

# Äitilinjan (mtDNA) selvittämisen menetelmistä

Jaakko Häkkinen (11.1.2020)

Äitilinjansa eli mitokondriaalisen DNA:nsa testanneet ovat huomanneet, että tuloksissa ei päästä yhtä suureen tarkkuuteen kuin isälinjojen kohdalla. Vaikka olisi testannut koko mitokondriaalisen DNA:n (HVR1, HVR2 ja *coding region*), silti voi olla, että nollan ja yhden mutaation päässä on ihmisiä hyvin laajalti Euroopasta, eikä esiäitien ajoittaminen ja paikantaminen onnistu kovinkaan tarkasti.

Tämä johtuu siitä, että mitokondriaalinen DNA on kooltaan hyvin pieni, vain 16 569 emäsparia (*base pair*) pitkä. Esimerkiksi Y-kromosomin koko on peräti 59 miljoonaa emäsparia. Siksi myös mutaatioita tapahtuu mitokondriaalisessa DNA:ssa odotetusti paljon vähemmän. Y-kromosomissa eli isälinjassa on niin suuri tarkkuus, että mutaatiokupolvi lähestyy jo todellista sukupolvea – jos isä ja poika testattaisiin, niin uusin BigY, joka testaa miljoonia SNP:itä (nukleotideja kohdista, joissa esiintyy vaihtelua) ja yhteensä 838 STR-markkera, voisi löytää mutaatioeron jopa heidän välilleen.

Äitilinjoihinkin on mahdollista päästä tarkempaan tulokseen, mutta se vaatii vaivannäköä. Varsinainen selvitystyö vasta alkaa siitä, kun oman geenitestin tulos valmistuu. Oma tärkeä osansa siinä on tietysti perinteisellä sukututkimuksella, mutta siihen en puutu tässä tarkastelussa, vaan keskityn ainoastaan DNA-puoleen.

Havainnollistan vertailumenetelmiä omalla äitilinjänäytteelläni, joka kuuluu alahaploryhmään H5a1e.

## Oma tulos

FamilyTreeDNA:lla oma tulos annetaan suhteessa ihmislajin rekonstruoituun kantaäitiin (RSRS = *Reconstructed Sapiens Reference Sequence*) niin, että haplotyyppissä listataan ainoastaan kaikki erottavat mutaatiot. Osumalista kertoo etäisyyden mutaatioina, testatun henkilön (tässä leikattu poisyksityisyydensuojan vuoksi), varhaisimman tunnetun suoran äitilinjän esiäidin, haploryhmän ja osuman päivämäärän (milloin testi-tulos on valmistunut).

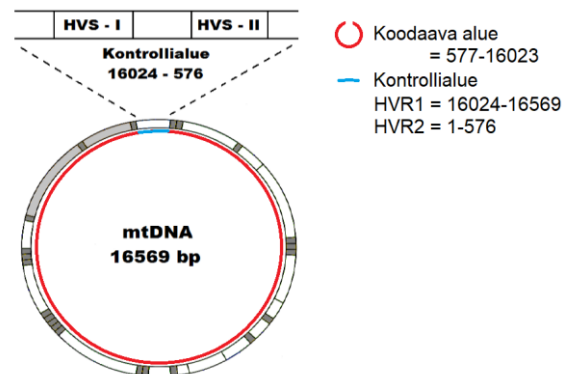
Extra Mutations					315.1C C16519T									
Missing Mutations														
HVR1 DIFFERENCES FROM RSRS					HVR2 DIFFERENCES FROM RSRS					CODING REGION DIFFERENCES FROM RSRS				
A16129G	A16166G	T16187C	C16189T	T16223C	G73A	C146T	C152T	C195T	A247G	A769G	A825t	A1018G	G2706A	A2758G
G16230A	T16278C	T16304C	C16311T	C16519T	315.1C	C456T				C2885T	T3594C	G4104A	T4312C	T4336C
										T7028C	G7146A	T7256C	A7521G	T8468C
										T8655C	G8701A	C9540T	G10398A	T10664C
										A10688G	C10810T	C10873T	C10915T	A11719G
										A11914G	T12705C	G13105A	G13276A	T13506C
										T13650C	T14766C	C15833T		

Osumalista (tiivistetty)			
Genetic Distance	Earliest Known Ancestor	mtDNA Haplogroup	Match Date
0		H5a1e	10/1/2019
0	Lisa Larsdotter b. 1787 d. 1860	H5a1e	5/13/2019
0		H5a1e	12/31/2018
0	Malin Jääskö, b. 1699, d. 1760	H5a1e	10/2/2018

Minulla ei siis ole tapahtunut yhtään sellaista mutaatiota, jota ei olisi myös monilla muilla suomalaisilla ja ruotsalaisilla saman alahaploryhmän edustajilla (etäisyys 0). Tilanne vaikuttaa näin ollen jopa epäsuotuisalta jatkoselvittelyjen kannalta.

Kuitenkaan läheskään kaikki eivät ole testanneet koko mtDNA:ta, joten eroja saattaisi hyvinkin esiintyä laajalla koodaavalla alueella (*coding region*) suhteessa niihin, jotka eivät ole sitä testanneet. Halvin testi kartoittaa vain HVR1:n, ja seuraava HVR1:n ja HVR2:n. Vasta kallein testi kartoittaa näiden lisäksi koodaavan alueen, joka kattaa ylivoimaisesti suurimman osan koko mitokondriaalisesta DNA:sta.



Överst, Sanni 2014: Äitilinjien fylogeneettinen hienorakenne suomalaisilla – metsästäjä-keräilijät vs. maamviljelijät. Pro gradu -tutkielma.

## GenBank: koko mtDNA:n näytteet

GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) on taho, johon tutkijat ja tutkijaryhmät tallentavat näytedataa. Äitilinjojen osalta siellä on jo runsaasti sekä suppeampia pätkiä (esim. HVR1) että koko mtDNA:n mittaisia näytteitä. Ne näyttävät tältä, aina kohtaan 16569 saakka:

```
1 gatcacaggt ctatcacctt attaaccact cacgggagct ctccatgcat ttggtatatt
61 cgtctggggg gtatgcagc gatagcattg cgagacgtg gagcgggagc accctatgct
121 gcagtatctg tcttggattc ctgctctatc ctattatatta tcgcacctac gttcaatatt
181 acagcgaac atacttacta aagtgtgta ataatatatt gctttagga cataataa
241 acaattgaat gtctgcacag ccgctttcca cacagacatc ataacaataa atttccacca
301 aacccccctt cccccgctt ctggccacag cacttaaaca catctctgcc aaaccccaaa
361 aaccaagaac cctaaccaca gcctaaccag atttcaaat ttatcttttg gcggtatgca
421 cttttaacag tcacccccca actaacacat tatttttccc tcccactccc atactactaa
481 tctcatcaat acaacccccg cccatctctac ccagcacaca caccgtctgt aacccctac
541 cccgaaccaa ccaaacccca aagacacccc ccacagtta tgtagcttac ctctcacaag
601 caatacactg aaaatgttta gaggggctca catcacccca taacaataa ggtttggctc
661 tagcctttct attagctctt agtaagatta cacatgcaag catccccgtt ccagtgagtt
721 caccctctaa atcaccacga tcaaaagggg caagcatcaa gcacgcagca atgcagctca
781 aaacgcttag cctagccaca cccccacggg aaacagcagt gattaacct tagcaataa
841 cgaagttta actaagctat actaacccca ggggtgtgca atttctgtcc agccaccgag
901 gtcacacgat taacccaagt caatagaagc cggcgtaaag agtgttttag atcacccctt
961 cccaataaa gctaaaactc acctgagttg taaaaaactc cagttgacac aaatagact
1021 acgaaagtgg ctttaacata tctgaacaca caatagctaa gacccaaact gggtagtagt
1081 accccactat gcttagccct aaacctcaac agttaaatca acaaaactgc tcgccagaac
1141 actacgagcc acagcttaaa actcaaaagga cctggcggtg cttcatatcc ctctagagga
1201 gcctgtctg taatcgataa acccggatca acctcaccac ctcttgctca gctatatac
```

Nukleotidit on ryhmitelty kymmenen nippuihin, eli esim. kohdassa 73 on A-nukleotidi (adeniini). Jos josain haploryhmässä on tässä kohdassa tapahtunut mutaatio, silloin yleisimmin A>G eli A on muuttunut G:ksi (guaniini). Harvinaisemmissa mutaatioissa A>C (sytoosiini) tai A>T (tyymiini). Tämä johtuu siitä, että puriinit muuttuvat helpoimmin toisiksi puriineiksi (A ↔ G) ja pyrimidiinit muuttuvat helpoimmin toisiksi pyrimidiineiksi (T ↔ C). Mutaatioita, jotka tapahtuvat näiden emäsparien välillä, sanotaan transitoioiksi (*transition*). Harvinaisempia mutaatioita, jotka tapahtuvat puriinin ja pyrimidiinin välillä (A ↔ C, A ↔ T, G ↔ C, G ↔ T) sanotaan transversioiksi (*transversion*).

Miten tämä sitten hyödyttää oman äitilinjansa selvittelijää? GenBankin hakukenttään voi kirjoittaa oman (ala)haploryhmänsä. Esimerkiksi minun alahaploryhmäni H5a1e on kuitenkin niin vastikään saanut oman luokituksensa, ettei sitä ole vielä osattu merkitä vuosia tai jopa vuosikymmeniä sitten syötettyihin näytteisiin. Vain yksi tunisialainen(!) näyte löytyy tällä haulla, ja siinäkin on vain HVR1. Kuitenkin haulla H5a1 löytyy jo paljon enemmän näytteitä, ja osa näistä on ainakin Suomessa ja Ruotsissa nimenomaan H5a1e:tä.

Miten sitten oman nuoremman alahaploryhmänsä voi noista bongata? Ensin pitää mennä PhyloTreehen (<https://www.phylotree.org/tree/>) ja oikean haploryhmän sivuille, käyttäen siellä hakua (CTRL+F) halutun haploryhmän paikantamiseksi. Sieltä käy ilmi, että H5a1e:tä luonnehtii mutaatio A16166G. Tämä merkitsee, että tuossa kohdassa A on muuttunut G:ksi (transitio A>G).

Ongelmia tulee siitä, että joissain GenBankin näytteissä on tapahtunut deleetioita (nukleotidi on kadonnut) tai insertioita (on ilmestynyt ylimääräinen nukleotidi). Nämä aiheuttavat sen, että kaikki nukleotidit tuon kohdan jälkeen ovat yhden verran väärässä paikassa. Jos siis kohdasta 309 katoaa nukleotidi, oikeasti kohdassa 310 oleva nukleotidi näyttääkin olevan nyt kohdassa 309, ja kohdassa 14095 näyttääkin olevan kohdassa 14094, ja niin edelleen. Insertioissa puolestaan siirtymä on askeleen verran eteenpäin.

Pitää siis varmistaa, ovatko nukleotidit oikeassa vai väärässä kohdassa. Avaimiksi olen valinnut satunnaisen loppupään kymmenryhmän, jossa on helposti hahmotettava eli ”sanamainen” nukleotidirimpu. Se on

kohdassa 16051–16060: accaccaag (moneen kertaan varmistettu eri haploryhmien tunnusmutaatiot paikantamalla). Jos rimpu alkaakin ccaccaag, on jossain (yleensä hyvin alkupäässä, HVR2:ssa) tapahtunut deleetio, ja oikea kohta on yhden verran vasemmalla laske-  
tusta kohdasta (eli aloitus-A on siirtynyt vasemmalle edelliseen kymmenikköön). Jos taas rimpu alkaa tac-  
caccaaa, on tapahtunut insertio (nekin yleensä jo hyvin alkupäässä), ja oikea kohta on yhden verran oikealla las-  
ketusta kohdasta (eli aloitus-A on siirtynyt oikealle).

## Suomalaisia koko mtDNA:n näytteitä

Ensiksi pitää tietenkin löytää suomalaisia näytteitä oikeasta haploryhmästä. Niitä löytyy GenBankista iso joukko koko mtDNA:n laajuudelta, liittyen tuoreeseen äitilinjatutkimukseen Neuvonen et al. (2015). Haploryhmittäin ne menevät numerojärjestyksessä näin:

H = AY339402–AY339432 (31 kpl)  
V = AY339433–AY339459 (27 kpl)  
W = AY339460–AY339496 (27 kpl)  
I = AY339497–AY339509 (13 kpl)  
X = AY339510–AY339513 (4 kpl)  
Z = AY339514–AY339522 (9 kpl)  
U5a = AY339523–AY339529 (7 kpl)  
U5b = AY339530–AY339543 (14 kpl)  
U+ = AY339544–AY339553 (10 kpl)  
K = AY339554–AY339565 (12 kpl)  
T = AY339566–AY339576 (11 kpl)  
J = AY339577–AY339593 (17 kpl)

Näistä voi sitten etsiä ne H5a1:t, joissa kohdassa 16166 onkin A:n tilalla G – tai mikä sitten onkaan etsijän kiinnostuksen kohteena oleva haploryhmä. Kunhan löytää yhdenkin tällaisen näytteen, voi käyttää GenBankin BLAST-työkälyä ja syöttää siihen kyseisen näytteen numeron.

Tämä ohjelma tuottaa listan osumista samankaltaisuusjärjestyksessä eli ylimmäs ne, joihin nähden eroja on vähiten. Noin pitkien rimpsujen silmämääräinen vertailu on kuitenkin äärimmäisen rasittavaa, joten kannattaa hyödyntää Excel-taulukkolaskentaohjelman ominaisuuksia.

BLAST-tuloslistaus näyttää tältä:

Sequences producing significant alignments Download Manage Columns Show 100

select all 100 sequences selected

[GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens isolate F31 (PS16) mitochondrion, complete genome</a>	30596	30596	100%	0.0	100.00%	<a href="#">AY339432.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens isolate K343 mitochondrion, complete genome</a>	30570	30570	100%	0.0	99.97%	<a href="#">JX128054.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens isolate LP143mt mitochondrion, partial genome</a>	30568	30568	99%	0.0	99.97%	<a href="#">MK059495.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens isolate patient-24B-DgB mitochondrion, complete genome</a>	30568	30568	100%	0.0	99.97%	<a href="#">MF497518.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens isolate O05 mitochondrion, complete genome</a>	30568	30568	100%	0.0	99.97%	<a href="#">KU867581.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens isolate 4795367 mitochondrion, complete genome</a>	30568	30568	100%	0.0	99.97%	<a href="#">JX153844.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens isolate Bla13 mitochondrion, complete genome</a>	30568	30568	100%	0.0	99.97%	<a href="#">KF523405.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens isolate 126943 mitochondrion, complete genome</a>	30568	30568	100%	0.0	99.97%	<a href="#">KF161664.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens isolate PS-41 mitochondrion, complete genome</a>	30568	30568	100%	0.0	99.97%	<a href="#">JX128069.1</a>

## Excel-vertailu

Vertailua varten olen kopioinut kaikki BLAST-haulla löytyneet suomalaisten H5a1e-näytteiden lähi-osumat Excel-tilukkuun. Tämä tapahtuu ”maalamalla” hiiren kursorilla koko mtDNA:n data, aina kohdasta 1 kohtaan 16569 saakka. Sitten se kopioidaan (CTRL+C) ja liitetään Excel-tilukkuun (CTRL+V). Data asettuu tilukkuun vieden tilaa kahdeksan sarakkeen leveydeltä ja kymmenien rivien pituudelta.

Ensimmäiselle ”riville” kopioin kaikki keskenään vertailukelpoiset näytteet, eli ne, joissa on edellä kuvattu tunnusjakso accaccaag kohdassa 16051–16060. Seuraavalle ”riville”, eli aiempien näytteiden alapuolelle, kopioin kaikki sellaiset näytteet, joissa tunnusjakso on taccaccaa (on tapahtunut insertio) – näitä voi verrata vain keskenään muttei ylemmän ”rivin” näytteisiin. Kolmannelle ”riville” kopioin kaikki sellaiset näytteet, joissa tunnusjakso on ccaccaag (on tapahtunut deleetio); näitäkin voi verrata vain toisiinsa.

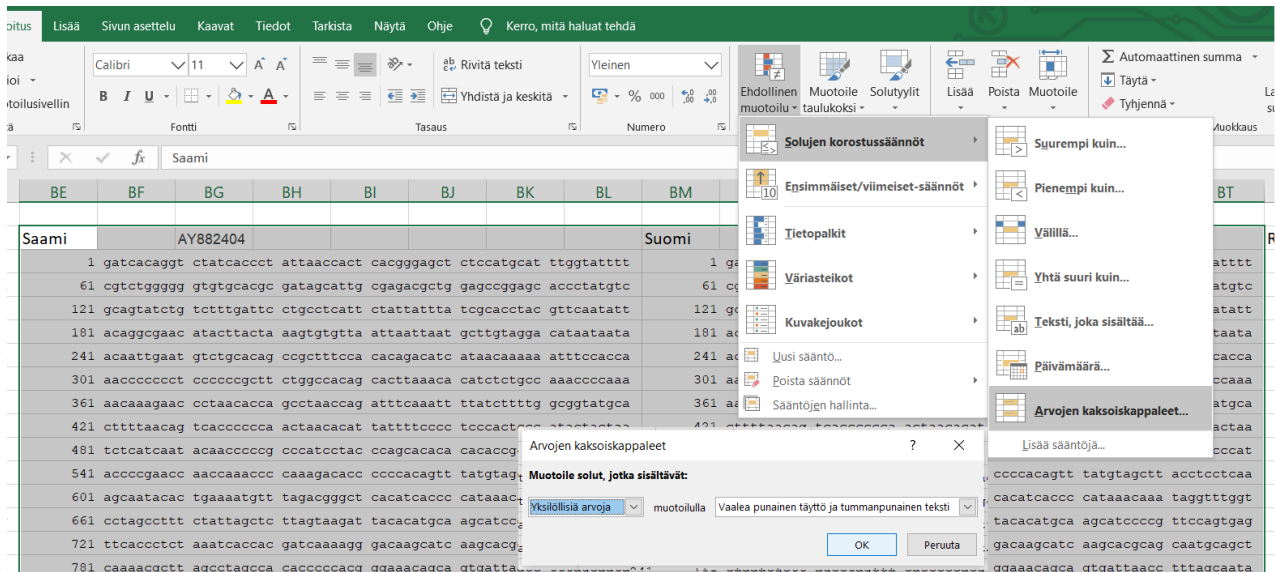
Vaikkei eri ”rivien” näytteitä voikaan verrata suoraan toisiinsa, ne asettuvat silti lähisukuisten näytteiden rykelmäksi niin, että ne voidaan suhteuttaa toisiinsa. Usein useammalle riville osuu näytteitä, joissa on tapahtunut mutaatio samassa kohdassa, joten ne voidaan ankkuroida toisiinsa. Lisäksi eri ”riveille” osuu yleensä näytteitä myös tunnusmutaation esimuotojen puolelta. Lisäksi silmämääräisesti voi tarkistaa, onko näytteessä tapahtunut mutaatioita samoissa kohdissa kuin toisen ”rivin” näytteissä (kunhan muistaa siirtymän vasemalle tai oikealle).

Sellaisen toiveen voisi esittää tulevaisuuden varalle, että vertailun helpottamiseksi näytteistä neutraloitaisiin kaikki insertiot ja deleetiot. Deleetion voi esittää esimerkiksi tyhjänä kohtana (acca ccaag), jolloin jälkeen tulevat nukleotidit eivät siirry väärään kohtaan. Insertion voi vain jättää pois ja mainita kuvaustiedoissa, että tietyssä kohdassa esiintyy ylimääräinen nukleotidi. Näin kaikki näytteet olisivat helposti toisiinsa vertailtavissa.

Käytännössä vertailu tapahtuu Excelissä näin:

1. Maalataan kaksi vierekkäistä näytettä alusta loppuun (kohdasta 1 kohtaan 16569).
2. Valitaan Aloitus-valikosta **Ehdollinen muotoilu**.
3. Valitaan siitä **Solujen korostussäännöt**.
4. Valitaan siitä **Arvojen kaksoiskappaleet**.
5. Valitaan siitä kohtaan **Muotoile solut, jotka sisältävät valinnaksi Yksilöllisiä arvoja**.
6. Painetaan OK: kaikki rivit, joilla nämä kaksi näytettä eroavat, muuttuvat vaaleanpunaisiksi.

Silmämääräisesti on vielä etsittävä se kohta, joka kullakin rivillä eroaa; accaccaag-näytteissä se on juuri se kohta kuin numero näyttääkin, mutta insertio- ja deleetio-riveillä se on viereisen numeron kohta. Muista peruuttaa värikoodaus painamalla CTRL-Z, koska tulet vertaamaan näytettä myös sen toisella puolella olevaan näytteeseen, eikä silloin saa olla vanhoja värjäyksiä häiritsemässä. Värjäykset voi poistaa myös **Ehdollinen muotoilu** -valikon käskyllä **Poista säännöt** (joko maatalulta alueelta tai koko dokumentista.)



## Haplotyyppisukupuu (mutaatiopuu)

**H5** (neol. KF523405)

**H5a** 4336 T>C

8557 G>A, 15175 C>T (da: KF161664)

**H5a8** 11969 G>A (fi: JX153844)

11807 A>G (fi: AY339430)

**H5a1** 15833 C>T (AY195747; JX128054)

514 G>A, 13674 T>C (sp? KU867581)

710 G>A, 722 T>C (JX128056)

4703 T>C (ru: KY671084)

4947 T>C (MK059495)

12414 T>C (sw: MF588555)

12630 G>A (da: KF162780)

14552 A>G (pe: KC911284)

16294 C>A (JX128069)

16343 A>T (sw: MF383616)

**H5a1e** 16166 A>G (fi: MF497518; tu: KJ187880;  
[FTDNA] fi: MINÄ 187856, MI22881, N27753,  
N263548, 127599, IN22836, 213050; sw: 81743,  
363479, 451405, 536851)

8265 T>C (JQ703620)

15880 A>G (fi: AY339432)

16344 C>T (fi: AY339431; [FTDNA] fi: 724160,  
N57113, MI29709; sw: 314525)

[cacccaag – ccaccaag – accaccaag – taccaccaa]

BLAST-listalta kerätyt GenBankin koko mtDNA:n tarkkuudella lähimmät osumat muodostavat yllä näkyvän sukupuun. Siihen on lisätty myös FamilyTreeDNA:n projekteista kerätyt H5a1e-näytteet, mukana oma tulokseni.

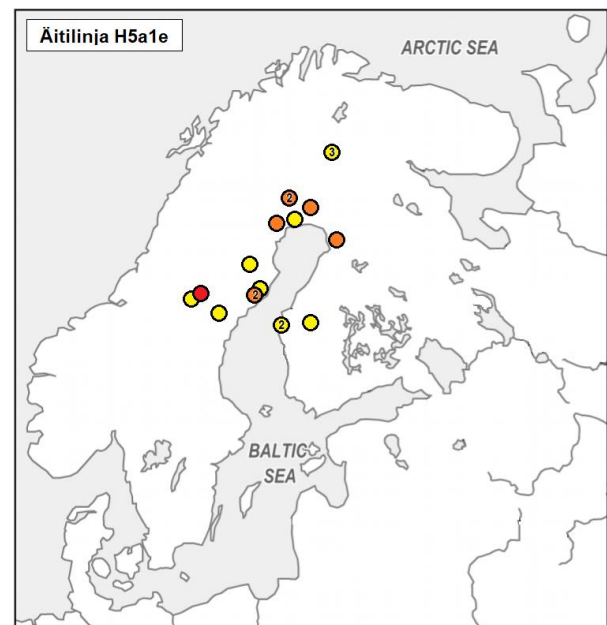
Haploryhmä H5:n leviäminen yhdistetään maanviljelijöiden leviämiseen Lähi-idästä neoliittisella kivikaudella, mutta jo H5a näyttää syntyneen Euroopassa, ja sitä on löytynyt jo varhaispronssikautisesta kellopikarikulttuurista (Roostalu et al. 2007; Brotherton et al. 2013).

## Alueellinen jakauma

Alahaploryhmä **H5a1e** jakautuu neljään ryhmään:

- (1) perustajahaplotyypin (vain tunnusmutaatio 16166 A>G)
- (2) jälkeläishaplotyypin (1:n lisäksi mutaatio 8265 T>C)
- (3) jälkeläishaplotyypin (1:n lisäksi mutaatiot 15880 A>G)
- (4) jälkeläishaplotyypin (1:n lisäksi mutaatio 16344 C>T)

Kun kartoitetaan näihin ryhmiin kuuluvien näytteiden alueellista jakaumaa, niin syntyy tämänkoinen kartta:



- Perustajahaplotyyppi (16166 A>G)
- Jälkeläishaplotyyppi (16344 C>T)
- Jälkeläishaplotyyppi (2 mutaatiota)

Lisäksi on monta näytettä, joista ei löydy paikkatietoja. Valitettavasti tämä koskee myös kahta jälkeläishaplotyypia

Koska koodaavan alueen mutaatiot eivät näy tietokannoissa, en voi erottaa 2- ja 3-tyyppejä toisistaan: ne näkyisivät minulle vain 1-etäisyyden osuina, joiden HVR1:ssä ja HVR2:ssa ei näy eroa minuun. Kartassa ei siksi erotu kuin 4-tyyppi.



Kartalla on vain 19 näytettä, vaikka minulla on 22 osunmaa 0-etiäisyydellä (perustajahaplotyyppiä kuten minäkin) ja 18 osunmaa 1-etiäisyydellä (jokin jälkeläishaplotyypeistä). Näistä valitettavasti vain noin puolet on ilmoittanut varhaisimman esiäitinsä asuinpaikan; osa ei edes kerro esiäitiään, jotta googlaamalla voisi löytää tietoa sukututkimustietokannoista.

Kaikki eivät myöskään ole liittyneet projekteihin (H5-projekti, Sweden-projekti, Finland-projekti), joten heidän haplotyyppiään ei pääse näkemään. On siis

mahdollista, että alahaploryhmässä H5a1e olisi vielä muitakin jälkeläishaplotyypejä kuin tässä tarkastelussa näkyviin saadut.

Projekteissa näytetään vain HVR1- ja HVR2-mutaatioita, koska koodaavan alueen mutaatioissa voi näkyä terveyteen vaikuttavia asioita. Silti jo niistäkin näkee erilaiset haplotyyppit, jotka voidaan nyt liittää koko mtDNA:n perusteella rakennettuun haplotyyppisuku-puuhun (ks. FTDNA-tunnukset edellisen sivun sukupuussa).

724160	Riitta Juhontr Talonpoika, b1711-d,1780, Kiiminki	Finland	H5a1e	A16129G, A16166G, T16187C, C16189T, T16223C, G16230A, T16278C, T16304C, C16311T, <b>C16344T</b> , C16519T	G73A, C146T, C152T, C195T, A247G, C456T, 315.1C
MI22881	Catharina Hansdtr Gabbels b. 1767 and d. 1851	Finland	H5a1e	A16129G, A16166G, T16187C, C16189T, T16223C, G16230A, T16278C, T16304C, C16311T, <b>C16344T</b> , C16519T	G73A, C146T, C152T, C195T, A247G, C456T, 315.1C

Näyttää selvältä, että H5a1e on syntynyt Perämeren ympäristössä: sitä tavataan Suomen puolella vain Pohjanmaalla ja Lapissa, ja Ruotsin puolella Norrbottenissa, Västerbottenissa, Ångermanlandissa ja Jämtlandissa.

Tunnusmutaatio eli H5a1e-kantaäiti on todennäköisesti syntynyt Ruotsin puolella, koska siellä sekä näytteiden lukumäärä että niiden diversiteetti on suurempi (ainakin kolme eri haplotyyppiä). On silti vaikea arvioida, mihin esihistorialliseen tai historialliseen ilmiöön (muuttoliikkeeseen tai yhteysverkostoon) äitilinjan leviäminen voisi liittyä. Vaihtoehtoja ovat ainakin:

- saamelaiset (rautakaudella)
- kveenit/kainulaiset (rautakauden lopulla)
- ruotsalaiset (keskiajalla)

Tutkimuksessa Översti et al. (2017) on laskettu H5a1e:n iäksi noin 2000 vuotta, mutta heillä on vain neljä näytettä, kaikki Suomesta. Ryhmän ikä saattaisi laajemmalla ja useampia haaroja kattavalla otoksella olla vieläkin vanhempi.

Toisaalta näyttää siltä, että perustajahaplotyyppi ja jälkeläishaplotyyppit (ainakin se, josta on paikkatietoja) olisivat ehkä levinneet samanaikaisesti samalle alueelle. Tällöin tietystikään leviäminen ei olisi voinut tapahtua heti alahaploryhmän synnyttyä, vaan vasta myöhemmin, kun jälkeläishaplotyyppieihinkin johtavat mutaatiot olivat jo tapahtuneet. Koska näitä ei ole ajoitettu, olemme toistaiseksi arvailujen varassa.

Todennäköisimmältä vaikuttaisi tässä tilanteessa jokin myöhäisemmistä ilmiöistä, suunnilleen viimeisen tuhat vuoden aikana tapahtunut. Ruotsalaisasutus saavutti Perämeren pohjukan suunnilleen 1300-luvulla, ja samaan aikaan Suomen puolelle Pohjanmaalle asettui ruotsalaisia. Tämä on yksi vaihtoehto äitilinjan leviämiselle.

Syvällä Suomen Lapissa on kolme näytettä, mutta ne polveutuvat kaikki samasta esiäidistä: Malin Jääskö (os. Orajarvi/Tepsa, syntynyt 1699; myös minun suora esiäitini). Malin näyttää kuuluneen Kemin Lapin vanhoihin lapinsukuihin (saamelaiskuihin), ja lapinmaathan avattiinkin suomalaisille uudisasuttajille vasta 1673.

On hyvin mahdollista, että yksittäinen suomalais- tai ruotsalaisperäinen nainen olisi päätenyt naimakaupan seurauksena keskiajalla tai viimeistään 1500-luvulla syvälle Kemin Lappiin saamelaisten keskuuteen. On kuitenkin teoriassa mahdollista sekin, että alahaploryhmä liittyisi kokonaisuudessaan saamelaisiin.

Toisaalta tätä selitystä heikentää se, ettei alahaploryhmää H5a1e kuitenkaan löydy nykysaamelaisilta. Silkekin on kuitenkin mahdollinen selityksensä: jo keskiajalla saamelaiset syrjäytettiin ja sulautettiin rannikkovyöhykkeeltä ruotsalaisten ja suomalaisten toimesta. Niinpä olisi täysin ymmärrettävää, ettei sisämaan saamelaisilta löydy sellaista alahaploryhmää, joka olisi syntynyt ja levinnyt rannikkovyöhykkeellä vasta joskus saamelaisten alkuperäistä levittäytymistä myöhemmin.

Saamelaisista on kyllä löydetty harvinaisena haplo-ryhmää H5 niin Norjasta (Delghandi, Utsi & Krauss 1998), Suomesta (Sajantila et al. 1996) kuin Venäjältäkin (Tambets et al. 2018). Koska näissä tutkimuksissa joko ei ollut koodaavan alueen (*coding region*) mutaatioita tai niissä oli sieltä vain harvoja ja valittuja kohtia, löydetty näytteet voisivat olla myös H5a:ta ja jopa H5a1:tä, joiden tunnusmutaatiot osuvat juuri koodaavalle alueelle eivätkä siksi näy suppeammassa tarkastelussa. Jos saamelaisissa sitten todella olisi H5a1:tä, voisi sen välitön jälkeläinen, alahaploryhmä H5a1e hyvinkin olla syntynyt juuri saamelaisissa.

Tässä yhteydessä voidaan viitata tuoreisiin genomilaajuisiin tutkimustuloksiin, joiden mukaan Eteläpohjanmaan Isonkyrön Levänluhdan keskirautekauden (n. 300–800 jKr.) suovainajat olivat perimältään hyvin samankaltaisia nykyisten saamelaisten kanssa ja poikkesivat selvästi suomalaisista (Lamnidis et al. 2018). Alueelta on löydetty myös saamelaisperäistä paikannimistöä (Rahkonen 2017).

Onkin mielenkiintoista, että juuri Etelä-Pohjanmaalta ja sen rannikolta (ruotsinkieliseltä Pohjanmaalta) ovat eteläisimmät paikannettavissa olevat H5a1e-esiäidit. Levänluhtalaisten perimää on selvitetty myös äitilinjojen osalta (Översti et al. 2019), ja nykysaamelaisissa on joitain tarkalleen samoja äitilinjahaplotyyppiä kuin levänluhtalaisissa, ainakin eräs U5b1b1a-haplotyyppi ja U5b1b1a1-perustajahaplotyyppi (Häkkinen, tulossa).

Kysymys jää toistaiseksi auki: meillä ei ole mitään keinoja arvioida luotettavasti, mihin väestöliikkeeseen tai yhteysverkostoon alahaploryhmä H5a1e:n leviäminen olisi todennäköisimmin liittynyt.

Oma mysteerinsä on tunisialainen näyte KJ187880. Se on GenBankissa vain HVR1:n laajuudelta, mutta siinäkin on tapahtunut mutaatio 16166 A>G. Tietysti tämä voisi olla myös itsenäinen rinnakkaismutaatio – sellaisia tiedetään tapahtuvan. Asian varmistamiseksi onkin katsottava kokonaiskuvaa: vastaako HVR1-haplotyyppi muuten H5a1e:tä vai ei?

Vastaus on myönteinen: mitään erottavia mutaatioita ei esiinny, eli tuolla tarkkuudella haplotyyppi todella on identtinen H5a1e-perustajahaplotyyppin kanssa. Jos se edustaisi jotain muuta haplo-ryhmää tai alahaploryhmää, siinä pitäisi näkyä niille ominaisia muutoksia ja siitä pitäisi puuttua H5a1e:hen johtavat mutaatiot.

HVR1:stä löytyy toisaalta lisäksi vain H5:n tunnusmutaatio 16304 (sekin on tunisialaisessa sama kuin H5a1e:ssä), mutta esimerkiksi monessa muussa H:n alahaploryhmässä olisi omat mutaationsa HVR1:ssä, kokonaan toisista haplo-ryhmistä puhumattakaan.

Silti sekin on yhä mahdollista (joskaan ei kauhean todennäköistä), ettei kyseessä näistä kahdesta samasta mutaatiosta huolimatta olisikaan aito H5a1e. Tunisiasta tai mistään sieltäpäinkään ei nimittäin tiettävästi ole löydetty H5a1e:tä tarkemmalla erottelutasolla, eikä GenBank-osumalistsauksissakaan niitä sieltä nouse esiin. Niitä ei ylipäättäen ole löydetty mistään Suomen ja Ruotsin ulkopuolelta. Tämäkin asia jää siis auki.

Vaikka Tunisiassa olisikin todellinen H5a1e, sen täytyy kaiken edellä esitetyn perusteella olettaa kulkeutuneen Pohjois-Afrikkaan täältä pohjoisesta. Milloin, miten ja miksi, se jäänee ikiajoiksi mysteeriksi.

## Tietokannat

Finland DNA – Suomi DNA. (FamilyTreeDNA)

<https://www.familytreedna.com/groups/finland/about/background>

GenBank.

<https://www.ncbi.nlm.nih.gov/genbank/>

H5'36, H5, and H36 mtGenome Haplogroups Project. (FamilyTreeDNA)

<https://www.familytreedna.com/groups/mt-dna-h5/about/background>

Sweden DNA – Sverigeprojektet. (FamilyTreeDNA)

<https://www.familytreedna.com/groups/sweden/about>

## Lähteet

- BLAST** = Zheng Zhang, Scott Schwartz, Lukas Wagner, Webb Miller 2000: A greedy algorithm for aligning DNA sequences. – *Journal of Computational Biology* **2000**; 7(1–2): 203–14.
- Paul **Brotherton**, Wolfgang Haak, Jennifer Templeton, Guido Brandt, Julien Soubrier, Christina Jane Adler, Stephen M. Richards, Clio Der Sarkissian, Robert Ganslmeier, Susanne Friederich, Veit Dresely, Mannis van Oven, Rosalie Kenyon, Mark B. Van der Hoek, Jonas Korfach, Khai Luong, Simon Y.W. Ho, Lluís Quintana-Murci, Doron M. Behar, Harald Meller, Kurt W. Alt, Alan Cooper & The Genographic Consortium **2013**: Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. – *Nature Communications*. DOI: 10.1038/ncomms2656
- Madjid **Delghandi**, Egil Utsi, Stefan Krauss **1998**: Saami Mitochondrial DNA Reveals Deep Maternal Lineage Clusters. – *Human Heredity* 48: 108–114.

- Häkkinen**, Jaakko (tulossa): Kokonaiskuva saamelaisten äitilinjoista.
- Thiseas C. **Lamnidis**, Kerttu Majander, Choongwon Jeong, Elina Salmela, Anna Wessman, Vyacheslav Moiseyev, Valery Khartanovich, Oleg Balanovsky, Matthias Ongyerth, Antje Weihmann, Antti Sajantila, Janet Kelso, Svante Pääbo, Päivi Onkamo, Wolfgang Haak, Johannes Krause, Stephan Schiffels **2018**: Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. – *Nature Communications* (2018) 9:5018 | DOI: 10.1038/s41467-018-07483-5
- Anu M. **Neuvonen**, Mikko Putkonen, Sanni Översti, Tarja Sundell, Päivi Onkamo, Antti Sajantila, Jukka U. Palo **2015**: Vestiges of an Ancient Border in the Contemporary Genetic Diversity of North-Eastern Europe. – *PLOS ONE* | DOI:10.1371/journal.pone.0130331
- PhyloTree** = Mannis van Oven, Manfred Kayser **2009**: Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. – *Human Mutation* 30(2):E386-E394. doi:10.1002/humu.20921 <http://www.phylotree.org>
- Rahkonen**, Pauli **2017**: Onomasticon of Levänluhta and Kälviälä region. *Suomalais-Ugrilaisen Seuran Aikakauskirja* 96. <https://journal.fi/susa/article/view/70231/31260>
- U. **Roostalu**, I. Kutuev, E.-L. Loogväli, E. Metspalu, K. Tambets, M. Reidla, E. K. Khusnutdinova, E. Usanga, T. Kivisild, R. Villems **2007**: Origin and Expansion of Haplogroup H, the Dominant Human Mitochondrial DNA Lineage in West Eurasia: The Near Eastern and Caucasian Perspective. – *Molecular Biology and Evolution* 24(2):436–448. doi:10.1093/molbev/msl173
- Antti **Sajantila**, Päivi Lahermo, Tiiu Anttinen, Matti Lukka, Pertti Sistonen, Marja-Liisa Savontaus, Pertti Aula, Lars Beckman, Lisbeth Tranebjaerg, Tobias Gedde-Dahl, Laurie Issel-Tarver, Anna Di Rienzo, Svante Pääbo **1995**: Genes and Languages in Europe: An Analysis of Mitochondrial Lineages. – *Genome Research* 5: 42–52.
- Kristiina **Tambets**, Bayazit Yunusbayev, Georgi Hudjashov, Anne-Mai Ilumäe, Siiri Rootsi, Terhi Honkola, Outi Vesakoski, Quentin Atkinson, Pontus Skoglund, Alena Kushniarevich, Sergey Litvinov, Maere Reidla, Ene Metspalu, Lehti Saag, Timo Rantanen, Monika Karmin, Jüri Parik, Sergey I. Zhadanov, Marina Gubina, Larisa D. Damba, Marina Bermisheva, Tuuli Reisberg, Khadizhat Dibirnova, Irina Evseeva, Mari Nelis, Janis Klovins, Andres Metspalu, Tõnu Esko, Oleg Balanovsky, Elena Balanovska, Elza K. Khusnutdinova, Ludmila P. Osipova, Mikhail Voevoda, Richard Villems, Toomas Kivisild, Mait Metspalu **2018**: Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. – *Genome Biology* (2018) 19: 139.
- Översti**, Sanni 2014: *Äitilinjojen fylogeneettinen hienorakenne suomalaisilla – metsästäjä-keräilijät vs. maanviljelijät*. Pro gradu -tutkielma: Turun yliopisto, Biologian laitos 20.3.2014.
- Sanni **Översti**, Päivi Onkamo, Monika Stoljarova, Bruce Budowle, Antti Sajantila, Jukka U. Palo **2017**: Identification and analysis of mtDNA genomes attributed to Finns reveal long-stagnant demographic trends obscured in the total diversity. – *Nature: Scientific Reports* 7. | DOI:10.1038/s41598-017-05673-7.
- Sanni **Översti**, Kerttu Majander, Elina Salmela, Kati Salo, Laura Arppe, Stanislav Belskiy, Heli Etu-Sihvola, Ville Laakso, Esa Mikkola, Saskia Pfrenge, Mikko Putkonen, Jussi-Pekka Taavitsainen, Katja Vuoristo, Anna Wessman, Antti Sajantila, Markku Oinonen, Wolfgang Haak, Verena J. Schuenemann, Johannes Krause, Jukka U. Palo, Päivi Onkamo **2019**: Human mitochondrial DNA lineages in Iron-Age Fennoscandia suggest incipient admixture and eastern introduction of farming-related maternal ancestry. – *Nature: Scientific reports* 9:16883 | <https://doi.org/10.1038/s41598-019-51045-8>.