

**TARTU ÜLIKOOL**  
**BIOLOOGIA-GEOGRAAFIATEADUSKOND**  
**MOLEKULAAR JA RAKUBIOLOOGIA INSTITUUT**  
**EVOLUTSIOONILISE BIOLOOGIA ÕPPETOOL**

**Kedy Kivirand**

**Ebatüüpiliste mtDNA kontrollregiooni HVS-I  
järjestusmotiividega haplotüüpide klassifitseerimine  
haplogrupp U näitel**

**MAGISTRITÖÖ**

**Juhendaja: PhD Ene Metspalu**

**Tartu 2006**

# Sisukord

<b>SISUKORD</b> .....	<b>2</b>
<b>KASUTATUD LÜHENDID</b> .....	<b>3</b>
<b>PÕHILISTE ANTUD MAGISTRITÖÖS KASUTATUD TERMINITE DEFINITSIOONID</b> .....	<b>3</b>
<b>SISSEJUHATUS</b> .....	<b>4</b>
<b>1. KIRJANDUSE ÜLEVAADE</b> .....	<b>5</b>
1.1. MITOKONDRI ISELOOMUSTUS .....	5
1.2. HOMO SAPIENS SAPIENS´I PÄRITOLU .....	9
1.3. MTDNA HAPLOGRUPPIDE NOMENKLATUUR JA FÜLOGEOGRAAFILINE JAOTUS .....	10
1.4. SUPERHAPLOGRUPP U .....	17
1.4.1. Haplogrupp U5 .....	20
1.5. MTDNA ANALÜÜSIMISE JA TULEMUSTE TÕLGENDAMISEGA SEOTUD PROBLEEMID .....	23
<b>2. MATERJAL JA METOODIKA</b> .....	<b>26</b>
2.1. KATSEMATERJAL .....	26
2.2. DNA AMPLIFITSEERIMINE .....	27
2.3. RFLP ANALÜÜS .....	29
2.4. AMPLIFITSEERITUD PRODUKTI PUHASTAMINE .....	30
2.5. DNA SEKVENEERIMINE .....	31
2.6. ANDMEANALÜÜS .....	32
<b>3. TULEMUSED JA ARUTELU</b> .....	<b>33</b>
3.1. TÖÖ EESMÄRK .....	33
3.2. MTDNA KODEERIVA PIIRKONNA DEFINEERIVATE NPS ANALÜÜS .....	33
3.3. HVS-I PIIRKONNA NUKLEOTIIDSETE JÄRJESTUSTE ANALÜÜS .....	35
3.4. MTDNA-DE TÄISJÄRJESTUSE MÄÄRAMINE JA FÜLOGENEETILISTE SÜNTEESIPUUDE ANALÜÜS .....	38
<b>4. KOKKUVÕTE JA JÄRELDUSED</b> .....	<b>44</b>
<b>SUMMARY</b> .....	<b>45</b>
<b>TÄNUAVALDUSED</b> .....	<b>46</b>
<b>KASUTATUD KIRJANDUS</b> .....	<b>47</b>
<b>LISAD</b> .....	<b>56</b>

## Kasutatud lühendid

(alam)hg	(alam) haplogrupp
AMH	anatomiliselt tänapäeva inimene ( <i>anatomically modern human</i> )
bp/kbp	aluspaar ( <i>base pair</i> )/tuhat (kilo) aluspaari
DNA	desoksüribonukleiinhape
HVS-I/II	mitokondriaalse DNA kontrollregiooni I/II hüpervarieeruv segment
LGM	viimane jääaja maksimum ( <i>the Last Glacial Maximum</i> )
MRCA	viimane ühine eellane ( <i>the most recent common ancestor</i> )
mtDNA	mitokondriaalne DNA
np(s)	nukleotiidi positsioon(id)
PCR	polümeraasne ahelreaktsioon ( <i>polymerase chain reaction</i> )
rCRS	Cambridge'i mtDNA järjestus ( <i>revised Cambridge reference sequence</i> )
RFLP	restriktsioonifragmentide pikkuse polümorfism ( <i>restriction fragment length polymorphism</i> )

## Põhiliste antud magistritöös kasutatud terminite definitsioonid

haplotüüp	iseloomulike polümorfismidega mtDNA järjestus; hõlmab kõiki identseid järjestusi
haplogrupp	haplotüüpide, mis jagavad antud iseloomulikke polümorfisme monofüleetiline klad
klad, klaster	monofüleetilised ühikud
koalestsentsi aeg	aeg MRCA-ni
fülogeograafia	uuring genealoogiliste liinide geograafilisest jaotusest
tähekujuline fülogenees	fülogeneetilise puu ühest haplotüübist/kesksest sõlmest toimub uute harude tähekujuline hargnemine

## Sissejuhatus

Mutatsioonid, mis tekivad inimese genoomis kannavad endas väärtuslikku informatsiooni, mille abil saab analüüsida inimpopulatsioonide ajalugu. Mitokondriaalne DNA (mtDNA) ja Y kromosoom on olnud ligi kaks aastakümnet inimese evolutsioonibioloogia huviorbiidis, kuna neile mõlemale on omane rekombinatsiooni puudumine, kõrge varieeruvus ja uniparentaalne pärandumine. Viimane tähendab seda, et Y kromosoom pärandub edasi vaid isaliinis ja mtDNA emaliini pidi. Haplogruppide leviku mustrite uurimine erinevates populatsioonides võimaldab konstrueerida fülogeneetilisi puid ja võrgustikke, mis annavad meile aimu inimpopulatsioonide eelajalooliste ja ka hilisemate geenivoolude kohta. Inimkonna fülogeneetilise puu baasil on võimalik mõista üksikutes populatsioonides leiduvate liinide asukohta globaalsel fülogeneetilisel puul, hargnemiste suundi ja toetudes koalestsentsiteooriale, ajalist sügavust. Molekulaarantropoloogia peamiseks küsimuseks on kus ja millal inimpopulatsioonide diversiteet tekkis.

MtDNA fülogeneetiline klassifikatsioon on siiani peamiselt baseerunud kontrollregiooni varieeruvuse ja kodeeriva regiooni polümorfisuse klassifikatsioonide sünteesil. Selleks, et täpsustada olemasolevat mtDNA nomenklatuuri ja lahendada seni ebaselgeid kohti fülogeneetilisel puul, on viimastel aastatel üha suuremat tähelepanu hakatud pöörama mtDNA genoomide täisjärjestustele, mis annavad täpsema fülogeneetilise resolutsiooni kui eelpool nimetatud lähenemisviisid.

Antud magistritöö keskendub mtDNA superhaplogrupi U alamharude haplogrupeerimisega seotud probleemidele ja ebaselgustele. Tähelepanu all on ebatüüpiliste HVS-I järjestusmotiividega mtDNA-d, mis seni on defineeritud U\*-na. Probleemi uurimiseks ja lahendamiseks kasutatakse erinevaid meetodikaid nagu restriktsoonifragmentide pikkuse polümorfismi ja kodeeriva regiooni defineerivate mutatsioonide analüüsi ning HVS-I piirkonna ja kodeeriva regiooni nukleotiidses järjestuses määramist.

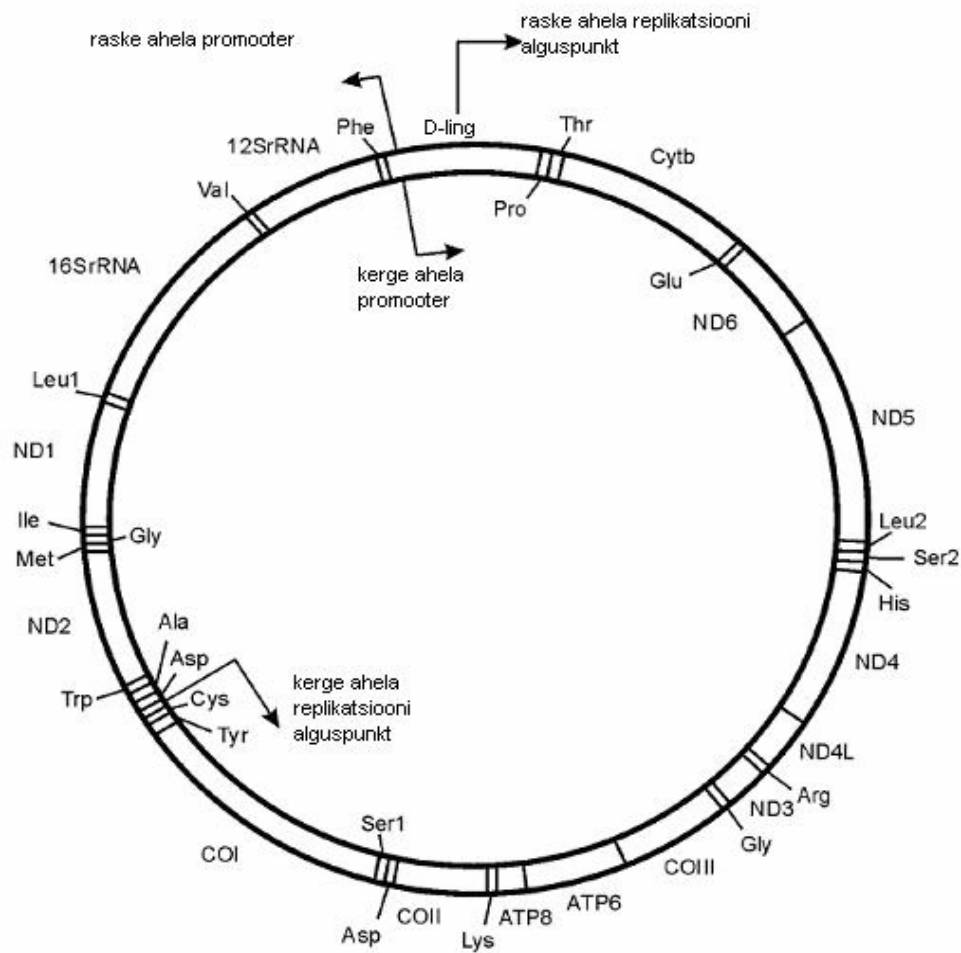
# 1. KIRJANDUSE ÜLEVAADE

## 1.1. Mitokondri iseloomustus

On mitmeid põhjuseid, miks inimese ja teiste kõrgemate loomade mitokondriaalne DNA on leidnud nii laialdast kasutust fülogeneetilistes uuringutes. Mitokondritel on olemas tuumaväline genoom – iga mitokondri maatriksis on keskmiselt 5 – 10 mtDNA molekuli, somaatilises rakus on neid ligikaudu 1000 – 10 000 (Lightowers *et al.* 1997). Enamasti on indiviidi mtDNA molekulid organismis identsed, st esineb homoplasmia. Imetajate mtDNA on tsirkulaarne kaheaheelaline molekul, mille pikkus erinevatel liikidel on vahemikus 15 000 - 17 000 bp. Inimese mtDNA pikkus on 16 569 bp (Anderson *et al.* 1981; Andrews *et al.* 1999)(joonis 1).

Mitokondriaalne DNA kodeerib 2 rRNA-d, 22 tRNA-d, tsütokroom C subühikuid ja ATPaasi subühikuid, mis põhiliselt on seotud mitokondri sisemembraanil toimuva elektrontranspordi ja oksüdatiivse fosforüleerimise rajaga ehk siis rakusisese energia tootmise süsteemiga (Anderson *et al.* 1981). Enamus hingamisahela ensüümkompleksi valkudest kodeeritakse raku tuumas ja transporditakse mitokondrisse tsütoplasmast. MtDNA-l eristatakse kahte ahelat – guanosiinirikas raske ehk H-ahel ja tsütosiinirikas kerge ehk L-ahel. Mitokondriaalselt DNA-lt kodeeritavast 37 geenist asub 28 H-ahelal ja 9 L-ahelal (Wallace *et al.* 1999). MtDNA on väga kompaktne ning erinevalt tuuma DNA-st 93% ulatuses kodeeriv. Puuduvad intronid ja mõlemat ahelat transkribeeritakse kogu ahela ulatuses (Anderson *et al.* 1981). Suurim mittekodeeriv regioon mtDNA-s on 1,1 kbp pikkune segment D-ling ehk kontrollregioon, mis sisaldab mtDNA replikatsiooni ja transkriptsiooni kontrollelemente (Lightowers *et al.* 1997)(joonis 1).

Uued mitokondrid tekivad jagunemise teel, kusjuures nende jagunemine ja kasv ei ole seotud tuuma jagunemisega. Fülogeneetiliste uuringute aspektist lähtudes on mtDNA väga sobiv uurimismaterjal mitmel põhjusel. Esiteks on mtDNA-le omane unilineaarne emaliinis päritavus (Giles *et al.* 1980), mille põhjuseks on see, et spermatoosidi mtDNA degradeeritakse ubiquitinüleerimise teel vahetult peale viljastamist (Hastings 1992;



**Joonis 1.** Inimese mtDNA (16,6 kb) geenikaart, kus on ära märgitud oksüdatiivse fosforüleerimise kompleksid ja nendega seotud geenide asukohad: kompleks I (ND1, 2, 3, 4, 4L, 5, 6); kompleks III (Cytb); kompleks IV (COI, COII, COIII). Kaardil on näidatud 12SrRNA, 16SrRNA ja 22 tRNA geenide paiknemine ning D-lingu piirkond. MtDNA kergelt ja raskelt ahelalt transkribeeritavad geenid on joonisel märgitud vastavalt sisemisel või välimisel ringjoonel. Nooled märgivad transkriptsiooni promooterite ja replikatsiooni alguspunktide asukohta. <http://www.mitomap.org>

Thomson *et al.* 2003; Sutovsky *et al.* 2004). MtDNA unilineaarne pärandumine annab meile võimaluse uurida inimpopulatsioonide emaliini ajalugu. Teiseks on mitokondriaalsele DNA-le omane rekombinatsiooni puudumine näiteks erinevalt tuuma DNA-st (Merrywether *et al.* 1991; Piganeau ja Eye-Walker 2004). See võimaldab säilitada erinevate mutatsioonide kombinatsioone mtDNA geenitiigis, sest mitokondriaalne DNA on geneetiliselt kui üks lookus, olenemata tema kõrgest

koopaarvust rakus. Lisaks iseloomustavad mtDNA-d suhteliselt kõrge mutatsioonide tekke kiirus ja intensiivse valiku puudumine enimvarieeruvatele piirkondadele.

### **MtDNA kontrollregioon**

Kontrollregioon ehk D-ling on mitokondriaalses DNA-s peamine regulatiivne piirkond, kus paiknevad mtDNA raske ahela replikatsiooni alguspunkt ja promootorid mõlema ahela transkriptsiooniks (Desjardins ja Morais 1990; Lightowers *et al.* 1997). Antud piirkond on mtDNA mittekodeeriv ala, millele on omane väga kõrge varieeruvus – 90% mitokondri genoomi varieeruvusest on kontsentreerunud just sellesse piirkonda (Greenberg *et al.* 1983). Kontrollregioon asub nukleotiidi positsioonide (nps) 16024 ja 576 vahel. Polümorfised piirkonnad kontrollregioonis on peamiselt koondunud kolme hüpervarieeruvasse segmenti (HVS) – HVS-I (nps 16024-16365), HVS-II (nps 73–340) ja HVS-III (nps 438–574), millest kõige kõrgema varieeruvusega on HVS-I (Lutz *et al.* 1998; Vigilant *et al.* 1991).

### **Mutatsioonide teke ja fikseerumine**

Keskmiselt akumul eeruvad mutatsioonid mtDNA-s 10 korda kiiremini, kui rakutuumas (Brown *et al.* 1979; Ingman ja Gyllenstein 2001). Mitokondriaalsete mutatsioonide tekkepõhjuseid arvatakse olevat kolm – esiteks hingamisahelast tekkivate vabade radikaalide oksüdatiivne mõju mtDNA-le, teiseks võrreldes tuumaga on mitokondriaalne reparatsiooni süsteem nõrgem ning ei suuda kõrvaldada kõiki tekkinud mutatsioone ning kolmandaks mtDNA-d kaitsvate valkude, nagu näiteks histoonide puudumine (Bogenhagen 1999). Selleks, et mutatsioon kanduks edasi järgnevasse põlvkondadesse, peab see tekkima organismi sugurakkudes.

Mutatsioonide tekke kiirus ja fikseerumine mitokondriaalses DNA-s sõltub mtDNA piirkonnast ja mutatsiooni iseloomust. On teada, et mitokondriaalse DNA kodeeriva regiooni mittesünonüümsete asenduste määr on sarnane tuuma genoomiga. MtDNA sünonüümsete asenduste määr ja mitokondriaalsete rRNA geenide mutatsiooni kiirus on kuni 20 korda ning tRNA geenide puhul isegi *ca* 100 korda kõrgem kui sarnase funktsiooniga geenidel tuumas (Pesole *et al.* 1999). Transitsioonid, kus puriin asendub

puriiniga või pürimidiin pürimidiiniga, toimuvad 12 - 37 korda sagedamini kui transversioonid, kus toimub puriin-pürimidiin asendumine (Meyer *et al.* 1999 koos viidetega).

Eriti kõrge mutatsioonikiirus on iseloomulik mtDNA mittekodeerivas alas paiknevatele HVS-I ja HVS-IIIle, kuna see piirkond ei ole suure selektiivse surve all ning tekkinud mutatsioonid saavad suure tõenäosusega fikseeruda. Inimese HVS-I piirkonna muteerumise määraks on arvatud 36% miljoni aasta kohta (Foster *et al.* 1996) ning kogu mtDNA genoomil 2-4% miljoni aasta jooksul (Cann *et al.* 1987; Torroni *et al.* 1994a). Fülogeneetilises uuringutes on peamiselt kasutatud koalestsentsiaja arvutamiseks mutatsioonimäära, mis väljendab aega, mille jooksul tekib HVS-I piirkonnas üks transitsioon. Selleks ajaks loetakse 20 180 aastat (Forster *et al.* 1996). Tõenäosus, et tekkinud mutatsioonid mtDNA-s fikseeruvad, sõltub sellest, kas mutatsioon on organismile kasulik, kahjulik või neutraalne.

### **Fülogeneetilised puud ja võrgustikud**

Kõige levinum viis kuidas väljendada erinevate geenide või organismide fülogeneetilisi suhteid, on seda teha fülogeneetilise puuna. Inimpopulatsioonide fülogeneesi uurimiseks ei ole aga antud meetod kõige optimaalsem. Selle põhjuseks on suhteliselt lühikesed geneetilised distantsid erinevate populatsioonide/indiviidide vahel, tihtipeale väga suured valimid, paralleelsed ja tagasipöörduvad mutatsioonid. Kõiki võrdselt võimalikke puid saab kokku võtta fülogeneetilisse võrgustikku, mis sisaldab ka lahendamata retikulatsioonid, kui alternatiivseid evolutsiooni teid. Väiksemate valimite jaoks on otstarbekas kasutada *reduced median* (Bandelt *et al.* 1995) ning suuremahuliste andmete töötlemiseks *median joining network*'i (Bandelt *et al.* 1999). Selleks, et näidata fülogeneetilise puu või võrgustiku hargnemiste ajalist kihistumist, peab sel olema juur. Kasutatakse ka nn välisgruppi (*outgroup*), milleks on tavaliselt viimasest ühisest eellasest (MRCA) eraldunud erinev liin. Inimese mtDNA analüüsil on selleks sageli šimpansi vastav järjestus (Vigilant *et al.* 1991; Ingman *et al.* 2000; Maca-Meyer *et al.* 2001) või neandertallase mtDNA kontrollregiooni järjestus (Klings *et al.* 1997; 1999a; Ovchinnikov *et al.* 2000). Kui välisgruppi pole võimalik määratleda, siis kasutatakse

keskpunkti juurimist (*midpoint rooting*). Keskpunkti juur paigutatakse nii, et distantss kõikidest puu terminaalsetest sõlmedest oleks minimaalne (Cann *et al.* 1987).

## 1.2. *Homo sapiens sapiens*'i päritolu

Antropoloogia, kui teadusharu, keskseks küsimuseks on alati olnud tänapäeva inimese (*Homo sapiens sapiens*) päritolu. Sellele küsimusele vastuse leidmiseks on välja pakutud kaks erinevat hüpoteesi. Esiteks multiregionaalne hüpotees (*Multiregional model*), mis põhineb fossiilsete leidude morfoloogiliste markerite interpretatsioonil. Multiregionaalse mudeli pakkus esmakordselt välja Franz Weidenreich, mis hiljem teiste poolt edasiarendamist leidis (Wolpoff *et al.* 1984; Wolpoff ja Caspari 1997). Vastavalt sellele teooriale, on kõik tänapäeva inimesed arenenud ühisest arhailisest esivanemast *Homo erectus*'est, kes lahkus Aafrikast ligikaudu 1,8 miljonit aastat tagasi. Erinevatel kontinentidel toimus *in situ* inimpopulatsioonide järk-järguline evolutsioon anatoomiliselt tänapäeva inimeseks (AMH) läbi pidevalt toimuva geenivoolu. "Multiregionalismile" vastupidine hüpotees on Aafrikast hiljutise väljarände (*Out of Africa*) ehk Eedeni aia mudelil (*Garden of Eden*), mille järgi AMH arenes välja Aafrikas suhteliselt hiljuti ning koloniseeris järk-järgult kogu maailma, tõrjudes välja eelnevalt eksisteerinud arhailised hominiidid (Cann *et al.* 1987; Stringer ja Andrews 1988).

Tuginedes viimasel aastakümnele tehtud mtDNA ja Y kromosoomi fülogeneetiliste uuringute tulemustele ning paleoantropoloogilistele leidudele (Foley 1998; Stringer 2000; 2003), on Aafrikast hiljutise väljarände teooria leidnud enim kinnitust. On näidatud, et Aafrika populatsioonides on mtDNA fülogeneetilise puu harude ajaline sügavus suurim ja mtDNA nukleotiidses järjestuses diversiteet kuni kaks korda kõrgem võrreldes teiste kontinentide populatsioonidega, kus diversiteet on palju madalam, puu harude ajaline sügavus väiksem ning puude struktuur valdavalt tähe-kujuline (Ingman *et al.* 2000; Yu *et al.* 2002).

Paleoantropoloogilised leidude järgi ilmub AMH esmakordselt Aafrikas ja ligikaudu 160 000 aastat tagasi (Clark *et al.* 2003). MtDNA ja Y kromosoomi liinide põhjal on

AMH oletatavaks koalestsentsiajaks arvatud ligikaudu 150 000 - 200 000 aastat tagasi (Underhill *et al.* 1997; Ingman *et al.* 2000; Maca-Meyer *et al.* 2001; Takahata *et al.* 2001). Kuigi mtDNA ja Y kromosoomi tänapäeva geenitiigi koalestsentsiajad ei pruugi tingimata ühtida liigi tegeliku koalestsentsiajaga, on paleo- ja molekulaar-antropoloogiliste andmete sarnasus märkimisväärne. AMH Aafrikast väljarände alguse ajaks loetakse 50 000 – 80 000 aastat tagasi (Watson *et al.* 1997; Ingman *et al.* 2000; Maca-Meyer *et al.* 2001).

### 1.3. MtDNA haplogrupid nomenklatuur ja fülogeograafiline jaotus

Esimesed inimpopulatsioonide mtDNA varieeruvuse uuringud põhinesid kas restriksioonifragmentide pikkuse polümorfismi analüüsil (RFLP) või kontrollregiooni hüpervarieeruvate segmentide (tavaliselt HVS-I piirkond) sekveneerimisel. Praegu kasutusel olev mtDNA nomenklatuur on kontrollregiooni varieeruvusel (Richards *et al.* 1996) ja RFLP analüüsil (Torroni *et al.* 1994b) põhinevate klassifikatsioonide süntees (Torroni *et al.* 1996; Richards *et al.* 1998; Macaulay *et al.* 1999).

Olemasolevate mtDNA järjestuste mitmekesisus on tekkinud mutatsioonide akumulatsioonide tõttu ja nende fikseerumisel läbi geenitriivi. Erinevad mtDNA järjestused ehk haplotüübid saab jagada monofüleetilistesse rühmadesse ehk haplogruppidesse (hg). Haplogrupid on tähistatud suurte tähtedega ja jagunevad tavaliselt edasi alam-haplogruppideks. Enamusele haplogruppidele on omane spetsiifiline geograafiline levikuala – võib öelda, et haplogrupid on kontinendi või etnose spetsiifilised. MtDNA fülogeneetilise puu ja nomenklatuuri täiustamine on pidevalt toimuv protsess. Viimastel aastatel on üha suuremat tähelepanu hakatud pöörama mtDNA-de täisjärjestustest saadavale informatsioonile, mis annab parema võimaluse emaliinide fülogeneetika lahendamiseks (Finnilä *et al.* 2001; Herrnstadt *et al.* 2002; Kivisild *et al.* 2002; 2005; Reidla *et al.* 2003; Palanichamy *et al.* 2004; Achilli *et al.* 2005).

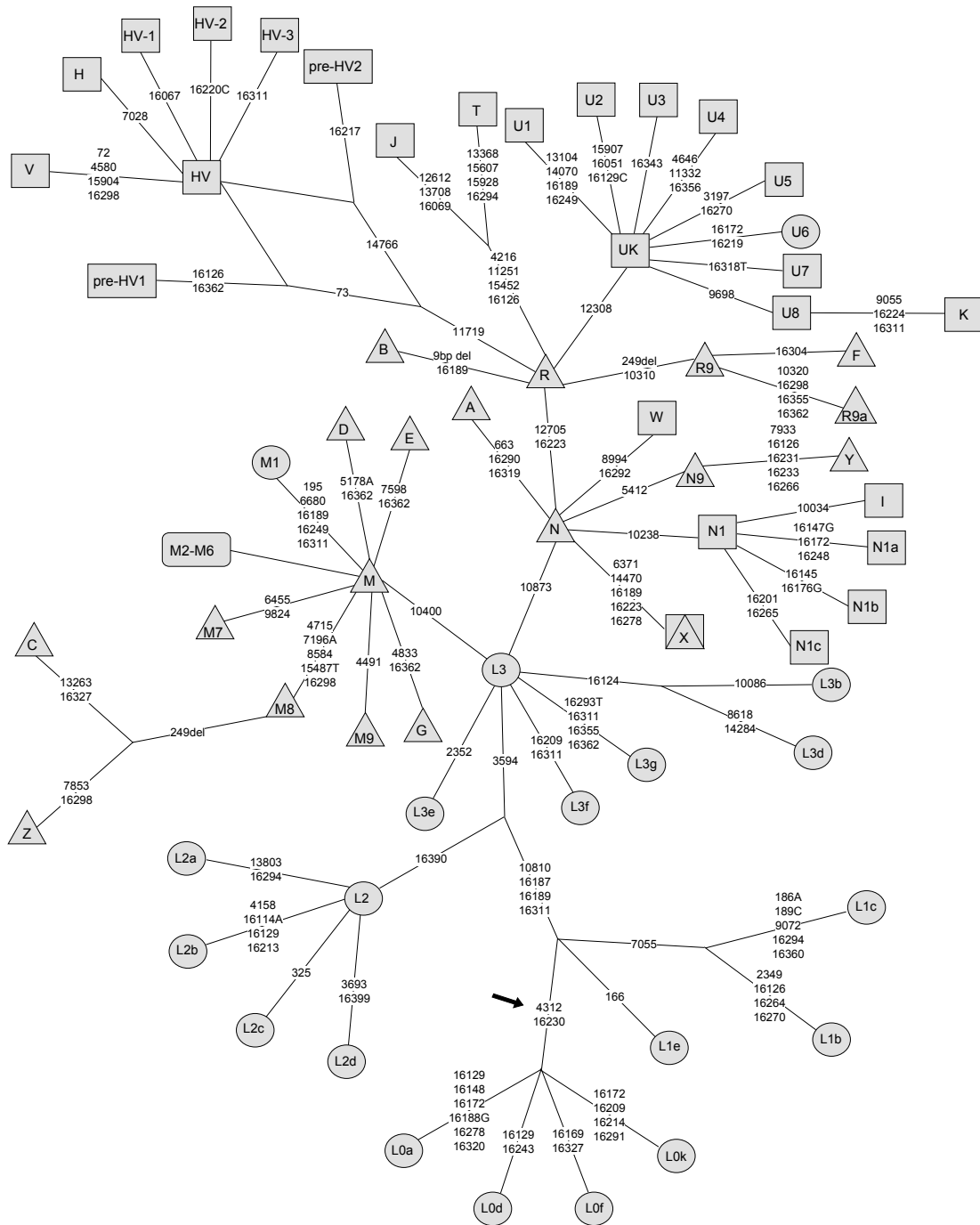
Globaalse inimpopulatsioonide mtDNA puu põhiline skeem on kujutatud joonisel 2. Puu juure asukoht on haplogruppide L0 ja L1 vahel, jagades fülogeneetilise puu kaheks suureks osaks – L0 ja ülejäänud.

### **Makrohaplogrupp L**

Kõige vanem inimese mtDNA haru on makrohaplogrupp L, mis jaguneb edasi neljaks põhiliseks haplogrupiks - L0, L1, L2 ja L3 (joonis 2). L0, L1 ja L2 jagavad restriktioonisaiti HpaI nps 3592 (3594T/C) ning on Aafrikale omased kolm peamist suur haplogruppi (Chen *et al.* 1995; Salas *et al.* 2002; Mishmar *et al.* 2003).

Kui haplogrupp L0, mida peetakse üheks kõige varasemaks L haruks (Chen *et al.* 1995; 2000) on valdavalt levinud Ida- ja Lõuna-Aafrikas, siis haplogrupp L1 on levinud üle kogu Aafrika. Alamhaplogrupp L1b on tõenäoliselt tekkinud Lääne-Aafrikas, kuid populatsioonide ekspansioonid ja migratsioonid on teda kandnud ka Põhja- ja Kesk-Aafrikasse, kus esineb ka alamhaplogrupp L1c. Vanimaks L1 alamhaplogruppidest peetakse L1e-d, mis on levinud Ida- ja Kesk-Aafrikas (Salas *et al.* 2002). Haplogrupi L2 ekspansiooniga peetakse L0 ja L1-st hilisemaks, koalestsentsiajaks on arvatud 59 000 – 78 000 aastat (Chen *et al.* 1995; 2000). Alamhaplogrupp L2a on kõige suurema sagedusega haplogrupp Aafrikas, hõlmates ühte kolmandikku kõigist Aafrika põlistest liinidest. Alamhaplogrupid L2b, L2c ja L2d on levinud peamiselt Aafrika lääneosas, vanimaks neist peetakse L2d (Chen *et al.* 2000; Pereira *et al.* 2001; Salas *et al.* 2002).

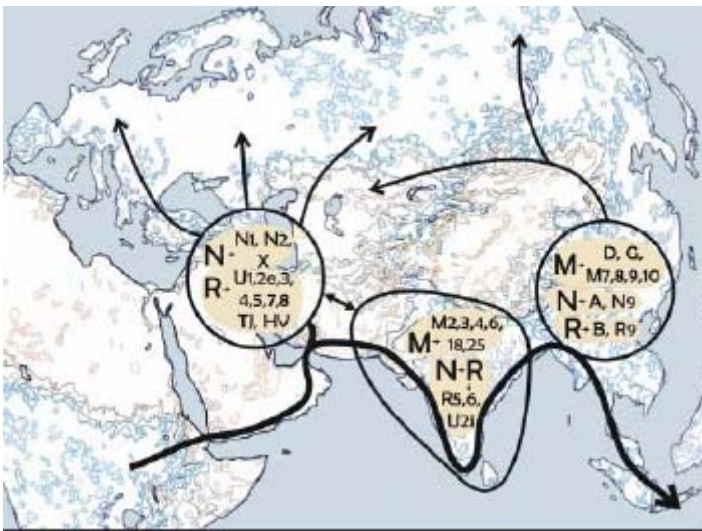
Aafrika mtDNA harud, millel puudub HpaI restriktioonisait nps 3592, kuuluvad superhaplogruppi L3 (joonis 2). Kõikidel Euraasia, Okeania ja Ameerika haplogruppidel puudub samuti eelnimetatud sait, seega on nad ekspandeerunud superhaplogrupist L3 (Chen *et al.* 1995; Maca-Meyer *et al.* 2001; Kivisild *et al.* 2002). L3 ekspansiooni algusajaks loetakse hinnanguliselt 50 000 – 70 000 aastat tagasi (Ingman *et al.* 2000; Maca-Meyer *et al.* 2001). Kuna L3 sagedus ja diversiteet on Ida-Aafrikas suhteliselt suur, peetakse just seda piirkonda tema tekkekohaks (Watson *et al.* 1996). L3 alamhaplogruppidest on kõige sagedasem L3e, mis tõenäoliselt hakkas ekspandeeruma



**Joonis 2.** Inimese mtDNA skemaatiline fülogeneesi puu, mille koostamisel on lähtutud HVS ja kodeeriva piirkonna defineerivatest mutatsioonidest. Ringid, kolmnurgad ja ruudud tähistavad vastavaid geograafilisi piirkondi nagu Aafrika, Lääne-Euraasia ja Ida-Euraasia/Ameerika. Puu harudel toodud numbrid tähistavad antud haru iseloomustavaid polümorfisme; numbril järel olev täht märgib transversiooni, transitsioone eraldi ei tähistata. Puu juure asukoht on joonisel tähistatud noolega, välisgrupina on kasutatud šimpansi vastavat järjestust (Torroni *et al.* 2006). Puu koostamiseks on kasutatud järgmisi allikaid: Chen *et al.* 2000; Richards *et al.* 2000; Bandelt *et al.* 2001b; Finnilä *et al.* 2001; Maca-Meyer *et al.* 2001; Herrnstadt *et al.* 2002; Kivisild *et al.* 2002; 2003b; Kong *et al.* 2003; Salas *et al.* 2002; 2004; Macaulay *et al.* 2005.

46 000 aastat tagasi Kesk- või Ida-Aafrikas ja haruneb omakorda mitmeteks alamklastriteks (Bandelt *et al.* 2001b).

Pikalt on vaieldud, kuidas toimus väljaränne Aafrikast. On pakutud vähemalt kahte iseseisvat migratsiooni (Kivisild *et al.* 2000), kuid samas on ka arvamusi, et toimus ainult üks väljaränne (Foster *et al.* 2001). Euraasias levinud haplogrupid kuuluvad põhiliselt kahte suurde makrohaplogruppi M ja N, millest viimasest haruneb omakorda makrohaplogrupp R. Makrohaplogrupid M ja N hargnevad superhaplogrupist L3, mille teised alamharud on levinud vaid Aafrikas. Seetõttu arvatakse, et makrohaplogruppide M ja N eellas-mtDNA variandid olid esindatud Euraasia esimeste koloniseerijate geenitiigis, kuid antud makrohaplogruppide divergeerumine toimus juba kohapeal (Metspalu *et al.* 2004)(joonis 3).



**Joonis 3.** Euraasia arvatav asustamine AMH poolt (Metspalu *et al.* 2004).

### **Makrohaplogrupp M**

Makrohaplogrupi M divergeerumise alguseks Ida- ja Lõuna-Aasias loetakse ligikaudu 50 000 aastat tagasi (Kivisild *et al.* 2002; Metspalu *et al.* 2004). M koondab enda alla haplogrupid M1-M9 ja E, G, D, C ning Z (joonis 2). Huvitava fülogeograafilise levikuga on haplogrupp M1, mis on levinud Ida- ja Põhja-Aafrikas (Rando *et al.* 1998; Quintana-Murci *et al.* 1999), Lähis-Idas (Macaulay *et al.* 1999) ja Indias (Kivisild *et al.* 1999a),

kus tema diversiteet on suurim. Selle alusel on arvatud, et haplogrupp M1 ekspandeerus Indias ning hiljem leidis aset tagasiränne Aafrikasse (Maca-Meyer *et al.* 2001). Haplogrupid M2 ja M3 haaravad enda alla enamuse Lõuna-Aasias levinud mtDNA liinidest (Kivisild *et al.* 2003b). Haplogrupid D ja C esinevad kõrge sagedusega Kirde-Aasias ja Ameerika indiaanlaste seas (Torrioni *et al.* 1993). Haplogrupid G ja Z on enamlevinud Kirde-Aasias ning haplogrupp E Kagu-Aasias (Schurr *et al.* 1999; Qian *et al.* 2001; Kivisild *et al.* 2002; Yao *et al.* 2002; Yao ja Zhang 2002).

### **Makrohaplogrupp N**

Makrohaplogrupp N, teine suur haru, mis pärineb superhaplogrupist L3, koondab enda alla enamuse Lääne-Euraasia haplogruppidest [R(H, V, T, J, U), I, W, X], aga on esindatud ka Ida-Aasia mtDNA geeniitiigis (A, B, R9,Y) ja Ameerika indiaanlaste seas (A, B, X)(joonis 2).

Haplogrupil A on sarnane levik eelpool mainitud haplogruppidega D ja C (Starikovskaya *et al.* 1998). Haplogrupid B ja R9 on levinud eelkõige Kagu-Aasias ning nende divergeerumine toimus arvatavasti Lõuna-Hiinas või/ja Kagu-Aasias (Yao *et al.* 2002). Hg B on suhteliselt sage ka Ameerika indiaanlaste seas (Torrioni *et al.* 1993; Kivisild *et al.* 2002; Yao *et al.* 2002; Yao ja Zhang 2002) ning teatud populatsioonides, nagu polüneeslased, on hg B peaaegu ainus esinev haplogrupp (Ballinger *et al.* 1992). Haplogrupp Y esineb kõrge sagedusega Ida- ja Kirde-Aasias ning sporaadiliselt üle kogu Aasia (Ballinger *et al.* 1992; Comas *et al.* 1996; Schurr *et al.* 1999; Bermisheva *et al.* 2002; Kivisild *et al.* 2002). Haplogrupid I ja W on esindatud väga madala sagedusega Euroopas (alla 2%)(Kivisild *et al.* 1999b; Richards *et al.* 2000), hg W ka Indias (Kivisild *et al.* 1999b). Huvitav on see, et soomlaste hulgas on hg W sagedus ligikaudu 11%, mis viitab kas asutaja efektile või on populatsioon läbinud pudelikaela (Meinila *et al.* 2001). Haplogrupile X on omane huvitav geograafiline levik – teda võib leida Euroopast (Torrioni *et al.* 1996), kuid ka Põhja- ja Ida-Aafrikast, Lähis-Idast, Siberist ning Ameerika indiaanlaste seast (Foster *et al.* 1996; Brown *et al.* 1998; Smith *et al.* 1999; Derenko *et al.* 2001).

## **Makrohaplogrupp R**

Makrohaplogrupist N harunev makrohaplogrupp R moodustab monofüleetilise sõlme, kust omakorda hargnevad enamuse Lääne-Euraasiale iseloomulikke mtDNA haplogruppe nagu H, V, T, J, I, U, W ja X ning lisaks eelpool mainitud haplogrupid B ja R9 (joonis 2).

Vaadates Euroopa haplogruppide fülogeneetilist pilti, on see üsnagi homogeenne – geneetilised distantid erinevate populatsioonide vahel on suhteliselt lühikesed, kuigi on ka mõned erandid nagu näiteks saamid, sardiinlased ja baskid (Cavalli-Sforza *et al.* 1994). Sügavam fülogeneetiline analüüs on siiski välja toonud haplogruppide levikumustrite erinevusi (Richards *et al.* 2000; 2003).

Haplogrupi H esialgne ekspansioon toimus hinnanguliselt 23 200 – 28 400 aastat tagasi Lähis-Idas (Richards *et al.* 2000). Euroopas levinud H vanus on tunduvalt noorem – ligikaudu 18 400 aastat (Achilli *et al.* 2004). Haplogrupp H ja tema alamharud on Euroopas enimlevinud emaliini klastrid (Achilli *et al.* 2004; Loogväli *et al.* 2004). Kõrgeim esinemissagedus (40-60%) on Lääne- ja Põhja-Euroopas, kuid antud haplogruppi leidub ka Lähis-Idas (Torroni *et al.* 1998; Richards *et al.* 2000), Põhja-Aafrikas (Rando *et al.* 1998) ja Kesk-Aasias (Metspalu *et al.* 1999; Quintana-Murci *et al.* 2004). Madalama sagedusega on H esindatud Lõuna-Aasias (Kivisild 1999a; 1999b; 2003b) ja Siberis (Torroni *et al.* 1998; Derbeneva *et al.* 2002a; Derenko *et al.* 2003; Fedorova *et al.* 2003).

Haplogrupiga H fülogeneetiliselt lähedased haplogrupid (preHV)1, HV1 ja HV2 on peamiselt esindatud Lähis-Idas ja Kaukaasias (Metspalu *et al.* 1999; Richards *et al.* 2002; 2003), kuid madala sagedusega ka Euroopa populatsioonides (Metspalu *et al.* 1999; Tambets *et al.* 2000a). HV1 klaad, mis on levinud Vahemere piirkonnas ja Kirde-Aafrikas, ekspandeerus tõenäoliselt Kaukaasias ligikaudu 30 000 aastat tagasi (Tambets *et al.* 2000).

Haplogrupp V on samuti H haplogrupile fülogeneetiliselt lähedane, kuid tema esinemissagedus on enamikes Euroopa populatsioonides vaid 1-6%, välja arvatud baskid

ja kataloonlased Ibeeria poolsaarel (sagedus 20%)(Torrioni *et al.* 1998; 2001) ning saamid skandinaavia põhjaosas (41,6%)(Tambets *et al.* 2004) Arvatakse, et hg V ekspansioon Ibeeria poolsaarel ja levimine Euroopas toimus 10 000 – 15 000 aastat tagasi peale viimast jääaja maksimumi (LGM)( Torrioni *et al.* 1998; 2001).

Haplogrupid T ja J on fülogeneetiliselt lähedased klaadid (joonis 2), mille esinemissagedus Euroopas on suhteliselt sarnane - ligikaudu 20%. Mõlemad haplogrupid arvatakse pärinevat Lähis-Idast, kus nad ekspandeerusid ja levisid neoliitikumi perioodil 9000 aastat tagasi Euroopa aladele (Richards *et al.* 1998; 2000).

Superhaplogrupp U koondab enda alla terve rea haplogruppe (U1 – U9 ja K) ning on valdavalt iseloomulik Lääne-Euraasia populatsioonidele, kuid mitte eranditult. Kuna antud magistritöö keskendub superhaplogrupile U ning detailsemat käsitlust leiab ka alamharu U5, siis erinevatest U alamhaplogruppidest tuleb põhjalikumalt juttu järgnevas peatükis.

### **Euroopa asustamine**

Euroopa esmane koloniseerimine AMH poolt toimus ligikaudu 40 000 – 50 000 aastat tagasi (Richards *et al.* 2000; Tambets *et al.* 2003), kuid Euroopa asustamise laineid oli kokku arvatavasti neli - varajases ülempaleoliitikumis, keskmises ülempaleoliitikumis, hilises ülempaleoliitikumis peale viimast jääaja maksimumi (LGM) ja neoliitikumis (Richards *et al.* 2000). Enne LGM-i, mis leidis aset 24 000 – 16 000 aastat tagasi, levisid Euroopas Aurignac'i kultuur, mis saabus koos Euroopa esmaasustajatega ning hiljem tekkinud Gravetti kultuur. Antud kultuuridele oli omane algeliste tööriistade valmistamine ning esimeste kunstiteose ilmumine nagu luuskulptuurid ja koopamaalingud (Churchill ja Smith 2000). LGM-i perioodil taandusid inimpopulatsioonid soojema kliimaga aladele ehk refuugiumitesse. Teada on mitmeid arvatavaid refuugiumi alasid Lõuna- ja Ida-Euroopas, Lõuna-Siberis, Balkani poolsaarel ja Musta mere rannikul (Tambets *et al.* 2003).

Kui omavahel kõrvutada Euroopa ja Lähis-Ida haplogruppe, siis on näha, et suures osas haplogrupid kattuvad. Eraldi peab välja tooma haplogrupid U5, V, H3 ja H1 mis on Euroopa sisese päritoluga. Diversiteedi arvutused Euroopa ja Lähis-Ida haplogruppidele on näidanud, et mõlemas piirkonnas esindatud haplogruppide diversiteet oli üldjuhul suurem Lähis-Idas, välja arvatud Euroopa sisese tekkega haplogrupid (Richards *et al.* 2000). Lähis-Ida on seega osutunud teatud mõttes sõlmpunktiks, kust toimus Euroopa esmane asustamine (Richards *et al.* 2000; Tambets *et al.* 2003).

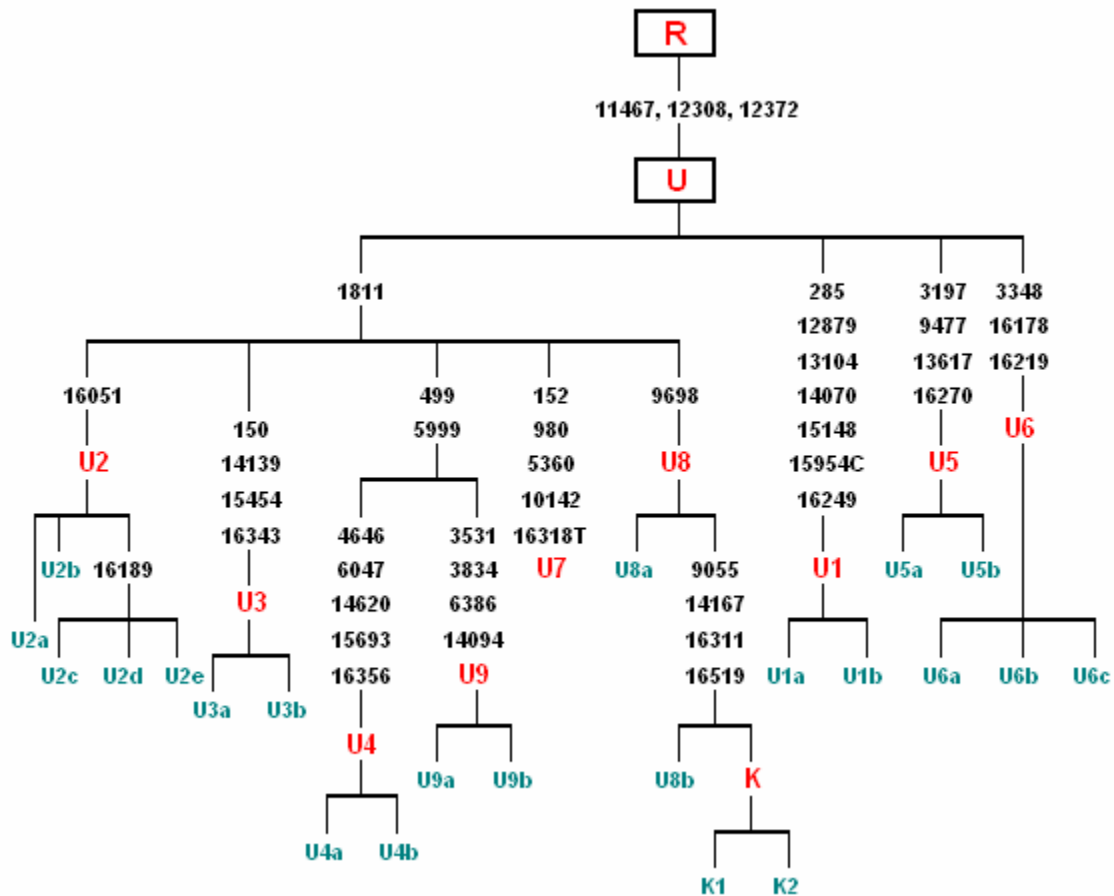
#### 1.4. Superhaplogrupp U

Superhaplogrupi U tekkepiirkonnaks peetakse Lähis-Ida regiooni, kus tema hinnanguline vanus on ligikaudu 50 000 aastat või isegi rohkem (Torrioni *et al.* 1996; Richards *et al.* 1998; 2000; Maca-Meyer *et al.* 2003). Euroopa kõikidest emaliinidest kuulub superhaplogruppi U ligikaudu 20% (Richards *et al.* 2000). U koondab enda alla terve rea erineva fülogeograafiaga haplogruppe nagu U1, U2, U3, U4, U5, U6, U7, U8, U9 ja K (Torrioni *et al.* 1996; Richards *et al.* 1998; 2000)(joonis 4).

Haplogrupp U1, mis jaguneb alamhaplogruppideks U1a ja U1b (joonis 4), on peamiselt levinud Lähis-Idas ja Euroopa Vahemere piirkonnas (Macaulay *et al.* 1999), kus tema arvatav ekspansioon toimus 20 000 – 30 000 aastat tagasi (Richards *et al.* 2000).

Haplogrupp U2 jaguneb alamhaplogruppideks U2a – U2e (Quintana-Murci *et al.* 2004; Palanichamy *et al.* 2004)(joonis 4), millest ainsana on Euroopas esindatud U2e (Macaulay *et al.* 1999; Richards *et al.* 2000). U2a, U2b ja U2c on põhiliselt levinud Pakistanis ja Indias. Indo-Pakistani ja Lääne-Euraasia U2-e ekspandeerumise ajaks on arvatud 53 000 aastat (Kivisild *et al.* 1999a). Alamhaplogruppi U2d on leitud peamiselt Lähis-Idast (Quintana-Murci *et al.* 2004).

Haplogrupp U3, mis jaguneb alamhaplogruppideks U3a ja U3b (joonis 4), on levinud Lähis-Idas, Põhja-Aafrikas ja Kaukaasia piirkonnas. Viimast peetaksegi U3 tekkepaigaks ning antud piirkonnale arvatud koalestsentsiaeg on 30 000 aastat (Krings *et al.* 1999b; Metspalu *et al.* 1999; Maca-Meyer *et al.* 2001).



**Joonis 4.** Superhaplogrupi U üldine topoloogia. Puu koostamisel on kasutatud HVS ja kodeeriva piirkonna defineerivaid mutatsioone. Puu harudel toodud numbrid tähistavad antud haru iseloomustavaid polümorfisme; numbri järel olev täht märgib transversiooni, transitsioone eraldi ei tähistata; insertioonid on tähistatud + märgi ja lisandunud nukleotiidiga. Puu koostamiseks on kasutatud järgmisi allikaid: Maca-Meyer *et al.* 2003; Quintana-Murci *et al.* 2004; Achilli *et al.* 2005; Kivisild *et al.* 2005.

Haplogrupp U4 on Euroopas U5 järel sageduselt teine haplogrupp ning tema vanuseks loetakse ligikaudu 25 000 aastat (Richards *et al.* 2000). U4 jaguneb kaheks alamhaplogruppiks U4a ja U4b (joonis 4). Kõige kõrgema sagedusega on U4 esindatud Lääne-Siberi rahvaste seas, ulatudes ugri keeli kõnelevate manside hulgas 16% (Tambets *et al.* 2003; Derbeneva *et al.* 2002a; 2002b). Euroopas on U4 sagedus kõrgem Kesk-, Lääne- ja Kirde-Euroopa aladel ning madalam Lõuna-Euroopa Vahemere piirkonnas (Tambets *et al.* 2003).

Haplogrupi U6 ekspansioon toimus 30 000 aastat tagasi arvatavalt Lähis-Ida piirkonnas, kust ta tagasirändega Põhja-Aafrikasse levis (Maca-Meyer *et al.* 2003). U6 jaguneb kolmeks alamhaplogruppiks U6a – 6c (Joonis 4). U6 on põhiliselt esindatud Kirde-, Põhja- ja Loode-Aafrikas ning Kanaari saartel (Rando *et al.* 1998; 1999; Krings *et al.* 1999b; Salas *et al.* 2002; Maca-Meyer *et al.* 2003). Euroopas on U6 levinud vaid Iberia poolsaarel (Pereira *et al.* 2000; Larruga *et al.* 2001).

Haplogrupp U7 on vähesel määral esindatud Ida- ja Põhja Euroopas, kuid sage Lähis- ja Kesk-Idas ning Lääne-Indias (Kivisild *et al.* 1999b; 2000). Haplogrupp U8 (joonis 4) on levinud peamiselt Euroopas, kuid U8b alamklasrit on leitud Lääne-Iraani kurdi populatsioonidest, kus ta esineb ligikaudu 10% sagedusega (Quintana-Murci *et al.* 2004). Haplogrupp K, mis haruneb haplogrupist U8 (joonis 4), jaguneb omakorda mitmeteks alamklastriteks. K on enam levinud Lääne- ja vähem Ida-Euroopas (Richards *et al.* 2000; Herrnstadt *et al.* 2002), samas leidub teda Lähis-Idas ja Anatoalias (tänapäeva Türgi)(Kivisild *et al.* 2003a). Kõrgeim haplogrupi K sagedus (30%) on leitud Ashkenazi ehk Euroopa juutidel (Behar *et al.* 2004). Haplogrupp U9 on suhteliselt haruldane ja hiljuti defineeritud haplogrupp, mida on leitud peamiselt Lõuna- ja Kagu-Pakistanist, kus ta hõlmab kuni 10% kõigist emaliinidest (Quintana-Murci *et al.* 2004).

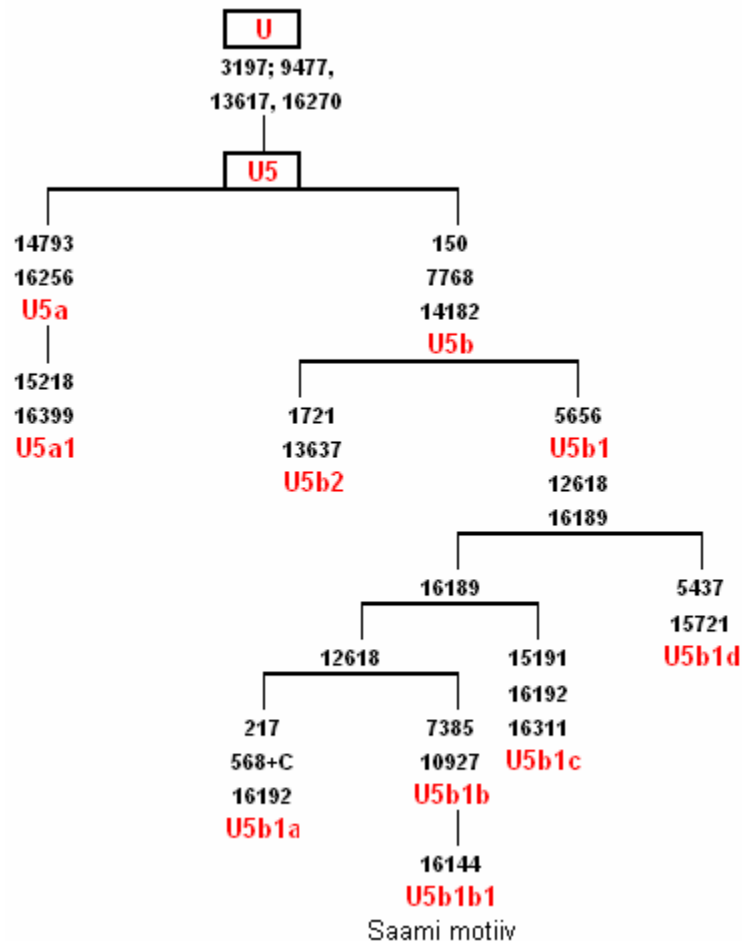
### 1.4.1. Haplogrupp U5

Haplogrupp U5 on superhaplogrupi U üks suurim ja divergeerunud haru, mis jaguneb omakorda mitmeteks alamhaplogruppideks ja klastriteks (Richards *et al.* 1998; 2000; Tambets *et al.* 2003; 2004)(joonis 5). Haplogrupi U5 koalestsentsiajaks on arvatud ligikaudu 45 000 – 55 000 aastat (Richards *et al.* 2000), mis näitab et U5 on suhteliselt vana haplogrupp. Kui enamus Lääne-Euraasiale iseloomulikke haplogruppe pärineb algselt Lähis-Ida piirkonnast, siis U5 on Euroopa sisese tekkega (Richards *et al.* 2000).

Kõige kõrgema sagedusega on haplogrupp U5 levinud Euroopas, kus ta hõlmab kuni 10% kõikidest emaliinidest (Richards *et al.* 1998; 2000). Kuid siingi ei ole levikumuster ühtlane, sest rohkem on U5 levinud Euroopa ida- ja põhjaosas (Orekhov *et al.* 1999; Tambets *et al.* 2003) ning vähem lõuna- ja läänepoolsetel aladel (Torroni *et al.* 1996). Kõrgeim U5 esinemissagedus on Põhja-Euroopa rahvastel nagu eestlased, karjalased, soomlased ja saamid (Sajantila *et al.* 1995; Richards *et al.* 1996; Villems *et al.* 1998; Tambets *et al.* 2003; 2004). Peale Euroopa on U5-t madala sagedusega leitud ka Lähis-Idast (Richards *et al.* 2000; Quintana-Murci *et al.* 2004), Kaukaasiast (Tambets *et al.* 2003; Quintana-Murci *et al.* 2004), Niiluse jõe orust Põhja-Aafrikast (Klings *et al.* 1999b), Aafrika lääneranniku Senegambia piirkonnast (Rosa *et al.* 2004) ning Siberi erinevatelt aladelt (Derbeneva *et al.* 2002a; 2002b; Derenko *et al.* 2003; Fedorova *et al.* 2003).

U5 jaguneb kaheks alamhaplogrupiks U5a ja U5b, mis omakorda jagunevad alamhaplogruppideks ja harudeks (joonis 5). Alamhaplogrupp U5a on levinud pigem Euroopa lõuna- ja vähemal määral loodeosas (Richards *et al.* 1998). U5a ja tema alamharu U5a1 vanuseks on arvatud vastavalt 40 000 ja 30 000 aastat, mis lubab arvata, et U5a ekspandeerus enne viimast jääaja maksimumi (Richards *et al.* 1998). U5a alamklaster U5a1a on ekspandeerunud suhteliselt hiljuti, ligikaudu viimase 2000 aasta jooksul. Antud alamklastril on väga madal diversiteet ning piiratud levikuala Loode- ja Kesk-Euroopas (Richards *et al.* 1998).

Alamhaplogrupp U5b ja tema alamharud on levinud madalal sagedusel üle kogu Euroopa - Ibeeriast kuni Uurali mägedeni (Sajantila *et al.* 1995; Richards *et al.* 1998; Tambets *et al.* 2004). Eraldi tuleks rääkida U5b alamharust U5b1b1, mille üks alamklaster moodustab saamide emaliinidest 47,6% (Tambets *et al.* 2004).



**Joonis 5.** Haplogrupi U5 topoloogia. Puu koostamisel on kasutatud HVS ja kodeeriva piirkonna defineerivaid mutatsioone, tähistus vt joonis 4 allkirjast. Kasutatud on järgmisi kirjandusallikaid: Finnilä *et al.* 2001; Herrnstadt *et al.* 2002; Tambets *et al.* 2004; Achilli *et al.* 2005.

Saamisid peetakse Põhja-Skandinaavia algseteks asutajateks. Tänapäeval elab Norra, Rootsi ja Soome põhjaosas ligikaudu 60 000 saamit (Haetta 1996). Klassikaliste geneetiliste markerite analüüs Cavalli-Sforza *et al.* (1994) näitas saamide ja teiste Euroopa rahvaste vahel väga suurt geneetilist distantssi ning seetõttu peeti saamisid pikka

aega geneetiliselt Euroopast eraldiseisvaks ja pakuti neile pigem mongoloidset päritolu. Kuid Saami mtDNA harude analüüs näitas, et valdav enamus neist kuulub siiski Euroopa mtDNA emaliinide alla (Torrioni *et al.* 1998; Villems *et al.* 1998; Tambets *et al.* 2004). Saamide populatsioonis on esindatud kaks põhilist haplogruppi - U5b1b1 alamklaster moodustab 47,6% ja haplogrupp V 41,6% (Richards *et al.* 1998; Tambets *et al.* 2004). Saamide U5b1b1 alamklaster ehk nn “saami motiiv“ sisaldab transitsioone nps 16189, 16144 ja 16270 (Sajantila *et al.* 1995). Alguses arvati, et antud motiiv on omane ainult saamidele, kuid seda on vähesel määral leitud ka naaberpopulatsioonidest (karjalased ja soomlased), mis on seletatav rahvaste segunemisega (Sajantila *et al.* 1995; Meinilä *et al.* 2001). Saamide ebatavalise mtDNA leviku mustri põhjuseks on populatsiooni ajaloos toimunud juhuslik geenitriiv (pudelikael/asutajaefekt), mis on mänginud peamist rolli selle väikese Euroopa populatsiooni mtDNA maastiku kujundamisel (Tambets *et al.* 2004).

U5b1b1 alamklaster on peale saamide levinud suhteliselt madala sagedusega ka Ida-Euroopa piirkonnas (Tambets *et al.* 2004). U5b1b1 eellasklastri U5b1b suurem diversiteet Lääne- Euroopas toetab teooriat, et U5b1b ekspansioon algas 8600 aastat tagasi Euroopa lääneosast (Achilli *et al.* 2005). U5b1b1 fülogeograafiline muster viitab sellele, et antud alamklaster on divergeerunud U5b1b eellasklastrist Ida-Euroopas. Seega jõudis alamklaster U5b1b1 Skandinaavia põhjaossa läbi Ida-Euroopa, mitte mööda Skandinaavia poolsaare Atlandi ookeani rannikut (Tambets *et al.* 2004). U5b1b1 liinide koalestsentsiajaks Euroopas on arvatud 4300 aastat, mis on vanem kui skandinaavia U5b1b1 liinide 1200 aastat (Tambets *et al.* 2004).

Haplogrupp U5 on teinud peale viimast jääaja maksimumi tõenäoliselt läbi ekspansioonifaasi, sest U5 fülogeneetiline puu näitab ligi tosina alam-asutaja haplotüübi (*subfounder*) olemasolu. Peaaegu kõigil neil on tähe-kujuline topoloogia ning koalestsentsiaeg jääb vahemikku 11 000 – 13 000 aastat (Tambets *et al.* 2003). On leitud, et väga erinevates geograafilistes piirkondades asuvad populatsioonid, nagu näiteks Skandinaavia saamid ja Põhja-Aafrika berberid, võivad jagada suhteliselt noore U5b1b

alamklastri mutatsioone (Achilli *et al.* 2005). Antud nähtus seletubki asjaoluga, et LGM-i ajal asus Euroopa edela osas Franco-Cantabrian'i refuugium, kus haplogrupp U5 tegi tõenäoliselt läbi ekspansioonifaasi ning peale viimase jääaja maksimumi lõppu taasasustasid antud refuugiumi populatsioonid Euroopa ja Põhja-Aafrika.

### 1.5. MtDNA analüüsimise ja tulemuste tõlgendamisega seotud probleemid

Teadustehnika kiire areng viimasel aastakümneil on loonud võimaluse analüüsida üha kiiremini ja vähesema vaevaga kümneid ja sadu mtDNA järjestusi suhteliselt väikese ajakuluga. mtDNA andmebaaside kvantitatiivne kasv on samas tõstatanud küsimuse ka kvaliteedist (Bandelt *et al.* 2001a; 2002; 2006; Parson *et al.* 2004). Bandelt *et al.* tehtud uurimused on näidanud, et avaldatud mtDNA andmetes on esinenud olulisi vigu, mis omakorda võivad mõjutada tulemuste lõppinterpretatsiooni. Ka antud magistritöö üheks eksperimentaalse töö osaks oli eelnevalt saadud andmete õigsuse kontrollimine ning täpsustamine. Seetõttu on oluline mõista DNA analüüsimise protsessi etappe ja teada põhjuseid, mis viivad ebaadekvaatsete lõpptulemusteni.

DNA analüüsimisel läbitakse laboris mitmed etapid, millest esimene on proovimaterjali kogumine. Proovimaterjalina kasutatakse kõige sagedamini juukseid, verd ja põse limaskestast epiteeli kaabet. Üldjuhul peaks iga prooviga kaasas olema mõningane informatsioon doonori kohta nagu näiteks tema etniline taust 2-3 põlve emaliinis tagasi. Doonori päritolu teadmine võimaldab lõppandmete interpreteerimisel õieti analüüsida erinevate populatsioonide päritolu ja populatsiooni sisest dünaamikat. Peale proovimaterjali kogumist järgneb DNA eraldamine ning seejärel PCR valitud praimerpaariga. Amplifitseeritud produktilt teostatakse sekveneerimisreaktsioon vastavalt valitud praimeriga ning saadud produkti nukleotiidne järjestus analüüsitakse sekvenaatoris. Viimases faasis toimub tulemuste interpreteerimine.

Probleeme, mis mõjutavad lõpptulemust, võib esineda kõikides eelpool nimetatud töö etappides (Bandelt *et al.* 2006). Esimene neist, DNA eraldamine, omab võtmerolli

edasiste eksperimentide õnnestumise või ebaõnnestumise suhtes. Kõige suurem oht antud töö etapis on proovi saastumine võõr-DNA-ga, mida saab hiljem kindlaks teha piisava hulga polümorfsete saitide võrdlemisel (Bandelt *et al.* 2006). Kui proovi ja võõr-mtDNA järjestused on suhteliselt sarnased, siis võidakse sekventsilt välja lugeda valepositiivsed heteroplasmiidid. Kui järjestused on suhteliselt erinevad, siis saadud sekvents on näha palju ebaselgeid nukleotiidi positsioone (märgitud kui N). Väga tugeva saastuse puhul on võimalik, et võõr-mtDNA nõ jookseb üle õige järjestuse ning saadud tulemus ei vasta tegelikult doonori proovile (Bandelt 2006). Kõigi nende vigade tekkimise vastu aitab hea labori praktika järgimine (*good laboratory practice*), mille peamine nõue on see, et töö käigus liigutaks analüüsitavaid proovidega igas etapis uuele alale ning mitte kunagi tagasi, st et proovide liikumine (*work flow*) oleks ühesuunaline. Soovituslik on üles seada laboritöötajate mtDNA järjestuste andmebaas, kus oleks ka hiljuti laboris analüüsitud proovid, et oleks võimalik jälgida ja kontrollida saadud tulemusi (Bandelt *et al.* 2001a).

PCR, sekveneerimisreaktsioon ja proovide analüüsimine sekvenaatoris on samuti aldis saastusele ning proovide segamini minekule, eriti kui on palju eraldiseisvaid töö etappe ning enamus neist tehakse käsitsi, mitte automatiseeritud kinnistes süsteemides (Bandelt *et al.* 2006). Sekveneerimiskeemia ja meetodikate areng on jätnud oma jälje ka mtDNA sekveneerimise tulemustesse. Tsüklilise sekveneerimise kommertsiaalsed komplektid on tunduvalt parandanud järjestuste elektroferogramme ehk fluorestsentssignaalide kvaliteeti, kuid osade komplektide puhul on kaasnevaks probleemiks nõ fantoommutatsioonide teke (Bandelt *et al.* 2002). Antud mutatsioonide anomaalne tekkeprotsess näib olevat põhjustatud suuresti analüüsi spetsiifilistest biokeemilistest tingimustest. Näiteks vanema *dye terminator sequencing* kiti puhul on probleemiks nukleotiidses järjestuses 3'otsa esimene adensiiniga (A) külgnev guansiin (G), mis elektroferogrammil on kujutatud väga madala piigina (Parson *et al.* 1998). Teatud tingimustel, nagu pikkade järjestuste lugemine, võib G piik täiesti olematuks kaduda. Oluline on sekveneerimisreaktsiooni produktide hoolikas puhastamine enne sekvenaatoriga analüüsimist, selleks et eemaldada mitteseondunud terminaatorid, praimerid jms, mis muidu jääks segama produktide analüüsimist, varjutades esimese

paarikümne nukleotiidi fluorestsents signaali ning on samuti soodsaks pinnaseks fantoom-mutatsioonidele (Bandelt *et al.* 2002).

MtDNA analüüsi viimaseks etapiks on saadud andmete tõlgendamine. Sekvenerimisprotsessi jooksul toimub arvuti tarkvara poolt automaatne nukleotiidi positsioonide nimetamine (*automated basecalling*), mis aga ei pruugi tulenevalt programmi seadistustest alati kõigi nukleotiidi positsioonide puhul adekvaatne olla (Bandelt *et al.* 2006). Seepärast on oluline esmased andmed enne tabelitesse ja andmebaasidesse kandmist üle vaadata ning kontrollida detekteeritud mutatsioonide õigsust ning paikapidavust. Selline lähenemine on küll aeganõudvam, kuid välistab arvutitarkvara poolt tehtud vigade edasikandumist lõppjäreldestesse. Kehva kvaliteediga järjestus, kus on suhteliselt palju ebaselgeid positsioone, peaks loetama kehtetuks, sest referentsjärjestuse (rCRS)(Andrews *et al.* 1999) najal analüüsitud halva kvaliteediga järjestus on aldis andma valepositiivseid ja -negatiivseid tulemusi.

Andmete dokumenteerimisel ja tabelitesse/andmebaasidesse kandmisel võivad samuti tekkida tähelepanematuses vead (Bandelt 2006) - mutatsioon jääb märkamata, toimub numbrite segiajamine või äravahetamine, C märgitakse G-ks või vastupidi jne. Kuna andmetabeleid koostatakse käsitsi, siis on võimalik ka ridade ja/või veergude segamini minek.

Kui vaadata mtDNA andmeid, siis tähelepanelikuks peaks muutma haruldaste mutatsioonide esinemine erinevates järjestustes, ka hüpervariaabel piirkondades, mis võib viidata sekvenerimise artefaktidele või muudele probleemidele järjestuse lugemisel. Ka transversioonide ning haruldaste indelite liigsage esinemine peaks muutma ettevaatlikuks antud andmete adekvaatsuse suhtes (Bandelt *et al.* 2001a).

## 2. MATERJAL JA METOODIKA

### 2.1. Katsematerjal

Antud magistritöö eksperimentaalne osa põhineb 46 mtDNA nukleotiidsel polümorfisuse määramisel ja analüüsil. Kõik antud magistritöös analüüsitud mtDNA-d olid eelnevalt meie labori töörühma poolt np12308 transitsiooni alusel määratud superhaplogruppi U kuuluvaks, kuid HVS-I piirkonna järjestuse põhjal ei kuulunud nad ühtegi U alamharusse. MtDNA proovid pärinesid erinevatest geograafilistest piirkondadest: Eesti (1 indiviid); Küpros (2); Läti (3); Poola (3); Prantsusmaa (2); Rootsi (2); Rumeenia (4); Sitsiilia (2); Slovakkia (1); Altai piirkond (1); Kasahstan (1); Ukraina (1); Volga-Uraali piirkond (2); Venemaa edela piirkond (9); Albaania (1); Egiptus (1); Iraan (1); Jordaania (1); Maroko (2); Kuveit (1); Süüria (3) ja Türgi (2). Valimi koostamiseks analüüsiti meie töörühma andmebaasis 36 populatsiooni, kus esines superhaplogruppi U ja tema alamharusid. Analüüsitud andmete kogum sisaldas 9383 mtDNA haplogrupeerimise tulemusi, millest 1801 olid määratud superhaplogruppi U erinevatesse alamhaplogruppidesse ja 46 U\*-ks.

Viimastel aastatel lisandunud teave mtDNA kodeeriva regiooni defineerivate nps kohta on võimaldanud täpsustada haplogruppide sisetopoloogiat. Kuna antud valimi puhul ei olnud HVS-I regiooni järjestuste informatsioon piisav alamhaplogrupeerimiseks, siis analüüsiti antud valimit kodeeriva regiooni U alamhaplogruppe defineerivate nps suhtes. Selleks kasutati RFLP analüüsi ja nps sekveneerimist (tabel 1 ja 2). Samuti kontrolliti HVS-I piirkonna järjestuste õigsust teistkordse HVS-I piirkonna sekveneerimisega. Kuuele mtDNA-le antud valimist teostati lisaks eelnevatele analüüsidele ka genoomi täisjärjestamine. Kõik nps ja RFLP saidid on alljärgnevates tabelites ära toodud rCRS-i suhtes (Andrews *et al.* 1999).

Haplogruppi U5 ja U3 täisjärjestuste sünteesipuu koostamisel kasutati lisaks antud magistritöö käigus sekveneeritud 6 täisjärjestusele andmeid järgmistest kirjandusallikatest: Finnilä *et al.* 2001 (22 täisjärjestust); Herrnstadt *et al.* 2002 (24);

Palanichamy *et al.* 2004 (1); Kivisild *et al.* 2005 (12); Achilli *et al.* 2005 (18). U3 ja U5 puude koostamiseks kasutati kokku 83 täisjärjestust. Fülogeneetilised puud koostati käsitsi, vastavalt *median-joining* põhimõtetele (Bandelt *et al.* 1995; 1999). U5b2 alamharu koalestsentsiaja arvutamiseks kasutati statistilist väärtust  $\rho$  (Foster *et al.* 1996; Saillard *et al.* 2000) ning arvutamisel kasutati kodeeriva regiooni (nps 577-16023) mutatsioone (va indelid). Ühe mutatsiooni tekkeajaks loeti vastavalt 5139 aastat (Mishmar *et al.* 2003). Standardviga arvutati Saillard *et al.* 2000 järgi.

## 2.2. DNA amplifitseerimine

DNA kogus amplifikatsiooni reaktsioonisegus oli 10-20 ng ning polümeraasne ahelreaktsioon (PCR) viidi läbi 25 $\mu$ l mahus. Kõigi DNA amplifitseerimisel kasutatud praimerite vahelahuse kontsentratsioon oli 10 pmol/ $\mu$ l. PCR-i reaktsioonid viidi läbi termoblokis “Biometra UNO II”.

<u>lõppkontsentratsioon</u>	<u>PCR-i komponent</u>
1/10	Tartrazine puhver (750 mM Tris-HCl, pH 8.8, 200 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 0.1% Tween 20; 5% Ficoll 400)
2,5 mM	MgCl <sub>2</sub> (25 mM)
0,1 mM	dNTP mix (dATP, dCTP, dGTP, dTTP) (10 mM)
0,05U/ $\mu$ l	Taq DNA polümeraas FIREPol 5 U/ $\mu$ l (Solis Biodyne) ja/või LongExtention Taq 2 U/ $\mu$ l (Fermentas)
0,2 pmol/ $\mu$ l	F praimer (10 pmol/ $\mu$ l)
0,2 pmol/ $\mu$ l	R praimer 10 pmol/ $\mu$ l
	1-2 $\mu$ l DNA-d
	deioniseeritud vesi

**DNA amplifikatsiooni tingimused:**

* denaturatsioon	94°C	3 min	}	40-44 tsüklit
* denaturatsioon	94°C	25 s		
* praimerite seondumine	52-61°C	20 s		
* ekstensioon	72° C	1,5 – 2 min		
* lõplik ekstensioon	72° C	3 min		

Amplifikatsiooni tsüklite arv ja praimerite seondumise temperatuur sõltus praimerite spetsiifisusest ja DNA kvaliteedist. Defineerivate nps sekveneerimiseks ja RFLP analüüsiks kasutatud praimerite nimekiri ja nukleotiidne järjestus on ära toodud tabelis 1 ja 2. MtDNA HVS-I piirkonna ja kogu genoomi täisjärjestuse sekveneerimiseks kasutatud praimerid koos nukleotiidse järjestusega on ära toodud antud töö lisades.

PCR-i produkti olemasolu kontrolliti 2% horisontaalses agarosgeelis [0,5 kordses TBE puhvis (45mM Tris-boraat, 0,05M EDTA)], kasutati konstantset pinget 140 V, elektrodide distants 32 cm. DNA visualiseerimiseks ultraviolettkiirguses lisati geelile etiidumbromiidi lõppkontsentratsiooniga 0,5µg/ml. Geelile kanti 4µl PCR-i produkti.

**Tabel 1.** MtDNA kodeeriva regiooni U haplogruppe defineerivad nps ja nende sekveneerimiseks kasutatud praimerid. Täht np järel märgib transversiooni. #

hg	defineeriv np	viide	amplifitseerimiseks kasutatud praimerid	
<b>U2a</b>	16206C	5	F	5'- 15879 - AATGGGCCTGTCCTTGTAGTAT - 3'
			R	5'- 16545 - AACGTGTGGGCTATTTAGGCT - 3'
<b>U2d</b>	13720	1	F	5'- 13338 - CTTCAAAGCCATACTATTTATGTG - 3'
			R	5'- 13843 - CTAGGGCTGTTAGAAGTCCTAGGA - 3'
<b>U3</b>	14139	2	F	5'- 14000 - TAACCTGACTAGAAAAGCTATTA - 3'
			R	5'- 14268 - CGGGAGGATCCTATTGGTGCGGG - 3'
<b>U5</b>	13617	6	F	5'- 13338 - CTTCAAAGCCATACTATTTATGTG - 3'
			R	5'- 13843 - CTAGGGCTGTTAGAAGTCCTAGGAAAGT - 3'
<b>U7</b>	980	2	F	5'- 543 - CCGAACCAACCAACCCCAAAGA - 3'
			R	5'- 1220 - TATCGATTACAGAACAGGCTCCTCTAGAG - 3'
<b>U8</b>	9698	6	F	5'- 9230 - TCATATAGTAA AACCCAGCCCATG - 3'
			R	5'- 10130 - TAGTCAAATGTAATAATTATTAGTAG - 3'

### 2.3. RFLP analüüs

Restriktsiooni ensüümid ehk restriктаasid on võimelised teatud nukleotiidses järjestuses ehk äratundmissaidi esinedes lõikama DNA-d spetsiifilistest kohtadest. Mutatsioonid DNA järjestuses võivad põhjustada restriktionisaidi kao või tekke. RFLP analüüsil kasutatud restriктаasid on ära toodud tabelis 3. Restriktsiooni reaktsiooniks lisati 20µl PCR produktile 0,3 U restriктаasi (Fermentas) ja 2µl vastavat puhvrit. Restriktsiooni tingimused valiti vastavalt tootja poolsetele soovitudele. Restriktsioonifragmendid lahutati agarosgeel-elektroforeesil. Kõik ebaselged RFLP tulemused kontrolliti üle vastava nps sekveneerimisega.

**Tabel 2.** U haplogruppe defineerivad nps ja RFLP analüüsil kasutatud praimerid. Täht np järel märgib transversiooni. #

hg	defineeriv np	viide	amplifitseerimiseks kasutatud praimerid	
U	12308	4	F	5'- 12104 - CTCAACCCCGACATCATTACCGGGTTT - 3'
			R	5'- 12388 - GGGGAATTAGGGAAGTCAG - 3'
U1	13104	3	F	5'- 12541 - GCCACAACCCAAACAACCCA - 3'
			R	5'- 13403 - ATATCTTGTTTCATTGTAAAGG - 3'
U2b	15049	1	F	5'- 14815 - CCCATCCAACATCTCCGCATGATGAA - 3'
			R	5'- 15431 - CGTCTTTGATTGTGTAGTAAG - 3'
U2c	5790A	1	F	5'- 5424 - CATAAAAACCCACCCC ATTCCTCCCC - 3'
			R	5'- 5920 - CAACGGTCGGCGAACATCAG - 3'
U2e	13734	1	F	5'- 13583 - CCTCCCTG ACAAGCGCCT ATAGCA - 3'
			R	5'- 13843 - CTAGGGCTGTTAGAAGTCCTAGG - 3'
U4	4646	3	F	5'- 4308 - GGAGCTTAAACCCCTTATTTCTAG - 3'
			R	5'- 4739 - GGTAGTATTGGTTATGGTTCATTGTC - 3'
U6	3348	7	F	5'- 3153 - TCACAAA GCGCCTTCCC CCGTAAATGA - 3'
			R	5'- 3426 - TACAACGTTGGGGCCTTTGCTAG - 3'
K	9055	1	F	5'- 8829 - CT CATCACCCAACATAAAAATAT - 3'
			R	5'- 9184 - GGTTTACTAGAAGTGTGAAAACGTAG - 3'
U9	6386	1	F	5'- 6242 - CGCATCTGC TATAGTGGAG GCCGGAG - 3'
			R	5'- 6699 - CTTTTTTTCCGGAGTAGTAAGTTAC - 3'

# Tabelis tähistab: F ja R vastavalt mtDNA kerge ning raske ahela praimereid; number praimerjärjestuse ees praimeri 5'otsa np. Viide 1 – Quintana-Murci *et al.* 2004; 2 - Palanichamy *et al.* 2004; 3 – Macaulay *et al.* 1999; 4 - Torroni *et al.* 1996; 5 – Kivisild *et al.* 1999b; 6 – Finnilä *et al.* 2001; 7 – Torroni *et al.* 1999.

**Tabel 3.** U haplogruppide RFLP analüüsil kasutatud restriктаasid ja restriktsooni saidid. + märgib restriktsoonisaidi teket ja - kadu. Viide 1 – Quintana-Murci *et al.* 2004; 3 – Macaulay *et al.* 1999; 4 - Torroni *et al.* 1996; 7 – Torroni *et al.* 1999.

hg	defineeriv np	RFLP motiiv	äratundmis-motiiv	PCR-i produkti pikkus	RFLP tulemus		viide
					+	-	
U	12308 A→G	12308 HinfI +	5'G ↓ANTC 3'	284 bp	40+138+30	40+168	4
U1	13104 A→G	13104 MboI+	5'↓GATC 3'	862 bp	563+299	862	3
U2b	15049 C→T	15047 HaeIII-	5'GG ↓CC 3'	615 bp	231+384	615	1
U2c	5790A C→A	5789 TaqI+	5'T↓CGA 3'	496 bp	131+365	496	1
U2e	13734 T→C	13730 HinfI+	5'G ↓ANTC 3'	260 bp	113+147	260	1
U4	4646 T→C	4643 RsaI+	5'GT↓AC 3'	432 bp	157+178+96	157+274	3
U6	3348 A→G	3348 MboI+	5'↓GATC 3'	273 bp	195+78	273	7
K	9055 G→A	9052 HaeII-	5'GCGC↓ 3'	356 bp	228+128	356	1
U9	6386 C→T	6383 HaeIII-	5'GG ↓CC 3'	457 bp	19+123+315	19+438	1

## 2.4. Amplifitseeritud produkti puhastamine

Amplifitseeritud DNA sekveneerimise puhul on oluline PCR-i produkti puhastamine praimeritest ning ahelasse lülitamata jäänud vabast nukleotiididest. Selleks lisati PCR-i produktile 1 µl puhastussegu, mis koosnes:

- \* eksonukleas I (10U/µl)(ExoI *E. coli*, Amersham Pharmacia Biotech) 1 osa
- \* Shrimp aluseline fosfataas (1U/µl)(SAP, Amersham Pharmacia Biotech) 9 osa

### Töötlemine toimus termoblokkis järgmistel tingimustel:

- \* vabade nukleotiidide defosforüleerimine ja üheaahelise DNA degradeerimine 37<sup>0</sup>C 20 min
- \* ensüümide inaktiveerimine 80<sup>0</sup>C 15 min

## 2.5. DNA sekveneerimine

MtDNA primaarjärjestuse määramiseks kasutati Sangeri didesoksüterminatsiooni meetodit. DNA sünteesitakse *in vitro* DNA polümeraasi poolt, mis kasutab matriitsina üheaahelalist DNA-d. Reaktsioonisegus sisalduvad kindlas kontsentratsioonis didesoksünukleotiidtrifosfaadid (ddNTP-d), mille kaasamisel sünteesireaktsiooni lõpeb konkreetse DNA ahela süntees. Eri pikkusega DNA fragmentide detekteerimiseks on ddNTP-dega seotud neli erinevat fluorestseeruvat värvi.

### Sekveneerimisreaktsioon DYE

* sekveneerimispraimer (5 pmol/μl)	1 μl
* puhastatud PCR-i produkt	5 μl
* sekveneerimissegu	4 μl

Sekveneerimissegu koosneb 1 μl DYE Sequencing Premix (Amersham Dyeamic™ ET Terminator Cycle Sequencing Kit; Amersham Pharmacia Biotech, USA) ja 3 μl lahjenduspuhvrast (puhver B 100 μl: 750 mM Tris-HCl pH8.9, 200 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 01% Tween 20; MgCl<sub>2</sub> 30 μl (25mM); ddH<sub>2</sub>O 270 μl).

### Sekveneerimisreaktsiooni tingimused DYE Sequencing Premix`ile

* DNA denaturatsioon	94°C 20 sek	} 40 – 44 tsüklit
* praimerite seondumine	50 - 56°C 15 sek	
* ekstensioon	60°C 60 sek	

Tsüklite arv ja praimerite seondumise temperatuur sõltus praimerite spetsiifikast ja DNA kvaliteedist.

## Sadestamine

- 10 µl produktile lisati 2 µl lahust, mis koosnes 1,5 M CH<sub>3</sub>COONH<sub>4</sub> (ammooniumatsetaat), [pH>8]/EDTA (250mM) ja 20mg/ml punane dekstraan suhtes 1:1
- lisati 30 µl külma (-20<sup>0</sup>C) 96% etanooli
- proove hoiti -20<sup>0</sup>C juures ~30 min
- tsentrifuugiti 15 min 13 000 rpm, eemaldati supernatant
- pesti 200 µl 70% etanooliga
- tsentrifuugiti 8 min 13 000 rpm, eemaldati supernatant
- pesti 200 µl 70% etanooliga
- tsentrifuugiti 8 min 13 000 rpm, eemaldati supernatant
- lasti sademel kuivada +37<sup>0</sup>C 10 min
- proovid suspendeeriti 10 µl MegaBACE geelile kandmise lahuses (70% formamiid, 1 mmol/l EDTA) (Amersham Pharmacia Biotech, USA)

Reaktsiooni produktid analüüsiti kapillaarsekvenaatoriga MegaBACE<sup>TM</sup> 1000 (Amersham Biosciences).

## 2.6. Andmeanalüüs

Sekveneeritud mtDNA järjestused järjestati käsitsi ja analüüsiti vastavalt Cambridge referents järjestusele (rCRS) (Andrews *et al.* 1999), kasutades selleks GCG programmipaketti, versioon 10,0 (Genetics Computer Group, Madison, Wisconsin) ja selle alamprogrammi SeqLab. Mutatsioonide detekteerimiseks kasutati MITOMAP'i andmebaasi (A Human Mitochondrial Genome Database)(<http://www.mitomap.org>). Aluspaariliste muutuste poolt põhjustatud aminohappe teisendusi analüüsiti andmebaasis MitoAnalyzer (National Institute of Standards and Technology, Gaithersburg, MD, USA) (<http://www.cstl.nist.gov/biotech/strbase/mitoanalyzer.html>).

### 3.TULEMUSED JA ARUTELU

#### 3.1. Töö eesmärk

Antud magistritöö eesmärkideks oli:

- 1) analüüsida ebatüüpiliste HVS-I järjestusmotiividega U haplotüüpe, mis eelnevalt olid defineeritud kui U\*, kasutades selleks kodeeriva piirkonna defineerivaid nps.
- 2) kontrollida eelnevalt sekveneeritud HVS-I piirkonna nukleotiidses järjestuses õigsust.
- 3) määrata genoomi täisjärjestus teatud hulga mtDNA-de antud valimist, eesmärgiga täpsustada nende asukohta fülogeneetilisel puul.

#### 3.2. MtDNA kodeeriva piirkonna defineerivate nps analüüs

MtDNA-de haplogrupeerimine põhines pikka aega peamiselt mtDNA HVS-I piirkonna polümorfisusel ja üksikutel kodeeriva regiooni defineerivatel nps. Antud valimisse kuulunud 46 mtDNA-d olid eelnevalt defineeritud np 12308 transitsiooni alusel superhaplogruppi U kuuluvaks, kuid HVS-I piirkonna ebatüüpiliste järjestusmotiivide tõttu ei paigutunud nad ühtegi U alamhaplogruppi ning olid defineeritud U\*-na.

Kasutades hetkel teadaolevat informatsiooni mtDNA kodeeriva regiooni defineerivate nps kohta, analüüsiti antud töös kõiki valimi mtDNA-sid vastavate U alamharude markerite suhtes. Kui võimalik, kasutati nps detekteerimiseks RFLP analüüsi, vastava restriksioonisaidi puudumisel kasutati np sekveneerimist. Kõik U alamharude määramiseks kasutatud restriksioonid ja defineerivad nps on ära toodud tabelis 1 ja 3 materjalid ja metoodika osas. Analüüsi tulemused on ära toodud tabelis 4, kust on välja jäetud antud valimis mitte esinenud haplogruppide RFLP/nps. Kõik tulemused ja mutatsioonid on ära toodud rCRS-i suhtes (Andrews *et al.* 1999).

Kõikidele mtDNA-dele tehti uuesti 12308 HinfI+ analüüs, mis kinnitas eelnevalt saadud tulemusi. Edasi analüüsiti 14 erinevat U alamharusid defineerivat markerit, mis defineerisid haplogruppe U1, U2a-U2e, U3, U4, U5, U6, U7, U8/K ja U9. Haplogrupiline

**Tabel 4.** RFLP analüüsi tulemused.

geogr. piirkond	proovi tähis	U Hinf I+ 12308	U1 Mbo I+ 13104	U4 Rsa I+ 4646	U6 Mbo I+ 3348	K/U8b Hae II- 9055	U3 14139	U5 13617	U7 980	U8 9698
Eesti	Est 361	+						C		
Küpros	Kyp 41	+	+							
	Kyp 187	+	+							
Läti	Lat 26-II	+		+						
	Lat 49-II	+		+						
	Lat 30-II	+		+						
Poola	PI 35	+		+						
	PI 45	+		+						
	PI 109	+		+						
Prantsusmaa	P1	+						C		
	Up Rh 6331	+				-				
Rootsi	Ro g1 74	+						C		
	Ro g3 40	+						C		
Rumeenia	Rom 36x	+		+						
	Rom 381	+						C		
	Rom 429	+		+						
	Rom 500	+		+						
Sitsiilia	S87	+	+							
	M4165	+				-				
Slovakkia	Slov 90	+								C
Altai	Alt 26m	+		+						
Kazahstan	Kaz 111	+						C		
Ukraina	Ukr 716	+						C		
Volga-Uraal	Udm 105	+						C		
	Udm 24	+		+						
Venemaa	kst 121	+		+						
	kzl 124	+				-				
	pin 132	+						C		
	pin 136	+						C		
	BGM 176	+				-				C
	bg 201	+		+						
	BB 225	+		+						
	bg 229	+				-				C
	sm 320	+		+						
Albaania	Alb 7	+		+						
Egiptus	Egip 785	+					G			
Iraan	Ira 207	+		+						
Jordaania	Jord 64	+				-				
Maroko	Mar 36E	+						C		
	Mar(E) B32	+			+					
Kuveit	Kuw 240	+		+						
Süüria	Syr 1368	+							C	
	Syr 1300	+	+							
	Syr 519	+	+							
Türgi	Tyr 346	+						C		
	Tyr 172	+					G			

jaotuvus valimis oli järgmine – nii haplogruppi U6 kui U7 kuulus vaid 1 mtDNA. Järgnes haplogrupp U3, kuhu kuulus 2 ja haplogrupp U1, kuhu kuulus 5 mtDNA-d. Haplogruppi U8/K kuulus kokku 9 proovi ning kõrgeima sagedusega olid antud valimis esindatud haplogrupid U4 ja U5 – vastavalt 18 ja 12 mtDNA-d. Valimi hulgas ei olnud ühtegi proovi, mis kuulunuks haplogruppi U2 või U9.

Antud analüüs näitas, et kõik eelnevalt U\*-ks haplogrupeeritud mtDNA-d asetsid kodeeriva regiooni defineerivate nps alusel superhaplogrupi U alamharudesse.

### 3.3. HVS-I piirkonna nukleotiidsete järjestuste analüüs

Kõikidel U haplogruppidel on teada HVS-I piirkonnas defineerivad nps (tabel 6). Väikese erandi moodustavad haplogrupid U8/K ja U9, kus HVS-I piirkonna defineerivad mutatsioonid ei määra mitte haplogruppi, küll aga alamhaplogruppe. Antud töös analüüsitud valimil ei olnud eelneva HVS-I piirkonna sekveneerimise põhjal detekteeritud vastatvaid U haplogruppe defineerivaid mutatsioone. Kuna mtDNA kodeeriva regiooni defineerivate nps analüüs alamhaplogrupeeris kõik valimi mtDNA-d, siis järgmise etapina viidi läbi HVS-I piirkonna nukleotiidse järjestuse korduv sekveneerimine, et kontrollida olemasolevate andmete õigsust. Võrdlevad tulemused on toodud tabelis 5.

46 mtDNA HVS-I piirkonna sekveneerimistulemuste võrdlemisel selgus, et kordussekveneerimine andis sama tulemuse 61% ehk 28 järjestust 46-st kattusid. 18 proovil ehk 39% oli eelnevalt sekveneeritud HVS-I piirkonna nukleotiidne järjestus erinev sellest, mis saadi teistkordsel sekveneerimisel (tabel 5). Seega oli antud mtDNA valimis sekveneerimistulemuste korratavus 61%. Samas, kui võtta arvesse algse valimi suurus, millest lähtudes eksperimentaalset tööd alustati (1801 U alamharudesse haplogrupeeritud proovi), siis moodustab antud töö katsete mittekorratavus algsest valimist 1%.

**Tabel 5.** HVS-I piirkonna võrdlevad sekveneermise tulemused. Tabelis olevate nps eest on ära jäetud numbrikombinatsioon 16.

geogr. piirkond	proovi tähis	HVS-I eelnevad andmed	HVS-I kontrollitud järjestus	määratud haplogrupp
Eesti	Est 361	93, 129, 189	93, 129, 189, 209, 270	<b>U5</b>
Küpros	Kyp 41	129, 189	129, 189, 249, 288	<b>U1</b>
	Kyp 187	93, 129, 189	93, 129, 183, 189, 249, 274, 288	<b>U1</b>
Läti	Lat 26-II	214	214	<b>U4</b>
	Lat 49-II	104, 294	104, 294	<b>U4</b>
	Lat 30-II	CRS	CRS	<b>U4</b>
Poola	PI 35	129	129, 362	<b>U4</b>
	PI 45	CRS	CRS	<b>U4</b>
	PI 109	255	255	<b>U4</b>
Prantsusmaa	P1	239, 311	239, 311	<b>U5</b>
	Up Rh 6331	CRS	182, 224, 311	<b>K</b>
Rootsi	Ro g1 74	271, 174, 189	271, 174, 189	<b>U5</b>
	Ro g3 40	271, 174, 189	271, 174, 189	<b>U5</b>
Rumeenia	Rom 36x	119	119	<b>U4</b>
	Rom 381	189	189, 325	<b>U5</b>
	Rom 429	CRS	119	<b>U4</b>
	Rom 500	CRS	CRS	<b>U4</b>
Sitsiilia	S87	189	189, 381	<b>U1</b>
	M4165	189, 209, 259	189, 209, 259, 234	<b>U8b</b>
Slovakkia	Slov 90	146,180	146, 180, 342	<b>U8a</b>
Altai	Alt 26m	CRS	356	<b>U4</b>
Kasahstan	Kaz 111	286	256, 270	<b>U5</b>
Ukraina	Ukr 716	189, 192, 234, 256, 311, 362	189, 192, 234, 256, 311, 362	<b>U5</b>
Volga-Uraal	Udm 105	CRS	CRS	<b>U5</b>
	Udm 24	221	134	<b>U4</b>
Venemaa	kst 121	CRS	CRS	<b>U4</b>
	kzl 124	259, 354	224, 311	<b>K</b>
	pin 132	311	144, 189, 270	<b>U5</b>
	pin 136	CRS	CRS	<b>U5</b>
	BGM 176	129, 189, 215, 320	129, 182, 189, 215, 320	<b>U8b</b>
	bg 201	263	263	<b>U4</b>
BB 225	BB 225	290, 294	290, 294	<b>U4</b>
	bg 229	66, 129, 189, 234	66, 129, 189, 234	<b>U8b</b>
	sm 320	93	93	<b>U4</b>
Albaania	Alb 7	93,183,254	93, 183, 254	<b>U4</b>
Egiptus	Egip 785	192, 244, 311	192, 244, 311, 343	<b>U3</b>
Iraan	Ira 207	119	119	<b>U4</b>
Jordaania	Jord 64	189, 311, 324	189, 311, 324	<b>U8b</b>
Maroko	Mar 36E	189, 278	189, 270	<b>U5</b>
	Mar(E) B32	129, 169, 172, 189	129, 169, 172, 189	<b>U6</b>
Kuveit	Kuw 240	134, 266, 355	134, 266, 356	<b>U4</b>
Süüria	Syr 1368	126, 148, 309	126, 148, 309	<b>U7</b>
	Syr 1300	189, 223	189, 249	<b>U1</b>
	Syr 519	CRS	249, 327	<b>U1</b>
Türgi	Tyr 346	311	311	<b>U5</b>
	Tyr 172	86, 247	86, 247	<b>U3</b>

**Tabel 6.** U haplogruppe defineerivad HVS-I piirkonna mutatsioonid. Hg U9 puhul ei ole andmete vähesuse tõttu veel täpselt teada, kas kõik toodud mutatsioonid on defineerivad.

hg	HVS-I def. mutatsioonid	hg	HVS-I def. mutatsioonid
<b>U1</b>	16249	<b>U5</b>	16270
<b>U1a</b>	16189, 16249	<b>U5a</b>	16270, 15256
<b>U1b</b>	16111, 16311, 16249, 16327	<b>U5b1b</b>	16189, 16270
<b>U2</b>	16051	<b>U5b1c</b>	16189, 16270, 16311
<b>U2a</b>	16051, 16206C	<b>U6</b>	16172, 16219
<b>U2c</b>	16051, 16234	<b>U7</b>	16318T
<b>U2e</b>	16051, 16129C, 16189, 16362	<b>U8a</b>	16146, 16342
<b>U3</b>	16343	<b>U8b/K</b>	16311,
<b>U3a</b>	16343, 16390	<b>U9a</b>	16051, 16261, 16278, 16311
<b>U4</b>	15356	<b>U9b</b>	16242

Analüüsid 18 mtDNA-d, mille järjestused kahel sekveneerimisel ei ühtinud, siis neist 15 paigutus peale teist sekveneerimist HVS-I piirkonna põhjal mõnda U haplogruppi. Kolmel juhul andis teistkordne sekveneerimine küll erineva tulemuse, kuid antud mtDNA-d jäid HVS-I põhjal siiski alamhaplogrupeerimata. Maroko B32 proovi puhul oli ilmselt tegemist eelneva haplogrupeerimise veaga, sest np 16172, mis defineerib haplogruppi U6, oli detekteeritud ka esimesel sekveneerimisel, kuid tõenäoliselt märkamata jäänud.

28 mtDNA proovi puhul, mille HVS-I nukleotiidne järjestus oli kahel sekveneerimisel sama, võib detekteerida mitmeid mutatsioonide tagasipöördumisi. Näiteks positsioon 16270, mis defineerib haplogruppi U5, on kõigil antud valimi ebatüüpiliste HVS-I järjestusmotiividega U5-tel tagasipöördunud. Sama võib näha ka haplogrupi U4 puhul, kus tagasi on pöördunud defineeriv mutatsioon 16356.

Analüüsid tabelis 5 toodud sekveneerimistulemuste erinevusi, saab välja tuua mitmeid oletatavaid põhjuseid. Näiteks Eesti 361 ja Küprose proovide puhul on näha, et järjestuse esimene osa kattub, kuid tagumine mitte. Siin võib põhjus olla selles, et esimese sekveneerimise käigus ei ole õieti kokku viidud kerge ja raske ahela järjestusi. Kui sekvents on täiesti erinev, siis on toimunud proovide segimine kas praktilise töö käigus või hilisemal järjestuste analüüsimisel. Kui mõni mutatsioon puudub nukleotiidest järjestusest võrreldes teise sekveneerimisega, on tegemist ilmselt tähelepanematusega

sekventside analüüsimisel ning kui on nõ üle, võib tegu olla fantoom-mutatsiooniga, mille tekkepõhjusteks peetakse sekveneerimiskeemia iseärasusi.

MtDNA kodeeriva regiooni defineerivate nps ja HVS-I piirkonna sekveneerimistulemuste analüüsi põhjal on tabelis 5 määratud kõigi antud valimi mtDNA-de haplogrupiline kuuluvus.

Antud tulemusi kokku võttes peab välja tooma kahte olulist asja. Esiteks, nagu tehtud analüüs selgelt näitab, on mtDNA kodeeriva regiooni defineerivad mutatsioonid haplogrupeerimisel määrava tähtsusega. Kiirelt muteeruva HVS-I piirkonna informatsioon ei pruugi olla alati piisav, eriti kui on toimunud defineerivate mutatsioonide tagasipöördumine. Selleks, et vältida üleliigse müra teket fülogeneetilistes puudes, on mtDNA-de haplogrupeerimisel oluline põhjalik ja kompleksne analüüs. Teiseks tõi antud töö välja katsete korratavuse probleemi, millele põhjustele tuleks kindlasti pöörata suuremat tähelepanu.

### 3.4. mtDNA-de täisjärjestuse määramine ja fülogeneetiliste sünteesipuude analüüs

Selleks, et täpsustada antud töös analüüsitud ebatavaliste HVS-I järjestusmotiividega U haplogruppide asukohta fülogeneetilisel puul, valiti antud proovide hulgast välja 6 mtDNA-d ning täissekveneeriti. 5 proovi olid kodeeriva regiooni defineerivate nps alusel määratud haplogruppi U5 ja 1 haplogruppi U3. U5 mtDNA-del oli teada np 16270 ja U3 mtDNA-l np16343 tagasipöördumine. Sekveneerimiseks valitud proovid olid: U5 - Rumeenia 381, Prantsusmaa P1, Rootsi 74, Rootsi 40, Türgi 346; U3 – Türgi 172 (vt tabel 5). MtDNA-dele teostati genoomi täisjärjestamine ning saadud tulemuste ja kirjandusallikate põhjal koostati erinevad sünteesipuud. Detekteeritud mutatsioonid on ära toodud tabelis 7.

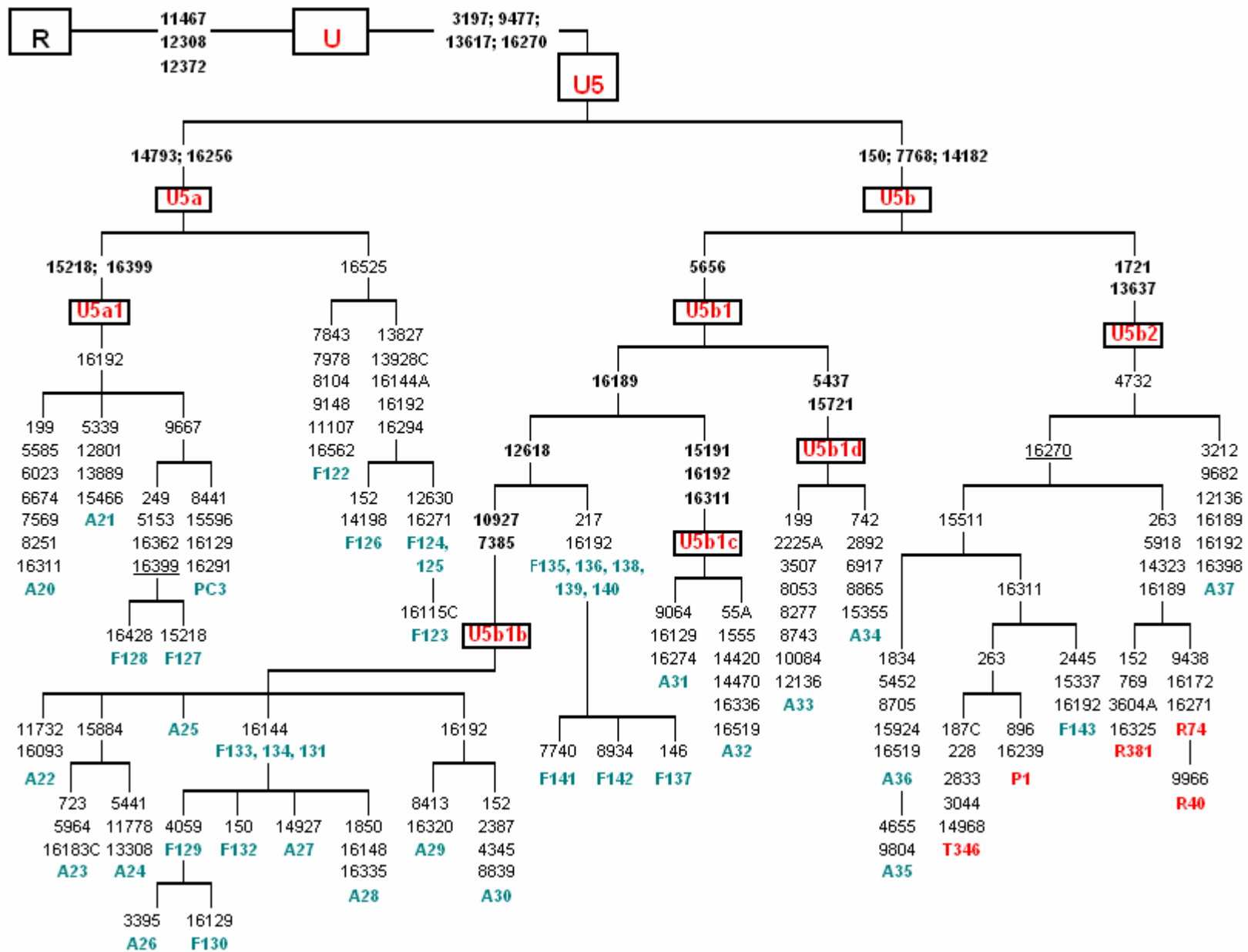
**Tabel 7.** MtDNA-de genoomi täisjärjestamisel detekteeritud mutatsioonid (va HVS-I piirkond). Sinises kirjas on märgitud rCRS-st tulenevad mutatsioonid (rCRS kuulub hg H2a)(Andrews *et al.* 1999).

vahemik	Rum. 381	Türgi 346	Rootsi 74	Rootsi 40	Prants. P1	Türgi 172	mutatsiooni mõju am.h. koodonile
16574	73	73	73	73	73	73	(kontrollregioon)
-	150	150	150	150	150	150	(kontrollregioon)
16569	152					152	(kontrollregioon)
		187C					(kontrollregioon)
		228					(kontrollregioon)
1 -	263	263	263	263	263	263	(kontrollregioon)
1999	769						-
					896		-
1000 -	1438	1438	1438	1438	1438	1438	-
	1721	1721	1721	1721	1721		-
						1811	-
2000 -	2706	2706	2706	2706	2706	2706	-
		2833					-
3000 -		3044					-
	3197	3197	3197	3197	3197		-
						3546A	sünon. m.
	3604A						-
4000 -						4188	sünon. m.
						4562	sünon. m.
						4640A	mittesünon. m.
						4654	mittesünon. m.
	4732	4732	4732	4732	4732		mittesünon. m.
	4769	4769	4769	4769	4769	4769	sünon. m.
5000 -						5465	sünon. m.
	5918		5918	5918			sünon. m.
7000 -	7028	7028	7028	7028	7028	7028	sünon. m.
	7768	7768	7768	7768	7768		sünon. m.
8000 -						8812	mittesünon. m.
	8860	8860	8860	8860	8860	8860	mittesünon. m.
						8778	sünon. m.
9000 -			9438	9438			mittesünon. m.
	9477	9477	9477	9477	9477		mittesünon. m.
						9656	sünon. m.
				9966			mittesünon. m.
11000	11467	11467	11467	11467	11467	11467	sünon. m.
-	11719	11719	11719	11719	11719	11719	sünon. m.
12000	12308	12308	12308	12308	12308	12308	-
-	12372	12372	12372	12372	12372	12372	sünon. m.
13000	13617	13617	13617	13617	13617		sünon. m.
-	13637	13637	13637	13637	13637		mittesünon. m.
						13743	sünon. m.
14000						14139	sünon. m.
-	14182	14182	14182	14182	14182		sünon. m.
	14323		14323	14323			sünon. m.
	14766	14766	14766	14766	14766	14766	mittesünon. m.
		14968					sünon. m.
15000	15326	15326	15326	15326	15326	15326	mittesünon. m.
-						15454	sünon. m.
16024		15511			15511		sünon. m.
						15629	sünon. m.

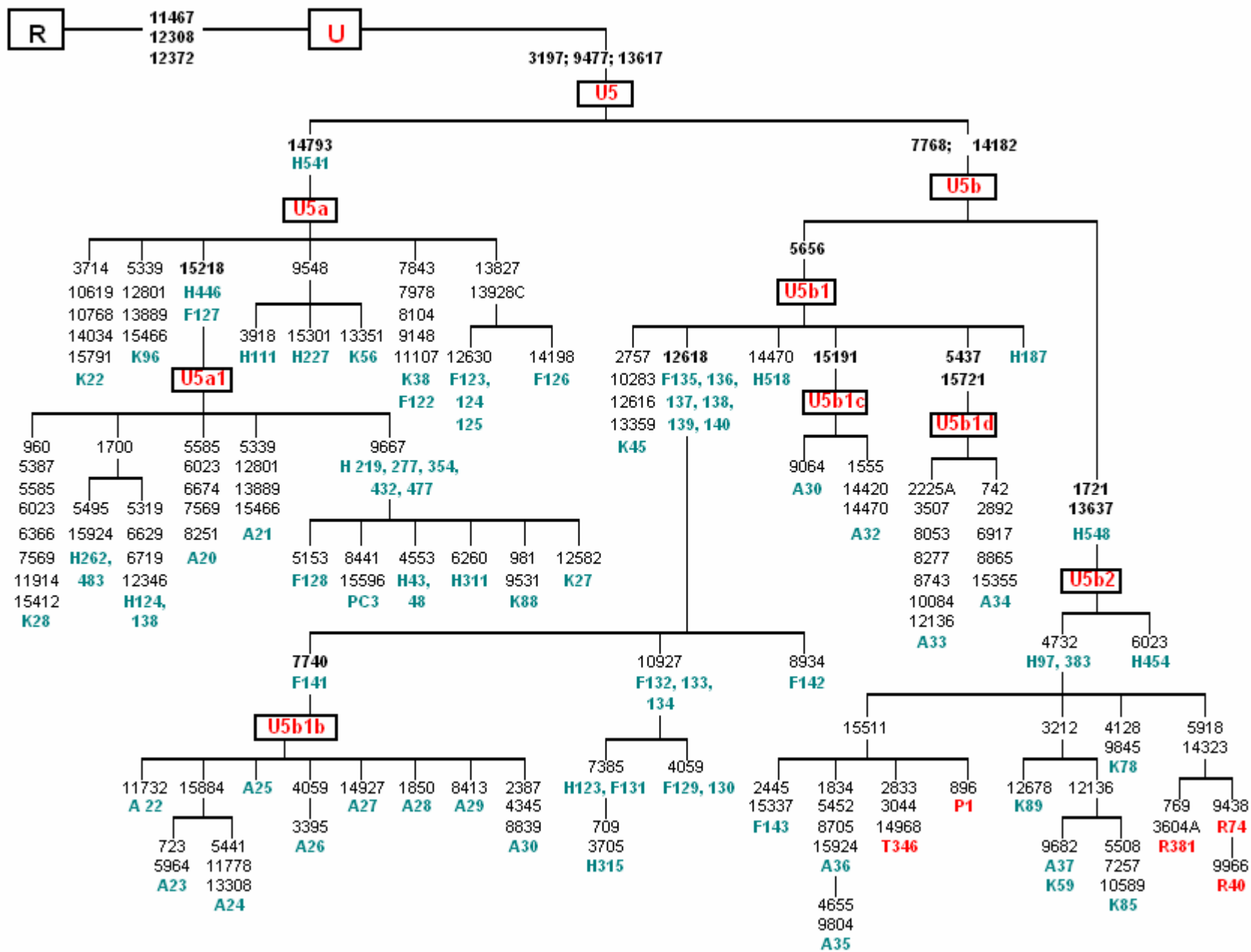
Joonisel 6 on haplogrupi U5 täisjärjestuste sünteesipuu ja joonisel 7 kodeeriva regiooni sünteesipuu. Antud töös sekveneeritud U5 mtDNA-d kuulusid alamhaplogruppi U5b2. U3 mtDNA paigutus U3b alamhaplogruppi ning täisjärjestuse tulemused on välja toodud joonisel 8.

U5 kodeeriva regiooni sünteesipuu alusel (joonis 7) arvutati U5b2 alamharu koalestentsiaeg, kuna antud harusse kuulusid ka antud töös sekveneeritud U5 mtDNA-d. Arvutamisel võeti arvesse kõik kodeeriva regiooni mutatsioonid (va indelid) ning ühe mutatsiooni tekkeajaks loeti 5139 aastat (Mishmar *et al.* 2003). U5b2 alamharu koalestentsiajaks saadi  $19\,000 \pm 6800$  aastat. Võrreldes saadud tulemust kirjanduse andmetega, siis Achilli *et al.* 2005 on arvutanud sama meetodit aluseks võttes U5 koalestentsiajaks  $41\,400 \pm 9200$  ja U5b1b  $8600 \pm 2400$  aastat. Alamharu U5b2 arvutatud vanus on noorem kui haplogrupil U5 ja vanem kui U5b1b-l, mis on igati ootuspärane tulemus. Koalestentsiaja arvutamisel saab aluseks võtta ka ainult kodeeriva regiooni sünonüümseid transitsioone ning sel juhul loetakse ühe mutatsiooni tekkeajaks 6764 aastat (Kivisild *et al.* 2005). Vanemate haplogruppide/alamharude puhul annavad Mishmar *et al* ja Kivisild *et al* suhteliselt sarnase koalestentsiaja, kuid nooremate mtDNA harude puhul on erinevus suhteliselt suur. Põhjus on mittesünonüümsetes transitsioonides, mida nooremates mtDNA harudes esineb suuremal hulgal, kui vanemates, kust puhastav seleksioon on suure osa mittesünonüümseid transitsioone eemaldanud. Võttes aluseks Kivisild *et al.* (2005) lähenemise saadi U5b2 alamharu koalestentsiajaks  $9100 \pm 2200$  aastat. Saadud tulemust ei olnud võimalik kirjanduse allikatega võrrelda, sest vastavasisulisi andmeid ei ole veel avaldatud.

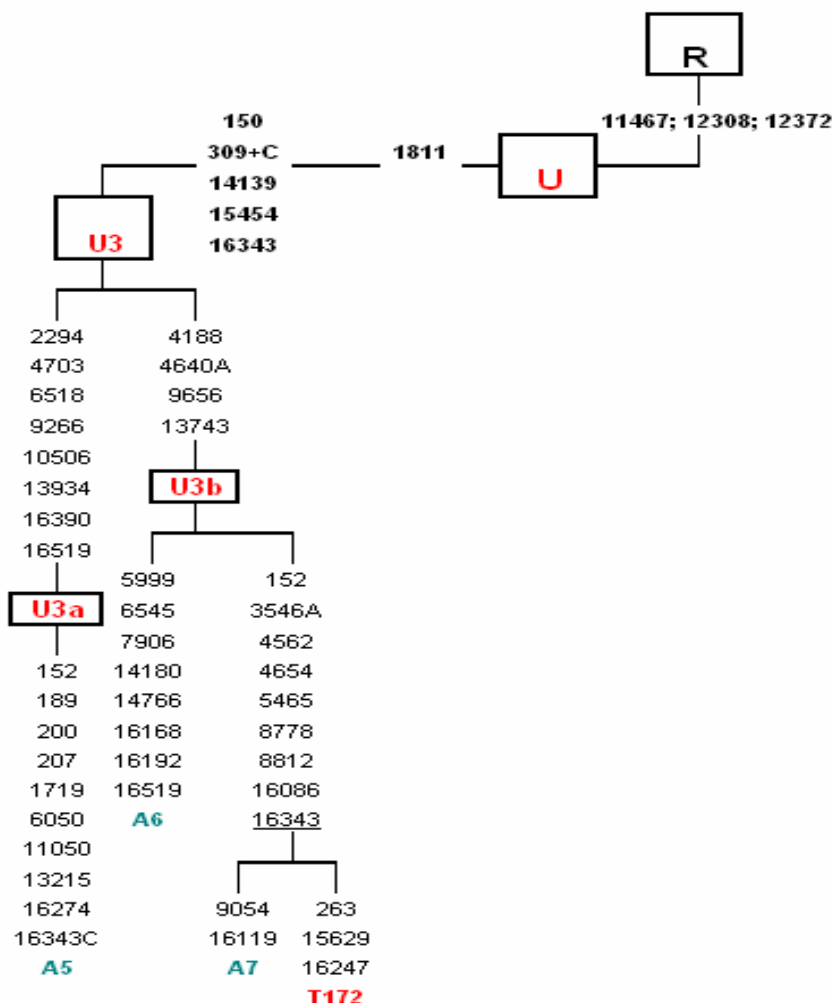
Kuue täisjärjestatud mtDNA nukleotiidse varieeruvuse analüüsi näitas, et antud individid kuulusid kõik U haplogruppide eelnevalt defineeritud alamharudesse. Sellest saab järeldada, et harvaesinevad nps kombinatsioonid HVS-I piirkonnas on enamasti antud piirkonna kiire muteerumise tulemus, mitte seni defineerimata alamharud. HVS piirkonna ebaselguste lahendamiseks on valik kodeeriva regiooni defineerivate nps kompleksne analüüs, sest HVS piirkonna informatsioon ei ole antud olukorras piisav.



Joonis 6. U5 täisjärjestuste sünteesipuu \*



Joonis 7. U5 kodeeriva regiooni sünteesipuu □



**Joonis 8.** Haplogrupi U3 täisjärjestuse sünteesipuu. Antud töös sekveneeritud Türgi 172 mtDNA on joonisel näidatud punase tähisega T172. Np 16343 tagasipöördumine on tähistatud allajoonitult. Puu koostamisel on kasutatud kirjanduse allikat Achilly *et al.* 2005.

\* Antud töös sekveneeritud mtDNA täisjärjestused on sünteesipuul näidatud siniste tähistega: Rumeenia 381(R381); Türgi 346(T346), Rootsi 74(R74), Rootsi 40(R40), Prantsuse P1(P1). Kirjanduse allikatest võetud järjestused on puul tähistatud rohelist värvi autori nime esitähedega: Achilli *et al.*2005 (A); Palanichamy *et al.*2004 (P); Finnälä *et al.* 2001 (F). Kokku on kirjanduse allikatest pärit 41 täisjärjestust lisaks antud töö 5 täisjärjestust. Sünteesipuu koostamisel ei kasutatud indeleid. Tagasipöörduvad mutatsioonid on allajoonitud.

□ Antud töös sekveneeritud mtDNA täisjärjestused on sünteesipuul tähistatud siniste tähistega: Rumeenia 381(R381); Türgi 346(T346), Rootsi 74(R74), Rootsi 40(R40), Prantsuse P1(P1). Kirjanduse allikatest võetud järjestused on puul tähistatud rohelist värvi autori nime esitähedega: Achilli *et al.*2005 (A) ; Kivisild *et al.*2006 (K); Palanichamy *et al.*2004 (P); Herrnstadt *et al.* 2002 (H); Finnälä *et al.* 2001 (F). Kokku on kirjanduse allikatest pärit 77 kodeeriva regiooni järjestust lisaks antud töö 5 täisjärjestust. Sünteesipuu koostamisel ei kasutatud indeleid ja kontrollregiooni informatsiooni.

## 4. KOKKUVÕTE JA JÄRELDUSED

MtDNA fülogeneetiline klassifikatsioon on siiani peamiselt baseerunud kontrollregiooni varieeruvuse ja kodeeriva regiooni polümorfisuse klassifikatsioonide sünteesil. Viimastel aastatel on üha suuremat tähelepanu hakatud pöörama mtDNA genoomide täisjärjestustele, sest nendest saadav põhjalik informatsioon aitab lahendada seni probleemseid kohti fülogeneetilisel pool.

Antud magistritöö keskendus mtDNA superhaplogrupi U alamharude haplogrupeerimisega seotud probleemidele ja ebaselgustele. Töö käigus analüüsiti 46 mtDNA-st koosnevat valimit, mis olid eelnevalt ebatüüpilise HVS-I regiooni järjestusmotiivi ja np 12308 alusel defineeritud kui U\*, see tähendab, et antud mtDNA-d kuulusid superhaplogruppi U, kuid samas ei asetunud ühtegi U alamhaplogruppi. Töö käigus analüüsiti põhjalikult antud valimi mtDNA kodeeriva regiooni U alamharusid defineerivaid nukleotiidipositsioone, määrati uuesti HVS-I piirkonna nukleotiidne järjestus ning kuuele proovile teostati genoomi täisjärjestuse määramine.

Saadud tulemused näitasid, et mtDNA kodeeriva regiooni defineerivad mutatsioonid on haplogrupeerimisel määrava tähtsusega, eriti juhul, kui HVS-I piirkonna alusel ei ole võimalik haplogrupilist kuuluvust määrata, sest on toimunud näiteks defineerivate mutatsioonide tagasipöördumine. Antud töö käigus sekveneeritud mtDNA täisjärjestused ei defineerinud ühtegi uut U alamharu, vaid asetsid kodeeriva regiooni põhjal teadaolevatesse U harudesse. Antud tulemused näitasid, et põhjalikum analüüs suutis lahendada kõigi seni U\*-na defineeritud mtDNA-de alamhaplogrupilise kuuluvuse.

Teine aspekt, mille käesolev töö välja tõi, oli katsete korratavuse ja mtDNA analüüsimise ning tõlgendamise probleem. Võrreldes eelnevalt teostatud HVS-I piirkonna sekveneerimise andmeid ja antud töös saadud tulemusi, oli sekveneeritud järjestuste kattuvus 61%. Teistkordsel sekveneerimisel saadud erinevad tulemused võimaldasid 83% juhtudel erinevalt algandmetest määrata mtDNA haplogrupilise kuuluvuse HVS-I regiooni põhjal. See näitab, kui oluline on proovide hoolikas käsitlemine laboris ja algandmete korrektne analüüs, sest tehtud vead võivad mõjutada hilisemaid tulemusi ja järeldusi.

## SUMMARY

The phylogenetic classification of mtDNA has been based on the combinational system taking into account the variability of the control region and the polymorphisms of the coding region. During the last few years, more attention has been paid to the entire mtDNA genome sequences, because this information can help better solve the problematic issues on the phylogenetic tree.

This study focuses on the problems and obscurity involved with the haplogrouping of the mtDNA superhaplogroup U sub-branches. 46 mtDNA-s were analyzed, which were previously defined as U\* based on the np 12308 transition and HVS-I region sequencing results. It means that these mtDNA-s belonged to the superhaplogroup U, but did not fall into place with any of the U sub-haplogroups. During the research of chosen mtDNA-s polymorphisms, the coding region's defining mutations were comprehensively analyzed, the sequence of HVS-I region was re-determined and the fullsequence of 6 samples was determined.

The obtained results depict clearly that the defining mutations in the mtDNA coding region are very important in haplogrouping, especially when the classification can not be determined using the HVS-I region. The sequenced mtDNA-s did not define any new U sub-branches, but fell into already known U branches. The results illustrate that a complex analysis was able to solve the sub-haplogroupic classification of all the mtDNA-s previously defined as U\*.

Another aspect, which was brought up by this research, was the problems with experiments repeatability and mtDNA analysis results interpretation. Comparing the results with the previous sequencing of the HVS-I region, the observed sequence homology was not excellent. Different results obtained from the second sequencing enabled most of cases to determine the haplogroupic classification based upon the HVS-I region. This shows that the correct analysis of original data is important, because the mistakes during the analysis can affect successive results and conclusions.

## TÄNUAVALDUSED

Kõigepealt ma sooviksin tänada oma juhendajat Ene Metspalu, kes oli käesoleva magistritöö valmimisel oma nõu ja jõuga igati abiks. Ma soovin tänada ka prof. Richard Villemsit, tänu kellele said minu magistrantuuriõpingud üldse võimalikuks.

Mul oli hea meel viibida nende aastate jooksul meie õpetooli toredate inimeste ja kaastudengite keskel, kes oskasid luua meeldiva ja rõõmsameelse tööõhkkonna. Suur aitäh Ille Hilpusele ja Jaan Lindile, kelle kaasabil sai valmis antud magistritöö eksperimentaalne osa.

Olen tänulik oma kaaskolleegide positiivsele ja toetavale suhtumisele Tartu Ülikooli Kliinikumist ning siiras tänu Kadri Raudsepale, Rain Lehtmele ja Anu Tammele, kelle mõistev suhtumine tegi kergemaks minu õpingute lõpetamise.

Suurimad tänud minu kodustele, kes mind alati on toetanud ja minu tegemistele kaasa elanud ning Martinile, kes oli alati minu kõrval.

Käesolev magistritöö on saanud toetust ETF grantidest 6040KT ja 5807EM.

## KASUTATUD KIRJANDUS

- Achilli A, Rengo C, Magri C, Battaglia V, *et al.* (2004). The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75:910-918
- Achilli A, Rengo C, Battaglia V, Pala M, *et al.* (2005). Saami and Berbers – an unexpected mitochondrial DNA link. *Am J Genet* 76:000-000
- Anderson S, Bankier AT, Barrell BG, *et al.* (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genet* 23:147
- Ballinger SW, Schurr TG, Torroni A, Gan YY, *et al.* (1992). Southeast Asian mitochondrial DNA analysis reveals genetic continuity of ancient mongoloid migrations. *Genetics* 130:139-152
- Bandelt HJ, Forster P, Sykes BC, Richards M (1995). Mitochondrial portraits of human populations using median networks. *Genetics* 141:743-753
- Bandelt HJ, Forster P, Röhl A (1999). Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48
- Bandelt HJ, Lahermo P, Richards M, Macaulay V (2001a). Detecting errors in mtDNA data by phylogenetic analysis. *Int J Legal Med* 115:64-69
- Bandelt HJ, Alves-Silva J, Guimaraes PEM, *et al.* (2001b). Phylogeography of the human mitochondrial haplogroup L3e: a snapshot of African prehistory and Atlantic slave trade. *Ann Hum Genet* 65:549-563
- Bandelt HJ, Quintana-Murci L, Salas A, Macaulay V (2002). The fingerprint of fantom mutations in mitochondrial DNA data. *Am J Hum Genet* 71: 1150-1160
- Bandelt HJ, Kivisild T (2006). Quality assessment of DNA sequence data: autopsy of a mis-sequenced mtDNA population sample. *Ann Hum Genet* 70(Pt 3):314-26.
- Behar DM, Hammer MF, Garrigan D, Villemis R, Bonne-Tamir B, Richards M, Gurwitz D, *et al.* (2004). MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. *Eur J Hum Genet* 12:355-364
- Bermisheva MA, Tambets K, Villemis R, Khusnutdinova EK (2002). Diversity of mitochondrial DNA haplogroups in ethnic populations of the Volga-Ural region of Russia. *Mol Biologiya* 36:990-1001
- Bogenhagen DF (1999). Repair of mtDNA in vertebrates. *Am J Hum Genet* 64:1276-81.
- Brown WM (1979). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc Natl Sci USA* 77:3609–3609

- Brown MD, Hosseini SH, Torroni A, Bandelt HJ, Allen JC, Schurr TG, Scozzari R, *et al.* (1998). MtDNA haplogroup X: an ancient link between Europe/Western Asia and North America? *Am J Hum Genet* 63:1852-1861
- Cann RL, Stoneking M, Wilson AC (1987). Mitochondrial DNA and human evolution. *Nature* 325:31–36
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994). *The History and Geography of Human Genes*. Princeton University Press, Princeton
- Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC (1995). Analysis of mtDNA variation in African populations reveals the most ancient of all human continent -specific haplogroups. *Am J Hum Genet* 57:133-149
- Chen YS, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, Wallace DC (2000). MtDNA variation in the South African Kung and Khwe-and their genetic relationships to other African populations. *Am J Hum Genet* 66:1362-1383
- Churchill SE, Smith FH (2000). Makers of the Early Aurignacian of Europe. *Yearbook of Physical Anthropol* 49:61-115
- Clark JD, Beyene Y, Woldegabriel G, *et al.* (2003). Stratigraphic, chronological and behavioural contexts of Pleistocene Homo sapiens from Middle Awash, Ethiopia. *Nature* 423:747 52
- Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bertranpetit J (1996). Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol Biol Evol* 13:1067-1077
- Derbeneva OA, Starikovskaya EB, Volodko NV, Wallace DC, Sukernik RI (2002a). Mitochondrial DNA variation in the Kets and Nganasans and its implications for the initial peopling of northern Eurasia. *Genetika* 38:1554-1560
- Derbeneva OA, Starikovskaya EB, Wallace DC, Sukernik I (2002b). Traces of early Eurasians in the Mansi of Northwest Siberia revealed by mitochondrial DNA analysis. *Am J Hum Genet* 70:1009-1014
- Derenko MV, Grzybowski T, Malyarchuk BA, Czarny J, Miscicka-Sliwka D, Zakharov IA (2001). The presence of mitochondrial haplogroup X in Altaians from South Siberia. *Am J Hum Genet* 69:237-241
- Derenko MV, Grzybowski T, Malyarchuk BA, *et al.* (2003). Diversity of mitochondrial DNA lineages in South Siberia. *Ann Hum Genet* 67:391-411
- Desjardins P, Morais R (1990). Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. *J Mol Biol* 212: 599–634
- Fedorova SA, Bermisheva MA, VILLEMS R, Maksimova NR, Khusnutdinova EK (2003). Analysis of Mitochondrial DNA Lineages in Yakuts. *Mol Biologiya* 37:643-653

- Finnilä S, Lehtonen MS, Majamaa K (2001). Phylogenetic network for European mtDNA. *Am J Hum Genet* 68:1475-1484
- Foley R (1998). The context of human genetic evolution. *Genome Res* 8:339-347
- Forster P, Harding R, Torroni A, Bandelt HJ (1996). Origin and evolution of Native American mtDNA variation: a reappraisal. *Am J Hum Genet* 59:935-945
- Forster P, Torroni A, Renfrew C, Röhl A (2001). Phylogenetic star contraction applied to Asian and Papua mtDNA evolution. *Mol Biol Evol* 18:1864-1881
- Giles RE, Blanc H, Cann HM, Wallace DC (1980). Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci USA* 77:6715-6719
- Greenberg BD, Newbold JE, Sugino A (1983). Intraspecific nucleotide sequence variability surrounding the origin of the replication in human mitochondrial DNA. *Gene* 21:33-49
- Haetta OM (1996). The Saami: an indigenous people of the Arctic. Davvi Girji, Karasjohka/Karasjoki, Vaasa.
- Hastings IM (1992). Population genetic aspects of deleterious cytoplasmic genomes and their effect on the evolution of sexual reproduction. *Genet Res* 59:215-225
- Herrnstadt C, Elson JL, Fahy E, Preston G, Turnbull DM, Anderson C, *et al.* (2002). Reduced-median-network analysis of mitochondrial DNA coding-region sequences for the major African, Asian and European haplogroups. *Am J Hum Genet* 70:1152-1171
- Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000). Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708-713
- Ingman M, Gyllensten U (2001). Analysis of the complete human mtDNA genome: methodology and inferences for human evolution. *J Hered* 92:454-461
- Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, *et al.* (1999a). Deep common ancestry of Indian and western-Eurasian mitochondrial DNA lineages. *Curr Biol* 9:1331-1334
- Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha SS, Villems R (1999b). The place of the Indian mtDNA variants in the global network of maternal lineages and the peopling of the Old World. In: *Genomic Diversity*. Kluwer Academic/Plenum Publishers, pp 135-152
- Kivisild T, Papiha SS, Rootsi S, *et al.* (2000). An Indian ancestry: a key for understanding human diversity in Europe and beyond. In: Renfrew C, Boyle K (eds) *Archaeogenetics: DNA and the population prehistory of Europe*. McDonald Institute Monographs, Cambridge, pp 267-275
- Kivisild T, Tolk HV, Parik J, Wang Y, Papiha SS, Bandelt HJ, R Villems (2002). The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* 19:1737-1751
- Kivisild T, Rootsi S, Metspalu M, Metspalu E, Parik J, Kaldma K, Usanga E, *et al.* (2003a). The Genetics of Language and Farming Spread in India. In: Bellwood P, Renfrew C (eds) *Examining the*

- farming/language dispersal hypothesis. The McDonald Institute for Archaeological Research, Cambridge, pp 215-222
- Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, *et al.* (2003b). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet* 72:313-332
- Kivisild T, Shen P, Wall DP, Do B, Sung R, *et al.* (2005). The role of selection in the human mitochondrial genomes. *Genetics* 172:373-387
- Kong QP, Yao YG, Sun C, Bandelt HJ, Zhu CL, YP Zhang (2003). Phylogeny of East Asian mitochondrial DNA lineages inferred from complete sequences. *Am J Hum Genet* 73:671-6
- Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997). Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19-30
- Krings M, Geisert H, Schmitz RW, Krainitzki H, Pääbo S (1999a). DNA sequence of the mitochondrial hypervariable region II from the neandertal type specimen. *Proc Natl Acad Sci USA* 96:5581-5585
- Krings M, Salem AH, Bauer K, Geisert H, Malek A, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, *et al.* (1999b). MtDNA analysis of Nile River valley populations: A genetic corridor or a barrier to migration? *Am J Hum Genet* 64:1166-1176
- Larruga MJ, Diez F, Pinto MF, Flores C, Gonzalez AM (2001). Mitochondrial DNA characterisation of European isolates: the Maragatos from Spain. *Eur J Hum Genet* 9, 708-716
- Lightowers RN, Chinnery PF, Turnbull DM, Howell N (1997). Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends Genet* 13:450-455
- Loogväli EL, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, *et al.* (2004). Disuniting uniformity: a pied cladistic canvas of mtDNA haplotype H in Eurasia. *Mol Biol Evol* 21(11): 2012-2021
- Lutz S, Weisser HJ, Heizmann J, Pollak S (1998). Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 111:67-77
- Maca-Meyer N, Gonzalez MA, Larruga MJ, *et al.* (2001). Major genomic mitochondrial lineages delineate early human expansions. *BMS Genetics* 1471– 2156/2/13
- Maca-Meyer N, Gonzalez AM, Pestano J, Flores C, Larruga JM, Cabrera VM (2003). Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. *BMC Genetics* 4:15
- Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonne-Tamir B, Sykes B, Torroni A (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control region sequences and RFLPs. *Am J Hum Genet* 64:232–249
- Macaulay V, Hill C, Achilli A, Rengo C, *et al.* (2005). Single rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science* 308: 1034-1036

- Meinilä M, Finnilä S, Majamaa K (2001). Evidence for mtDNA admixture between the Finns and the Saami. *Hum Hered* 52:160-170
- Merrywether DA, Clark AG, Ballinger SW, Schurr TG, Soodyall H, Jenkins T, Sherry ST, Wallace DC (1991). The structure of human mitochondrial DNA variation. *J Mol Evol* 3:543-555
- Metspalu E, Kivisild T, Kaldma K, Parik J, Reidla M, Tambets K, Villems R (1999). The Trans-Caucasus and the expansion of the Caucasoid-specific human mitochondrial DNA. In: Deka R, Papiha S, Chakraborty R (eds) *Genomic Diversity: Applications in Human Population Genetics*. Kluwer Academic/Plenum Publishers, New York, p 121-133
- Metspalu M, Kivisild T, Metspalu E, Parik J, *et al.* (2004). Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet* 5:26
- Meyer S, Weiss G, Haesler A (1999). Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152:1103-1110
- Mishmar D, Ruiz-Pesini E, Golik P, *et al.* (2003). Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100:171-176
- Orekhov V, Poltoraus A, Zhivotovsky LA, Spitsyn V, Ivanov V, Yankovsky N (1999). Mitochondrial DNA sequence diversity in Russians. *FEBS Letters* 445:197-201
- Ovchinnikov I, Götherström A, Romanova GP, Kharitonov VM, Liden K, Goodwin W (2000). Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404:490-493
- Palanichamy M, Sun C, Agrawal S, Bandelt HJ, *et al.* (2004). Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet* 75:966-978
- Parson W, Parsons TJ, Scheithauer R, Holland MM (1998). Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int J Legal Med* 111: 124-132
- Parson W, Brandstätter A, Alonso A, *et al.* (2004). The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. *Forensic Sci Int* 139:215-226
- Pereira L, Prata MJ, Amorim A (2000). Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Eur J Hum Genet* 64, 491-506
- Pereira L, Macaulay V, Torroni A, Scozzari R, Prata MJ, Amorim A (2001). Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. *Ann Hum Genet* 65:439-458

- Pesole G, Gissi C, De Chirico A, Saccone C (1999). Nucleotide substitution rate of mammalian mitochondrial genomes. *J Mol Evol* 48:427-434
- Piganeau G, Eyre-Walker A (2004). A reanalysis of the indirect evidence for recombination in human mitochondrial DNA. *Heredity* 92:282-288
- Qian YP, Chu ZT, Dai Q, Wei CD, Chu JY, Tajima A, Horai S (2001). Mitochondrial DNA polymorphisms in Yunnan nationalities in China. *J Hum Genet* 46:211-220
- Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, *et al.* (1999). Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. *Nature Genet* 23:437-441
- Quintana-Murci L, Chaix R, Wells S, *et al.* (2004). Where West meets East: The complex mtDNA landscape of the Southwest and Central Asian corridor. *Am J Hum Genet* 74:827-845
- Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM, Bandelt HJ (1998). Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. *Ann Hum Genet* 62:531-550
- Rando JC, Cabrera VM, Larruga JM, Hernandez M, Gonzalez AM, Pinto F, Bandelt HJ (1999). Phylogeographic patterns of mtDNA reflecting the colonisation of the Canary Islands. *Ann Hum Genet* 63, 413-428
- Reidla M, Kivisild T, Metspalu E, *et al.* (2003). Origin and diffusion of mtDNA haplogroup X. *Am J Hum Genet* 73:1178-1190
- Richards M, Côté-Real H, Forster P, Macaulay V, *et al.* (1996). Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59(1):185-203
- Richards MB, Macaulay VA, Bandelt HJ, Sykes BC (1998). Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62:241-260
- Richards M, Macauley V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, *et al.* (2000). Tracing European founder lineages in the near Eastern mtDNA pool. *Am J Hum Genet* 67:1251-1276
- Richards M, Macaulay V, Torroni A, Bandelt HJ (2002). In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet* 71:1168-1174
- Richards M, Rengo C, Cruciani F, Gratrix F, Wilson J, Scozzari R, Macaulay V, *et al.* (2003). Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am J Hum Genet*. 72:1058-1064
- Rosa A, Brehm A, Kivisild T, Metspalu E, Villems R (2004). MtDNA profile of West Africa Guineans: towards a better understanding of the Senegambia region. *Ann Hum Genet* 68, 340-352

- Saillard J, Forster P, Lynnerup N, Bandelt HJ, Norby S (2000). MtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am J Hum Genet* 67:718-726
- Sajantila, A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, Beckman L, Tranebjaerg L, Geddedahl T, Isseltarver L, Di Rienzo A, Pääbo S (1995). Genes and languages in Europe – an analysis of mitochondrial lineages. *Genome Res* 5:42-52
- Salas A, Richards M, De la Fe T, Lareu MV, Sobrino B, Sánchez-Diz P, Macaulay V, Carracedo A (2002). The making of the African mtDNA landscape. *Am J Hum Genet* 71: 1082-1111
- Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* 74:454-465
- Schurr TG, Sukernik RI, Starikovskaya YB, Wallace DC (1999). Mitochondrial DNA variation in Koryaks and Itelmen: population replacement in the Okhotsk Sea–Bering Sea region during the Neolithic. *Am J Phys Anthropol* 108:1-39
- Smith DG, Malhi RS, Eshleman J, Lorenz JG, Kaestle FA (1999). Distribution of mtDNA haplogroup X among Native North Americans. *Am J Phys Anthropol* 110:271-284
- Starikovskaya YB, Sukernik RI, Schurr TG, Kogelnik AM, Wallace DC (1998). MtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *Am J Hum Genet* 63:1473-1491
- Stringer CB, Andrews P (1988). Genetic and fossil evidence for the origin of modern humans. *Science* 239:1263-1268
- Stringer C (2000). Coasting out of Africa. *Nature* 405:24-25, 27
- Stringer C (2003). Human evolution: Out of Ethiopia. *Nature* 423:692-3, 695
- Sutovsky P, Van Leyen K, McCauley T, Day BN, Sutovsky M (2004). Degradation of paternal mitochondria after fertilization: implications for heteroplasmy, assisted reproductive technologies and mtDNA inheritance. *Reprod Biomed Online* 8:24-33
- Takahata N, Lee SH and Satta Y (2001). Testing multiregionality of modern human origins. *Mol Biol Evol* 18:172-183
- Tambets K, Kivisild T, Metspalu E, *et al.* (2000). The topology of the maternal lineages of the Anatolian and Trans-Caucasus populations and the peopling of Europe: some preliminary considerations. In: Renfrew C, Boyle K (eds) *Archaeogenetics: DNA and the Population Prehistory of Europe*, Cambridge, p 219-235
- Tambets K, Tolk HV, Kivisild T, *et al.* (2003). Complex signals for population expansions in Europe and beyond. In: Bellwood P, Renfrew C (eds). *Examining the farming/language dispersal hypothesis*. McDonald Institute Monographs, Cambridge, pp 449-457

- Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogväli EL, Tolk HV, *et al.* (2004). The Western and Eastern Roots of the Saami--the Story of Genetic "Outliers" Told by Mitochondrial DNA and Y Chromosomes. *Am J Hum Genet* 74:661-682
- Thompson WE, Ramalho-Santos J, Sutovsky P (2003). Ubiquitination of prohibitin in mammalian sperm mitochondria: possible roles in the regulation of mitochondrial inheritance and sperm quality control. *Biol Reprod* 69:254-260
- Torroni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, Comuzzie AG, Wallace DC (1993). MtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native Americans. *Am J Hum Genet* 53:591-608
- Torroni A, Neel JV, Barrantes R, Schurr TG, Wallace DC (1994a). Mitochondrial DNA 'clock' or the Amerinds and its implications for timing their entry into North-America. *Proc Natl Acad Sci USA* 91: 1158–1162
- Torroni, A, Lott MT, Gabell MF, Chen YS, Lavergne L, Wallace DC (1994b). MtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am Hum Genet* 55:760–776
- Torroni A, Huoponen K, Francalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, *et al.* (1996). Classification of European mtDNAs from an analysis of three European populations. *Genetics* 144:1835–1850
- Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savantaus ML, Bonne-Tamir B, Scozzari R (1998). MtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137-1152
- Torroni A, Cruciani F, Rengo C, *et al.* (1999). The A1555G mutation in the 12S rRNA gene of human mtDNA: recurrent origins and founder events in families affected by sensorineural deafness. *Am J Hum Genet* 65:1349-1358
- Torroni A, Bandelt HJ, Macaulay V, *et al.* (2001). A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* 69:844-852
- Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006). Harvesting the fruit of the human mtDNA tree. *Trends Genet* in press
- Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997). Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 7:996-1005
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991). African populations and the evolution of human mitochondrial DNA. *Science* 253: 1503–1507

- Villems R, Adojaan M, Kivisild T, Metspalu E, Parik J, Pielberg G, Rootsi S *et al.* (1998). Reconstruction of maternal lineages of Finno-Ugric speaking people and some remarks on their paternal inheritance. In: Wiik K, Julku K (eds) *The roots of people and languages of Northern Eurasia I*. Societas Historiae fenno-Ugricae, Turku, pp 180-200
- Wallace DC, Brown MD, Lott MT (1999). Mitochondrial DNA variation in human evolution and disease. *Gene* 238:211-230
- Watson E, Bauer K, Aman R, Weiss G, Haeseler A, Pääbo S (1996). MtDNA sequence diversity in Africa. *Am J Hum Genet* 59:437-444
- Watson E, Forster P, Richards M, Bandelt HJ (1997). Mitochondrial footprints of human expansion in Africa. *Am J Hum Genet* 61:691–704
- Wolpoff M, Wu XZ, Thorne AG (1984). Modern *Homo sapiens* origins: A general theory of hominid evolution involving the fossil evidence from East Asia. In: Smith FH, Spencer F (eds) *The origin of modern humans: A world survey of the fossil evidence*. Alan L. Riss, New York, pp 411-483
- Wolpoff M, Caspari R (1997). *Race and Human Evolution*. Simon & Schuster, New York
- Yao YG, Kong QP, Bandelt HJ, Kivisild T, Zhang YP (2002). Phylogeographic differentiation of mitochondrial DNA in Han Chinese. *Am J Hum Genet* 70:635-651
- Yao YG, Zhang YP (2002). Phylogeographic analysis of mtDNA variation in four ethnic populations from Yunnan Province: new data and a reappraisal. *J Hum Genet* 47:311-318
- Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, Patthy L, Ramsay M, Jenkins T, Shyue SK, Li WH (2002). Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* 161:269-274

## LISAD

**Tabel 1. HVS-I amplifitseerimiseks ja sekveneerimiseks kasutatud praimerid**

F	15974	5' ACTCCACCATTAGCACCCAAAGCT 3'
R	16420	3' CTGGTGGTAGGAGGCACTTTAGT 5'

**Tabel 2. MtDNA täisjärjestuste amplifitseerimiseks kasutatud praimerid**

praim. paar 1	F	330	5' CACTTAAACACATCTCTGCCA 3'
	R	2216	3' AATTCTTTTCGCAAGTTTCGAGTTGT 5'
2	F	2063	5' GAACCCTCTAAATCCCCTTGTAAT 3'
	R	4251	3' TAATGTTAGAGGTCGTAAGG G 5'
3	F	4166	5' ACCTCCTATGAAAAACTTCCTACC 3'
	R	6220	3' GTATTGGAAGACTGAGAATGG 5'
4	F	6242	5' CGCATCTGCTATAGTGGAGGCCG 3'
	R	8366	3' TTGTGGAGAAATGTCACCTTACGG 5'
5	F	7925	5' GCGGACTAATCTTCAACTCCTA 3'
	R	9884	3' AAGTAGGCGGTTGATTATAAAGTG 5'
6	F	9767	5' CATTCCGACGGCATCTACGGCTC 3'
	R	11748	3' AGGAGTAATGATAAGACGGATCG 5'
7	F	11614	5' CATTGCATACTTTCAATCAGCCAC 3'
	R	13638	3' GGATTGTCCAGTTGGAGCGAAGG 5'
8	F	13539	5' ATCATACACAAACGCCTGAGCC 3'
	R	15431	3' TTAGTTTCTGCGAATGATGTG 5'
9	F	15149	5' GAGGCCAAATATCATTCTGAGGG 3'
	R	743	3' AGTGGGAGATTTAGTGGTGCTAG 5'

**Tabel 3. MtDNA täisjärjestuste sekveneerimiseks kasutatud praimerid**

Number esimeses veerus tähistab produkti amplifitseerimiseks kasutatud praimerpaari.

praim. paar 1	F	436	5' CCCAACTAACACATTATTTTCCCCT 3'
	F	1291	5' AGTGGGAGATTTAGTGGTGCTAG 3'
	F	1615	5' AACACAAAGCACCCAACTTACACT 3'
	R	743	3' AGTGGGAGATTTAGTGGTGCTAG 5'
	R	1080	3' CGATTCTGGGTTTGACCCTAATCT 5'
	R	1568	3' TTGTAAAATATATCTCCTCTGTTCA 5'
2	F	2245	5' AACTGAACTCCTCACA CCAATTGG 3'
	F	2891	5' CAATTGATCCAATAACTTGACCA 3'
	F	3388	5' CTAGGCTATATACAACACTACGCAA 3'
	F	4057	5' TCCCCTGAACTCTACACAACATAT 3'
	R	3717	3' AGCCGCGTGACGCTCGTCATCGG 5'
3	F	4308	5' GGAGCTTAAACCCCTTATTTCT 3'
	F	4925	5' CCCAAGGCACCCCTCTGACATCC 3'
	F	5548	5' AGCCCTCAGTAAGTTGCAATACT 3'
	R	4739	3' GTTACTTGGTATTGGTTATGATGG 5'

	R	5920	3' GTGACTACAAGCGGCTGGCAAC 5'
4	F	6890	5' TCGTTATACTAATGATCTGCTG 3'
	F	7458	5' GAATCGAACCCCCCAAAGCTGG 3'
	F	7793	5' GCCATCATCCTAGTCCTC ATCGC 3'
	R	6699	3' ATTGAATGATGAGGCCTTTTTTTC 5'
	R	7467	3' TCTGTTTTTTCCTTCCTTAGCTTGG 5'
5	F	8188	5' AGCAAACCACAGTTTCATGCCCA 3'
	F	9230	5' TCATATAGTAAAACCCAGCCCAT 3'
	R	8806	3' GAGGACGGAGTGAGTAAATGTG 5'
	R	9403	3' GCGCTACATTGTGCTCTTTCGTG 5'
	R	9848	3' TTGTGAAGTGCAGTAATAACCGA 5'
6	F	9911	5' CGAAGCCGCCGCCTGATACTGG 3'
	F	10672	5' CCTGCGAAGCAGCGGTGGGCCT 3'
	F	11295	5' TCACTCTCACTGCCCAAGAACTA 3'
	R	10289	3' GGAGGAAAATGGGGATGGTACT 5'
	R	10959	3' GGATTGTTGGGGGGAGGATTA 5'
7	F	11692	5' CGCAGTCATTCTCATAATCGCCC 3'
	F	12210	5' AAAGCTCACAAGAACTGCTAA 3'
	F	13172	5' GCTTAGGCGCTATCACCACTCTG 3'
	R	12818	3' GAGTAGTCAACTACTATGCGGGC 5'
	R	13403	3' TACTTGTTCTATAAGCTTTTTTAT 5'
8	F	13583	5' CCTCCCTGACAAGCGCCTATAG 3'
	F	14365	5' CACCAATCCTACCTCCATCGCTA 3'
	F	14815	5' CCCATCCAACATCTCCGCATGAT 3'
	R	13843	3' GGATCCTGAAGATTGTCTGGGATC 5'
	R	14562	3' AGTCTTATTATTGTGTGGGCTG 5'
9	F	15409	5' CCCTTACTACACAATCAAAGAC 3'
	F	15968	5' TCTTTAACTCCACCATTAGCACC 3'
	F	29	5' CTCACGGGAGCTCTCCATGCAT 3'
	R	15701	3' TTGTTTCGTATTATAAAGCGG 5'
	R	16515	3' AGGATGAAGTGACATAGGCTGTA 5'