetty Euroopasta ja ne ovat noin 27–36 000 vuotta vanhoja. Muualla maailmassa fossiilien ikä vaihtelee 9000 ja 13 000 vuoden välillä. Aasian vanhimmat fossiilit ovat 33 000 vuotta vanhoja. (Freedman & Wayne, 2016). Koirien jäänteiden tunnistaminen arkeologisista tallenteista on ollut vaikeaa. Tämä johtuu domestikaation alkuvaiheessa olevan koiran ja suden morfologian samankaltaisuudesta, mikä vaikeuttaa koiran alkuperän hahmottamista maantieteellisen sijainnin pohjalta. Koira-eläinten samankaltaisuus, fossiilisten näytteiden vähäisyys ja niiden hajanaisuus vaikuttavat koiran alkuperän kartoittamiseen. Arkeologisten kaivausten vähäisyys joissain maanosissa vääristää domestikaation historian kartoitusta, mutta koirien hautaaminen niiden omiin tai ihmisten hautausmaihin on tehostanut koirien arkeologista näkyvyyttä (Larson ym., 2012.) Germonprén ym. 2009 tekemässä tutkimuksessa belgialaisilla suurilla koira-eläimillä havaittiin selvää geneettistä diversiteettiä. Ensimmäisillä koirilla oli mahdollisesti sudenkaltainen geenisekvenssi, minkä seurauksena ne eivät ole geneettisesti tunnistettavissa koiriksi. Domestikaation aikaisessa vaiheessa valikoiva jalostaminen aiheutti pullonkaulan, jonka vaikutuksesta susissa ja koirissa voidaan havaita erilaistumista vasta muutaman tuhannen vuoden jälkeen. Sen seurauksena, että ihmiset hoitivat domestikaation jälkeen sudesta kesytettyjä koiria ne eivät olleet täysin negatiivisen valinnan alaisena, joten epäedullisten piirteiden oli mahdollista siirtyä eteenpäin.

Belgiassa löydettyjen suurien koira-eläinten jäänteiden sekvenssit olivat uniikkeja ja ne eivät vastanneet minkään koirarodun genotyyppiä, minkä vuoksi niiden oletettiin olevan susien jäänteitä. Koiran morfologiset erot suden morfologiaan verrattuna ilmestyivät nopeasti domestikaation vaikutuksesta ja siten ne fiksoituivat koirapopulaatioon. Useiden muinaisten paleoliittisten koirien kallot osoittivat morfologista

samankaltaisuutta suden kallojen kanssa. Vanhimmat näistä kalloista olivat Aurignacin kulttuurin aikaisia, joten on mahdollista, että koiran domestikaatio oli jo alkanut ~ 31 700 vuotta sitten (Germonpré ym., 2009).

## 2.2. Genominen data

Tässä kappaleessa esitellään kaksi koiran koko genomisekvensointia, joista ensimmäinen Lindblad-Toh ym. 2005 koostama bokseri-koiran genomi on käytössä pääasiassa koiran referenssigenomina. Toinen läpikäytävä koiran genomi on Kim ym. 2012 koostama koreanjindonkoiran genomi.

### 2.2.1. Bokseri

Lindblad-Toh ym. 2005 kokosivat koiran genomista korkealaatuisen sekvenssin, jossa on noin 99% koko koiran eukromaattisesta genomista. Tämän bokseri-koirasta koostettu genominen datan kokonaisuudessa ei ollut puutteita ja sen nukleotidien sekvensoinnin tarkkuus, sekvenssin yhtäjaksoisuus ja sen pitkän etäisyyden liittyneisyys (long-range connectivity) ovat hyvälaatuisia. Samassa tutkimuksesta luotiin vedos koirapopulaatioiden yhden nukleotidien polymorfismin (SNP) datasta ja verrattiin 11 koirarotua bokserinsekvenssiin. Tämän datan avulla tutkittiin koiran genomien rakennetta, sen geenien evoluutiota ja sen haplotyyppien rakennetta sekä koiran fylogeneettista taustaa.

Nisäkkäiden genomien muovautuminen eroaa eri sukujuurissa siten, että koiralla on pienin keskiarvoinen transposonien insertoituminen kaikista nisäkkäistä. Tämä on seurausta pienemmästä endogeenisten retrovirusten ja DNA:n transposonien toiminnasta, sekä koiran lyhyiden sekoittuneiden elementtien (short interspersed element, SINE) pituudesta. Tämän seurauksena koiran genomissa on pienempi määrä toistuvia elementtejä. Useimmilla nisäkkäillä on samoja funktionaalisia elementtejä, jotka ovat puhdistavan valinnan alaisena. Noin 50 % eniten konservoidussa eikoodaavassa sekvenssissä havaittiin merkittävää ryhmittymistä noin 200 geeniköyhässä alueessa, joissa on avainroolina solun identiteetin ylläpitäminen tai luominen, kuten transkriptiotekijät tai aksonin ohjausreseptorit (Lindblad-Toh ym., 2005).

Koirilla on ollut kaksi merkittävää pullonkaulaa: aikainen domestikaatio ja nykyinen rotujen synty. Geneettisten sairauksien vaikutukset saattavat olla rotujen välillä jaettu piirre, mikä johtuu haplotyyppien rakenteesta ja niiden kyvystä laajentua pitkiä etäisyyksiä. Nykyinen SNP:n kartoitus antaa tarpeeksi hyvän kuvan rotujen sisäisestä polymorfian määrästä ja mahdollistaa kiinnostavien ominaisuuksien kartoittamisen. Bokserin genomisekvenssiä voidaan käyttää koiraeläinten evoluutiopuun kartoittamiseen (Lindblad-Toh ym., 2005).

Koko genomi on kooltaan 2.41 gigaemäsparia (Gb), josta noin 2.38 Gb on nukleotidi sekvenssiä ja 1% on aukko kohtia. Genomin kokoonpano on erittäin yhtäjaksoinen. Genomin jatkumon keskiarvo on N50, joka määritetään lyhyimmistä jatkumosta, joka on 50 %:a genomisekvenssin kokonaisuudesta, on kooltaan noin 180 kiloemäsparia ja N50 kehukset (scaffold) ovat 45.0 megaemäsparia. Tämä tarkoittaa, että suurimmassa osassa genejä ei ole aukkoja sekvenssissä, minkä seurauksena suurin osa koiraeläinten kromosomeista on kokonaan järjestäytynyt yhteen tai kahteen kehukseen. Sekvenssistä noin 97 % on järjestäytynytromosomeissa. Noin 3 % genomin kokoonpanosta on suureksi osaksi toistoa mukaan lukien kahdeksan kehystä, jotka koostuvat satelliittisekvensseistä. Nukleotidien määrä ja genomin kattavuus kokoonpanossa ovat korkeita (Lindblad-Toh ym., 2005).

Suurimmat koiran geeniperheistä ovat histoni H2B-geeniperhe ja a-interferonit, jotka jakautuvat monofyleettisiin kladeihin. A-interferonit esiintyvät kuudella lajilla, joista jokaisella ne ovat monofyleettisiä. Nykykoira jakautuu useaan erilliseen rotuun, jotka eroavat toisistaan sairauksiensa, morfologiansa ja käyttäytymisensä pohjalta.

Muinaisella koirapopulaatiolla oli lyhyttä kytKentäepätasapainoa, ja nämä kytKentäepätasapainon alueet olivat lyhyempiä kuin ihmisellä havaitut. Rotujen synnyn seurauksena koiraroduilla on ilmentynyt useita pullonkauloja. Tämä näkyy muinaisten populaatioiden jättämistä haplotyyppien yhdistelmistä, joista kehittyivät nykyisten rotujen spesifit haplotyyppit. Kuitenkaan koiraroduilla ei ole tapahtunut täyttä fiksaatiota. Koirarodut jakavat keskenään noin 100 kiloemäsparin verran haplotyyppijä. Saman haplotyyppin havaitseminen kahden eri rodun kromosomeista on epätodennäköistä (Lindblad-Toh ym., 2005).

### 2.2.2. Koreanjindonkoira

Kim ym., 2012 tehdyssä genomien sekvensoinnissa käytettiin koreanjindonkoiraa genomien pohjana. Tutkimuksessa käytettiin uudenlaista sekvensointi tekniikkaa (Illumina HiSeq 2000) ensimmäistä kertaa koiran genomien kartoitukseen. Saadusta 45-kertaisesta sekvenssidatan sekvensointisyvyydestä tehtiin annotaatio koreanjindonkoiran mitokondriaalisesta genomista. Näin saatiin syvää ymmärrystä SNP-karttaan, bokseri-koiran referenssigenomien sekvenssin aukkojen alueisiin (gap region), koiran hajuaistin geeniperheeseen ja fylogeniaan.

Koreanjindonkoiran genomien vastasi noin 94.02 % Bokserin genomista. Sen lukusyvyyden ja GC-sisällön välinen yhteys koko genomien läpi osoittaa jindon genomien vastaavan melkein koko GC-sisällön pituutta ja varsinkin sen lukusyvyyden kattavuus on 15 - 47 kertaisena genomien alueilla, joissa GC-sisältö on 11 - 77 % välillä. Koreanjindonkoiran genomista havaittiin 1 976 029 homozygootista ja 2 699 408 heterozygootista SNP:ia, joista bokseriin verrattuna koreanjindonkoiralla on 3 346 058 uutta ja 1 329 379 jaettua SNP:ta. 1065 jindon SNP:ta on geenien sisäisiä (Kim ym., 2012).

Koirilla on tunnetusti ihmistä vahvempi hajuaisti, joka koreanjindonkoiralla on varsin voimakas. Ihmisellä on havaittu ~900 ja koiralla 1094 hajuaisti geeniä, jotka ihmisellä ovat suureksi osin pseudogenejä. Koreanjindonkoiralla havaittiin 2299 SNP:ta hajuaistigeeneissä, joista 773 on homozygootisia ja 1526 on heterozygootisia SNP:ta. Bokserin genomissa esiintyy samoja heterozygootisia alleleja kuin koreanjindonkoiralla, mutta ilman homozygootisia SNP:ta. Kim ym. kohdistivat tutkimuksen koreanjindonkoiran hajuaistigeenien homozygootisiin SNP:n, jotka esiintyvät koreanjindonkoiralla, mutta puuttuvat bokseringenomista. Näistä hajuaistigeenien SNP:sta 236 on ei-synonyymisiä ja 137 on synonyymisiä. Ei-synonyymisten homotsygoottisten SNP:n olemassaolo koreanjindonkoiralla, mutta niiden puuttuminen bokserin hajuaistireseptorigeeneistä viittaa siihen, että koreanjindonkoiran hajuaistikyky saattaa merkittävästi erota bokserin hajuaistikyvystä (Kim ym., 2012).

Koirarotujen erottelu funktionaalsiin ja fenotyypisiin ryhmiin ei kaikkien roturyhmittelyjen kannalta toimi. Jotkin koirarodut ovat heterogeenisiä ja vastaavasti toiset ovat vähemmän heterogeenisiä. Nämä koirarotujen välisten suhteiden mallit viittaavat siihen, että koiran domestikaation diversiteetin mekanismi on selkeä. Eritoten

yksittäisten mutaatioiden fiksaatio ja ilmentyminen johtavat uusiin rotufenotyyppeihin, jotka risteytyvät toisiin roturyhmiin. Tämän seurauksena uudet mutaatiot siirtyvät eteenpäin ja tehostavat koirarotujen diversiteettiä kasvattaen koiran morfologiaa entisestään (Wayne & von-Holdt, 2012).

## 3. Suden genomiikka

### 3.1. Suden genomi

Nykyisessä tutkimuksessa on puute yleisistä menetelmistä ja geenimerkeistä, jotka ovat universaalisti verrattavissa tutkimusten välillä. Susipopulaatioiden analyysille genomiikan nopea kehitys vaikuttaa lupaavalta, mutta kuitenkin ei ole selvää, mitkä menetelmät ovat datan laadun ja edullisuuden kannalta sopivia (Hindrikson ym., 2016). Suuressa osassa susien ja koirien väliseen suhteeseen liittyvissä genomiikan tutkimuksissa on käytetty joko Lindblah-Toh ym., 2005 rakentamaa bokserin genomia tai tuloksia on verrattu SNP matriiseilla koiran genomien variaatioon. On mahdollista, että näissä tutkimuksissa on saattanut jäädä huomaamatta tai on kadotettu tärkeitä suden genomien variaation rakenteita, kun sitä on kartoitettu koiran referenssigenomiin tai kohdistettu koiran variaation SNP:llä. Gopalakrishan ym. 2017 koostivat ensimmäisen suden referenssigenomin, minkä avulla on mahdollista tutkia suden ja muiden koira-eläinten suhdetta ja genomiikkaa.

Suden genomi on kooltaan 2.34 Gb ja se koostuu 8747 kehikosta, joista 8569 on suurempia kuin 1 Kb. Genomin pisin kehikko on 12.66 Mb, N50-kehys on 1.56 Mb ja N80 on 512 Kb. Vastaavasti genomien N50 jatkumo on 94 Kb ja N80 on vastaavasti 34 Kb. Verratessaan toistuvien elementtien runsasta määrää suden genomien ja bokseri-koiran välillä Gopalakrishan ym. 2017 havaitsivat 902 Mb toistuvia elementtejä sudella, mikä vastaa 39.8 % kokoonpanon aukottomasta (non-gapped) sekvenssistä, kun vastaavasti bokserin genomista havaittiin 42.1 %. Kun toistuvat elementit jaetaan niiden superperheisiin, suden ja koiran kokoonpanoissa esiintyy samankaltaista runsautta toistuvissa elementeissä, poissulkien satelliittisekvenssit (Gopalakrishnan ym., 2017).

Hindrikson ym., 2016 mukaan aineiston kerääminen ei ole aina riittävä, kun otetaan huomioon otoskoot ja maantieteellinen peitto ja varsinkin, kun otetaan huomioon se, että analyysit kohdistuvat kuolleisiin yksilöihin, jotka ovat kuolemansa seurauksena

eronneet populaatiosta. Idealisessa tilanteessa tutkimuksissa saataisiin susipopulaation nykytilanteen parhaiten kuvaava katsaus, jossa mahdollisesti olisi mukana populaation yksilöidensukupuu. Tämän lisäksi populaatiota seurattaisiin kauemmin siinä tapahtuvan kehityksen ymmärtämiseksi. Äidinpuoleisen perimän analysointiin tarvitaan tulevaisuudessa kokonaisten mitokondriaalisten genomien sekvensointia, kun vastaavasti isänpuoleisen kytkennän analyysistä varten tarvitaan suuria määriä polymorfisia Y-kromosomin lokuksia, kuten SNP:ta ja mikrosatelliitteja. Autosomaalisten parentaalimarkkereiden analyysissä voidaan käyttää koko genomisekvensointia, SNP-siruja tai NGS-pohjaista mikrosatelliittien genotyypitystä. SNP-sirujen ja NGS-pohjaisten genotyyppien määrittäminen on ekonomisesti kannattavaa verrattuna koko genomisekvensoinnin tekniikoihin. Kuitenkin koko genomisekvensoinnista saadaan tarkempaa dataa ja se mahdollistaa autosomien, mitokondriaalisen genomien ja Y-kromosomin havainnoinnin (Hindrikson ym., 2016).

## 4. Koiran ja suden genomien vertailu

Pääasiallisesti suurin huolenaihe verrattaessa koiran ja suden referenssigenomeja toisiinsa on niiden laatu. Lindblad-Toh ym. 2005 bokserin genomi on tarkempi kattavuudessaan verrattaessa Gopalakrishnan ym. 2017 suden genomiin, minkä seurauksena niiden vertaaminen toisiinsa saattaa vaikuttaa tutkimuksen tulokseen.

Gopalakrishnan ym. 2017 tutkimuksessaan vertasivat suden ja koiran referenssigenomeja toisiinsa käyttäen harhan estona *Canis aureus* eli kultasakaalin genomia. Joissain tutkimuksissa kultasakaalin tapaisen ulkoryhmän genomien käyttäminen ei ole mahdollista kustannusten tai ulkoryhmän puuttumisen takia. Useimmiten referenssigenomit eivät ole yhtä läheisiä, kuin susi ja koira ovat, joten toisistaan etäisten näytteiden kohdalla on huomioitava mahdollinen harha tutkimuksen tuloksista. Koska susi ja koira ovat erittäin läheisiä, referenssigenomin käyttö vaikuttaa useisiin analyysihin, kuten PCA ja fylogenian tarkasteluun. Kuitenkin admixture-analyysistä havaittiin referenssigenomin vaikutus useassa rykelmässä. Tämän seurauksena suden referenssigenomin käyttö on suositeltavaa susien ja koirien suhteiden tarkastelussa (Gopalakrishnan ym., 2017).

## 4.1. Suden ja koiran introgressio

Randin ja Lucchinin, 2002 mukaan bayesilaiset admixture-analyysit mahdollistavat yksilöiden tunnistamisen sekoittuneesta syntyperästä, ja sitä voidaan käyttää arvioitaessa koiran geenien introgression laajuutta susipopulaatioon. Näiden metodien käytöllä voidaan kehittää strategioita susipopulaatioiden suojeluun. Hindrikson ym. 2016 mukaan Euroopassa ei ole tarpeeksi tutkimusta suden ja koiran hybridisaatiosta ja koiran geenien introgressiosta susipopulaatioihin. On tärkeää tunnistaa susipopulaatiot, joissa introgressiosta on uhkaa populaation eheyteen.

### 4.1.1. Tiibetin mastiffi ja susi

Tiibetin ylängöllä on jo kauan elänyt koiran villi sukulainen, Tiibetin susi, minkä seurauksena se saattaa olla Tiibetin mastiffin hypoksia-adaptaation geneettinen lähde. Miao ym., 2016 eivät havainneet merkittävää introgressiota genomien laajuiselta tasolta, mutta paikalliselta tasolta genomista havaittiin, että Tiibetin mastiffin hypoksia-adaptaatiolle tärkeät EPAS1 ja HBB lokukset saattavat tulla Tiibetin sudelta. Molemmat lokukset osoittivat, että Tiibetin mastiffi on ollut valintapyyhkäisy (selective sweep) alaisena verrattaessa alankokoiriin, mistä voidaan olettaa, että introgressiolla on adaptiivinen rooli. On mahdollista, että susien ja koirien sekoittumisen jälkeen ylänkökoirien jalostamisen myötä osa sudelta periytyneestä DNA:sta menetettiin ja ylängöllä elämää edistänyt osa DNA:sta pysyi koirissa.

On epätodennäköistä, että Tiibetin suden ja mastiffin sukujuuret ovat säilyttäneet haplotyyppi rakenteensa ilman rekombinaatiota polymorfian seurauksesta. Suden ja mastiffin lokuksien sekvenssiendivergenssin määrä on huomattavasti vähentynyt verrattaessa niiden genomiseen perimään. Miao ym. 2016 tutkimuksen tulokset osoittavat siihen, että Tiibetin mastiffin haplotyyppi rakenne on todennäköisemmin seurausta introgressiosta kuin muinaisesta polymorfista. Koiraeläinten genomeihin perustuvat haplotyyppien verkostot osoittavat, että ylänkökoirat perivät nämä lokukset ylänkösusilta (Miao ym., 2016).

### 4.1.2. Italian susi

Italian sudet hybridisoituivat, joko susipopulaation 1970-luvun pullonkaulan aikana tai sen jälkeen. On myös mahdollista, että susipopulaatio on sekoittunut Apenniinien

alueella viimeisen 40 vuoden aikana. Galaverni ym. 2017 määrittivät 118 suden, 31 kyläkoiran ja 72 hybridi-suden genotyypin. He käyttivät näiden lisäksi 456 julkisesti saatavilla olevaa koiran genotyyppiä ja niiden genomien laajuisia SNP:ta sekä hyödynsivät haplotyyppien lohkoista saatua informaatiota tunnistamaan sekoittumisen ajankohdan sekoittuneista genotyypeistä, jotka ovat vanhempia kuin sukupolvet, joissa on tapahtunut vasta muutama takaisinristeytys. Sen lisäksi Galaverni ym. käyttivät syntyperän rekonstruointimenetelmiä tunnistamaan susihybridien genomista satunnaisesta periytymisestä eroavia alueita. Suden ja koiran hybridisaatio saattaa näkyä fenotyypillisellä tasolla.

Galaverni ym. 2017 arvioivat poikkeavat SNP:t ketjuttamalla toisiinsa liittyvät SNP:t ja rekonstruoimalla niiden haplotyyppien lohkot. 90 mustan turkin väriin vaikuttavat SNP:t osoittivat, että mustan turkin väri perustuu yhteen haplotyyppiin villisusissa, kun taas kuusi eri haplotyyppiä ovat peräisin koirasta. 41 merkittävässä lohossa kaikilla mustaturkkisilla susilla oli yksi koirasta lähtöisin oleva haplotyyppi.

Galaverni ym. 2017 Italian susipopulaation hybridisaation genomien laajuisen arvion tuloksista voidaan havaita, että noin 8 % suden ja koiran hybridien genomissa on enemmän koiralta perittyjä geenejä kuin suden geenejä, mikä todennäköisemmin on seurausta 1.-3. sukupolvea sitten tapahtuneesta lajien sekoittumisesta. Suden ja koiran populaatioissa on ollut toistuvasti lajien sekoittumista, mutta niiden vaikutukset ovat ajan myötä heikentyneet takaisinristeytyksen seurauksesta ja odotettavasti susipopulaatioiden koiranperimä tulee katoamaan noin seitsemän takaisinristeytys sukupolven jälkeen (Galaverni ym., 2017).

Galaverni ym. 2017 tutkimuksen tuloksista voidaan havaita, että suden valkoiset kynnet ovat mahdollisesti hybridisaation seurausta tai ympäristöllisten muutosten seurausta. Muutamaa mustan turkinvärin omaavaa sutta ei voida havaita hybridiksi edes genomien laajuisella tutkinnalla. Tämän seurauksena suden hybridisaatio tapahtui kauan aikaa sitten ja suurin osa koiran alleleista on kadonnut johtaen melkein puhtaisiin susigenomeihin.

## 4.2. Suden ja koiran populaatiogenetiikka

Gopalakrishan ym. 2017 mukaan käytettäessä Lindblad-Toh ym. 2005 sekvensoitua koiran referenssigenomia sudet ja koirat ovat monofyleettisiä. Vastaavasti suden

referenssigenomia käytettäessä koirat ovat monofyleettisiä susien suhteen, mutta sudet eivät ole monofyleettisiä koiran suhteen. Käytettäessä kahta syntyperätekijää (ancestry component) koirien ja suden referenssigenomeista riippumatta koirat ja sudet jakautuvat kahteen ryhmittymään, joissa molemmissa koirassa on sudenperimää 20 %.

Arvioitaessa sekoittumista neljällä syntyperätekijäryhmällä referenssigenomin valinnalla on vaikutusta. Valittaessa suden genomi havaitaan, että kultasakaalin näyte erottuu muista näytteistä, kun taas valittaessa bokserin genomi sudesta havaitaan uusia rakenteita ja kultasakaali on yhdistettynä yhteen suden ryhmistä. Koiran referenssigenomilla suuremmalla määrällä syntyperätekijöitä voidaan selittää koirien varianssia ja vastaavasti suden referenssigenomista saadaan esiin uusia rakenteita (Gopalakrishnan ym., 2017).

Gray ym. 2009 vertasivat suurten koiraeläinten populaatioiden kytkentäepätasapainon (LD) kantamaa koirarotujen estimaatteihin, LD:n suhdetta ja demograafista historiaa vertaamalla LD:n estimaatteja tunnettuihin populaatiohistorioihin. Demograafisesta mallinnuksesta saatiin evidenssiä kohtuullisesta populaation kutistumisesta noin 15 000 vuotta sitten ja suuresta kutistumisesta rotujen synnyn aikaan. Rotujen synnyn kutistuma oli voimakkaampi kuin domestikaation aiheuttama. Nukleotididiversiteetti estimaattien perusteella domestikaatiosta seurasi 5 % vähentymä diversiteettiin, kun vastaavasti rotujensynnystä seurasi 35 % vähentymä diversiteetissä. Tämä on seurausta rodun jatkuvasta sisäsiitaisuudesta, vaikka populaatiokoko on kasvanut rotujen synnystä.

Lisäksi villien koiraeläinpopulaatioiden ja koirarotujen alueellisten frekvenssien spektrit havaittiin yhdenmukaisiksi LD:n estimaattien ja tunnetun populaation kanssa. Villeissä populaatioissa populaatioiden koot laskivat Espanjan ja Israelin susipopulaatioissa, kuten historiasta voi olettaa. Kuitenkaan kojootin, Alaskan ja Yellowstonen susipopulaatioissa ei tapahtunut muutosta. Koiraroduista kiinanpalatsikoira ja berninpaimenkoira olivat kärsineet suurimmasta populaation kutistumasta. Vähäisempiä kutistumia vastaavasti havaittiin kultaisista- ja labradorinnoutajista. LD:n ulottuvuus, demograafinen mallinnus ja tunnettu demograafinen historia tukevat LD:n käyttöä populaatiohistorian päättelyssä sekä malli- että villipopulaatioiden kohdalla (Gray ym., 2009).

Grey ym. 2009 estimoivat 11 luonnollisen harmaasusipopulaation, yhden kojoottipopulaation ja 18 koirarodun LD:n. Populaatiohistorian ja LD:n välillä on kausaalinen suhde, jonka avulla voidaan päätellä demograafisia ja evolutiivisia tapahtumia villeissä ja kesyissä koiraeläimissä. Koirilla on esiintynyt populaatioiden kutistumisia domestikaation seurauksena, mutta kuitenkin geneettinen variaatio on säilynyt. Seurauksena koirien variaatio on jakautunut koirarotuihin, joissa on korkeaa ja vaihtelevaa määrää LD:ta. Susipopulaatioissa LD:n korkea taso viittaa mahdolliseen luonnollisten populaatioiden piirteiden kartoitukseen.

## 5. Pohdinta ja johtopäätökset

Tämän tutkielman tarkoituksena oli käsitellä aikaisempaan tutkimukseen pohjautuen miten suden ja koiran genomit eroavat toisistaan, miten referenssigenomin valinta vaikuttaa tutkimustuloksiin ja miten suden ja koiran lähisukulaisuus vaikuttaa kyseisten lajien geneettiseen diversiteettiin.

Koiran eroaminen sudesta on ollut suuren kiistelyn alaisena, mikä on johtunut siitä, että suuressa osassa tutkimuksia suden ja koiran jäänteet ovat olleet yllättävän samakaltaisia ja molemmat lajit ovat läheisyytensä seurauksesta saattaneet hybridisoitua keskenään. Tutkimuksissa kuitenkin on auttanut se, että koiran arkeologinen näkyvyys on suurta ihmisen ja koiran maantieteellisen läheisyyden seurauksena. Kuitenkin suuressa osassa tutkimuksia koiran domestikaation alku saadaan arvioitua 30 000 – 33 000 vuosien välille.

Koiran genomisen data on koostettu useaan kertaan ja verrattu Lindblad-Toh ym. 2005 koostettuun bokserin referenssigenomiin. Samalla ollaan saatu kartoitettua koiran domestikaation seurauksena kehittyneille roduille omalaatuisten tautien historiaa, morfologisia eroja ja koiran käyttäytymisen genetiikkaa. Kim ym. 2012 tutkimus antoi esimerkin rotujen eroavaisuudesta verrattaessa koreanjindonkoiran ja bokserin hajuaihia keskenään. Vuoden 2017 kesällä Gopalakrishan ym. sekvensoima suden referenssigenomi antaa vastaisuudessa mahdollisuuden suden ja koiran eroavaisuuksien hahmottamiseen, mikä voi siten edistää ymmärrystä domestikaatiosta, koiran syntyperästä sekä mahdollistaa suden ja muiden koiraeläinten vertailun.

Suden ja koiran eroavaisuudet referenssigenomin valinnalla riippuvat taustatekijöiden määrästä, esimerkiksi suuret määrät antavat selviä tuloksia, mutta tulokset ovat

vastaavasti huomaamattomia pienellä määrällä. Referenssigenomien käyttäminen tutkimuksissa antaa uusia näkökulmia tutkimukseen ja mahdollistaa siten koiran alkuperän tarkemman tutkimisen.

## Lähteet

1. Freedman, A. H. & Wayne, R. K. (2017). Deciphering the origin of dogs: From fossils to genomes. *Annual Review of Animal Biosciences*, 5, 281-307.
2. Galaverni, M., Caniglia, R., Pagani, L., Fabbri, E., Boattini, A. & Randi, E. (2017). Disentangling timing of admixture, patterns of introgression, and phenotypic indicators in a hybridizing wolf population. *Molecular Biology and Evolution*, 34(9), 2324-2339. doi:10.1093/molbev/msx169
3. Germonpré, M., Sablin, M. V., Stevens, R. E., Hedges, R. E. M., Hofreiter, M. et al. Després, V. R. (2009). Fossil dogs and wolves from palaeolithic sites in belgium, the ukraine and russia: Osteometry, ancient DNA and stable isotopes. *Journal of Archaeological Science*, 36(2), 473-490. doi:10.1016/j.jas.2008.09.033
4. Gopalakrishnan, S., Castruita, J. A. S., Sinding, M. S., Kuderna, L. F., Räikkönen, J. et al. Marques-Bonet, T. (2017). The wolf reference genome sequence (*canis lupus lupus*) and its implications for *canis* spp. population genomics. *BMC Genomics*, 18(1), 495.
5. Gray, M. M., Granka, J. M., Bustamante, C. D., Sutter, N. B., Boyko, A. R. et al. Wayne, R. K. (2009). Linkage disequilibrium and demographic history of wild and domestic canids. *Genetics*, 181(4), 1493-1505. doi:10.1534/genetics.108.098830
6. Hindrikson, M., Remm, J., Pilot, M., Godinho, R., Stronen, A. V. et al. Saarma, U. (2017). Wolf population genetics in Europe: A systematic review, meta-analysis and suggestions for conservation and management. *Biological Reviews*, 92(3), 1601-1629. doi:10.1111/brv.12298
7. Kim, R. N., Kim, D. -, Choi, S. -, Yoon, B. -, Kang, A. et al. Park, H. -. (2012). Genome analysis of the domestic dog (korean jindo) by massively parallel sequencing. *DNA Research*, 19(3), 275-287. doi:10.1093/dnares/dss011

8. Larson, G., Karlsson, E. K., Perri, A., Webster, M. T., Ho, S. Y. et al. Lindblad-Toh, K. (2012). Rethinking dog domestication by integrating genetics, archeology, and bio-geography. *Proceedings of the National Academy of Sciences of the United States of America*, 109(23), 8878-8883. doi:10.1073/pnas.1203005109 [doi]
9. Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B. et al. Zody, M. C. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, 438(7069), 803.
10. Miao, B., Wang, Z. & Li, Y. (2017). Genomic analysis reveals hypoxia adaptation in the tibetan mastiff by introgression of the gray wolf from the tibetan plateau. *Molecular Biology and Evolution*, 34(3), 734-743. doi:10.1093/molbev/msw274
11. Randi, E. & Lucchini, V. (2002). Detecting rare introgression of domestic dog genes into wild wolf (*canis lupus*) populations by bayesian admixture analyses of microsatellite variation. *Conservation Genetics*, 3(1), 29-43.
12. Serres-Armero, A., Povolotskaya, I. S., Quilez, J., Ramirez, O., Santpere, G. et al. Marques-Bonet, T. (2017). Similar genomic proportions of copy number variation within gray wolves and modern dog breeds inferred from whole genome sequencing. *BMC Genomics*, 18(1) doi:10.1186/s12864-017-4318-x
13. Skoglund, P., Ersmark, E., Palkopoulou, E. & Dalén, L. (2015). Ancient wolf genome reveals an early divergence of domestic dog ancestors and admixture into high-latitude breeds. *Current Biology*, 25(11), 1515-1519. doi:10.1016/j.cub.2015.04.019
14. Vilà, C., Walker, C., Sundqvist, A., Flagstad, Ø., Andersone, Z. et al. Ellegren, H. (2003). Combined use of maternal, paternal and bi-parental genetic markers for the identification of wolf–dog hybrids. *Heredity*, 90(1), 17.
15. Wayne, R. K. & Vonholdt, B. M. (2012). Evolutionary genomics of dog domestication. *Mammalian Genome*, 23(1-2), 3-18. doi:10.1007/s00335-011-9386-7