

Myytti suomalaisten sisäsiittoisuudesta

Jaakko Häkkinen, 4.11.2013

Tavallisen kansan tietoisuuteen on jostain syystä iskosnutun käsitys, että suomalaiset olisivat jotenkin sisäsiittoista kansaa. Tässä kirjoituksessa osoitan geenitutkimusten tulosten perusteella, että tämä käsitys on täysin virheellinen.

Sisäsiittoisuus merkitsee tilannetta, jossa henkilön vanhemmat ovat sukua toisilleen. Tietysti kaikki meistä ovat sukua toisilleen jonkun esivanhemman kautta, kun mennään tarpeeksi kauas menneisyyteen, ja tietyllä alueella pitkään asuneilla suvuilla on varmasti yhteisiä esivanhempia. Tämä koskee kuitenkin aivan kaikkia maapallon kansoja ja ihmisiä, koska avioliittokenttä on aina käytännön syistä rajoittunut omaan lähialueeseen, joten syvemmän tason sukulaisuutta ei ole mielekästä käsitellä sisäsiittoisuutena. Käytännössä sisäsiittoisuudella tarkoitetaan ns. haitallista sisäsiittoisuutta eli tilannetta, jossa henkilön vanhemmat ovat läheistä sukua toisilleen – esimerkiksi ensimmäisiä serkkuja (yhteiset isovanhemmat).

Sisäsiittoisuus on mahdollista nähdä geneeistä, ja seuraavassa käsitelen suomalaisia erilaisten geenitasojen valossa.

1. Genetiikan menetelmät

Modernissa geenitutkimuksessa ns. genomilajainen tarkastelutaso kattaa jopa miljoona yksittäistä emästä sellaisista kohdista genomia, joissa ihmisväestöissä esiintyy eroja. Tarkastelu kattaa 22 autosomiparia eli kaikki muut kromosomit paitsi X- ja Y-kromosomit. Geenitutkimuksessa sisäsiittoisuutta voidaan jäljittää muutamilla eri tavoilla.

1.1. Yleinen homotsygoottisuus

Homotsygoottisuus merkitsee, että molemmilta vanhemmilta on peritty sama alleeli (emäs). Emäsparit GG, AA, TT ja CC ovat homotsygoottisia, kun taas GA, AG, TC ja CT ovat heterotsygoottisia. Yleisimmin mutaatiot tapahtuvat juuri näin pareittain, eli jossakussa henkilössä G on muuttunut A:ksi tai T on muuttunut C:ksi (tai päinvastoin), ja kun tämä mutaatio on sitten periytynyt osalle jälkeläisistä, väestö on muuttunut tämän kohdan suhteen polymorfiseksi eli monimuotoiseksi.

Suomalaisilla ja eurooppalaisilla yleinen homotsygoottisuus on yleensä vähän päälle 70 %, riippuen vähän geenitestistä ja kuinka paljon yksimuotoisia kohtia siihen osuu suhteessa monimuotoisiin kohtiin. Toisin sanoen meillä kaikilla useammassa kuin kahdessa kohdassa kolmesta kautta koko genomia on molemmilta vanhemmilta peritty täsmälleen sama alleeli. Korkeakaan homotsygoottisuus ei kuitenkaan ole var-

ma merkki sisäsiittoisuudesta, koska muutkin prosessit voivat johtaa kohonneeseen homotsygoottisuuteen.

Esimerkiksi geneettinen ajautuminen merkitsee tilannetta, jossa väestö on niin pieni, että sattuman vaikutuksesta harvinaisempia alleeleja katoaa: jossain sukupolvessa tulee tilanne, että kukaan väestössä (G2 eli toinen sukupolvi) ei ole sattunut perimään vanhemmiltaan (G1 eli ensimmäinen sukupolvi) harvinaisempaa alleelia C, vaan kaikilla on molemmilta vanhemmilta peritty yleisempi alleeli T. Näin väestöstä on tullut kyseisen kohdan osalta jälleen yksimuotoinen, eli kaikki ovat homotsygootteja.

Geneettinen ajautuminen: harvinainen alleeli C katoaa

G1: TT -- TC TT -- TT TC -- TT TT -- TT
G2: TT,TT TT,TT,TT TT TT,TT

Genetic drift: rare allele C disappears

Erityisesti hyvin pienissä ja eristyneissä väestöissä homotsygoottisuus voi olla tavallista suurempaa. Suurissa väestöissä harvinaisenkaan alleeli ei katoa yhtä helposti, ja sekoittuneet väestöt ovat luonnollisesti saaneet uusia alleeleja muilta väestöiltä.

On muistettava, että homotsygoottisuuden vaihtelu on suurempaa yksilöiden välillä kuin väestöjen välillä: jos vanhemmat sattuvat edustamaan samaa geenipohjaa, yksilö on homotsygoottisempi kuin sellainen, jonka vanhemmat edustavat eri geenipohjaa eli tulevat saman väestön kaukaisemmista osista.

1.2. ROH

ROH (region/run of homozygosity = homotsygoottinen jakso) merkitsee geenijaksoa, jonka kaikki kohdat ovat homotsygoottisia (molemmilla kromosomin puoliskoissa on sama emäs). ROH katkeaa heterotsygoottiseen kohtaan, jossa siis eri vanhemmilta on peritty eri alleelit. Rekombinaation vuoksi ROHit pilkkoutuvat joka sukupolvi lyhyemmiksi ja ovat pitkiä vain geneettään samanlaisten jälkeläisillä, joten pitkät ROHit (yli 10 Mb) ovatkin osoittautuneet parhaaksi yksilön sisäsiittoisuuden osoittajaksi (Kirin et al. 2010).



Lyhyet ROHit (alle 1,5 Mb) puolestaan kertovat perustajaväestön pienuudesta, ja niitä on sitä enemmän, mitä kauempana Afrikasta väestö on "syntynyt".

1.3. Väestön sisäsiittoisuus F

Väestön sisäsiittoisuus F lasketaan suhteuttamalla väestönsisäinen havaittu heterotsygoottisuus (H_{obs}) odotettuun heterotsygoottisuuteen (H_{exp}). Odotettu heterotsygoottisuus tarkoittaa alleelifrekvenssien perusteella laskettua tilannetta: jos alleelin A frekvenssi väestössä on 65 % ja alleelin a frekvenssi 35 %, on heterotsygootteja $Aa + aA$ odotuksenmukaisesti $0,65 \times 0,35 \times 2 = 0,46$ eli 46 %, jolloin homotsygootteja on 54 % ($AA: 0,65 \times 0,65 = 42$ % ja $aa: 0,35 \times 0,35 = 12$ %).

Jos väestössä on tilastollisesti odotettua enemmän homotsygoottisuutta, esimerkissä vaikkapa 60 %,

johtuu se pääsääntöisesti siitä, että osa väestöön kuuluvista yksilöistä on lisääntynyt sukulaistensa kanssa. Kaavalla $F = (H_{exp} - H_{obs}) / H_{exp}$ saadaan väestön sisäsiittoisuustekijäksi $(0,46 - 0,40) / 0,46 = 0,13$. Sisäsiittoisuus F on sitä suurempi, mitä suurempi on havaittu homotsygoottisuus suhteessa odotettuun; tai kääntäen, mitä pienempi on havaittu heterotsygoottisuus suhteessa odotettuun.

Tutkimuksessa Salmela et al. (2008) on esitetty länsisuomalaisien (FIW), itäsuomalaisien (FIE), ruotsalaisien (SWE), pohjoissaksalaisien (GER) ja englantilaisten (BRI) heterotsygoottisuus (= $1 - \text{homotsygoottisuus}$) ja sisäsiittoisuus F :

population	abbreviation	n	sample het1	marker het1	MAF2	F3
Western Finland	FIW	141	0,2783	0,3050	0,1929	0,0077
Eastern Finland	FIE	139	0,2768	0,3022	0,1906	0,0057
Sweden	SWE	113	0,2779	0,3009	0,1927	0,0037
Northern Germany	GER	256	0,2793	0,3047	0,1941	0,0031
Great Britain	BRI	296	0,2799	0,3069	0,1943	0,0005
1 Median heterozygosity						
2 Median minor allele frequency						
3 Median inbreeding coefficient						

- Otoksen heterotsygoottisuus oli pienin (eli homotsygoottisuus oli suurin) itäsuomalaisilla, sitten ruotsalaisilla ja länsisuomalaisilla.
- Markkerin heterotsygoottisuus oli pienin ruotsalaisilla, sitten itäsuomalaisilla ja pohjoissaksalaisilla.
- Harvinaisemman alleelin mediaanifrekvenssi MAF (*minor allele frequency*) oli pienin itäsuomalaisilla, sitten ruotsalaisilla ja länsisuomalaisilla. Pieni keskimääräinen tai mediaaninen MAF kertoo geneettisestä ajautumisesta (ks. 1.1).
- Väestön sisäsiittoisuus F oli suurin länsisuomalaisilla, sitten itäsuomalaisilla ja ruotsalaisilla. Suuruusluokka oli kuitenkin sama kaikilla tutkituilla väestöillä paitsi briteillä, eli käytännössä ero odotetun ja havaitun heterotsygoottisuuden välillä oli erittäin pieni kaikilla väestöillä, edellä ollutta esimerkkiä jatkaen noin 0,2 %: $(0,460 - 0,458) / 0,46 = 0,0043$. Yksikään näistä väestöistä ei siis ole sisäsiittoinen, mikä tietysti johtuu siitä, että lisääntymistä sukulaisten kanssa on täällä päin maailmaa aina vältetty.

Tuoreessa tutkimuksessa pohjoisitalialaisista isolaateista (eristyneistä [osa]väestöistä) on vertailussa mukana myös muita kansoja (Esko et al. 2013). Keskieurooppalaisten väestöjen F on yleisesti alle 0,01, kun taas pohjoisempina se on hieman yli 0,01. Ruotsalaisilla se oli 0,0114, venäläisillä 0,0122 ja helsinkiläisillä 0,0163. Vielä hieman suurempi se oli kasvavassa järjestyksessä virolaisilla, pohjoisitalialaisilla, liettualaisilla, espanjalaisilla, latvialaisilla ja puolalaisilla (0,0194). Yli 0,02 arvo oli tunnetuilla isolaateilla eli kuusamolaisilla, sardinialaisilla ja baskeilla. Neljässä koillisitalialaisessa kylässä F oli peräti yli 0,03. (Tässä tutkimuksessa oli erilainen geenivalikoima kuin aiemmin mainitussa, siksi luvuissa on suuruusluokkaero.)

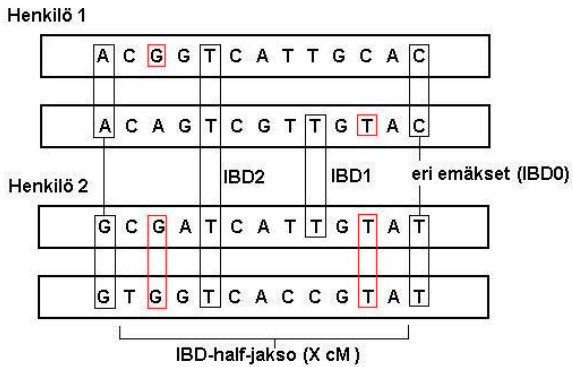
On huomattava, että väestön sisäsiittoisuuteen ei vaikuta se, tarkastellaanko pientä kylää vai suurta pääkaupunkia, johon on tullut ihmisiä eri puolilta, vaiko peräti koko maan kattavaa otosta, koska F lasketaan suhteuttamalla havaittu heterotsygoottisuus kyseisen tarkasteluryhmän *omien* alleelifrekvenssien pohjalta laskettuun odotettuun heterotsygoottisuuteen. Sisäsiittoisuusarvot ovat siis vertailukelpoisia, ja luonnollisesti kaupunkiin muuttaneella on pienempi todennäköisyys osua pariutumaan etäisenkään sukulaisten kanssa kuin suvun kotiseudulla asuvalla, ja tiheään asutussa maassa asuvalla on pienempi todennäköisyys osua pariutumaan etäisenkään sukulaisten kanssa kuin harvaan asutussa maassa asuvalla.

Suomalaiset eivät ole sen sisäsiittoisempia kuin naapurikansatkaan. Pohjoisen harvaan asuttuihin erämaihin asettuneet kuusamolaiset ovat tyypillinen isolaatti ja vastaavat esim. sardinialaisia ja baskeja. Erämaat, meret ja vuoret rajoittavat ihmisten liikkumista ja lisäävät endogamiaa (puoliso otetaan oman ryhmän sisältä).

Tämä on kuitenkin vielä aivan eri asia kuin sisäsiittoisuus, jossa pariudutaan tunnettujen sukulaisten kanssa.

1.4. IBD-half

Tutkimuksessa Henn et al. (2012) on tutkittu sisäsiittoisuutta mm. *IBD-half*-menetelmän avulla. Siinä lasketaan sellaisia emäsjaksoja, joissa kahdella eri henkilöllä on samoissa kohdissa yksi (IBD1) tai kaksi (IBD2) samaa emästä.



Saman väestön yksilöiden jakamat *IBD-half*-jaksot edustavat kolmea eri suuruusluokkaa:

1. yhteinen *IBD-half* enintään kymmeniä centiMorganeja
2. yhteinen *IBD-half* enintään satoja centiMorganeja
3. yhteinen *IBD-half* enintään tuhansia centiMorganeja

Näistä kolmanteen ryhmään kuuluu pieniä ja eristyneitä Etelä-Amerikan väestöjä, joiden keskuudessa sukulaisavioliitot ovat yleisiä eli jotka ovat oikeasti sisäsiittoisia. Toiseen ryhmään kuuluu (suurempia?) Etelä-Aasian ja Etelä-Amerikan väestöjä, joilla sukulaisavioliitot ovat yleisiä, sekä pieniä ja eristyneitä väestöjä kuten papualaiset, beduiinit, pygmit, jakuutit ja orkneyläiset. Ensimmäiseen ryhmään kuuluu pieniä eurooppalaisia ja itäaasialaisia väestöjä.

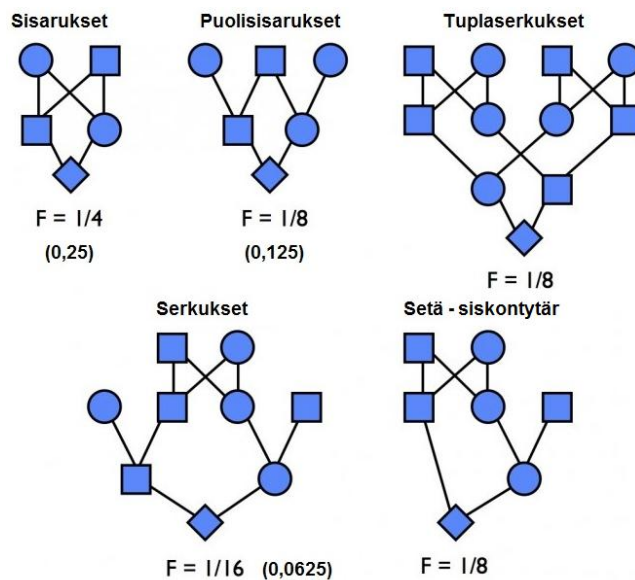
Samaan väestöön kuuluvat yksilöt jakavat keskimäärin yli 10 cM vain neljässä eurooppalaisessa väestössä (suomalaiset 10 cM, sardinialaiset 12,4 cM, islantilaiset 12,8 cM ja orkneyläiset 54,3 cM), kaikkien

muiden tällaisten väestöjen ollessa Aasiasta, Afrikasta tai Amerikasta. Silti *ROHeja* oli näillä neljällä eurooppalaisella väestöllä vain 2–3 % eli saman verran kuin muillakin eurooppalaisilla väestöillä, kun niitä etelä-amerikkalaisilla väestöillä oli 11–22 %. Keskimääräinen *ROHin* pituus oli islantilaisilla 2,09 cM, suomalaisilla 2,16 cM, sardinialaisilla 2,35 cM ja orkneyläisillä 2,40 cM verrattuna eteläamerikkalaisten väestöjen yleiseen 3–4 cM:iin. Monilla eurooppalaisilla väestöillä (esim. ruotsalaisilla, venäläisillä, virolaisilla, tanskalaisilla jne.) *ROHin* keskimääräinen pituus oli suurempi kuin suomalaisilla.

Suomalaisilla pelkästään *IBD-half* on muihin eurooppalaisiin verrattuna korkeampi, joten kyseessä ei ole sisäsiittoisuus, joka näkyy parhaiten *ROHeista* ja väestön laskennallisesta sisäsiittoisuudesta *F*. Suomalaisen korkea *IBD-half* selittyikin perustajavaikutuksesta ja/tai pullonkaulasta: kun geenipoolissa on vähemmän harvinaisia alleleja eli tavallista useampi kohta genomissa on yksimuotoinen (kaikilla väestöön kuuluvilla on sama alleeli), *IBD-half*-jaksot katkeavat harvemmin eli ovat pidempiä.

1.5. Yksilön sisäsiittoisuus *F_{in}*

Yksilön sisäsiittoisuus *F_{in}* lasketaan vanhempien todellisesta sukulaisuusasteesta, eli tätä tasoa ei näe geneistä. Se antaa kuitenkin hyvän vertailukohtan geneistä laskettavalle väestön sisäsiittoisuudelle. Kun yksilö perii puolet geneistään isältään ja puolet äidiltään, on sisaruksilla keskimäärin puolet geneistä yhteisiä.



1.6. *FST*-taso

Väestöjen välistä geneettistä etäisyyttä kuvataan *FST*-arvolla (*fixation index: subpopulation to total population*), joka lasketaan väestön markkerikohtaisista alleelifrekvensseistä ja joka perustuu väestön odotettuun heterotsygoottisuuteen suhteessa kokonaisuuden odotettuun heterotsygoottisuuteen.

Kahden väestön välinen suuri *FST*-arvo merkitsee, että väestöt ovat kehittyneet eri suuntiin vailla keskinäistä geenivirtaa, kun taas pieni *FST*-arvo merkitsee, että väestöjen välillä on ollut geenivirtaa.

Suomalaiset eivät ole yksi väestö, vaan voidaan erottaa ainakin seitsemän suomalaista väestöä, joiden etäisyys toisistaan on vähintään *FST* 20 (0,002 kerrottuna 10 000:lla). Tämä etäisyys on suurempi kuin esimerkiksi ruotsalaisten, saksalaisten ja englantilaisten välillä. Suurin suomalaisväestöjen välinen *FST*-arvo on lähes 70, eikä sille vedä vertoja Euroopan kansojen sisäisistä geenietäisyyksistä kuin sardinialaisten ja pohjoisitalialaisten välinen etäisyys. (Salmela et al. 2008; Jakkula et al. 2008; yhteenveto Häkkinen 2011a.)

IBS-tasolla (*identity by state*; markkerikohtainen alleelien samanlaisuus) itä- ja länsisuomalaiset eivät ole suhteessa muihin eurooppalaisiin niin kaukana toisistaan, mikä osoittaa että geenipohja on suuressa määrin sama, vaikka keskinäinen eristyneisyys ja geenivirta eri suunnilta ovatkin erilaistaneet suomalaisväestöjä.

FST-tasolla näkyy (ainakin IBS-tasoon verrattaessa) erityisesti isolaattien kuten itäsuomalaisten, baskien ja sardinialaisten geneettinen ajautuminen: nämä väestöt ovat alleelifrekvenssien osalta kauimpana muista eurooppalaisista väestöistä. Niiden suuret *FST*-etäisyydet muihin väestöihin merkitsevät, ettei sekoittumista ole tapahtunut. Euroopan keskivyohykkeellä puolestaan väestöt ovat niin lähellä toisiaan, että niitä voidaan pitää oikeastaan väestöjatkumona (*FST* alle 10). Siksi niiden perinnölliset sairaudetkin ovat yhteisiä eivätkä rajoitu vain yhteen kansaan niin kuin eristyneemmillä väestöillä.

Joka tapauksessa suomalainen kansa on laajalla alueella eläneiden pienten väestöjen yhteenliittymä; etelämpänä väestöt vastaavat jossain määrin vanhoja heimoja (lounaissuomalaiset = varsinaissuomalaiset + hämäläiset; pohjalaiset; itäsuomalaiset = savolaiset + karjalaiset?), mutta Lapin kolme suomalaisväestöä heijastanevat jo saamelaisilta perittyä väestörakennetta, koska Lapin suomalaistuminen on niin nuori ilmiö ja on tapahtunut suuressa määrin entisten metsäsaamelaisien suomalaistumisena.

Suomalaiset eivät voi olla sisäsiittoinen väestö, koska suomalaiset eivät ole yksi väestö.

1.7. Perinnölliset sairaudet

Suomalaisten tautiperimää on tutkittu paljon, samoin kuin juutalaisten. Eräs merkittävä syy tähän on näiden väestöjen tietynasteinen sulkeisuus, jopa isolaatio (eris-

tyneisyys): niillä on sellaisia omia sairauksia, joita ei muilla väestöillä ole. Kolikon kääntöpuolena on, että näiltä väestöiltä myös puuttuu tai esiintyy vain harvinaisina useita sellaisia sairauksia, jotka ovat yleisiä muualla Euroopassa. Nimittäin väestön pienuudesta ja eristyneisyydestä johtuva geneettinen ajautuminen voi yhtä hyvin aiheuttaa haitallisten alleelien katoamista kuin niiden yleistymistä; sattuma ratkaisee.

On tärkeää ymmärtää, että tautiperimän rajoittuminen tiettyyn väestöön ei voi todistaa tulevaa geenivirtaa vastaan – se voi todistaa ainoastaan lähtevää geenivirtaa vastaan. Suomalainen tautiperimä ei leviäisi Suomen ulkopuolelle, vaikka tänne muuttaisi miljoonia uutta ihmistä. **Tautiperimä ei siksi voi kertoa mitään suomalaisväestöjen sisäsiittoisuudesta.**

Tarinan opetus

Kattokäsitteen "homogeenisyys" (tässä: väestön yksilöillä usein samat alleelit) alle mahtuvat ainakin seuraavat tasot:

1. Suppea väestöpohja / pieni väestö

Tämä näkyy vähentyneenä geneettisenä diversiteetinä eli kasvaneena yleisenä homotsygoottisuutena (1.1). Lisäksi lyhyiden *ROH*ien (1.2) määrä on korkea ja väestöön kuuluvien keskenään jakamat *IBD-half*-jaksot (1.4) ylittävät 10 cM.

Suomalaiset väestöt kuuluvat tähän luokkaan.

2. Väestön eristyneisyys / endogamia

Tämä näkyy edellisten lisäksi jonkin verran kohonneena väestön sisäsiittoisuutena *F* (1.3), kasvaneena (lähemmäs 100 cM) väestöön kuuluvien jakamina *IBD-half*-jaksojen kokonaisuutena (1.4) sekä korkeina *FST*-etäisyyksinä naapuriväestöihin (ks. 1.6).

Itäsuomalaiset, kuusamolaiset ja itälappilaiset kuuluvat tähän luokkaan (ks. Häkkinen 2011b).

3. Väestön sisäsiittoisuus

Tämä näkyy edellisten lisäksi selvästi kohonneena väestön sisäsiittoisuutena *F* (1.3) sekä pitkien *ROH*ien (1.2) suurempana määränä. Lisäksi väestöön kuuluvilla yhteisten *IBD-half*-jaksojen kokonaispituus on useita satoja tai jopa tuhansia centimorganeita (1.4).

Suomalaiset väestöt eivät kuulu tähän luokkaan.

4. Yksilön sisäsiittoisuus

Tämä varmistetaan sukuselvityksellä (1.5), mutta se näkyy myös erittäin pitkänä *ROH*eina (1.2).

Yksilön sisäsiittoisuus liittyy sukulaisasiittoihin, joita Suomessa on vieroksuttu. Siksi varmaankin erittäin harva suomalainen on sisäsiittoinen.

Tulos: Suomalaiset eivät ole sisäsiittoisia.

Lähteet

Esko, Tõnu et al. 2013: Genetic characterization of northeastern Italian population isolates in the context of broader European genetic diversity. – *European Journal of Human Genetics* 21.

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3658181/>

Henn, Brenna M. et al. 2012: Cryptic Distant Relatives Are Common in Both Isolated and Cosmopolitan Genetic Samples. – *PLoS ONE* volume 7, issue 4.

<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0034267>

Häkkinen 2011a: Seitsemän suomalaista väestöä. (Käsikirjoitus)

<http://www.elisanet.fi/alkupera/Suomalaisvaestot.pdf>

Häkkinen, Jaakko 2011b: Suomalaisväestöjen asutushistoriallinen rakenne. (Käsikirjoitus)

<http://www.elisanet.fi/alkupera/Asutushistoria.pdf>

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

[http://www.cell.com/AJHG/abstract/S0002-9297\(08\)00590-9](http://www.cell.com/AJHG/abstract/S0002-9297(08)00590-9)

Kirin, Mirna et al. 2010: Genomic Runs of Homozygosity Record Population History and Consanguinity. – *PLoS ONE* volume 5, issue 11.

<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0013996>

Salmela, Elina et al. 2008: Genome-Wide Analysis of Single Nucleotide Polymorphisms Uncovers Population Structure in Northern Europe. – *PLoS ONE* volume 3, issue 10.

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0003519>