

# Suomalaisväestöjen asutushistoriallinen rakenne

Jaakko Häkkinen, 15.7.2011

Edellisessä kirjoituksessani kaivoin genomilaajuisista tutkimuksista esiin seitsemän suomalaista väestöä (<http://www.mv.helsinki.fi/home/jphakkin/Suomalaisv%C3%A4est%C3%B6t.pdf>), ja nyt yritän katsoa mitä suomalaisten asutushistoriasta voidaan saada selville, eli kuinka on päädytty tilanteeseen, jossa nykysuomalaiset muodostavat seitsemän erillistä väestöä.

Olennainen kysymys on, missä määrin havaitut suomalaisväestöt ovat seurausta suomalaisen asutuksen leviämisestä ja missä määrin sekoittumisesta saamelaisiin (tai maan muihin alkupe räisasukkaisiin). Tähän kysymykseen yritän löytää vastauksia suomalaisia ja saamelaisia käsittelevistä genomilaajuisista tutkimuksista.

## 1. Suomalaisväestöjen ajautuminen

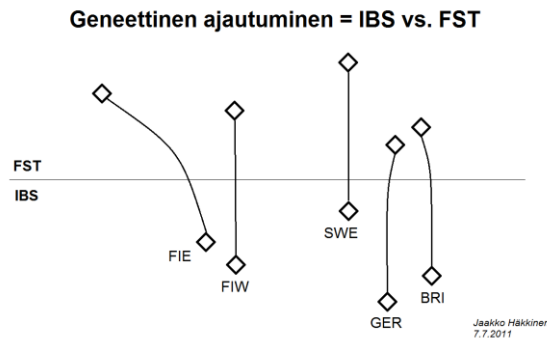
Geneettistä ajautumista (*genetic drift*) voidaan yrittää jäljittää vertaamalla väestöjen yhteisten geenien määrää (IBS = *identity by state*) niiden geneettiseen etäisyyteen (FST). Tässä ”ajautuminen” sisältää tarkemmin erittelemättä kaikenlaiset muutokset näiden kahden tason välillä: se mikä irrallisesti näyttää ajautumiselta, voi oikean verrokkiväestön löytyessä osoittautuakin sekoittumiseksi. Samankaltaisia näkemyksiä IBD:n (*identity by descent*) ja FST-etäisyyksien välisestä suhteesta on esitetty tutkimuksessa Atzmon et al. 2010. Testaan tätä menetelmää eurooppalaisten väestöjen avulla parilla esimerkillä; taulukko tutkimuksesta Salmela et al. 2008:

**Table 1.** Pairwise  $F_{ST}$ 's (lower diagonal) and the median IBS (upper diagonal) between population pairs.

	SWE	FIW	FIE	GER	BRI
SWE		0.7997	0.7990	0.7997	0.7997
FIW	0.0030		0.8005	0.7994	0.7993
FIE	0.0072	0.0032		0.7985	0.7982
GER	0.0021	0.0033	0.0084		0.8002
BRI	0.0024	0.0042	0.0094	0.0005	

Yläoikealla olevista pareittaisista IBS-arvoista nähdään, että eniten yhteisiä geenejä on itä- ja länsisuomalaisilla (0,8005), toiseksi eniten saksalaisilla ja briteillä (0,8002), kolmanneksi eniten ruotsalaisten ja länsisuomalaisten, saksalaisten sekä brittien välillä (0,7997) jne. Alavasemmalla olevien pareittaisten FST-arvojen perusteella kuitenkin lyhin etäisyys on saksalaisten ja englantilaisten välillä (5), toiseksi lyhin saksalaisten ja ruotsalaisten välillä (21), kolmanneksi lyhin ruotsalaisten ja

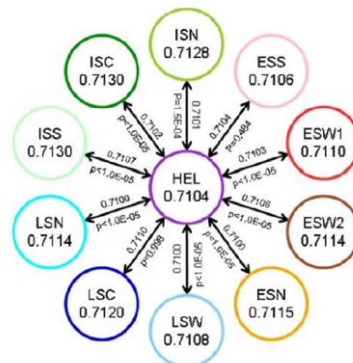
englantilaisten välillä (24), neljänneksi lyhin ruotsalaisten ja länsisuomalaisten välillä (30) ja vasta viidenneksi lyhin etäisyys länsi- ja itäsuomalaisten välillä (32). Tällä perusteella voidaan esittää näiden väestöjen kehitys graafisesti:



Islantilaisten osaväestöjen välillä on havaittu yhtä suuria FST-etäisyyksiä kuin itä- ja länsisuomalaisten välillä, vaikka saari on ollut asuttu vasta runsaat tuhat vuotta (Helgason et al. 2004). Yhtäältä eri alueilta (Norjasta, Britteinsaarilta) muuttaneet ovat saattaneet asettua eri osiin saarta, ja toisaalta Islannin väestö on aina ollut pieni ja asutus harvaa, mikä on mahdollistanut suuren geneettisen ajautumisen ja eristymisen. Ajautuminen riittää siis selittämään myös itäsuomalaisten suuret FST-etäisyydet muihin väestöihin.

### 1.1. Suomalaisen osaväestöjen ajautuminen

Myös suomalaisten osaväestöjen ajautumista voidaan tarkastella samalla tavalla, koska tutkimuksessa Jakkula et al. 2008 on esitetty väestöjen sisäiset IBS-arvot sekä IBS-arvo suhteessa helsinkiläisiin, muttei valitettavasti kaikkien osaväestöjen välisiä IBS-arvoja.

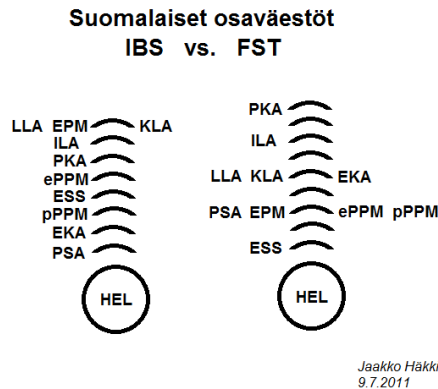


Myös FST-etäisyydet löytyvät samasta tutkimuksesta:

**Table 2. Fst Values for Subpopulations**

	CEU	SWE	HEL	ESS	ESW1	ESW2	ESN	LSW	LSN	LSC	ISS	ISC	ISN
CEU		0.001	0.007	0.006	0.007	0.008	0.008	0.007	0.009	0.009	0.012	0.012	0.011
SWE	0.001		0.005	0.004	0.005	0.006	0.006	0.005	0.007	0.007	0.010	0.010	0.009
HEL	0.007	0.005		0.000	0.001	0.001	0.002	0.001	0.002	0.001	0.002	0.004	0.003
ESS	0.006	0.004	0.000		0.001	0.001	0.003	0.002	0.002	0.001	0.003	0.004	0.004
ESW1	0.007	0.005	0.001	0.001		0.001	0.003	0.002	0.003	0.001	0.003	0.004	0.004
ESW2	0.008	0.006	0.001	0.001	0.001		0.003	0.002	0.002	0.001	0.002	0.003	0.003
ESN	0.008	0.006	0.002	0.003	0.003	0.003		0.003	0.002	0.004	0.005	0.006	0.005
LSW	0.007	0.005	0.001	0.002	0.002	0.002	0.003		0.004	0.003	0.005	0.006	0.005
LSN	0.009	0.007	0.002	0.002	0.003	0.002	0.002	0.004		0.003	0.004	0.004	0.002
LSC	0.009	0.007	0.001	0.001	0.001	0.001	0.004	0.003	0.003		0.002	0.003	0.004
ISS	0.012	0.010	0.002	0.003	0.003	0.002	0.005	0.005	0.004	0.002		0.003	0.004
ISC	0.012	0.010	0.004	0.004	0.004	0.003	0.006	0.006	0.004	0.003	0.003		0.004
ISN	0.011	0.009	0.003	0.004	0.004	0.003	0.005	0.005	0.002	0.004	0.004	0.004	

Koska IBS-arvoja ei ole kuin suhteessa helsinkiläisiin, esitän tulokset kaksiulotteisen kuvaajan sijasta yksiakselisena ”kiertoratakuvaajana” eli osaväestöjen suhteellisena etäisyytenä helsinkiläisiin. FST-kuvassa rajoittunut tarkkuus saa monet osaväestöt sijoittumaan samaan kohtaan; tätä kompensoin venyttämällä FST-kuvaa niin että joka toinen kohta jää tyhjäksi.



Yhteisen geeniosuuden (IBS) perusteella helsinkiläiset ovat lähimpänä itäsuomalaisia osaväestöjä: pohjoissavolaisia, eteläkainuulaisia ja pohjoisia pohjoispohjalaisia. Vasta sitten tulevat eteläsuomalaiset (lounaissuomalaiset), joiden jälkeen tulevat loputkin itäsuomalaisväestöt ennen lappilaisia ja eteläpohjalaisia. Pohjoiskainuulaiset näyttävät ryhmittyvän lähemmäs lappilaisia, eteläkainuulaiset taas savolaisten kanssa.

FST-etäisyyksissä suurimmat nousijat ovat eteläpohjalaiset, eteläsuomalaiset sekä länsi- ja keskilappilaiset, jonkin verran nousevat myös eteläiset pohjoispohjalaiset. Suurimpia putoajia ovat etelä- ja pohjoiskainuulaiset, mutta myös savolaiset putoavat jonkin verran.

FST-etäisyyksissä putoavat odotetusti eniten ne osaväestöt, joilla on suhteellisesti pienin sisäinen diversiteetti (EKA, PKA, PSA; ILA ei putoa vaikka diversiteetti on pieni), eli joilla geneettinen ajautuminen on ollut voimakkaampaa. Vastaavasti eniten nousevat ne osaväestöt, joilla on suhteellisesti suurin sisäinen diversiteetti (ESA, EPM, ePPM, LLA, KLA). Tämä voi selittyä kahdella tavalla: joko näiden osaväestöjen diversiteetti on kasvanut nimenomaan sekoittuessaan ”helsinkiläisiin” (= muihin suomalaisiin), tai sitten diversiteetti on ollut aina korkea, mikä on pitänyt ajautumisen vähäisenä. Jälkimmäisessä vaihtoehdossa itäsuomalaiset osaväestöt olisivat siis vain ajautuneet kauemmas helsinkiläisistä läntisten osaväestöjen (ja pohjoispohjalaisten joilla itäsuomalaisuudestaan huolimatta on korkeahko diversiteetti) säilyessä entisellä etäisyydellään.

## 1.2. Saamelaisten, baskien ja sardinialaisten ajautuminen

Tutkimuksessa Huyghe et al. 2011 on esitetty monien kansojen väliltä sekä IBS- että FST-arvot, joista olen koostanut seuraavan Eurooppa-keskeisen taulukon; diagonaalilla kulkee väestön sisäinen IBS-arvo, sen yläpuolella väestöjen väliset IBS-arvot ja alapuolella väestöjen väliset FST-etäisyydet.

FST\IBS	Saa	Rus	Orc	Fre	Bas	ItN	ItT	Ady	Uig	Sar	Haz	Pat
Saa	<b>0,7506</b>	0,7434	0,7408	0,7403	0,7388	0,7387	0,7382	0,7375	0,7369	0,7362	0,7357	0,7344
Rus	98 (1)	<b>0,7469</b>	0,7446	0,7444	0,7431	0,7437	0,7431	0,7416	0,7361	0,7409	0,7354	0,7374
Orc	144 (3)	17	<b>0,7475</b>	0,7458	0,7446	0,7449	0,7438	0,7418	0,7333	0,7432	0,7319	0,7363
Fre	151 (4)	17	~0	<b>0,7461</b>	0,7456	0,7461	0,7451	0,7424	0,7332	0,7448	0,7325	0,7365
Bas	216 (9)	83	53	29	<b>0,7505</b>	0,7459	0,7447	0,7409	0,7313	0,7454	0,7304	0,7344
ItN	154 (5)	18	~0	~0	12	<b>0,7469</b>	0,7462	0,7428	0,7328	0,7462	0,7320	0,7363
ItT	129 (2)	~0	~0	~0	~0	~0	<b>0,7468</b>	0,7429	0,7328	0,7455	0,7318	0,7365
Ady	181 (7)	54	48	39	103	~0	~0	<b>0,7453</b>	0,7410	0,7353	0,7350	0,7388
Uig	169 (6)	131	170	175	238	149	101	112	<b>0,7467</b>	0,7303	0,7449	0,7373
Sar	263(11)	130	86	50	81	~0	~0	11	264	<b>0,7503</b>	0,7293	0,7337
Haz	220(10)	184	233	231	294	205	166	163	~0	319	<b>0,7457</b>	0,7366
Pat	194 (8)	99	111	109	181	86	45	44	57	205	104	<b>0,7401</b>

Algoritmin epätarkkuus tuottaa geneettisesti läheisten väestöjen välille jopa negatiivisia arvoja, jotka olen pyöristänyt nolliin.

Saamelaisten jälkeen ovat järjestyksessä ne väestöt, joiden kanssa he jakavat eniten yhteisiä geenejä (IBS). FST-arvon perässä on suluissa sijoitus geneettisen etäisyyden mukaan, lähin on 1. Eristyneillä populaatioilla (saamelaiset, baskit, sardinialaiset) on pienin diversiteetti (korkein väestönsisäinen IBS-samankaltaisuus). Näiden väestöjen oletetaan ajautuneen suhteessa eniten, ja näin asia näyttääkin olevan: IBS-järjestykseen verrattuna suurin putoaja ovat baskit (4. → 9.), mikä selittyy näiden kahden eristyneen väestön ajautumisella kauemmas toisistaan. Sama koskee toista FST-etäisyyksissä pudonnutta (9. → 11.) eristynyttä väestöä sardinialaisia.

Suurin nousija (6. → 2.) ovat toscanalaiset, mikä genomilaajuksen analyysin kyseessä ollessa voi tuskin johtua sattumasta – saamelaisilla ja toscanalaisilla täytynee olettaa olevan jokin yhteinen väestökomponentti. Myös pataanit (11. → 8.) ja uiguurit (8. → 6.) nousevat FST-arvoissa; aiemmin esitin saamelaisten yhden juuren viittaavan näihin keskiaasialaisiin väestöihin jo pelkästään FST-etäisyyksien perusteella (ks. Etusivu uusiksi: saamelaisten yksi juuri Keski-Aasiasta?), ja nyt siis IBS-arvoihin vertaaminen vahvistaa tämän oletuksen ja osoittaa vaikutuksen myöhäisehköksi: saamelaisväestön perustajissa ei tätä keskiaasialaista osatekijää vielä ollut.

Muutamia havaintoja muiden kansojen välisistä IBS-suhteista: venäläiset jakavat saamelaisia enemmän geenejä kaikkien muiden kansojen kanssa paitsi uiguurien ja hazaroiden. Baskit jakavat ranskalaisten ja pohjoisitalialaisten jälkeen eniten geenejä sardinialaisten kanssa, mikä heijastanee yhteistä lounaiseurooppalaista perimää. Toisaalta baskeilla ei näytä olevan odotettua enempää yhteyttä Kaukasuksen adygeiden kanssa, vaikka eräät etävertailijat ovat yrittäneet todistaa baskin kieltä kaukasialaisten kielten sukulaiseksi.

Pohjoisitalialaiset ja toscanalaiset ovat hyvin läheisiä: näille kansoille yhteisiä geenejä on lähes yhtä paljon kuin kansojen sisäisiä geenejä. Pieniä eroja kuitenkin löytyy suhteessa muihin kansoihin: pohjoisitalialaiset jakavat enemmän yhteisiä geenejä sardinialaisten ja baskien kanssa, vaikka toscanalaiset asuvat lähempänä sardinialaisia. Tämäkin viittaa toscanalaisten vieraaseen komponenttiin; Toscanan alueella eläneet etruskit onkin perinteisesti johdettu Vähästä-Aasiasta.

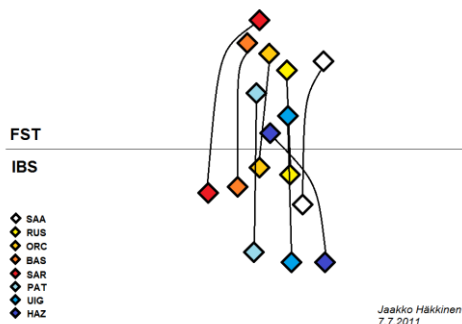
Adygeit jakavat toiseksi eniten yhteisiä geenejä pataanien kanssa (heti hazaroiden jälkeen) kuin mikään muu kansa, mutta hazaroiden kanssa heitä enemmän geenejä jakavat saamelaiset ja venäläiset. Sardinialaisten ja hazaroiden väliltä löytyy pienin yhteisten geenien osuus. Hazarat jakavat selvästi enemmän yhteisiä geenejä uiguurien kanssa kuin minkään muun kansan kanssa. Pataanit ovat lähimpänä adygeita, sitten venäläisiä ja uiguureja. Uiguurit jakavat eniten geenejä hazaroiden, adygeiden, pataanien ja saamelaisten kanssa.

FST-etäisyyksissä mielenkiintoista on toscanalaisten risteysasema: heidän etäisyytensä on olematon peräti seitsemään väestöön käsitellyistä yhdestätoista. Toscanalaiset ovat lisäksi lähempänä pataaneja ja uiguureja kuin saamelaisia, ja heidän läheisyytensä Kaukasuksen adygeihin tukee isälinjojen pohjalta muodostettua kuvaa: toscanalaisissa on vähäaasialainen komponentti, joka periytynee etruskeilta.

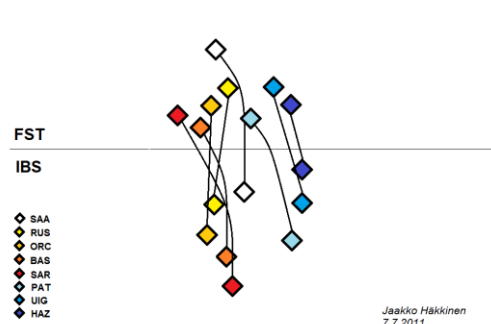
Olen ottanut seuraavaan kuvaajaan mukaan laitaväestöt ja jättänyt pois joitain keskelle sijoitettavia lähekkäisiä väestöjä, koska näiden FST-etäisyydet menevät taulukossa negatiivisiksi, mikä johtuu algoritmin vähäisestä epätarkkuudesta. Muuten väestöjen väliset IBS-arvot vastaavat kuvaajan etäisyyksiä, mutta venäläisten ja pataanien läheistä suhdetta ei saanut mahtumaan kaksiulotteiseen kuvaajaan.

IBS- ja FST-etäisyyksien perusteella voidaan piirtää seuraavanlaiset kuvaajat; sijainnit ovat molemmissa kuvissa samat, mutta väestörykelmiä on oikeanpuoleisessa kuvassa kierretty 90 astetta vasemmalle. Ensimmäisessä kuvassa eniten ajautuneita ovat sardinialaiset (SAR), hazarat (HAZ), saamelaiset (SAA) ja baskit (BAS), toisessa kuvassa sardinialaiset (SAR), saamelaiset (SAA), pataanit (PAT) ja baskit (BAS). Pataanit ovat ajautuneet lähemmäs eurooppalaisia (mikä tarkoittaa sekoittumista), samoin uiguurit jossain määrin, kun taas hazarat ovat ajautuneet omaan suuntaansa (tai sekoittuneet sellaisiin väestöihin joita kuvaajassa ei ole mukana). Sardinialaiset ja saamelaiset ovat ajautuneet kauimmas muista väestöistä, baskit eivät aivan yhtä suuressa määrin.

Geneettinen ajautuminen = IBS vs. FST



Geneettinen ajautuminen = IBS vs. FST



Venäläisten, orkneyläisten, uiguurien ja pataanien väliset erot näyttävät FST-arvoissa muuttuneen vähiten IBS-arvoista; lähinnä nämä väestöt ovat hieman lähentyneet toisiaan, tai sitten reunaväestöjen ajautuminen kauemmas vain vääristää mittakaavaa niin että keskusväestöt näyttävät lähentyneen toisiaan.

### 1.3. Ajautumisen osoittaminen

Menetelmä vaikuttaisi antavan uskottavia tuloksia: eristyneitä väestöjä pidetään alttiimpina geneettiselle ajautumiselle, ja niiden pienempi diversiteetti osoittaakin ajautumista tapahtuneen. Siksi eristyneiden väestöjen (SAA, BAS, SAR, FIE) muita suuremmat muutokset IBS-kartan ja FST-kartan välillä voidaan tulkita geneettisestä ajautumisesta johtuviksi.

## 2. Suomalaiset eurooppalaisina

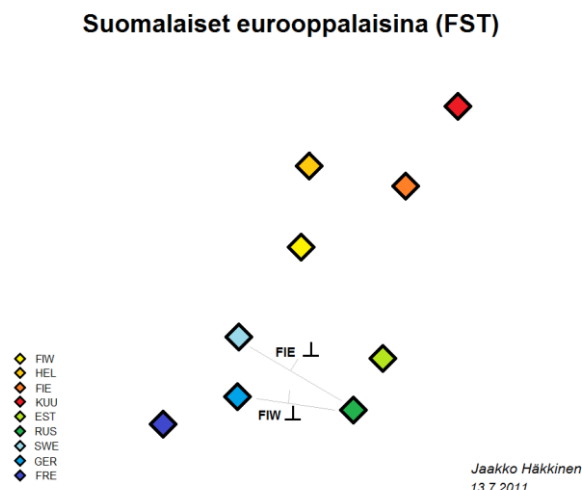
Suomalaisten ja eurooppalaisten väliset geneettiset etäisyydet ovat hajallaan eri tutkimuksissa, mutta kokonaiskuva on silti mahdollista saavuttaa. Muihin väestöihin on verrattu vain länsi- ja itäsuomalaisia (Salmela et al. 2008) tai helsinkiläisiä ja kuusamolaisia (Jakkula et al. 2008; Nelis et al. 2009). Taulukon suluissa olevat arvaukset ovat omiani ja perustuvat siihen, että helsinkiläiset näyttävät olevan suunnilleen länsi- ja itäsuomalaisten välissä ja että kuusamolaiset ovat itäsuomalaisista edelleen ajautunut osaväestö.

	FIW <sup>1</sup>	HEL <sup>2</sup>	FIE <sup>1</sup>	KUU <sup>2</sup>
virolaiset	(20?)	40 <sup>3</sup>	(60?)	90 <sup>3</sup>
ruotsalaiset	24	50	70	100
venäläiset	34	(60?)	69	(100?)
saksalaiset <sup>3</sup> /CEU <sup>2</sup>	35	60/70	85	120
ranskalaiset	60	(95?)	113	(150?)

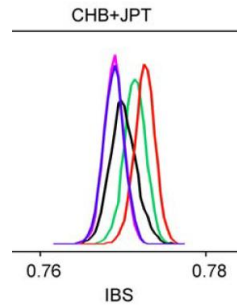
<sup>1</sup> = Salmela et al. 2008; <sup>2</sup> = Jakkula et al. 2008; <sup>3</sup> = Nelis et al. 2009

	virolaiset	ruotsalaiset	venäläiset	saksalaiset	ranskalaiset
virolaiset	---	30	10	30	50
ruotsalaiset	30	---	30	10	20
venäläiset	10	30	---	20	50
saksalaiset	30	10	20	---	10
ranskalaiset	50	20	50	10	---

Länsisuomalaisten ja helsinkiläisten etäisyys venäläisiin on samaa luokkaa kuin etäisyys saksalaisiin (suurempi kuin ruotsalaisiin), kun taas itäsuomalaisten etäisyys venäläisiin on samaa luokkaa kuin etäisyys ruotsalaisiin (pienempi kuin saksalaisiin). Vaikka itäsuomalaisten suuret etäisyydet muihin väestöihin selittyvätkin suureksi osaksi geneettisellä ajautumisella, osoittavat länsi- ja itäsuomalaisten erilaiset etäisyysuhteet muihin kansoihin myös erilaista sekoittumista tai geenipohjaa; tämä näkyy kuvaajaan merkityistä kohtisuorista.



Samaan viittaa se, että itäsuomalaiset (punainen käyrä) ovat geenipohjansa osalta (IBS) hieman lähempänä itäaasialaisia kuin länsisuomalaiset (vihreä käyrä) ja muut lähiväestöt (kuvaaja tutkimuksesta Salmela et al. 2008).



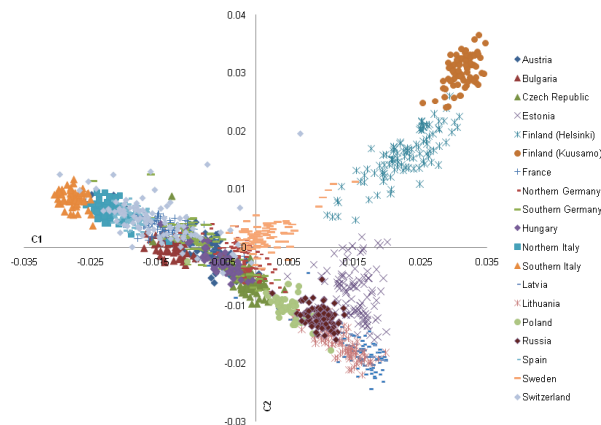
### *Suomalaisten etäisyys baltteihin*

Latvialaisia ja liettualaisia on harvoin tutkittu yhdessä suomalaisten kanssa, mutta tutkimuksessa Nelis et al. 2009 on helsinkiläisten ja kuusamolaisten FST-etäisyydet baltteihin. Yllättäen venäläiset, puolalaiset ja saksalaiset ovat lähempänä suomalaisia kuin latvialaiset ja liettualaiset; jopa unkarilaiset ovat hieman lähempänä helsinkiläisiä kuin baltit. Ranskalaisetkin ovat melkein yhtä lähellä suomalaisia kuin baltit, mikä on erikoista.

	EST	LAT	LIT	RUS	POL	SWE	GER	HUN
HEL	40	70	70	60	60	50	60	60
KUU	90	130	130	120	120	110	120	130

Kuitenkin latvialaiset, liettualaiset, virolaiset, venäläiset ja puolalaiset muodostavat läheisen rykelmän, kaikkien parien välisen keskinäisen etäisyyden ollessa vain 10–20. Keskieurooppalaiset (saksalaiset, itävaltalaiset, unkarilaiset, tšekit, sveitsiläiset, ruotsalaiset, ranskalaiset ja jopa bulgarialaiset) muodostavat toisen klusterin, jonka parien väliset etäisyydet ovat yhtä pieniä. Klusterien väliset erotkin ovat melko pieniä (pienimmillään saksalaiset–puolalaiset 10). Espanjalaiset, italialaiset ja suomalaiset muodostavat kukin oman ryhmänsä, joiden sisällä etelä- ja pohjoisitalialaiset ovat selvästi erillisiä väestöjään ja suomalainen ryhmä jakautuu peräti seitsemään erilliseen väestöön (ks. Seitsemän suomalaista väestöä).

Tutkimuksessa Nelis et al. 2009 on PCA-kuvaaja, joka osoittaa baltit geneettisesti yhdeksi eurooppalaisista ääripäistä (suomalaisten ja eteläitalialaisten ohella), huolimatta balttien keskeisestä maantieteellisestä sijainnista:

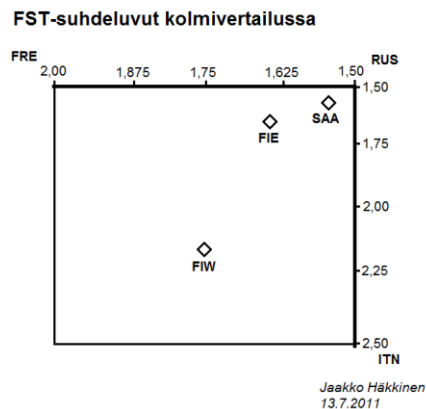


Ehkä balttilainen tekijä edustaa jotakin ikivanhaa eurooppalaista geenikomponenttia, joka on säilynyt yksinäisenä saarena ”yleiseurooppalaisten” geenien meressä. Virolaiset näyttäisivät olevan balttien ja suomalaisten välissä, slaavit sijoittuvat balttien ja keskieuropalaisten väliin ja ruotsalaiset keskieuropalaisten ja suomalaisten väliin.

### *Suomalaisten etäisyys saamelaisiin*

Saamelaisia on tutkittu genomilaajuisesti vasta yhdessä tutkimuksessa (Huyghe et al. 2011), eikä siinä ollut valitettavasti mukana verrokkiväestöinä suomalaisia eikä ruotsalaisia; edellä listatuista verrokkiväestöistä mukana olivat vain venäläiset ja ranskalaiset (vaikka mukana muuten oli paljon väestöjä). Suomen saamelaisten FST-etäisyys venäläisiin on 98 ja ranskalaisiin 151, näiden suhteen ollessa 1,54-kertainen. Itäsuomalaisilla se on 1,64-kertainen (113 : 69) ja länsisuomalaisilla 1,76-kertainen (60 : 34). Tällaisesta yksinkertaisesta ”kolmiomittauksesta” ei luonnollisesti voi vetää kummoisiakaan johtopäätöksiä, mutta jonkinlaisen länsi–itä-akselin siinä voi hahmottaa.

Toinen kolmiomittaus saadaan saamelaisten, helsinkiläisten ja kuusamolaisien välille. Saamelaisten FST-etäisyys venäläisiin on 98 ja pohjoisitalialaisiin 154, näiden suhteen ollessa 1,57-kertainen. Helsinkiläisillä suhde on peräti 2,17-kertainen (130 : 60) ja kuusamolaisilla 1,67-kertainen (200 : 120). Tällä akselilla itäsuomalaiset ovat paljon lähempänä saamelaisia kuin länsisuomalaisia. Olettaen, että helsinkiläiset muistuttaisivat länsisuomalaisia, saataisiin aikaan seuraava kaksiakselinen kuvaaja:



On korostettava, että tämä kuvaaja ei osoita saamelaisten läheisyyttä venäläisiin vaan saamelaisten läheisyyttä suhdelukuun 1,50 kummallakin akselilla. Kuvaaja ei myöskään osoita, että itäsuomalaiset olisivat lähempänä saamelaisia kuin länsisuomalaisia, vaan se osoittaa että itäsuomalaisten suhdeluvut kummallakin akselilla ovat lähempänä saamelaisten kuin länsisuomalaisten suhdelukuja.

On siis mahdollista, että itäsuomalaisten ja saamelaisten geenipohjassa on paljon samaa ja että itäsuomalaiset ovat länsisuomalaisten ja saamelaisten sekoitus. Pidän todennäköisenä, että yhteisten geenien (IBS) osalta itäsuomalaiset saattavat olla lähes yhtä lähellä saamelaisia kuin länsisuomalaisia, vaikka geneettisen etäisyyden (FST) osalta he varmasti ovat selvästi kauempana saamelaisista kuin länsisuomalaisista, johtuen yhtäältä geneettisen ajautumisen suuruudesta sekä itäsuomalaisissa että saamelaisissa ja toisaalta itä- ja länsisuomalaisten jatkuvista keskinäisistä kontakteista ja sekoitumisesta. Tämä selviää kuitenkin vasta tulevaisuuden geenitutkimuksissa.



## *Suomalaisten etäisyys ruotsalaisiin*

Tutkimuksen Salmela et al. 2011 täydennystiedoissa on esitetty myös ruotsalaisten ja suomalaisten osaväestöjen etäisyydet:

### **Ruotsalaiset (SWE)**

ruotsalaiset – ruotsinkieliset pohjalaiset: 13–33

ruotsalaiset – varsinaissuomalaiset: 19–33

ruotsalaiset – satakuntalaiset: 29–48

ruotsalaiset – eteläpohjalaiset: 33–55

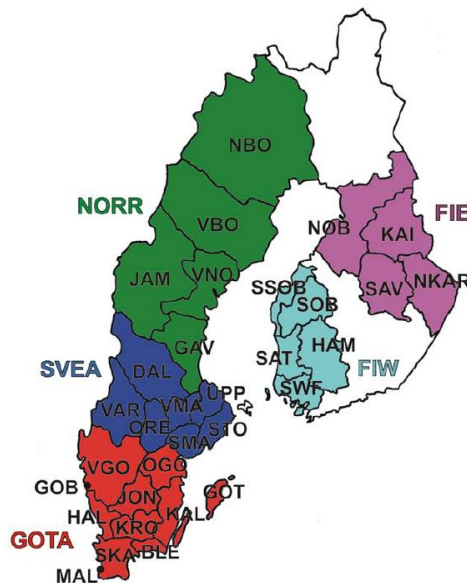
ruotsalaiset – hämäläiset: 36–58

ruotsalaiset – pohjoispohjalaiset: 60–79

ruotsalaiset – pohjoiskarjalaiset: 66–91

ruotsalaiset – savolaiset: 75–99

ruotsalaiset – kainuulaiset: 80–105



Kaikkien suomalaisten osaväestöjen geneettinen etäisyys on keskimäärin pienin sveanmaalaisiin (keskiruotsalaisiin), toiseksi pienin norrlantilaisiin (pohjoisruotsalaisiin) ja suurin götanmaalaisiin (eteläruotsalaisiin). Myös osalla itäsuomalaisista osaväestöistä pienimmät etäisyydet ovat keskiruotsalaisiin (savolaiset, pohjoiskarjalaiset), mutta kainuulaisilla (KAI) ja pohjoispohjalaisilla (NOB) pienin etäisyys on länsinorrlantilaisiin (VNO). Mielenkiintoista kyllä, ruotsinkielisten pohjalaisten (SSOB) pienin yksittäinen etäisyys on myös länsinorrlantilaisiin, vaikka keskimäärin etäisyys onkin pienempi keski- kuin pohjoisruotsalaisiin.

Tämä viittaa siihen suuntaan, että varhaisempia kontakteja on ollut erityisesti Merenkurkun yli. Tämä länsinorrlantilainen tekijä ulottuu ruotsinkielisistä pohjalaisista pohjoispohjalaisiin ja kainuulaisiin, erottaen ne eteläisemmistä osaväestöistä (niin länsi- kuin itäsuomalaisissakin). Tekijä ei kuitenkaan näy suomalaisten osaväestöjen sisäisessä vertailussa (ks. Seitsemän suomalaista väestöä) vaan vasta verrattaessa ruotsalaisiin.

Ruotsalainen vaikutus on muutenkin heikkoa eikä erotu suomalaisten osaväestöjen keskinäisessä vertailussa: ruotsinkielisten pohjalaisten ja eteläpohjalaisten välinen etäisyys on vain 7, vaikka eteläpohjalaisten etäisyys ruotsalaisiin on keskimäärin kaksinkertainen ruotsinkielisten pohjalaisten etäisyyteen verrattuna. Ruotsalaisvaikutus rajoittuu länsisuomalaisissakin rannikkovyöhykkeeseen (SWF, SAT, SSOB), sisämaan osaväestöjen (HAM, SOB) ollessa jo selvästi kauempana ruotsalaisista.

Kaikki länsisuomalaiset osaväestöt ovat lähempänä ruotsalaisia kuin itäsuomalaiset. Länsisuomalaisten etäisyys ruotsalaisiin vaihtelee ruotsinkielisten pohjalaisten arvoista 13–33 hämäläisten arvoihin 36–58. Yllättävää on, että eteläpohjalaiset ovat yhtä kaukana ruotsalaisista kuin hämäläiset, vaikka heidän etäisyytensä ruotsinkielisiin pohjalaisiin on pieni. Tämä voisi viitata siihen, että ruotsalaisten osuus ruotsinkielisten pohjalaisten geeniperimään on sittenkin kohtuullisen pieni. Toisaalta suurimmat ruotsalaisten väliset etäisyydet ovat noin 20, joten ruotsinkieliset pohjalaiset ja yllättäen jopa varsinaissuomalaiset ovat suhteellisen lähellä ruotsalaisia.

Läntisimmät itäsuomalaiset eli pohjoispohjalaiset ovat odotetusti lähimpänä ruotsalaisia, vaikkakin selvästi kauempana kuin länsisuomalaiset (huolimatta havaitusta norlantilaisesta tekijästä), mikä jälleen vahvistaa heidän itäisen juurensa voimakkuuden. Yllättävänä voidaan pitää sitä, että pohjoiskarjalaiset ovat lähempänä ruotsalaisia kuin heidän länsipuolellaan asuvat savolaiset; he ovat myös lähempänä kuin savolaiset kaikkiin muihin länsisuomalaisiin osaväestöihin nähden paitsi hämäläisiin. Tämä voi johtua Karjalankannaksen ja Viipurin alueen merkityksestä karjalaisten pohjoisessa ekspansiossa: länsikannaksen murteissa on selvä länsisuomalainen vaikutus. Toisaalta nykyiset pohjoiskarjalaiset ovat suurelta osin savolaisia, jotka asettuivat alueelle alkuperäisten pohjoiskarjalaisten paetessa Venäjälle Stolbovan rauhan jälkeen 1617. On arvioitu, että noin puolet silloisista pohjoiskarjalaisista muutti pois.

### 3. Suomalaisväestöjen ominaisuudet

Geenien eri tasot paljastavat erilaisia ominaisuuksia suomalaisväestöistä ja osaväestöistä.

- IBS vs. FST kertoo väestön geneettisen ajautumisen suhteellisen määrän (ks. luku 1).
- Diversiteetti kertoo väestön suurimman sisäisen erilaisuuden; matala diversiteetti kertoo homogeenisesta ja korkea diversiteetti heterogeenisesta väestöstä. Korkea diversiteetti voi johtua väestön ”syntymisestä” sekoittumisen tuloksena.
- LD  $r^2$  on väestösisäisessä parivertailussa suuri eristyneillä väestöillä ja pieni sekoittuneilla väestöillä.
- LD *map length* kertoo väestön suhteellisen iän viimeisen pullonkaulan (joka voi olla myös väestön perustaminen) jälkeen, mutta se näyttäisi liittyvän jossain määrin myös väestön avoimuuteen tai eristyneisyyteen: eristyneillä väestöillä LD *map* on lyhyempi (Angius et al. 2008).
- Homotsygoottisuus on suurta pienillä ja eristyneillä väestöillä, joilla samat geenit kiertävät.
- ROH-jaksojen suuri pituus liittyy endogaamisiin väestöihin (puoliso otetaan saman ryhmän sisältä), ja erittäin suuret pituudet kertovat sisäsiittoisuudesta.
- Lyhyiden ROH-jaksojen määrä kertoo perustajaväestön suhteellisen koon: mitä vähemmän näitä jaksoja on, sitä suurempi on aikoinaan ollut väestön efektiivinen koko.
- Yksimuotoiset markkerit kertovat väestön geneettisestä ajautumisesta.
- MAF kertoo väestön ajautumisesta.
- Sisäsiittoisuustekijä kertoo sukulaisavioliittojen suhteellisen osuuden.

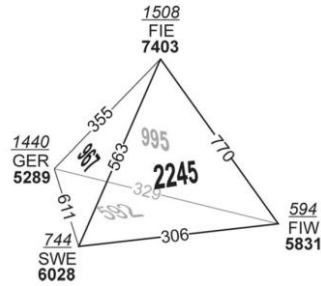
Vaikka tarkastelu rajoitettaisiin vain suomalaisten geneihin ilman ulkopuolisia verrokkiväestöjä, on silti mahdollista saada viitteitä siihen, onko jokin suomalaisväestö syntynyt perustajavaikutuksen tai pullonkaulan aiheuttaman geneettisen ajautumisen kautta vai ei. Seuraavaksi käsittelem näitä tekijöitä: osaa niistä on sovellettu vain länsi- ja itäsuomalaisten tarkkuudella, osaa osaväestöjen tarkkuudella.

#### 3.1. Suomalaiset suhteessa eurooppalaisiin

Aluksi tarkastelen länsi- ja itäsuomalaisia suhteessa ruotsalaisiin, pohjoissaksalaisiin ja britteihin tutkimuksen Salmela et al. 2008 pohjalta.

##### *Yksimuotoiset (monomorfiset) markkerit*

Yksimuotoisia SNP:itä (joissa ei esiinny lainkaan väestösisäistä vaihtelua) on enemmän ajautuneilla väestöillä (Angius et al. 2008), mikä kertoo väestön pienestä koosta. Kun väestö on pieni, siinä tapahtuu geneettistä ajautumista, jonka seurauksena harvinaisia alleleja menetetään ja joistain snipeistä (SNP = *single-nucleotid polymorphism*) tulee yksimuotoisia: kaikilla saman väestön edustajilla on tällaisessa markkerissa sama alleeli. Seuraava kuvaaja on tutkimuksesta Salmela et al. 2008; väestöjen yksimuotoiset markkerit on lihavoitu, ja vain näissä väestöissä yksimuotoisina esiintyvät markkerit on kursivoitu ja alleviivattu.



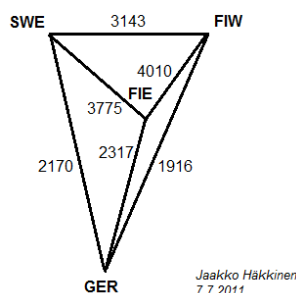
Eniten yksimuotoisia markkereita on itäsuomalaisissa (FIE) ja vähiten saksalaisissa (GER); ruotsalaisissa (SWE) niitä on hieman enemmän kuin länsisuomalaisissa (FIW). Saksalaisissa on kuitenkin toiseksi eniten vain tähän väestöön rajoittuvia yksimuotoisia markkereita, mikä johtunee siitä, että se on selvästi erillään muista väestöistä: kolmelle muulle väestölle (FIE, FIW, SWE) yhteisiä yksimuotoisia markkereita on 2 245, yli kaksi kertaa enemmän kuin millään muulla kolmen yhdistelmällä. Tämä saattaa viitata ikivanhaan eroon Fennoskandian ja manner-Euroopan välillä: kummallakin taholla on kehittynyt omia yksimuotoisia markkereitaan. Kuitenkin kaikille neljälle väestölle yhteisiä yksimuotoisia markkereita (ei kuvassa) on peräti 21 000; nämä edustanevat vanhinta yhteistä pohjoiseurooppalaisuutta ja periytyvät esihistoriallisista pullonkauloista tai muista ajautumista aiheuttaneista tekijöistä.

Kahdenkeskisissä luvuissa kiinnittää huomiota, että itä- ja länsisuomalaiset jakavat eniten yksimuotoisia markkereita, ruotsalaiset ja saksalaiset toiseksi eniten. Itäsuomalaiset ja ruotsalaiset jakavat yllättävästi kolmanneksi eniten, ja itäsuomalaiset ja saksalaiset jakavat neljänneksi eniten yksimuotoisia markkereita. Länsisuomalaiset ja saksalaiset jakavat viidenneksi eniten ja länsisuomalaiset ja ruotsalaiset vasta kuudenneksi eniten yksimuotoisia markkereita. Ennen kuin tästä voi vetää johtopäätöksiä, on laskettava kaikkien väestöjen yhteensä jakamat yksimuotoiset markkerit:

	GER	SWE	FIW	FIE
GER		2170	1916	2317
SWE	2170		3143	3775
FIW	1916	3143		4010
FIE	2317	3775	4010	

Eniten yhteisiä yksimuotoisia markkereita on itä- ja länsisuomalaisilla, toiseksi eniten itäsuomalaisilla ja ruotsalaisilla, kolmanneksi eniten länsisuomalaisilla ja ruotsalaisilla, neljänneksi eniten itäsuomalaisilla ja saksalaisilla, viidenneksi eniten ruotsalaisilla ja saksalaisilla ja kuudenneksi eniten länsisuomalaisilla ja saksalaisilla. Suhteet voidaan esittää graafisesti, kun muutetaan yhteisten markkerien määrä kääntäen etäisyydeksi (vähän = pitkä etäisyys):

**Yhteiset yksimuotoiset (monomorphic) markkerit**



Jaakko Häkkinen  
7.7.2011

Kokonaiskuvassa säilyy se odottamaton tilanne, että itäsuomalaiset näyttävät tällä tasolla olevan selvästi lähempänä ruotsalaisia kuin länsisuomalaiset, ja miltei yhtä lähellä ruotsalaisia kuin länsisuomalaisia. Yhtä yllättävästi itäsuomalaiset näyttävät olevan lähempänä saksalaisia kuin ruotsalaiset ja länsisuomalaiset. Tämä tulos on selvästi ristiriidassa IBS- ja FST-arvojen kanssa: niiden kummankin perusteellahan länsisuomalaiset ovat lähempänä ruotsalaisia ja saksalaisia kuin itäsuomalaiset.

Yksimuotoisten markkerien erikoinen tulos johtuneekin vääristymästä, johon vaikuttaa kaksi tekijää. Ensiksikin itäsuomalaisilla on peräti 1 400 yksimuotoista markkeria enemmän kuin seuraavalla väestöllä, ja toiseksi pohjoiseurooppalaisilla ovat todennäköisesti olleet harvinaisia enimmäkseen samojen markkerien alleelit. Niinpä harvinaiset alleelit ovat voineet kadota eri väestöissä itsenäisesti, ja koska itäsuomalaisilla on niin paljon yksimuotoisia markkereita, heiltä löytyy vastine useammalle ruotsalaisten ja saksalaisten markkereille kuin länsisuomalaisilta.

Tämän vääristymän vuoksi yksimuotoisten markkereiden perusteella ei voikaan tutkia väestöjen geneettistä läheisyyttä, mutta geneettisen ajautumisen paljastajana sitä voidaan pitää luotettavana: eniten näyttävät ajautuneen geneettisesti itäsuomalaiset, seuraavaksi eniten ruotsalaiset ja sitten vasta länsisuomalaiset.

### **MAF (minor allele frequency)**

Edellisen kaltaiseen tulokseen viittaa se, että MAF (*minor allele frequency*) ja heterotsygoottisuus (*sample*) on ruotsalaisilla toiseksi pienintä heti itäsuomalaisten jälkeen ja heterotsygoottisuus (*marker*) kaikkein pienintä; viime mainittu on länsisuomalaisilla jopa suurempi kuin pohjoissaksalaisilla. Pieni MAF johtuu siitä, että harvinaisemmat alleelit ovat jo ehtineet kadota väestöstä ajautumisen seurauksena. Taulukko tutkimuksesta Salmela et al. 2008:

	median heterozygosity (sample)	median heterozygosity (marker)	median MAF	median inbreeding coefficient
FIW	0,2783	0,3050	0,1929	0,0077
FIE	0,2768	0,3022	0,1906	0,0057
SWE	0,2779	0,3009	0,1927	0,0037
GER	0,2793	0,3047	0,1941	0,0031
BRI	0,2799	0,3069	0,1943	0,0005

Länsisuomalaisten ruotsalaisia suurempi heterotsygoottisuus (eli pienempi homotsygoottisuus) voi kuitenkin johtua siitä, että länsisuomalaiset voidaan perustellusti jakaa kahteen eri väestöön, lounaissuomalaisiin ja pohjalaisiin (ks. Seitsemän suomalaista väestöä).

### **Sisäsiittoisuus (inbreeding)**

Homotsygoottisuus ei ole automaattisesti merkki sisäsiittoisuudesta, vaikka sisäsiittoisuus lisääkin homotsygoottisuutta huomattavasti. Kulttuureissa, joissa harrastetaan serkusavioliittoja, on homotsygoottisuus jopa 11 %, kun se suomalaisilla osaväestöillä vaihtelee välillä 0,9–2 % (Jakkula et al. 2008). Homotsygoottisuus on korkeampi itäsuomalaisilla kuin länsisuomalaisilla (ks. edellinen taulukko), mutta siitä huolimatta länsisuomalaisilla on hieman korkeampi sisäsiittoisuusaste, kuten edellä olevasta taulukosta nähdään. Kohonnut homotsygoottisuus ei siis matalien arvojen osalta

kerro sisäsiirtoisuudesta vaan ainoastaan eristyneestä väestöstä johon ei tule kovinkaan paljon uusia geenejä.

Länsisuomalaisten sisäsiirtoisuusaste on hieman korkeampi kuin muilla väestöillä, vaikka erot eivät merkittäviä olekaan: väestön tasolla ei voidakaan vielä puhua todellisesta sisäsiirtoisuudesta. Tämä kuitenkin osoittaa, että pienet, eristyneet ja ajautuneet väestöt (kuten itäsuomalaiset) eivät ole automaattisesti muita sisäsiirtoisempia.

### 3.2. Suomalaisten osaväestöjen keskinäiset erot

Seuraavat suomalaisten osaväestöjen ominaisuudet perustuvat tutkimukseen Jakkula et al. 2008.

#### *Vähentynyt diversiteetti*

Diversiteetti saadaan selville IBS-analyysillä (*identity by state*). Yksilöitä verrataan pareittain, ja keskiarvo tai mediaani on väestön IBS-arvo. Mitä suurempi IBS-arvo on, sitä samankaltaisempia geenipohjaltaan ovat väestöön kuuluvat yksilöt eli sitä pienempi on väestön sisäinen diversiteetti. Korkeimmat arvot ovat listassa ylimpänä.

#### Jakkula et al. 2008:

ISC, pohjoiskainuulaiset: 0,7130

ISS, eteläkainuulaiset: 0,7130

ISN, itälappilaiset: 0,7128

LSC, pohjoissavolaiset: 0,7120

ESN, länsilappilaiset: 0,7115

LSN, keskilappilaiset: 0,7114

ESW2, pohjoiset pohjoispohjalaiset: 0,7114

ESW1, eteläiset pohjoispohjalaiset: 0,7110

LSW, eteläpohjalaiset: 0,7108

ESS, eteläsuomalaiset: 0,7106

Diversiteetti on pienin myöhään asutetuilla I-alueilla, mikä on odotuksenmukaista. Keskipientä diversiteetti on hieman varhemmin asutetuilla L-alueilla ja suurinta varhain asutetuilla E-alueilla, kahden poikkeusta lukuun ottamatta: länsilappilaisilla diversiteetti on odotettua pienempi (L-ryhmän tasolla) ja eteläpohjalaisilla odotettua suurempi (E-ryhmän tasolla). Käytännössä pohjoisilla pohjoispohjalaisilla diversiteetti ei ole sen suurempi kuin L-ryhmällä: se on sama kuin keskilappilaisilla. Eteläisillä pohjoispohjalaisilla diversiteetti on jo selvästi suurempi.

Eteläpohjalaisten diversiteetti on lähes yhtä suuri kuin eteläsuomalaisilla, joihin tutkimuksessa on sisällytetty ainakin kolme osaväestöä: varsinaissuomalaiset, satakuntalaiset ja hämäläiset. Kun huomioidaan lisäksi osaväestöjen väliset IBS-arvot suhteessa verrokkiväestönä toimiviin helsinkiläisiin, on ero odotetusti suurin helsinkiläisten ja lappilaisten välillä (länsi- ja keskilappilaiset 0,7100 ja itälappilaiset 0,7101), mutta eteläpohjalaiset ovat yllättäen jaetulla ykkössijalla (0,7100), kun taas heitä pohjoisemmilla pohjoispohjalaisilla ja kainuulaisilla osaväestöillä arvo suhteessa hel-

sinkiläisiin on pienempi (pohjoiskainuulaiset 0,7102, eteläiset pohjoispohjalaiset 0,7103, pohjoiset pohjoispohjalaiset 0,7106 ja eteläkainuulaiset 0,7107).

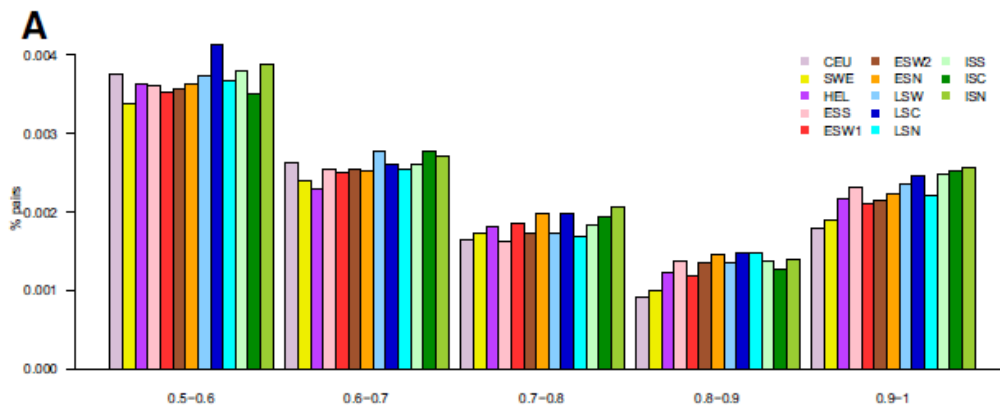
Eteläpohjalaiset eivät siis diversiteettianalyysin perusteella edusta myöhäistä asutusta vaan pikemminkin nykysuomalaisten vanhinta, kun otetaan huomioon että he ovat vain yksi osaväestö ja ainoa heitä suuremman sisäisen diversiteetin omaava väestö eteläsuomalaiset koostuu vähintään kolmesta erillisestä osaväestöstä.

Mainittakoon vielä, että IBS-arvot olivat suurempia jaoteltujen osaväestöjen sisällä kuin niiden välillä, mikä tutkijoiden mielestä on osoitus siitä että jako väestöihin vastaa todellisuutta eikä ole mielivaltainen.

### ***KytKentäepätasapaino (LD)***

Rekombinaatio on ilmiö, joka lyhyesti ja epätarkasti sanottuna silppuaa kromosomien eri vanhempaislinjoja pitkin perityt geenit hajalleen sitä pienemmiksi palasiksi, mitä enemmän aikaa on kulu-  
nut. KytKentäepätasapaino (LD, *linkage disequilibrium*) kuvaa tilannetta, jossa tätä silppuuntumista ei vielä ole merkittävässä määrin tapahtunut, mistä johtuen pitkät geenipätkät kromosomeissa periytyvät yhdessä nipussa.

Seuraavassa kuvassa (tutkimuksesta Jakkula et al. 2008) näkyy LD  $r^2$  kromosomi 22:n osalta. Pidemmät jaksot (vaaka-akselilla 0,7–1) painottuvat tässä tekijässä. Suurinta LD  $r^2$  on itälappilaisissa (ISN), savolaisissa (LSC) sekä etelä- (ISS) ja pohjoiskainuulaisissa (ISC). Mielenkiintoista kyllä, pienintä LD  $r^2$  on pohjoispohjalaisilla (ESW1, ESW2).



### ***LD-kartan pituus***

LD-kartan pituus (*Length of LD map*) näyttää olevan verrannollinen väestön ikään (aikaan viimeisimmästä pullonkaulasta): mitä lyhyempi se on, sitä nuorempi väestö on kyseessä. Myös eristyneillä väestöillä pituus on lyhyempi. Eteläpohjalaiset (LSW) ja savolaiset (LSC) näyttävät tämän perusteella olevan vanhimpia osaväestöjä heti eteläsuomalaisten jälkeen. Melko vanhalta näyttävät myös länsilappilaiset (ESN). Nuorimpia osaväestöjä ovat pohjois- ja eteläkainuulaiset.

**Table 3. Properties of LD Maps**

Population	Length of LD Map (LDU)	LD Holes (number)	LD Holes (kb span)
CEU	939.7	70	1239
SWE	778.1	32	1536
HEL	606.8	10	580
ESS <sup>a</sup>	743.3	23	1178
ESW1 <sup>a</sup>	524.2	4	535
ESW2 <sup>a</sup>	527.4	5	684
ESN <sup>a</sup>	560.0	9	390
LSW <sup>b</sup>	686.9	30	732
LSC <sup>b</sup>	693.2	45	1141
LSN <sup>b</sup>	529.6	4	369
ISS <sup>c</sup>	443.9	5	817
ISC <sup>c</sup>	393.1	1	338
ISN <sup>c</sup>	494.7	6	384

LD holes are defined as a gap of >2.5 LDU between adjacent markers.

Population abbreviations are the same as those used in Figure 1.

<sup>a</sup> Early settlement.

<sup>b</sup> Late Settlement.

<sup>c</sup> Isolate.

## Homotsygoottisuus

Homotsygoottisuus kuvaa tilannetta, jossa locusten alleelit vastinkromosomeissa ovat samanlaiset, eli molemmilta vanhemmilta peritään sama muoto. Tässä taulukossa homotsygoottisuus on keskimäärin (*mean*) korkein I-ryhmällä, sitten L-ryhmällä ja pienin E-ryhmällä, sillä erotuksella että aivan kuten diversiteetinkin kohdalla, vaihtavat länsilappilaiset (ESN) ja eteläpohjalaiset (LSW) paikkaa. Mediaanin kohdalla länsilappilaiset ja eteläpohjalaiset ovat samalla tasolla, kun taas pohjoiset pohjoispohjalaiset putoavat L-ryhmän tasolle.

Population	Genome homozygosity (%)			
	Mean	Median	Min	Max
SWE	0.59	0.45	0.08	6.09
HEL	1.17	0.92	0.25	7.08
ESS	0.89	0.82	0.21	3.12
ESW1	1.10	1.00	0.27	3.22
ESW2	1.18	1.14	0.34	3.36
ESN	1.23	1.05	0.31	5.51
LSW	1.16	1.05	0.43	2.82
LSC	1.37	1.30	0.49	2.95
LSN	1.34	1.21	0.19	4.48
ISS	1.82	1.63	0.61	6.38
ISC	2.03	1.85	0.41	5.96
ISN	1.59	1.33	0.52	3.94

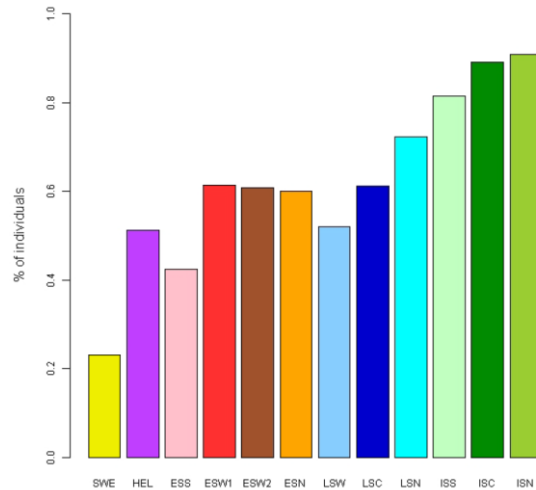
## ROH-segmentit

ROH (*region/run of homozygosity*) on kokonainen homotsygoottinen jakso. Lyhyempien ROH-jaksojen (0,5–4 Mb) runsaus kertoo eristyneestä endogaamisesta väestöstä (puoliso otetaan oman ryhmän sisältä), ja erittäin pitkien ROH-jaksojen (> 10 Mb) runsas esiintyminen kertoo sisäsiittoisuudesta (puolisot toisia tai ensimmäisiä serkkuja). Suomessa serkusavioliittoja ei ole suosittu, joten sisäsiittoisuus on harvinaista.

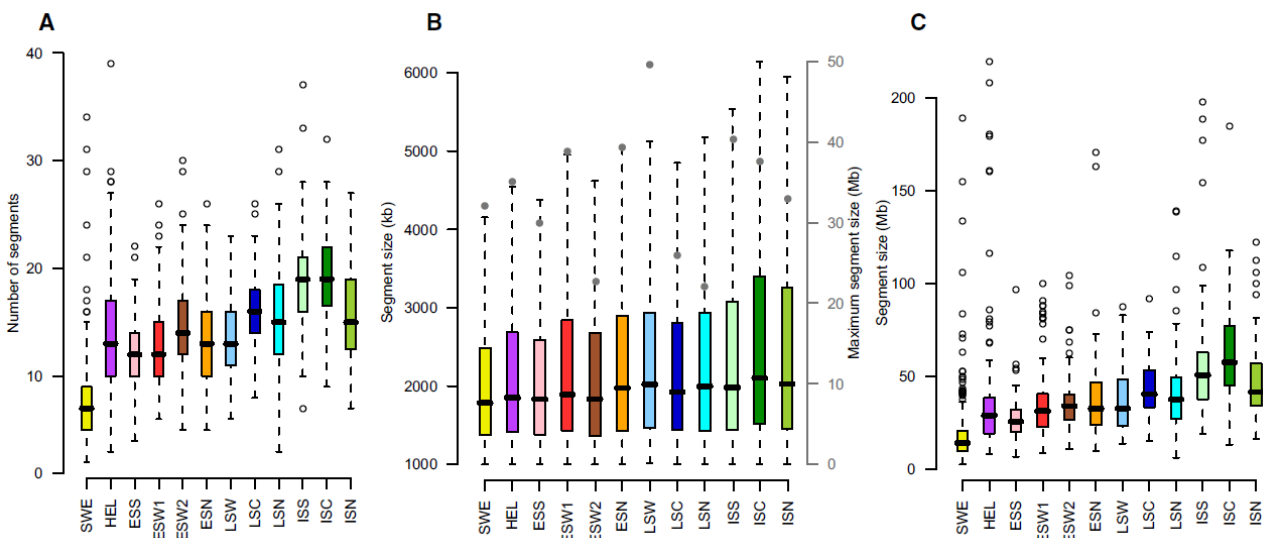
Seuraavassa kuvassa esitetään prosentteina yksilöt, joilla on ainakin yksi ROH > 5 Mb. Pienin osuus on eteläsuomalaisilla (ESS) ja eteläpohjalaisilla (LSW), pienehkö eteläisillä (ESW1) ja pohjoisilla pohjoispohjalaisilla (ESW2), länsilappilaisilla (ESN) ja savolaisilla (LSC), suurehko keski-



lappilaisilla (LSN) ja eteläkainuulaisilla (ISS) ja suuri osuus pohjoiskainuulaisilla (ISC) ja itälappilaisilla (ISN).



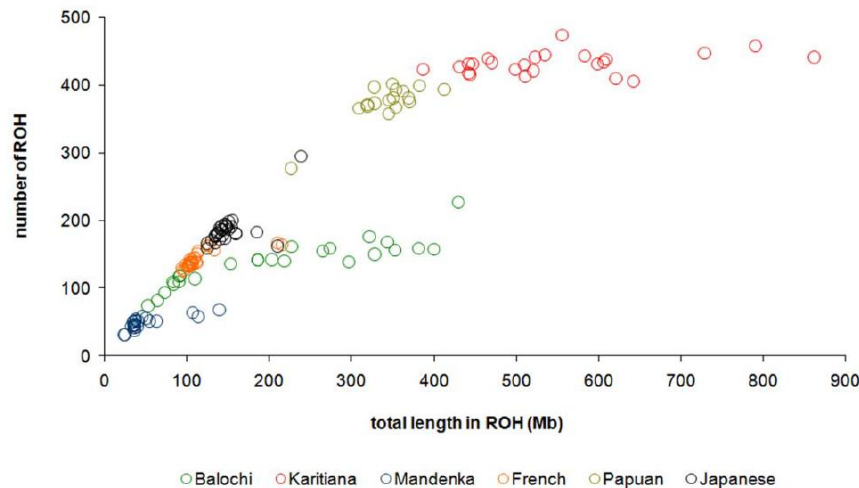
Kuvassa A on esitetty yli 1 Mb:n pituisten ROH-segmenttien lukumäärä osaväestöittäin; mitä pidempi palkki, sitä enemmän yksilöiden välistä vaihtelua. Vaakalinja merkitsee mediaania. Kainuulaiset (ISS, ISC) ovat selvästi homotsygoottisimpia, savolaisten (LSC) tullessa seuraavina ja keski- (LSN) ja itälappilaisten (ISN) sen jälkeen. Pohjoiset pohjoispohjalaiset (ESW2) kohoavat myös hieman, kun taas eteläpohjalaiset (LSW), länsilappilaiset (ESN), eteläiset eteläpohjalaiset (ESW1) ja eteläsuomalaiset (ESS) ovat lähellä toisiaan.



### ROH-koko

Vähentämällä ROH-segmenttien kokonaisuudesta yli 1 Mb:n pituisten ROH-segmenttien osuus saadaan selville alle 1 Mb:n pituisten ROH-segmenttien suhteellinen osuus. Tällaisten lyhyiden ROH-segmenttien osuuden on havaittu olevan sitä pienempi mitä suurempi miespuolisen väestön efektiivinen koko (~ isälinjojen diversiteetti) on ollut, ja näiden osuus myös kasvaa Afrikan koillis-

rajan tuntumasta kauemmas edettäessä (Kirin et al. 2010). Keskimäärin pidempien ROH-segmenttien läsnäolo viittaa geenipoolin pienuuteen ja ääritapauksissa sisäsiittoisuuteen.



Kuvaajassa diagonaalista oikealle venyvillä väestöillä on keskimäärin paljon pidempiä ROH-segmenttejä, ja niissä on myös runsaasti sukulaisavioliittoja (Etelä-Amerikan karitianat, Pakistanin balochit eli belutšit).

Valitettavasti Jakkulan tutkimuksessa ROH-segmentin alarajaksi on määritelty 1 Mb, joten tätä lyhyempien segmenttien määrää ei kuvaajista voi päätellä. Sen sijaan lukumäärän ja ROH-jaksojen kokonaispituuden suhde eli mediaanipituus näkyy suoraan edellä olleesta kuvasta B.

A-kuvan perusteella pisteet jaetaan seuraavasti (pienimmästä arvosta isoimmat pisteet):

1 piste: ISS, ISC; 2 pistettä: LSC, LSN, ISN; 3 pistettä: ESW2, ESN, LSW; 4 pistettä: ESS, ESW1.

B-kuvan perusteella pisteet jaetaan seuraavasti (pienimmästä arvosta isoimmat pisteet):

1 piste: ISC, ISN; 2 pistettä: ESN, LSW, LSN, ISS; 3 pistettä: ESW1, LSC; 4 pistettä: ESS, ESW2.

Yhteensä tästä tasosta tulee pisteitä edellisten keskiarvon mukaan alaspäin pyöristäen:

1 piste: ISC, ISS, ISN; 2 pistettä: LSN, LSC, ESN, LSW; 3 pistettä: ESW1, ESW2; 4 pistettä: ESS.

Tämä taso liittyy väestön kokoon ja avoimuuteen: savolaiset (LSC) ja pohjoiset pohjoispohjalaiset (ESW2) muistuttavat papualaisia siinä, että heillä on paljon ROH-segmenttejä mutta ne ovat melko lyhyitä. Tällainen viittaa väestöhistoriaan, jossa perustajia on ollut suhteellisen vähän mutta he ovat asuneet hajallaan ja puoliso on todennäköisesti otettu kauempaa, oman kylän tai ryhmän ulkopuolelta (eksogamia).

Pohjoiskainuulaisilla (ISC) ROH-segmenttejä on paljon ja ne ovat pitkiä, mikä viittaa pieneen perustajaväestöön joka on joko ollut jo valmiiksi homogeeninen tai sitten puoliso on otettu läheltä oman ryhmän piiristä (endogamia). Eteläkainuulaisilla (ISS) tilanne on muuten samanlainen, mutta ROH-segmentit ovat hieman lyhyempiä.

Eteläsuomalaisilla (ESS) ja eteläisillä eteläpohjalaisilla (ESW1) ROH-segmenttejä on vähän ja ne ovat lyhyitä, mikä viittaa siihen että perustajaväestö on ollut suhteellisen suuri ja heterogeeninen, eikä puolisoa ole otettu läheltä.

Eteläpohjalaisilla (LSW) ja länsilappilaisilla (ESN) ROH-segmenttejä on vähän ja ne ovat pitkiä, mikä viittaa siihen että perustajaväestö on ollut suurehko mutta eristynyt, ja puoliso on mahdollisesti otettu läheltä. Keski- (LSN) ja itälappilaisilla (ISN) ROH-segmenttejä on kohtalaisesti ja ne ovat kohtalaisen pitkiä.

## Yhteenveto

Käytän seuraavassa taulukossa jokaisen tekijän kohdalla neliportaista asteikkoa; joidenkin tekijöiden kohdalla luku on keskiarvo (esimerkiksi homotsygoottisuuden kohdalla *mean + median* : 2).

	diversiteetti (C)	LD r <sup>2</sup> (B, D)	LD map (A)	homotsygotia (B, C, D)	ROH (määrä) (B)	ROH (pituus) (B)	ROH-koko (D)
eteläs. (ESS)	suuri 4	pienellä 3	suuri 4	pieni 4	pieni 4	pieni 4	suuri 4
et.pohj. (ESW1)	suurehko 3	pieni 4	pienellä 2	pieni 4	pieni 4	pienellä 3	suuri 4
po.pohj (ESW2)	suurehko 3	pieni 4	pienellä 2	pienellä 3	pienellä 3	pienellä 3	suurehko 3
lä.lappil. (ESN)	suurehko 3	suurehko 2	suurehko 3	pienellä 3	pienellä 3	pienellä 3	suurehko 3
eteläpoh. (LSW)	suuri 4	suurehko 2	suuri 4	pienellä 3	pienellä 3	pieni 4	pienellä 2
savol. (LSC)	pienellä 2	suuri 1	suuri 4	suuri 1	suurehko 2	pienellä 3	pienellä 2
ke.lappil. (LSN)	suurehko 3	suurehko 2	pienellä 2	suurehko 2	suurehko 2	suurehko 2	pienellä 2
et.kainuul. (ISS)	pieni 1	suurehko 2	pieni 1	suuri 1	suuri 1	suurehko 2	pieni 1
po.kainuul. (ISC)	pieni 1	suurehko 2	pieni 1	suuri 1	suuri 1	suuri 1	pieni 1
itälappil. (ISN)	pienellä 2	suuri 1	pienellä 2	suuri 1	suurehko 2	suuri 1	pieni 1

A. vanhuus–nuoruus: LD map length

B. eristyneisyys–avoimuus: LD r<sup>2</sup>, ROH, homotsygoottisuus

C. yksijuurisuuus–monijuurisuuus: diversiteetti, homotsygoottisuus

D. suuruus–pienuus: ROH-koko (varhainen koko), LD r<sup>2</sup>, homotsygoottisuus (myöhäinen koko)

E. säilyminen–ajautuminen: IBS vs. FST, diversiteetti, (MAF, yksimuotoiset markkerit)

Seuraavassa taulukossa kuhunkin väestörakenteelliseen ominaisuuteen vaikuttavien tekijöiden summa on jaettu tekijöiden määrällä, jotta maksimipisteet olisivat aina 4.

	A. Vanhuus	B. Avoimuus	C. Monijuurisuuus	D. Koko (perustajat)	E. Säilyneisyys
eteläsuom. (ESS)	4	4–	4	3,5 (4)	4
et.p-pohj. (ESW1)	2	4–	3,5	4 (4)	3
po.p-pohj (ESW2)	2	3+	3	3,5 (3)	3
länsilappil. (ESN)	3	3–	3	2,5 (3)	3
eteläpohj. (LSW)	4	3	3,5	2,5 (2)	4
savolaiset (LSC)	4	2–	1,5	1 (2)	2
keskilappil. (LSN)	2	2	2,5	2 (2)	3
et.kainuul. (ISS)	1	1,5	1	1,5 (1)	1
po.kainuul. (ISC)	1	1+	1	1,5 (1)	1
itälappil. (ISN)	2	1+	1,5	1 (1)	2

- Eteläisimmät osaväestöt (eteläsuomalaiset, eteläpohjalaiset ja savolaiset) näyttävät olevan vanhimpia, pohjoispohjalaiset ja lappilaiset keskivanhoja ja kainuulaiset osaväestöt nuorimpia.
- Vähiten eristyksissä ovat olleet eteläsuomalaiset, pohjoispohjalaiset, eteläpohjalaiset ja länsilappilaiset. Eristyneempiä ovat olleet savolaiset ja keskilappilaiset, ja kaikkein eristyneimpiä itälappilaiset ja kainuulaiset.
- Monijuurisimpia tai varhain sekoittuneita osaväestöjä ovat eteläsuomalaiset, eteläpohjalaiset, pohjoispohjalaiset ja länsilappilaiset, sitten keskilappilaiset; vähemmässä määrin savolaiset ja itälappilaiset ja kaikkein vähiten kainuulaiset.
- Suurimpia osaväestöjä ovat olleet pohjoispohjalaiset ja eteläsuomalaiset, keskisuuria eteläpohjalaiset ja länsi- ja keskilappilaiset. Pienempiä ovat olleet kainuulaiset ja kaikkein pienimpiä savolaiset ja itälappilaiset.

- Eteläsuomalaisilla ja eteläisillä pohjoispohjalaisilla näyttäisi olleen suurin perustajaväestö; seuraavina tulevat pohjoiset pohjoispohjalaiset ja länsilappilaiset. Pienin perustajaväestö näyttää olleen I-ryhmän isolaateilla.
- Eniten ajautuneita osaväestöjä näyttävät olevan kainuulaiset, sitten savolaiset ja itälappilaiset. Vähemmän ajautuneita ovat länsi- ja keskilappilaiset sekä pohjoispohjalaiset, ja kaikkein vähiten ajautuneita eteläsuomalaiset ja eteläpohjalaiset.

### ***Ryhmittely osaväestön rakenteen perusteella***

Jos aineistossa on osaväestöjä, jotka muistuttavat suuresti toisiaan, saattaa niillä olla asutushistoriallinen yhteys. Selvä jako näyttää menevän viiden läntisen (ESS, ESW1, ESW2, ESN, LSW) ja viiden itäisen osaväestön välissä (LSC, LSN, ISS, ISC, ISN); ainoat poikkeamat ovat savolaisten vanhuus ja keskilappilaisten säilyneisyys. Väestörakenteellisesti siis suomalaisten sisäinen itä-länsi-jako näyttää ulottuvan Lappiin saakka.

Itäsuomalaiset voidaan edelleen jakaa kahteen ryhmään: keskiseen (LSC, LSN) ja itäiseen (ISS, ISC, ISN), vaikka itälappilaiset ovatkin yhtä lähellä myös keskistä ryhmää. Länsisuomalaisia ei sen sijaan voi jakaa kahteen ryhmään; kolmeen ryhmään heidät voisi jakaa, mutta ryhmät eivät olisi toisensa poissulkevia vaan ne olisivat lomittain.

## **4. Asutushistoria**

### **4.1. Suomalaisten väestöjen geenipohja**

Suomalaisten geenipohja on pääosin pohjoiseurooppalainen, vaikka pieni mutta selvä itäinen juurikin on havaittavissa. Erityisesti itäsuomalaisissa voi olla saamelaista perimää, joka IBS-etäisyyksien perusteella näyttää poikenneen muista eurooppalaisista (myös venäläisistä) eräiden itä- ja keski-aasialaisten väestöjen suuntaan. Toistaiseksi on mahdoton sanoa, juontuuko tämä saamelaisten itäisyys jo Koillis-Euroopan alkuausuttajista (jolloin se edustaisi Suomen varhaisimpia asuttajia) vai onko se vasta myöhäisempien väestöliikkeiden seurausta (kuten eräät itäiset äitilinjat; ks. Ingman & Gyllensten 2007), jolloin Suomen alkuausuttajat olisivat muistuttaneet enemmän muita pohjoiseurooppalaisia.

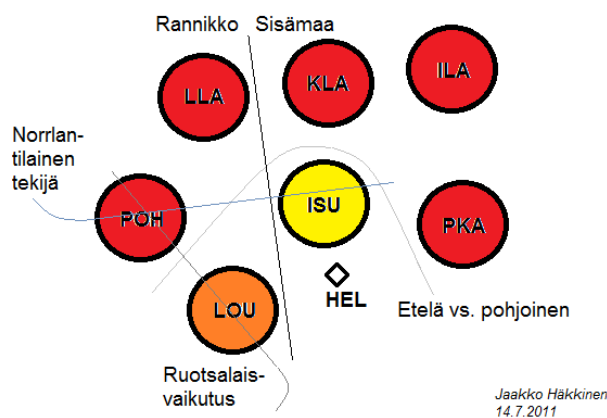
FST-etäisyyksien perusteella suomalaisten lähimpiä geneettisiä sukulaisia ovat virolaiset, ruotsalaiset ja muut Itämeren alueen kansat (venäläiset, puolalaiset ja saksalaiset). Länsisuomalaiset ovat selvästi lähempänä sekä ruotsalaisia että venäläisiä kuin itäsuomalaiset, mikä johtunee itäsuomalaisten geneettisestä ajautumisesta (tai sekoittumisesta saamelaisiin), mutta toisaalta se voi johtua osittain myös merenkulun mahdollistamasta Itämeren alueen rantaväestöjen (joihin länsisuomalaiset kuuluvat) keskinäisestä sekoittumisesta. Suomalaisten etäisyydet baltteihin ovat yllättävän pitkiä – tämä saattaa johtua pienten pohjoisten väestöjen ajautumisesta eri suuntiin, mutta geenipohja voi olla alun alkaenkin erilainen. Oikean vaihtoehdon osoittaisi vain näiden väestöjen välinen IBS-analyysi.

Joka tapauksessa länsi- ja itäsuomalaisten IBS-arvot suhteessa ruotsalaisiin, saksalaisiin ja britteihin sekä yhteisten yksimuotoisten markkerien lukumäärä osoittavat, että pohjoiseurooppalai-

silla väestöillä (vs. keskieurooppalaiset) on suurelta osin yhteinen geenipohja ja vanha yhteinen väestöhistoria (yksimuotoiset markkerit). Itä-länsi-suunnassa läntisten väestöjen näkökulmasta syvin raja kulkee länsi- ja itäsuomalaisten välissä (IBS), mutta länsi- ja itäsuomalaisten näkökulmasta suomalaisten ja läntisempien väestöjen välissä. Yksimuotoisten markkerien perusteella länsi- ja itäsuomalaiset edustavat yhdessä ruotsalaisten kanssa samaa väestöpohjaa, saksalaisten erotessa jo selvemmin näistä.

Väestörakenteellisin perustein suomalaisväestöt voidaan jakaa läntiseen (rannikon) ja itäiseen (sisämaan) ryhmään, mutta osaväestöjen IBS-suhde helsinkiläiseen verrokkiväestöön antaa perusteita kolmijakoon: eniten yhteistä geenipohjaa helsinkiläisten kanssa on itäsuomalaisilla (savolaiset, eteläkainuulaiset, eteläiset pohjoispohjalaiset), seuraavaksi eteläsuomalaisilla (sis. lounaissuomalaisen väestön eli hämäläiset, varsinaissuomalaiset, satakuntalaiset) ja vähiten läntis-pohjoisilla väestöillä (eteläpohjalaiset, pohjoiskainuulaiset, länsilappilaiset, keskilappilaiset ja itälappilaiset).

### Suomalaiset väestöt (IBS)



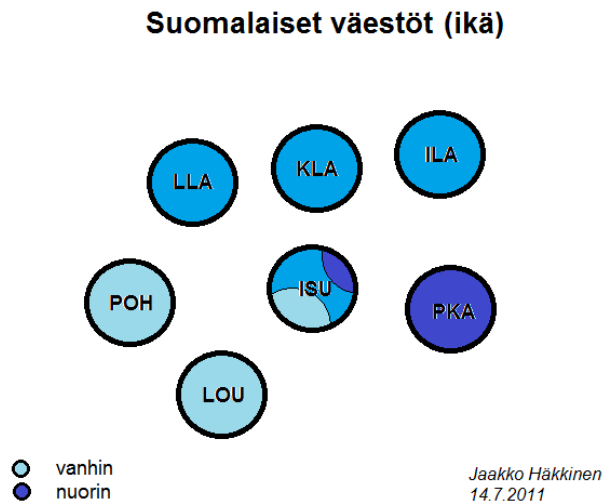
Mielenkiintoista kyllä, tässä tasossa selvin raja (vaaleanharmaa käyrä) erottaa eteläiset väestöt (itäsuomalaisen väestön taustalla oleva savolaisekspansio alkoi Etelä-Savosta) pohjoisista.

Jako rannikkoon ja sisämaahan (musta suora) kuvastaa pikemminkin asutushistoriallisten mekanismien ja olosuhteiden samankaltaisuutta kuin yhteistä geenipohjaa. IBS-arvot kuvastavat todellista yhteistä geenipohjaa mutta vain suhteessa helsinkiläisiin; toivottavasti kuva tarkentuu tulevien tutkimusten avulla myös suomalaisten väestöjen keskinäisillä IBS-arvoilla. Toistaiseksi suomalaisväestöjen keskinäinen samankaltaisuus ja erilaisuus perustuvat FST-etäisyyksiin, jotka ovat alttiita ajautumiselle.

Ruotsinkieliset pohjalaiset, varsinaissuomalaiset ja vielä satakuntalaisetkin ovat selvästi lähempänä ruotsalaisia kuin eteläpohjalaiset ja hämäläiset (tummanharmaa suora). Silti ruotsalaisvaikutus on niin heikkoa, ettei se selitä suomalaisten väestöjen ja väestöryhmien eroja. Norlantilainen tekijä erottaa ruotsinkieliset pohjalaiset, pohjoispohjalaiset ja kainuulaiset eteläisemmistä osaväestöistä (siniharmaa suora), mutta tämäkin tekijä on heikko eikä selitä suomalaisten väestöjen eroja. Tekijän läsnäolo pohjoiskainuulaisissa ja lappilaisissa on selvittämättä.

## 4.2. Suomalaisen väestön historia

Suomalaiset väestöt ja osaväestöt ovat ”syntyneet” suhteellisen myöhään, viimeisten vuosikymmentien kuluessa, perustajavaikutuksen tai muunlaisten pullonkaulojen, geneettisen ajautumisen sekä naapurialueilta mahdollisesti tulleiden geenivirtojen vaikutuksesta. Vanhimmat suomalaisväestöt (ne joiden ”perustamisesta” tai viimeisestä pullonkaulasta on eniten aikaa) sijaitsevat etelässä ja lännessä: eteläsuomalaiset (lounaissuomalaiset), eteläpohjalaiset ja savolaiset. Seuraavassa vaiheessa näyttävät syntyneen pohjoispohjalaiset sekä länsi-, keski- ja itälappilaiset väestöt, ja viimeisessä vaiheessa kainuulaiset.



Lapin suomalaisväestöt näyttävät syntyneen yllättävän varhain ja nopeasti, kun taas itäsuomalaiset osaväestöt ovat syntyneet pidemmän ajan kuluessa. Ero voi johtua siitä, että ainakin sisämaan lappilaisissa on melko vähän suomalaista vaikutusta, koska he ovat pääasiassa suomalaistuneita saamelaisia (Enbuske 2008). Niinpä väestöt oli jo ”perustettu” ennen Lapin suomalaistumista. Itäsuomalaiset osaväestöt puolestaan ovat suuremmissa määrin suomalaisasutuksen leviämisen seurausta.

Vanhimpien väestöjen (savolaiset ovat itäsuomalaisten ”kantamuoto”) kohdalla väestörakenteellisten tekijöiden vertailulla ei saada näkyviin mahdollista sekasyntyisyyttä, vaikka länsilappilaisissa ainakin voisi olettaa jonkinasteisen saamelaisen juuren. Tämä selviää vasta kun saamelaisia ja suomalaisväestöjä verrataan genominlaajuisesti. Savolaisten vähäinen avoimuus ja monijuurisuuden puute viittaa siihen, että mahdollinen saamelaisvaikutus edeltää ajallisesti väestön ”syntyä”.

Pohjoispohjalaiset näyttävät muodostavan yhden ainoan osaväestön, niin pieniä ovat eteläisten ja pohjoisten pohjoispohjalaisten väliset rakenteelliset erot. Kuitenkin IBS-suhde helsinkiläisiin on näillä osaväestöillä erilainen, mikä kertoo hieman erilaisista lähtöaluepainotuksista. Väestörakenteellisesti pohjoispohjalaiset voitaisiin sujuvasti selittää eteläsuomalaisten pohjalta, ja eteläiset pohjoispohjalaiset ovat myös IBS-arvoissaan helsinkiläisistä yhtä kaukana kuin eteläsuomalaiset. Kuitenkin pohjoispohjalaisten itäsuomalaisuus sekä avoimuus ja monijuurisuus suhteessa väestön nuoruuteen selittyvät monilähtöisyydellä: Pohjois-Pohjanmaan varhaisemman lounaissuomalaisen asutuksen päälle on tullut karjalaista ja paikoin jopa savolaista asutusta.

Keskilappilaisetkin ryhmittyvät väestörakenteellisesti itäsuomalaisiksi, mutta heidän avoimuutensa ja monijuurisuutensa on suurempaa kuin itäsuomalaisten ”kantaväestöllä” savolaisilla,

joten hekin näyttävät olevan sekoitus: ehkä länsilappilaisten ja savo-karjalaisten sekoitus. Tähän viittaisi heidän lyhyt FST-etäisyytensä itäsuomalaisiin.

Itälappilaiset muistuttavat väestörakenteellisesti monessa suhteessa savolaisia, ja sama koskee nuorimpia osaväestöjä etelä- ja pohjoiskainuulaisia. Rakenteellisesti etelä- ja pohjoiskainuulaiset näyttävät olevan hyvin samankaltaisia, vaikka ne nykyään eroavatkin geneettisesti toisistaan niin paljon että ne on syytä laskea eri väestöihin kuuluviksi. Suurelta osin tämä johtuu pohjoiskainuulaisten ajautumisesta kauemmas itäsuomalaisista (joihin eteläkainuulaiset kuuluvat), mutta IBS-etäisyys helsinkiläisiin sijoittaa pohjoiskainuulaiset ja itälappilaiset pohjoiseen väestöpohjaan ja eteläkainuulaiset eteläiseen väestöpohjaan.

Tämä osoittaa, että pienissä, näennäisen yksijuurisissa ja eristyneissäkin suomalaisväestöissä on säilynyt alueen vanhempaa, suomalaisasutusta edeltävää geenipohjaa. Suomalaisasutus ei siis ole levinnyt tyhjiössä, vaan se on eri alueilla sulauttanut itseensä aikaisempaa väestöä: itäsuomalaiset yleensä mahdollisesti jo varhain saamelaisista väestöä ja pohjoiskainuulaiset pohjoista geenipohjaa edustavaa väestöä. Lappilaiset kokonaisuudessaan ovat todennäköisesti suuressa määrin suomalais-tuneita saamelaisia, koska väestöt ovat selvästi vanhempia ja geneettisesti etäisempiä toisistaan kuin olisi suomalaisen asutushistorian perusteella odotettavissa.

Saamelaisten ja suomalaisväestöjen suhteet sekä IBS-tasolla että FST-etäisyyksinä selviävät toivottavasti tulevissa tutkimuksissa. Pidän todennäköisenä, että havaittu pohjoisten ja eteläisten väestöjen välinen raja liittyy geneettiseen saamelaisvaikutukseen. Sekoittumisen tarkkaa astetta ei kuitenkaan ole mahdollista arvioida vielä nykytiedon valossa.

## Lähteet

Angius et al. 2008: Patterns of Linkage Disequilibrium between SNPs in a Sardinian Population Isolate and the Selection of Markers for Association Studies. – *Human Heredity* 65.

<http://www.ncbi.nlm.nih.gov/pubmed/17652959>

Atzmon et al. 2010: Abraham's Children in the Genome Era: Major Jewish Diaspora Populations Comprise Distinct Genetic Clusters with Shared Middle Eastern Ancestry Selection of Markers for Association Studies. – *American Journal of Human Genetics* 86.

<http://www.cell.com/AJHG/abstract/S0002-9297%2810%2900246-6>

Enbuske, Matti 2008: *Vanhan Lapin valtamailla*. Bibliotheca Historica 113. Helsinki: Suomalaisen Kirjallisuuden Seura.

Helgason et al. 2009: Sequences From First Settlers Reveal Rapid Evolution in Icelandic mtDNA Pool. – *PLoS Genetics* 1 / 2009.

<http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000343>

Huyghe et al. 2011: A genome-wide analysis of population structure in the Finnish Saami with implications for genetic association studies. – *European Journal of Human Genetics* 3 / 2009.

<http://www.ncbi.nlm.nih.gov/pubmed/21150888>

Ingman & Gyllensten 2007: A recent genetic link between Sami and the Volga-Ural region of Russia. – *European Journal of Human Genetics* 15.

<http://www.nature.com/ejhg/journal/v15/n1/abs/5201712a.html>

Jakkula et al. 2008: The Genome-wide Patterns of Variation Expose Significant Substructure in a Founder Population. – *American Journal of Human Genetics* 83.

<http://www.cell.com/AJHG/abstract/S0002-9297%2808%2900590-9>

Kirin et al. 2010: Genomic Runs of Homozygosity Record Population History and Consanguinity. – *PLoS ONE* 11 / 2010.

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0013996>

Nelis et al. 2009: Genetic Structure of Europeans: A View from the North-East. – *PLoS ONE* 5 / 2009.

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0005472>

Salmela et al. 2008: Genome-Wide Analysis of Single Nucleotide Polymorphisms Uncovers Population Structure in Northern Europe. – *PLoS ONE* 10 / 2008.

<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0003519>

Salmela et al. 2011: Swedish Population Substructure Revealed by Genome-Wide Single Nucleotide Polymorphism Data. – *PLoS ONE* 2 / 2011.

<http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0016747>