

Simuloitua geenihistoriaa

Tarja Sundell ja Päivi Onkamo

Väestöjen alkuperää, muuttoreittejä sekä sukulaisuussuhteita on jo pitkään selvitetty erilaisin geneettisin menetelmin. Päätelmiä on tehty niin tutkimalla nykyväestön DNA:ta kuin aivan viime vuosina, myös muinais-DNA:ta.

Muinais-DNA

Suomen happamassa maaperässä orgaaninen aines säilyy huonosti, joten meidän mahdollisuutemme pystyä saamaan muinais-DNA:ta esihistoriallisista luista tai hampaista ovat rajallisia. Tiedyt ympäristön olosuhteet, kuten materiaalin nopea kuivuminen post mortem, alhainen lämpötila ja korkea suolapitoisuus mahdollistavat DNA:n paremman säilymisen jälkipolville tutkittavaksi. Arkeologisen materiaalin DNA-tutkimusta tehdään Suomessa Helsingin yliopiston Hjelt-instituutin muinais-DNA-laboratoriossa, Oulun yliopistossa sekä MTK:ssa Jokioisten tutkimuskeskuksessa.

Nykyväestön DNA

Modernien DNA-tekniikoiden kehittyminen on mahdollistanut populaatiogeneettisten tutkimusten räjähdysmäisen kasvun ja tuonut valtavasti uutta tietoa niin eläin- kuin ihmispopulaatioiden geneettisestä historiasta. Sen sijaan, että päätelmät perustuisivat yksittäisten veriryhmien tai geenien yleisyyksien vertailuihin väestöjen välillä, tehdään genotyyppaukset 2000-luvulla koko genomin laajuusina, sadoilla tuhansilla geenimerkeillä

kerrallaan (esim. Jakkula ym. 2008). Näitä menetelmiä on hyödynnetty myös suomalaisen geeniperimän tutkimisessa (Lappalainen ym. 2006 ja 2008; Hedman ym. 2007; Salme-la ym. 2008; Palo ym. 2009).

Suomalaisten väestöhistoriaa ja genejä tutkitaan myös monitieteisessä hankkeessa Argeopop. Finnish population history – an archaeological and population genetic approach. Argeopop-projektille on myönnetty Helsingin yliopiston kolmevuotinen tutkimusmääräraha vuosille 2008–2010 sekä Suomen Akatemian Biotieteiden ja ympäristön tutkimuksen toimikunnan myöntämä rahoitus vuosille 2010–2013. (Argeopopin kotisivu: <http://www.helsinki.fi/bioscience/argeopop/>)

Geneettiset simulaatiot

Erilaisten väestöhistoriaskenaarioiden vaikutuksia nykyisen väestön geenipooliin voidaan tutkia tietokoneavusteisesti, simuloimalla (Sundell 2008 ja 2009; Sundell & Onkamo 2010). Yksilöinä ovat kromosomit, joita koskevat kaikki samat evolutiiviset lainalaisuudet kuin mitkä perimäämme oikeastikin koskevat. Tavoitteena on arvioida tuottaako simulointi oletetulla demografisella mallilla lainkaan samankaltaista geneettisen variaation määrää ja rakennetta kuin mitä todellisuudessa nähdään.

Simulaatioita voidaan tehdä sekä ajassa eteenpäin että taaksepäin. Eteenpäin simuloitaessa määritellään ensin perustajaväestö, alipopulaatiot, syntyvyys ja kuolleisuus.

Malliin on mahdollista lisätä myös esim. alipopulaatioiden välisiä migraatiotodennäköisyyksiä tai vaikkapa populaatiopullonkauloja (Sundell ym. 2010; Heger 2011; Kammonen valmisteilla). Taaksepäin simulointia kutsutaan koalesenssimalliksi. Tässä mallissa pitää ottaa huomioon vain ne haplotyyppit, jotka ovat mukana otoksessa, eikä kaikkia mahdollisia populaatioissa esiintyviä haplotyyppijä (Jobling ym. 2004; Garrigan ja Hammer 2006). Koalesenssimallilla haetaan sitä rakennetta, millainen lähtötilanteen muinaisväestössä on täytynyt olla, jotta se tuottaisi nykyväestön geneettisen rakenteen. Sekä eteenpäin että taaksepäin simuloinnissa ajo toistetaan 1000–10000 kertaa, jotta sattuman vaikutus lopputuloksiin pystytään arvioimaan.

SimuPOP

Simulaatiot tehdään simuPOP-nimisellä forward-time -populaatiosimulaattorilla (Peng ja Kimmel 2005), joka on kokoelma erillisiä ohjelmamoduuleja. Simulaattoria käytetään kirjoittamalla oma Python-kielinen ohjelma, joka ajaa tarvittavia moduuleja tarkoituksen mukaisessa järjestyksessä. SimuPOPin tuottamia output-tiedostoja voidaan sitten muokata ja syöttää seuraavassa vaiheessa käytettäville ohjelmille, kuten esim. visualisointityökaluille.

Populaation simulointi

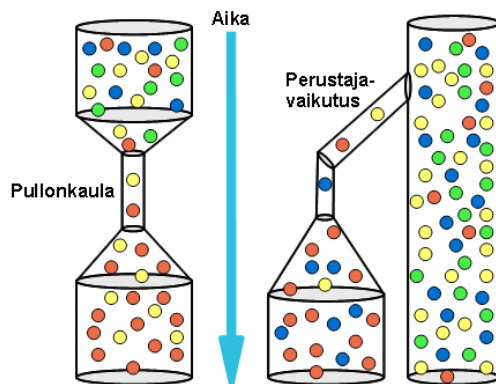
Simulointi aloitetaan luomalla lähtöpopulaatio, joka voidaan jakaa alipopulaatioihin haluttuina ajankohtina. Simulaatioon sisällytetään kaikki luonnolliselle populaatiolle ominaiset demografiset ominaisuudet, kuten väestönkasvu ja -väheneminen, migraatiot alipopulaatioiden välillä, yksilöiden eliniät, vaihteleva lisääntymismenestys ja muuttoreittien seuranta.

Tutkimushypoteesit

Tähän mennessä olemme keskittyneet kahteen tutkimushypoteesiin. Olemme simuloineet Suomen esihistoriallisia populaatiopullonkauloja sekä 1500-luvun savolaiskspansiota. Populaatiopullonkaulaksi (population bottleneck) kutsutaan tapahtumaa, jonka seurauksena huomattava osa alkuperäisestä populaatiosta on estynyt lisääntymästä ja populaatio pienenee huomattavasti. Pullonkaula voi olla hyvin nopea tapahtuma (nälänhätä, epidemia, sota, jne.) tai kestää kauan, useiden sukupolvien ajan. Sen seurauksena vain pienellä osalla alkuperäisestä populaatiosta on geneettistä kontribuutiota tuleville sukupolville. Perustajavaikutus (founder effect) taas kuvaa sitä, miten alkuperäisestä, suuremmasta populaatiosta irronneeseen uuteen pieneen populaatioon (esim. siirtokunta) on sattumalta valikoituneena pelkästään tietyt geenivariantit, eikä näiden kahden populaation välillä enää tapahdu merkittävää geenivirtaa. Molempien ilmiöiden geneettiset seuraukset ovat hyvin samantapaiset. (Kuva 1.)

Tutkimushypoteesi I: esihistorialliset pullonkaulat

Onko suomalainen väestö käynyt läpi kapean pullonkaulan joskus esihistoriansa aikana? Pullonkaula selittäisi esim. suomalaista tautiperintöä sekä Y-kromosomin alentunutta



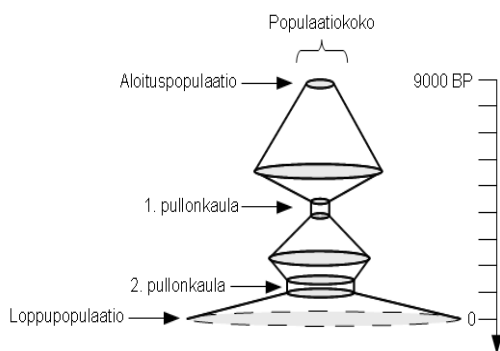
Kuva 1. Pullonkaula ja perustajavaikutus.

variaatiota. Suomalaiseen tautiperintöön katsotaan kuuluvaksi 36 perinnöllistä sairautta, jotka ovat yleisempiä Suomessa kuin muualla maailmassa. Taudit ovat monogeenisiä eli kukin johtuu mutaatiosta yhdessä geenissä. Ne periytyvät autosomaalisesti peittyvästi (32 kpl), vallitsevasti (2 kpl) tai x-kromosomaalisesti (2 kpl).

Tutkimuksessamme (Sundell ym. 2010) olemme aloittaneet pullonkaulasimulaatiot vuodesta 9000 BP. Lähtöpopulaation koko on kaikissa 1000 yksilöä. Erilaisia simulaatioita on tehty yhteensä 15 kappaletta (mallit A–O), joissa pullonkaulavaiheen väestömäärä on joko 5000, 1000 tai 200. Simulaatioihin on myös lisätty kaksi suurempaa migraatioaaltoa, tyypillisen kampakeramiikan ja nuorakeramiikan ilmestymisen aikoihin, sekä erikokoisia muuttotihkuja niin arkaaisesta eurooppalaisesta, arkaaisesta skandinaavisesta kuin saamelaisesta populaatiosta. (Kuva 2.)

Tutkimushypoteesi II: Early vs. Late Settlement

Vuonna 1542 Kustaa Vaasa julisti autiot erämaat vapaasti asutettaviksi kruunun maiksi (kuva 3). Erämaihin saatiin uudisasukkaita lupaamalla vapaavuosia verotuksesta sekä laajat kaski- ja riistamaat. Muuton myötä eräelinkeinot vaihtuivat osaksi kaskiviljelyyn



Kuva 2. Havainnekuva: pullonkaula. Kartion leveys esittää populaation suhteellisen koon kunakin aikakautena.

ja uusien kaskimaiden etsiminen sai aikaan mittavan muuttoliikkeen, jota kutsutaan savolaisekspansioiksi. Kruunun asutuspolitiikkaan liittyi pyrkimys lisätä verotalojen määrää ja samalla se havaittiin hyväksi keinoksi liittää uusia erämaita Ruotsi–Suomen valtakuntaan. (Pirinen 1982; Lehtosalohilander 1988; Vahtola 2003). Savolaisekspansio aiheutti perustajavaikutuksen ja on ainakin osittain vastuussa maantieteellisistä eroista geneettisessä monimuotoisuudessa. Selkeimmät erot ovat nähtävissä ennen savolaisekspansiota ja sen seurauksena asutettujen alueiden välillä. Näistä käytetään erityisesti geneetikkojen piirissä termejä ”Early Settlement Area” (ESA) ja ”Late Settlement Area” (LSA). Merkittävä osa aikaisemmasta geneettisestä tutkimuksesta onkin lähtenyt oletuksesta, että koko Itä-Suomi on asutettu vasta 1500-luvulta alkaen.

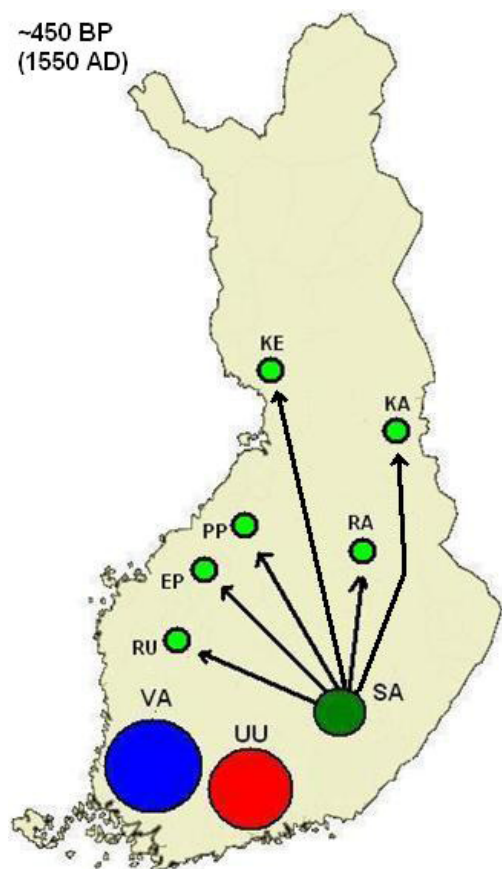
ESA-LSA simulaatiot aloitetaan vuodesta 900 AD kolmella alipopulaatiolla, Uusimaa ja Häme, Varsinais-Suomi sekä Savo. Myöhemmin, 1540-luvulla Savosta lähtee migranteja



Kuva 3. Kustaa Vaasa Jacob Binksin v. 1542 maalaamassa muotokuvassa. (Kuva: Wikipedia Commons.)

kuuteen uuteen alipopulaatioon, joiden sa-
vujen eli verolla olevien talojen lukumäärät
tunnetaan historiallisista lähteistä. Mallissa
simuloidaan mitokondrio- ja Y-STR-haplo-
tyyppejä aina 1950-luvulle saakka. Kaikkien
alipopulaatioiden välillä tapahtuu migraatiota
sukupuoli- ja alipopulaatiospesifisesti. On
todennäköisempää, että alipopulaatioon koh-
distuu migraatiota naapurialipopulaatiosta
kuin kauempaa ja tiedetään että kautta histo-
rian naiset ovat muuttaneet todennäköisem-
min kuin miehet (patrilokaalisuus). Kaiken
kaikkiaan simulaatioissa on pyritty historian-
kirjoituksen mukaiseen mahdollisimman rea-
listiseen demografiaan. (Kuva 4.)

Sukupolvenvälinä on käytetty 32 vuotta
(Tremblay ja Vézina 2000). Intuitiivisesti
tämä kuulostaa paljolta, mutta Tremblayn
mukaan useissa populaatiogeneettisissä tut-
kimuksissa käytetty 20–25 vuoden sukupol-
venväli on liian alhainen. Tremblay onkin



päätynyt 32 vuoteen vertailemalla etnogra-
fista aineistoa mm. Kanadassa olevaa Sa-
guenayn uudisasutusta 1830-luvulla. Uudis-
asuttajien perheissä ensimmäinen lapsi syntyi
äidin ollessa keskimäärin 22 ja isän 24 vuo-
den iässä. Viimeisin lapsi syntyi äidin ollessa
keskimäärin 39- ja isän 44-vuotias. Useim-
mat naiset pysyivät hedelmällisinä lähes
50-vuotiaiksi.

Simulaatiolähestymistavan rajoituk- sista

Lähestymistavan voidaan sanoa sisältävän
rakenteellisen heikkouden. Vaikka tietyl-
lä mallilla saadaan tuotettua todellisuutta
vastaava tulos, se ei takaa sitä, että mennei-
syyden väestö olisi todellisuudessa käyttäy-
tynyt juuri niin. Simulaatioiden tarkoitus ei
olekaan todistaa että yksi tietty väestöhista-
riamalli kuvaisi todellisuutta täydellisesti,
vaan niiden avulla on tarkoitus pystyä sulke-
maan pois vähiten todennäköiset vaihtoehdot.
Tähän mennessä saamiemme tulosten per-
ustella menneisyydessä tapahtuneiden pul-
lonkaulojen vaikutukset olisivat havaittavissa
nykyväestössä. Toisaalta jatkuva geenivirta
voi peittää alleen sekä suurien migraatioaal-
tojen että pullonkaulojen vaikutukset.

Kuva 4. Simulaation vaiheet. Simulaatio alkaa
kolmella alipopulaatiolla: Uusimaa ja Häme
(UU), Varsinais-Suomi (VA) sekä Savo (SA).
Savosta erkanee 1500-luvulla kuusi uutta
alipopulaatiota: Ruovesi (RU), Etelä-Pohjanmaa
(EP), Pohjois-Pohjanmaa (PP), Kemi (KE),
Rautalampi (RA) sekä Kainuu (KA).

Kirjallisuus

- Garrigan, D. ja Hammer, M.F. 2006: Reconstructing human origins in the genomic era. *Nature Reviews Genetics* 7, 669–680.
- Hedman, M., Brandstätter, A., Pimenoff, V., Sistonen, P., Palo, J.U., Parson, W. ja Sajantila, A. 2007: Finnish mitochondrial DNA HVS-I and HVS-II population data. *Forensic Science International* 172, 171–178.
- Heger, M. 2011: Population Genetic Simulations of the Savonian Expansion in Finland. Pro gradu -tutkielma. Helsingin Yliopisto, Biotieteiden laitos.
- Jakkula, E., Rehnström, K., Varilo, T., Pietiläinen, O.P.H., Paunio, T., Pedersen, N. L., deFaire, U., Järvelin, M.-R., Saharinen, J., Freimer, N., Ripatti, S., Purcell, S., Collins, A., Daly, M.J., Palotie, A. ja Peltonen, L. 2008: The Genome-wide patterns of variation expose significant substructure in a founder population. *American Journal of Human Genetics* 83, 787–794.
- Jobling, M.A., Hurles, M.E. ja Tyler-Smith, C. 2004: *Human Evolutionary Genetics. Origins, Peoples & Disease*. New York.
- Kammonen, J. valmisteilla: Making Inferences of Finnish Population History by Computer Simulations Applying Phylogenetic Methods and Pedigree Analysis. Pro gradu -tutkielma. Helsingin Yliopisto, Tietojenkäsittelytieteen laitos.
- Lappalainen, T., Koivumäki, S., Salmela, E., Huoponen, K., Sistonen, P., Savontaus, M.-L. ja Lahermo, P. 2006: Regional differences among Finns: A Y-chromosomal perspective. *Gene* 376: 207–215.
- Lappalainen, T., Laitinen, V., Salmela, E., Andersen, P., Huoponen, K., Savontaus, M.-L. ja Lahermo, P. 2008: Migration Waves to the Baltic Sea Region. *Annals of Human Genetics* 72:3, 337–348.
- Lehtosalo-Hilander, P.-L. 1988: *Savon Historia I*. Kuopio.
- Palo, J.U., Ulmanen, I., Lukka, M., Ellonen, P. ja Sajantila, A. 2009: Genetic markers and population history: Finland revisited. *European Journal of Human Genetics* 17, 1336–1346.
- Peng B ja Kimmel M. 2005: SimuPOP: a forward-time population genetics simulation environment. *Bioinformatics* 21, 3686–3687.
- Pirinen, K. 1982: *Savon historia II:1. Rajamaakunta asutusliikkeen aikakautena 1534–1617*. Pieksämäki.
- Salmela, E., Lappalainen, T., Fransson, I., Andersen, P.M., Dahlman-Wright, K., Fiebig, A., Sistonen, P., Savontaus, M.-L., Schreiber, S., Kere, J. ja Lahermo, P. 2008: Genome-Wide Analysis of Single Nucleotide Polymorphisms Uncovers Population Structure in Northern Europe. *PLoS ONE* 3:10, e3519. doi:10.1371/journal.pone.0003519.
- Sundell, T. 2008: Populaatiopullonkaulat kivikauden Suomessa – arkeologinen ja geneettinen tulkinta. Pro gradu -tutkielma. Helsingin Yliopisto, Filosofian, historian kulttuurin ja taiteiden tutkimuksen laitos, arkeologia.
- Sundell, T. 2009: Kadonneet geenit: Suomen kivikautisten populaatiopullonkaulojen tutkiminen geneettisin menetelmin. *Muinaistutkija* 4, 10–18.
- Sundell, T. ja Onkamo, P. 2010: Argeopop: keramiikasta geenivirtoihin? *Muinaistutkija* 1, 3–8.
- Sundell, T., Heger, M., Kammonen, J. ja Onkamo, P. 2010: Modelling a Neolithic Population Bottleneck in Finland: A Genetic Simulation. *Fennoscandia Archaeologica* XXVII, 3–19.

Tremblay, M. ja Vézina, H. 2000: New Estimates of Intergenerational Time Intervals for the Calculation of Age and Origins of Mutations. *American Society of Human Genetics* 66:2, 651–658.

Vahtola, J. 2003: Suomen historia – jääkaudesta Euroopan unioniin. Keuruu.

*FM Tarja Sundell on geneettisiin menetelmiin erikoistuva arkeologi. Hän tekee väitöskirjaa aiheesta Finnish population prehistory – an archaeological and genetic approach.
tarja.sundell@helsinki.fi*

*Dos. Päivi Onkamo työskentelee geneettisen bioinformatiikan yliopistonlehtorina Helsingin yliopiston biotieteiden laitoksella, perinnöllisyystieteen oppiaineessa. Hän myös vetää Argeopop-projektia.
paivi.onkamo@helsinki.fi*