

Suomalaisten sisäinen ja ulkoinen vertailu IBS-tasolla

Internal and external comparison of Finns at the IBS-level

Jaakko Häkkinen (päivitetty 22.10.2012 / updated 22nd October 2012)

1. Johdanto

Suomalaisia ei voi verrata toisiinsa suhteuttamatta heitä samalla naapurikansoihin. Tämä johtuu ensinnäkin siitä, että eri alueiden suomalaisten väliset geenierot ovat ainakin FST-tasolla eurooppalaisten suurimpia. Siksi tarvitaan tietoa muiden kansojen sisäisistä geenieroista, jotta voitaisiin asettaa oikeaan kohtaan kynnyks, jonka jälkeen voidaan puhua eri väestöistä. Toiseksi eri alueiden suomalaiset saattavat olla lähellä eri naapurikansoja, joten suomalaisten alkuperän selvittämiseksi on tärkeää saada vertailuaineistoa eri puolilta. Vertailun vuoksi mielenkiintoisia ovat myös kaukaisemmat näytteet.

Luonnollisesti analysoin ja kerrostan samalla myös muiden kansojen IBS-arvoja, joten tämä ei ole pelkästään suomalaisten projekti – tämä on pikemminkin suomalaiskeskeinen projekti. Toistaiseksi pystyn vertailemaan vain tiedostoja, joissa SNP:t ovat samassa järjestyksessä, eli voin verrata Family Finder -tiedostoja vain toisiinsa, **ja nyt myös alkuperäisiä Relative Finder -tiedostoja toisiinsa**. Tulokset näyttävät yhteismitallisilta, joten voin hyödyntää kummankinlaisia tiedostoja kokonaiskuvassa.

Kyseessä on jatkuvasti päivittyvä projekti, eli jos olet kiinnostunut osallistumaan, toimi seuraavasti:

1. Lähetä Family Finder tai Relative Finder -datasi mieluiten pakattuna (sellaisena kuin sen latsit FamilyTreeDNA:n sivuilta) sähköpostin liitetiedostona osoitteeseen hakkinenjaakko[at]gmail.com ; vaihtoehtoisesti voit lähettää minulle tunnuksesi ja salasanasasi, jolloin voin käydä lataamassa tiedoston itse koneelleni. (En pengo tiedostoa vaan annan laskurin verrata sitä muihin; mikäli haluan tulevaisuudessa laskea siitä joitain muita arvoja, kysyn lähettäjältä luvan.)
2. Kerro kansallisuutesi ja vanhempiesi tai mieluiten isovanhempiesi alkuperä/syntymäseutu.

1. Introduction

Finns cannot be compared to each other without relating them to the neighbouring peoples at the same time. First, this is because Finns from different area are genetically very different in the European frame, at least in the FST-level. Therefore we need information about the internal differences of other peoples, so that we could set in the proper level the threshold at which we can talk about different populations. Second, Finns from different areas may be close to different neighbouring peoples, so it is crucial for the origin of Finns to get reference data from different directions. Also the more remote samples are interesting for the comparative purposes.

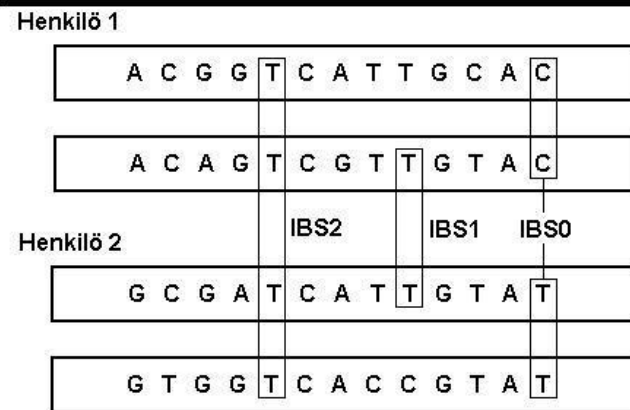
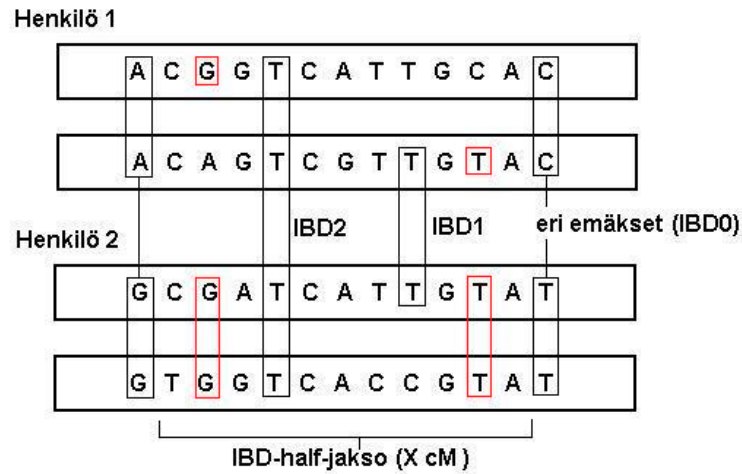
Naturally at the same time I will analyze and structure also the IBS-values of other peoples, so this is not merely a Finnish project – this is more like Finnocentric project. So far I can only compare files which have their SNP's in the same order, so I can compare Family Finder files only to each other, **and now also original Relative Finder files to each other**. The results look commensurate, so I can utilize both files in the whole picture.

This is a continuously updating project, so if you want to participate, follow the instructions:

1. Send your Family Finder or Relative Finder data preferably packed (like you downloaded it from FamilyTreeDNA's pages) as an attachment of e-mail to the address hakkinenjaakko[at]gmail.com ; alternatively you can send me your kit number and password, so I can download it myself. (I will not go through the file but I only give the calculator to compare it to the others; if I in the future want to calculate some other values, I will ask permission from the sender.)
2. Tell you nationality/ethnicity and the original place of your parents or preferably grandparents.

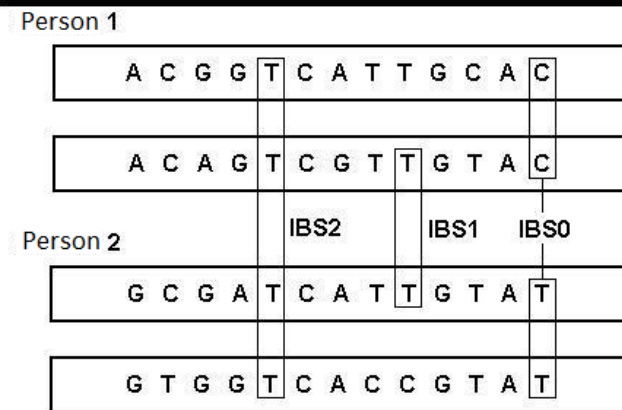
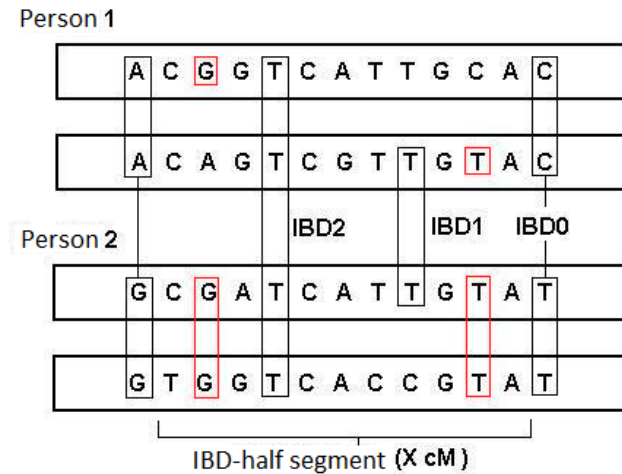
2. IBS – mikä ja miksi?

IBS on lyhenne ilmauksesta *identity by state*, mikä merkitsee sitä, että alleeli on muodollisesti samanlainen (identtinen) kuin toisella henkilöllä. Jokaisessa autosomaalisessa locuksessa (eli kromosomien 1–22 sijainnissa) on kaksi alleelia emäsjoukosta G, A, T ja C. Yleensä G ja A vaihtelevat keskenään sekä T ja C keskenään, mutta toisinaan tapahtuu mutaatioita myös näiden emäsparien välillä (esim. A > T).



2. IBS – what and why?

IBS comes from a phrase *identity by state*, which means that an allele is formally similar with an allele of another person. In every autosomal locus (or in every location of chromosomes 1–22) there are two alleles from the nucleotides G, A, T and C. Usually G and A alter with each other, and T and C with each other, but sometimes occur mutations also between the nucleotide pairs (for example A > T).



Sijaintikohtainen IBS-arvo on joko 2, 1 tai 0, kuten kuvan alaosasta näkyy. Kokonais-IBS on sitten näiden kaikkien keskiarvo jaettuna kahdella (jotta saadaan arvo väliltä 0–1). Kuvan sijaintien IBS-arvot ovat: 0, 1, 1, 1, 2, 2, 1, 1, 1, 2, 1, 2, 0. Näiden keskiarvoksi tulee 1,15 (summa 15 jaettuna lukumäärällä 13), ja lopulliseksi IBS-arvoksi 0,5750.

IBD eli *identity by descent* (peritysti samanlainen) on toinen tapa mitata alleelien samanlaisuutta. Irrallisesti ajatellen alleeli voi olla joko ”vain” IBS, eli se on samanlainen koska henkilöt kuuluvat esimerkiksi samaan kansaan, tai se voi olla IBD, eli se on samanlainen koska se on todella peritty yhteiseltä esivanhemmalta. On määrittelykysymys, missä näiden kahden välinen raja kulkee, mutta äskettäin on pystytty tunnistamaan niinkin lyhyitä kuin 2 cM mittaisia IBD-jaksoja (Browning & Browning 2010: High-Resolution Detection of Identity by Descent in Unrelated Individuals. The American Journal of Human Genetics 86), mikä vastaa esivanhemman aikasyvyyttä 21 sukupolven takana (noin 500–600 vuotta sitten).

Yleensä IBD:tä tarkastellaan geenijaksoina. Rekombinaation vuoksi yhdeltä esivanhemmalta perityt geenijaksot pilkkoutuvat joka sukupolvi pienemmiksi, joten mitä pidempiä IBD-jaksoja henkilöt jakavat, sitä lähempää sukua he ovat. Tällaisten jaksojen pituus ilmaistaan centimorganeina (cM). IBD-jakso päättyy kohtaan jossa IBD=0 eli yksikään emäs ei ole henkilöillä sama.

Family Finderissa ja Relative Finderissa käytetään IBD-half- menetelmää, jossa riittää että henkilö on perinyt yhteisen alleelin jommaltakummalta vanhemmaltaan (ks. kuvan yläosa). Tämä menetelmä ei kuitenkaan välttämättä kerro todellista sukulaisuuden astetta, koska joka toinen yhteinen alleeli on voitu periä isältä ja joka toinen äidiltä (kaikki vasemmanpuoleiset alleelit eivät siis ole yhdeltä vanhemmalta peräisin, vaan järjestys on sattumanvarainen). Tällöin mitään pitkää IBD-jaksoa ei todellisuudessa ole olemassakaan. Yksittäinen alleeli tai lyhyt jakso taas ei vielä kerro todellisesta sukulaisuudesta, koska suurimmalla osalla kyseiseen väestöön kuuluvista voi olla sama alleeli.

Todennäköisyys sille, että osa IBD-half-jaksosta onkin täydentynyt toiselta vanhemmalta peritystä kromosomista, on sitä suurempi mitä kaukaisemmasta sukulaisuudesta on kysymys (eli mitä enemmän on kulunut sukupolvia ja mitä lyhyempiä jaksot ovat). Sukulaisten välillä IBD-half-jakso onkin todennäköisesti todellinen IBD-jakso.

Locuswise IBS-value is either 2, 1, or 0, like it shows in the lower part of the figure. Total IBS is then the mean value of all these loci divided by 2 (to get a value between 0–1). The IBS-values of the loci in the figure are: 0, 1, 1, 1, 2, 2, 1, 1, 1, 2, 1, 2, 0. Their mean value is 1.15 (sum 15 divided by the number of loci 13), and the final IBS-value is 0.5750.

IBD (*identity by descent* = similar by inheritance) is another way to measure the similarity of alleles. An allele may be either “merely” IBS, when it is similar because the persons compared belong to the same population, or it can be IBD, when it is similar because it is truly inherited from a common ancestor. It is a matter of definition, where the border between the two goes, but recently even so short as 2 cM IBD-segments have been detected (Browning & Browning 2010: High-Resolution Detection of Identity by Descent in Unrelated Individuals. The American Journal of Human Genetics 86), which corresponds to the time-depth of 21 generations (ca. 500–600 years ago) for the common ancestor.

Usually IBD is analyzed as gene segments. Due to the recombination the segments inherited from certain ancestor get shortened every generation, so the longer segments the persons share, the closer their relatedness. The length of such segments is expressed in centimorgans (cM). IBD-segment ends in a locus where IBD=0, so neither nucleotide is similar in the persons compared.

Family Finder and Relative Finder use the IBD-half method, where it is enough that a person has inherited the shared allele from either parent (see the upper part of the figure). This method does not necessarily tell the actual level of relatedness, because every second shared allele may have been inherited from father and every second from mother (all the left alleles are not inherited from the same parent, but the order is arbitrary). In that case there exists no long IBD-segment at all. Single allele or short segment does not yet tell about the true relatedness, because the majority of certain population may have the same allele.

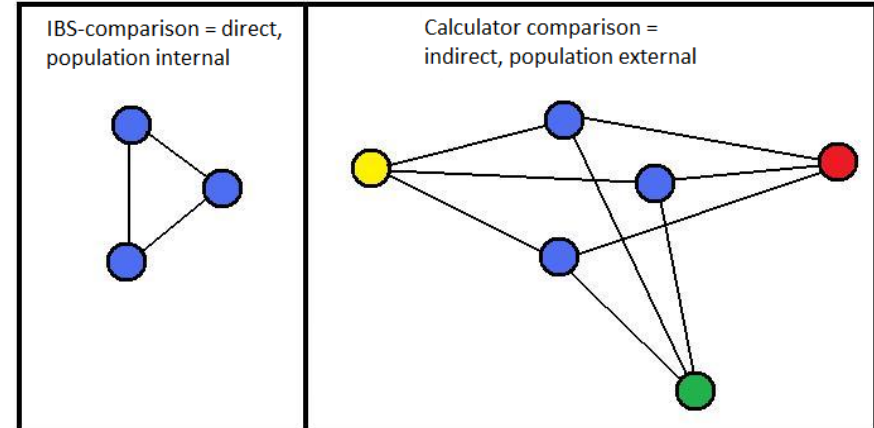
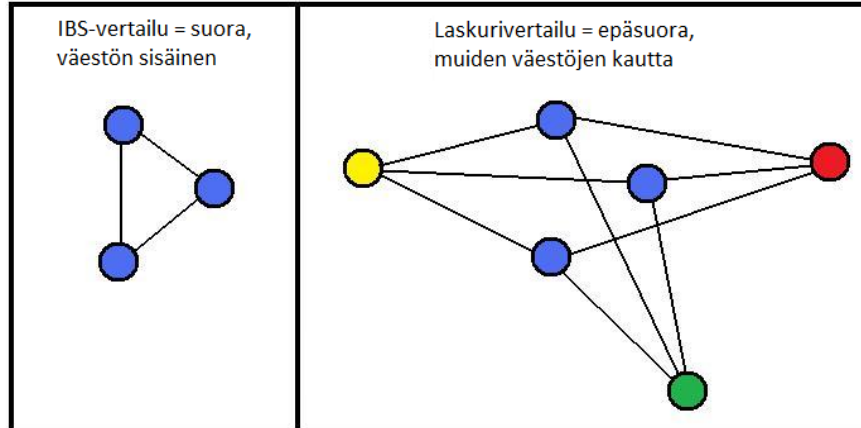
Probability for the occasion, where part of the IBD-half segment is complemented from the chromosome inherited from the other parent, is the greater the more remote is the relatedness (that is, the more generations have past and the shorter the segments are). Between relatives an IBD-half segment is probably true IBD-segment.

IBS kertoo todellisen sukulaisuuden luotettavammin kuin IBD-half, muttei yhtä täsmällisesti kuin todellinen IBD (phased data). Ei ole mahdollista arvioida kahdenvälisessä vertailussa, johtuuko kohonnut IBS-arvo genealogisesta sukulaisuudesta vai pelkästään samasta geenipoolista eli kyseisten henkilöiden kuulumisesta samaan osaväestöön. Tämä puute on kuitenkin mahdollista voittaa keräämällä laaja aineisto pareittaisia IBS-arvoja tarpeeksi tarkalla alueellisella erottelukyvällä. Silloin voidaan määritellä osaväestö (= henkilöt joiden keskinäiset arvot ylittävät tietyn kynnyksen) ja sen ”nollataso”, jonka yläpuolelle kohoavat arvot sitten edustavat todellista sukulaisuutta.

IBS-tasoa voidaan verrata myös sekoituslaskureiden tuloksiin (Dodecad, Eurogenes, Magnus Ducatus Lituaniae Project jne.). IBS-taso kertoo suoraan yhteisten alleelien määrän väestön sisäisenä verkostona, kun taas sekoituslaskurit kertovat epäsuorasti yhteisten segmenttien määrän väestön ulkoisiin suhteisiin perustuen. Laskuri etsii testattavilta tietyissä väestöissä taajaan esiintyviä geenijaksoja, eli testattavien yhteiset jaksot heijastetaan ulkopuolisten tahojen kautta. Onkin mielenkiintoista nähdä, millaisia eroja IBS-tason ja laskuritason välillä ilmenee.

IBS tells the true relatedness more trustworthy than IBD-half, but not as accurately as the true IBD (phased data) would. It is not possible to assess in the mutual comparison if the increase in the IBS-value is due to genealogical relatedness or only similar gene pool of the two persons from the same subpopulation. However, this weakness is possible to overcome by collecting large sample of pairwise IBS-values with high enough areal resolution. Then we can determine a subpopulation (= persons who exceed certain threshold in the intra-group values) and its “zero level”, above which we can detect true relatedness.

IBS-level can also be compared to the results from admixture calculators (Dodecad, Eurogenes, Magnus Ducatus Lituaniae Project etc.). IBS-level tells directly the measure of shared alleles in population internal network, while the calculators tell indirectly the measure of shared segments on the basis of population external network. In certain populations frequent gene segments are found out by calculator, so the shared segments of persons are reflected through third parties. It is interesting to see what kind of differences appear between the IBS-level and the calculator level.



3. Jäsenet / members 70 (62 FF + 8 RF*)

Suomi/Finland 44 (43 FF + 1 RF)

Ajvxyz (Varsinais-Suomi / Pohjois-Pohjanmaa)
ASespool (Satakunta)
Aune (Etelä-Pohjanmaa / ?)
Bertarido (Karjala // Pohjois-Savo / Lappi)
Erkki (Uusimaa x3 / Pohjois-Savo)
Esko (Etelä-Pohjanmaa / Pirkanmaa)
Helena (Häme / Satakunta // Pohjois-Savo / Keski-Suomi)
Jani (Häme / Kymenlaakso)
JaniM (Häme / Satakunta // Pohjois-Savo / Keski-Suomi)
Jaska (Pohjois-Karjala / Keski-Lappi)
Jola (Kainuu / Pohjois-Savo)
Kalifi (Pohjois-Savo)
Kalle (Pohjois-Savo // Karjalankannas /// Kymenlaakso / Keski-Suomi // Etelä-Savo x3 /
Etelä-Karjala)
Karoisä (Häme // Satakunta / Varsinais-Suomi)
*Karoliini (Häme x5 / Satakunta x2 / Varsinais-Suomi)
Karoäiti (Häme x3 / Satakunta)
Kirmukarmu (Häme)
Knaapi (Lounais-Suomi)
Laurukainen (Itä-Lappi)
Lea (Pohjois-Savo)
Levanki (Satakunta)
Loimo (Varsinais-Suomen sisämaa x3 / Satakunta // Pohjois-Karjala x3 / Etelä-Pohjanmaa)
Loimon ÄIs (Varsinais-Suomen sisämaa)
Loimoäiti (Varsinais-Suomen sisämaa x3 / Satakunta)
Maikki (Pohjois-Savo)
MariK ?????
Marja (Keski-Suomi)
Marjukka (Kanta-Häme // Uusimaa / Etelä-Pohjanmaa)
Marko (Häme)
Nina (Keski-Suomi / Pohjois-Pohjanmaa)
Ojaniemenkylä (Etelä-Pohjanmaa)
Pauli (Keski- ja Etelä-Pohjanmaa)
PekkaA (Etelä-Pohjanmaa / Itä-Suomi)
PekkaP (Pohjois-Pohjanmaa)
Pettersisko (Varsinais-Suomi / Häme / Uusimaa)

Petteräiti (Kanta-Häme / Varsinais-Suomi // Pirkanmaa x3 / Uusimaa)
Riikka (Keski-Suomi // Häme / Satakunta)
Risto (Etelä-Pohjanmaa)
Saikkala (Satakunta)
Senyho (Uusimaa)
Seppo (Keski-Suomi)
Tapio (Pirkanmaa)
Unto (Pohjois-Savo)
Veeruska (Keski-Suomi)
Äijälä (A-Tornio x2 / SWE / Satak. // E- Savo / Kainuu / Kemiö / Pohjanmaa)

Muut/others 26 (19 FF + 7 RF)

Albina (Russia)
Anatoly (Northern Russia)
Brian (England)
Carole (Finland / Greece)
Constantine (Western Latvia / Central Ukraine)
Dmitry (Russia / ?Tatar?)
Elena (Central Ukraine / South Russia)
ElenaCh (Russia)
Eugene (Finland // France / Canada)
Giedrius (Lithuania)
JAX (Finland / Italia // England / Ireland)
JAX dad/isä (Italia / Finland)
JAX grandfather/isoisä (Italia)
*JAX mom/äiti (England / Ireland)
*JAX uncle/setä (Italia / Finland)
Lest-K (Galicia = Southern Poland)
Lest-S (Galicia = Southern Poland)
*Linkus (Lithuania)
*Linkus mom (Lithuania / Aukštaitija)
*Linkus grandmother (Dzūkija = South Lithuania)
*Samogitian (Žemaitija = Northwest Lithuania)
Steve (Germany / England // Italia / Ireland / ? x2)
Stevemom (Germany / England)
*Sudovian (Southwest Lithuania)
Tomas (Aukštaitija = Northeast Lithuania)
Victor (Russia)

4. Menetelmät ja käytäntö

Jäsenten ryhmittelyssä olen siirtynyt käyttämään PAST-ohjelmaa (Hammer, Ø., Harper, D.A.T. and Ryan, P.D. 2001. PAST: Paleontological Statistics software package for education and data analysis. *Palaeontologica Electronica* 4(1): 9 pp.), koska se löytää luotettavasti lähisukuiset henkilöt samaan ryhmään senkin jälkeen kun näiden keskinäinen sukulaisuus on neutraloitu korvaamalla IBS-arvo lukemalla 0,7635.

Kun ryhmittely on tehty, on mahdollista määritellä (osa)väestön IBS-”nollataso”, kuten aiemmin määriteltiin. Tässä vaiheessa on jo havaittavissa useita toisistaan selvästi eroavia sukulaisuusasteita:

1. Vanhempi–lapsi: 0,85 (Stevemom–Steve, Marja–Riikka)
2. Isovanhempi–lapsenlapsi: 0,80 (Karoisä–Riikka)
3. Etäinen sukulaisuus / yhteinen osaväestö: 0,77 (Knaapi–Äijälä)
4. Yhteinen osaväestö / väestö: 0,76 (monet suomalaiset)
5. Eri väestö / geenipohja: alle 0,76 (kaukaisimmat suomalaiset)

Pareittaisten IBS-arvojen taulukko on jo liian laaja mahtuakseen kuvana tähän dokumenttiin, joten se löytyy erillisenä Excel-taulukkona osoitteesta: <http://www.mv.helsinki.fi/home/jphakkin/IBS.xlsx>

4.1. Väestökynnys ja väestöytimet

Tässä eräitä IBS-askeleita, joiden perusteella voidaan selvittää väestökynnystä:

1. Pienin suomalaisen ja muunmaalaisen välinen IBS-arvo: 0,7547.
2. Pienin kahden muunmaalaisen välinen IBS-arvo: 0,7566.
3. Pienin kahden suomalaisen välinen IBS-arvo: 0,7596.
4. Suurin suomalaisen ja muunmaalaisen välinen IBS-arvo: 0,7643.
5. Suurin kahden muunmaalaisen välinen IBS-arvo: 0,7656.
6. Suurin kahden ei-sukulaisen suomalaisen välinen IBS-arvo: 0,7666.

4. Results and practice

I have adopted the PAST program (Hammer, Ø., Harper, D.A.T. and Ryan, P.D. 2001. PAST: Paleontological Statistics software package for education and data analysis. *Palaeontologica Electronica* 4(1): 9 pp.) for clustering the members, because it can reliably find out related individuals even after replacing their mutual IBS-value with the neutral 0.7635.

When grouping is done, it is possible to define the “zero level” for IBS of a (sub)population, as defined earlier. At this point there already seem to be several distinguishable levels of relatedness:

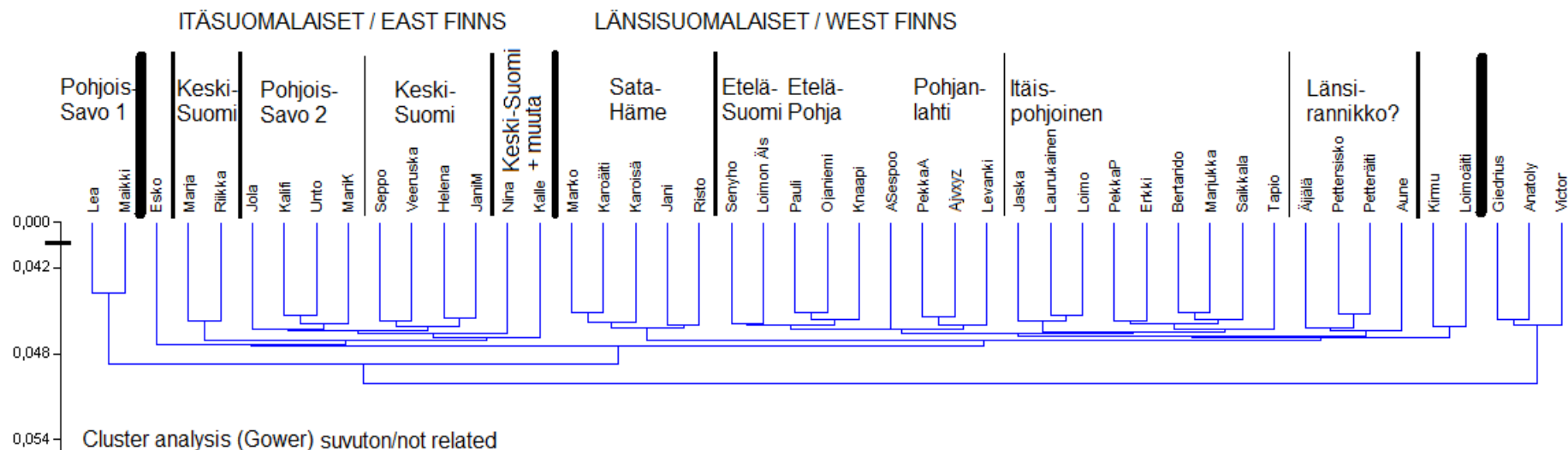
1. Parent–child: 0.85 (Stevemom–Steve, Marja–Riikka)
2. Grandparent–grandchild: 0.80 (Karoisä–Riikka)
3. Distant relation / shared subpopulation: 0.77 (Knaapi–Äijälä)
4. Shared subpopulation / population: 0.76 (many Finns)
5. Different population / gene base: below 0.76 (remote Finns)

The table of pairwise IBS-values is already too wide to fit in this document as a picture, so it can be found in a separate Excel-table in: <http://www.mv.helsinki.fi/home/jphakkin/IBS.xlsx>

4.1. Population threshold and population cores

Here are some IBS-steps, which may be used to find out the population threshold:

1. Lowest IBS-value between a Finn and a foreigner: 0.7575.
2. Lowest IBS-value between two foreigners: 0.7597.
3. Lowest IBS-value between two Finns: 0.7596.
4. Highest IBS-value between a Finn and a foreigner: 0.7643.
5. Highest IBS-value between two foreigners: 0.7656.
6. Highest IBS-value between two unrelated Finns: 0.7666.



5. Tulokset

Yllä olevassa ryhmittelyssä kaikki suomalaiset kuuluvat samaan haaraan. Ensimmäisinä suomalaisista eroaa erikoinen ryhmä Pohjois-Savo 1, johon kuuluvat serkukset Lea ja Maikki. Johtuuko heidän erikoisuutensa vieraasta perimästä vai geneettisestä ajautumisesta, siihen en osaa tässä vaiheessa ottaa kantaa. Seuraavaksi erottuvat toisistaan itä- ja länsisuomalaiset. Erikoista on, että länsisuomalainen Esko muodostaa yksin oman haaransa itäsuomalaisissa. Toinen erikoisuus on, että itäis-pohjoinen ryhmä kuuluu länsisuomalaisiin.

MDS-kuvaajassa itäsuomalaiset (punainen soikio) muodostavat hyvin tiiviin ryhmän suomalaisten keskelle; vain Riikka, Lea ja Maikki sijoittuvat kauemmas. Tämä heijastanee savolaista perustajavaikutusta. Länsisuomalaiset asettuvat kaikkialla itäsuomalaisten ympärille, eikä tilanne muutu vaikka ulkomaalaiset jätetään kokonaan pois. Tämä kertoo siitä, että länsisuomalaisten diversiteetti on huomattavan suurta.

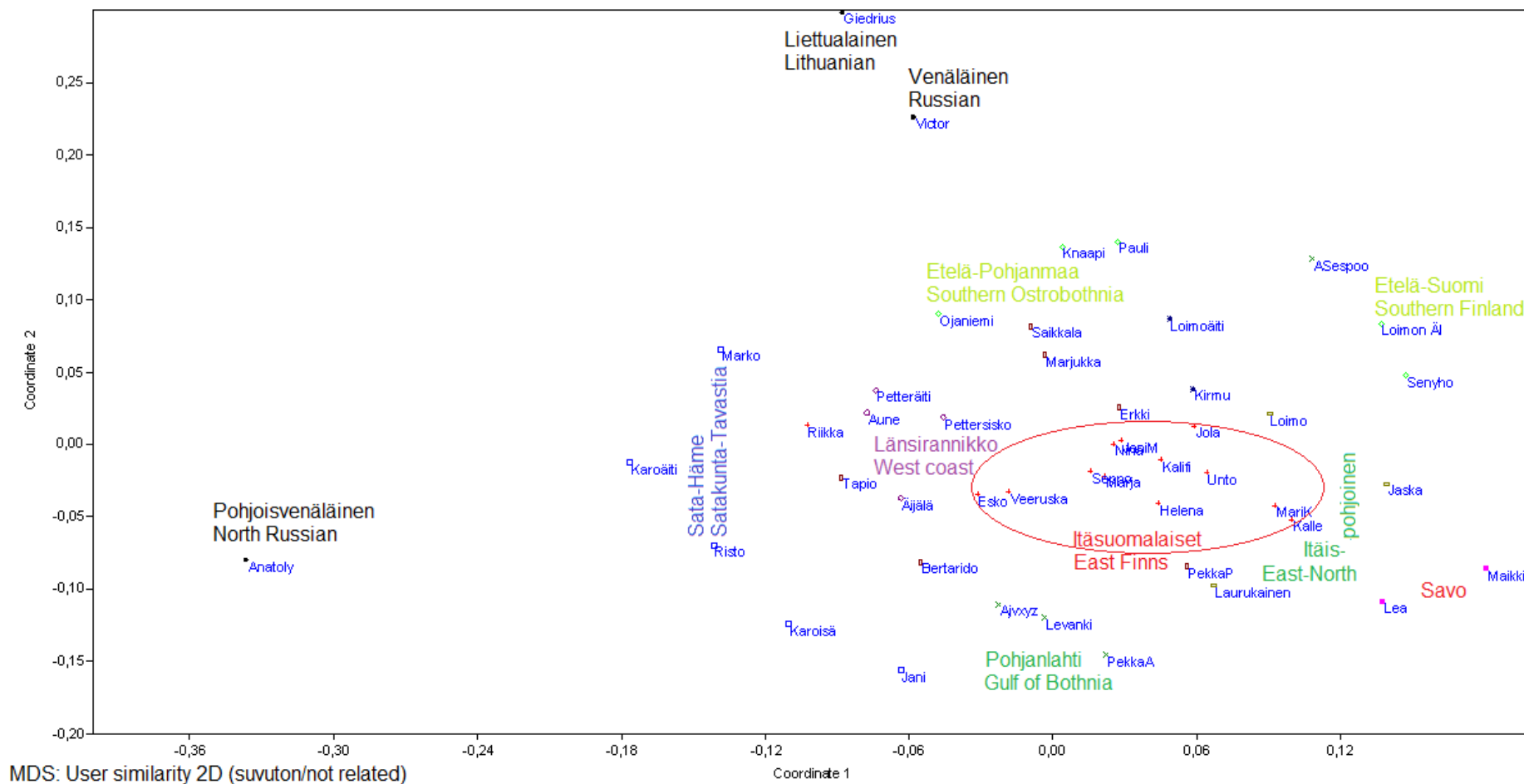
5. Results

In the cluster analysis above, all the Finns belong to the same branch. The first group to split off from the Finnish core is a peculiar group Northern (Pohjois-)Savo 1, including Lea and Maikki. At this point I cannot tell whether their special status is caused by some foreign influence or by the genetic drift. At the next stage East and West Finns become distinguished from each other. It is surprising that West Finn Esko forms a group on his own in the East Finnish branch. Another surprise is, that the East-North branch belongs to West Finns.

In the MDS plot East Finns (red oval) form a very tight cluster in the middle of Finns; only Riikka, Lea and Maikki are spread out of that area. This may reflect the Savonian founder effect. West Finns are everywhere around the East Finns, and this situation does not change even though the non-Finns are dropped from the sample. This tells that the diversity of West Finns is considerably great.

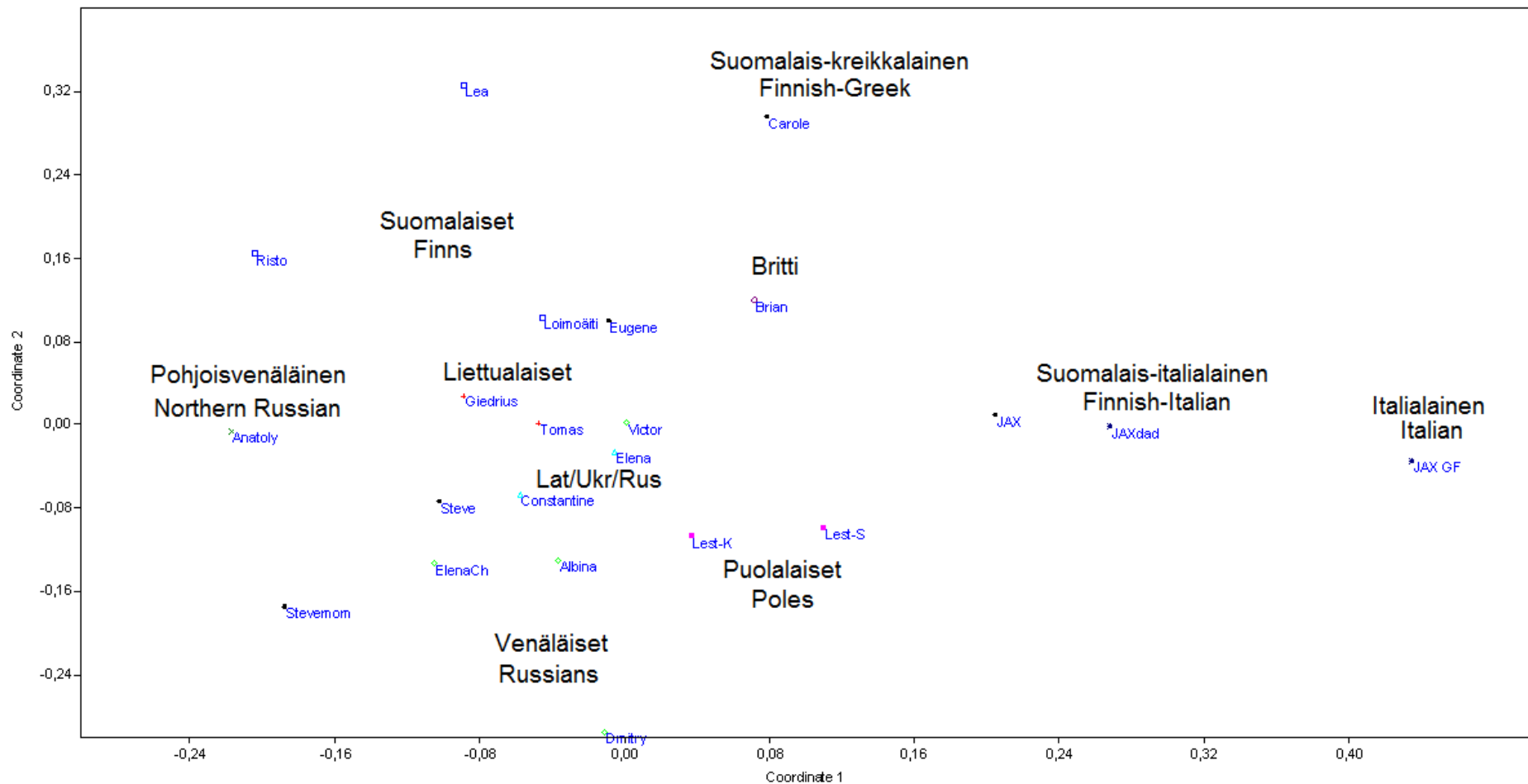
Lähimpänä pohjoisvenäläistä Anatolya näyttää olevan satakuntalais-hämäläinen ryhmä, joka alueellisin perustein edustanee "puhtainta" pohjoiskantasuomalaisuutta: länsisuomalaisen asutuksen leviäminen itään on alkanut juuri tältä alueelta noin 300 jaa. Satakuntalais-hämäläisten ja pohjoisvenäläisen läheisyys ei voi johtua myöhemmistä kontakteista, koska välissä ovat olleet itäsuomalaiset. Eteläpohjalaiset taas näyttävät olevan lähimpänä venäläistä Victoria ja liettualaista Giedriusta. Toisaalta osa länsisuomalaisista on hyvinkin kaukana näistä.

Closest to North Russian Anatoly is the Satakunta-Tavastian group, which could in the regional basis represent the "purest" Northern Proto-Finnicness: the spread of West Finnish settlement to the east has begun right in this area around 300 AD. The closeness of Satakunta-Tavastians cannot be due to any later contacts, because in between them there were East Finns. Southern Ostrobothnians seem to be closest to Russian Victor and Lithuanian Giedrius. On the other hand, part of West Finns are very distant from them.



Kun katsotaan kaikkia ulkomaalaisia suhteessa kolmeen keskenään mahdollisimman kaukaiseen suomalaiseen (Lea, Loimoäiti, Risto), asettuvat suomalaiset yhteen nurkkaan ja muut heidän ympärilleen. Pohjoisvenäläinen ja liettualaiset ovat lähimpänä suomalaisia. Venäläiset ja puolalaiset asettuvat lähekkäin liettualaisten taakse, paitsi Victor, jolla on kaukaisia esivanhempia myös Suomesta ja Latviasta. Carole sattuu olemaan Lean serkku, mutta Lean poistaminen ei muuta Carolen sijaintia.

When we look all the non-Finns compared to three Finns as distant from each other as possible (Lea, Loimoäiti, Risto), the Finns are found in one corner and others settle around them. Russians and Poles are together beyond Lithuanians, except Victor, who has distant ancestors also in Finland and Latvia. Carole happens to be a cousin for Lea, but removing Lea does not change the position of Carole in the figure.



Jonkinlainen maantieteellinen korrelaatio on havaittavissa: pohjoiset kansat ovat vasemmalla ja eteläiset oikealla. Erityisen mielenkiintoista on, että venäläisten ja pohjoisvenäläisten välille asettuvat länsieurooppalaistaustaiset Stevemom ja Steve.

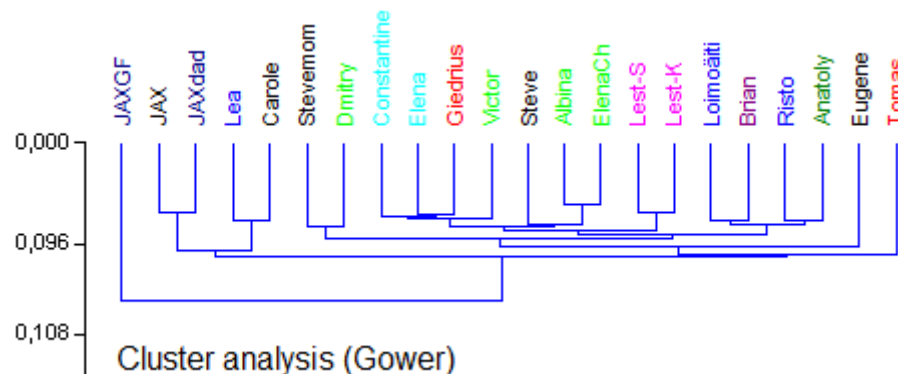
Ryhmittelyssä toistuu sama ilmiö kuin MDS-kuvaajissa: suomalaisista Risto on lähimpänä pohjoisvenäläistä Anatolya ja Lea on lähimpänä Carolea. Loimoäiti puolestaan ryhmittyy lähimmäs englantilaista Briania. Victorin länsivenäläinen ja latvialainen tausta näkyy siinä, että hän ryhmittyy liettualaisen Giedriuksen ja puoliukrainalaisten Constantinen ja Elenan kanssa.

On kuitenkin tärkeää ymmärtää, että tämänkaltainen ryhmittely pitää nähdä puhtaasti nykyiseen samankaltaisuuteen perustuvana; sitä ei voi pitää todisteena yhteisestä polveutumisesta. Kun ollaan tekemisissä automaalisten geenien kanssa, jotka peritään molemmilta vanhemmilta (toisin kuin äiti- tai isälinjat), mikään puumalli ei voi kuvata oikein polveutumishierarkiaa. Siihen tarkoitukseen verkostokuvaaja on paljon parempi havainnollistaja, koska se mahdollistaa useampia ulottuvuuksia, mutta tässä vaiheessa en ole vielä muuntanut pareittaisia IBS-arvoja SplitsTree-matriisiksi.

There is some geographic correlation: the northern peoples are on the left, and the southern peoples on the right. Especially interesting is that between Russians and Northern Russians there are West European -originating Stevemom and Steve.

In the cluster analysis we see the same as in the MDS plots: from the Finns Risto is closest to the North Russian Anatoly and Lea is closest to Carole. Loimoäiti is clusters together with the Englishman Brian. The distant Western Russian and Latvian background of Victor can be seen in his clustering together with the Lithuanian Giedrius and the half-Ukrainians Constantine and Elena.

However, it is important to understand that this kind of cluster analysis should be read only from the point of present similarity; it cannot be seen as a proof of shared descent. When we are dealing with autosomal genes, which are inherited from the both parents (unlike the maternal or paternal lineages), no tree model can correctly simulate the hierarchy of descent. For that purpose a network is far better illustration, as it takes into account more dimensions, but at this point I haven't yet converted the similarity matrix of all the pairwise IBS-values to a SplitsTree matrix.



Uudet näytteet ovat tervetulleita!

New samples are welcome!

Toistaiseksi Relative Finder -tiedostoja on liian vähän analysoitavaksi.

So far there are too few Relative Finder files to analyze them.