

TARTU ÜLIKOOL  
LOODUS- JA TÄPPISTEADUSTE VALDKOND  
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT  
EVOLUTSIOONILISE BIOLOOGIA ÕPPETOOL

Silja Aruhein

**Inimese Y-kromosoomi haplogrupi N3a alamharude N3a3 ja N3a4  
leviku ja sisestruktuuri analüüs eestlaste ja nende naaberrahvaste  
seas**

Bakalaureusetöö (12 EAP)

Bioloogia eriala

Juhendajad: Eesti Biokeskuse vanemteadur PhD Kristiina Tambets

Eesti Biokeskuse vanemteadur PhD Siiri Rootsi

TARTU 2017

## **Inimese Y-kromosoomi haplogrupi N3a alamharude N3a3 ja N3a4 leviku ja sisestruktuuri analüüs eestlaste ja nende naaberrahvaste seas**

Kõik inimese Y-kromosoomi variandid on omavahel seostatavad isaliinide ühtse fülogeneesipuu kaudu. Tänu erineva kiirusega evolutsioneeruvatele geneetilistele markeritele on võimalik uurida selle puu harusid ehk haplogruppe (hg) ja nende struktuuri. Töö eesmärgiks oli uurida Y-kromosoomi Põhja-Euroopas sagedase hg N3 alamklaadide N3a3 ja N3a4 levikut ja määrata nende sagedus eestlastel ja nende naaberpopulatsioonidel ning võrrelda saadud tulemusi populatsioonide keelelist ja ruumilist tausta arvestades. N3a3 sisestruktuuri täiendati uute markerite L550 ja VL39 abil. Analüüsi tulemused näitasid, et ligi kolmandik genotüpiseeritud eestlastest kuulub hg N3a, millest valdav enamus kuulub alamklaadi N3a3 harusse N3a3a. Sellesse kuulub ka enamik siinse valimi slaavi ja balti keele rääkijatest. Väiksem osa eestlastest kuulub koos soome-ugri keelte rääkijatega alamklaadidesse N3a3b ja N3a4.

**Märksõnad:** Y-kromosoom, haplogrupp, populatsioonigeneetika, geneetilised markerid, eestlased

**CERCS kood:** B220 Geneetika, tsütogeneetika

### **The distribution and internal structure of human Y chromosomal haplogroup N3a sub-clades N3a3 and N3a4 among Estonians and their neighbouring populations**

The phylogenetic tree of human paternal lineages connects all variants of human Y chromosomes. Genetic markers evolving with different speed allow the study of the structure of the branches of the tree – the haplogroups (hg). The aim of this study is to reveal the distribution of hg N3a3 and N3a4, sub-clades of hg N3 frequent in Northern Europe, among Estonians and their neighbouring populations, in the context of their linguistic and spatial background. The internal structure of hg N3a3 was updated by new markers L550 and VL39. The results showed that almost third of the genotyped Estonians belong to the hg N3a and majority of these to the N3a3a sub-branch. This sub-clade encompasses also the majority of the Baltic and Slavic speakers from the current sample. A smaller portion of Estonians belong together with Finno-Ugric speakers to the sub-clades N3a3b and N3a4.

**Keywords:** Y chromosome, haplogroup, population genetics, genetic markers, Estonians

**CERCS code:** B220 Genetics, cytogenetics

## Sisukord

Kasutatud lühendid ja mõisted .....	5
Sissejuhatus .....	6
1. Kirjanduse ülevaade .....	7
1.1. Inimese Y-kromosoomi spetsiifilised omadused .....	7
1.2. Y-kromosoom kui vahend populatsioonigeneetika uuringutes .....	8
1.3. Y-kromosoomi põhilised markerite tüübid .....	10
1.3.1. Inimese Y-kromosoomi SNP-d .....	10
1.3.2. Inimese Y-kromosoomi STR-id .....	11
1.4. Inimese Y-kromosoomi fülogeneesipuu ja põhilised Euroopas levinud isaliinid .....	11
1.5. Haplogrupi N üldine iseloomustus ja fülogeograafiline levik .....	15
1.5.1. Haplogrupi N alamhaplogrupi N3 iseloomustus .....	17
2. Töö eesmärgid .....	20
3. Materjal ja metoodika .....	21
3.1. Töös kasutatud DNA proovid .....	21
3.2. DNA amplifitseerimine .....	22
3.2.1. Polümeraasi ahelreaktsioon ehk PCR .....	22
3.2.2. Geelelektroforees .....	22
3.2.3. DNA produkti „puhastamine“ .....	23
3.2.4. Sekveneerimine .....	23
3.2.4.4. STR analüüs .....	25
3.2.5. Andmeanalüüs .....	26
4. Tulemused ja arutelu .....	28
4.1. N3a3 uute markerite L550 ja VL39 ning N3a4 esinemissagedus uuritud valimis .....	28
4.2. N3a3 ja N3a4 varieeruvus STR-haplotüüpide alusel .....	29
4.3. N3a3 ja N3a4 alamharude seos uuritud populatsioonide geograafilise ja keelelise taustaga .....	30
Kokkuvõte .....	32
Tänuavaldus .....	34
Kasutatud kirjandus .....	35
Kasutatud veebiaadressid .....	41
Lisad .....	42
Lisa 1. Töös kasutatud praimerid .....	42

Lisa 2. STR-markerid ja tandemkorduste arv uuritud indiviididel.....	43
Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks.....	44

## **Kasutatud lühendid ja mõisted**

ap – aluspaar (*base pair*)

DNA (*deoxyribonucleic acid*) – desoksüribonukleiinhape

EBK – Eesti Biokeskus

EDTA (*ethylenediaminetetraacetic acid*) – etüleendiamiintetra-äädikhape

hg (*haplogroup*) – haplogrupp (klaad)

Mb (Mbp, *mega base pairs*) – miljon aluspaari

MRCA (*most recent common ancestor*) – viimane ühine eellane

MSY (*male-specific region of the Y chromosome*) – Y-kromosoomi meessoo-spetsiifiline ala

mtDNA (*mitochondrial DNA*) – mitokondriaalne DNA

$N_e$  (*genetically effective population size*) – geneetiliselt efektiivne populatsioonisuurus

NGS (*next-generation sequencing*) – järgmise põlvkonna sekveneerimine

PAR (*pseudoautosomal region*) – pseudoautosomaalne regioon

PCR (*polymerase chain reaction*) – polümeraasi ahelreaktsioon

SAP (*shrimp alkaline phosphatase*) – krevetite aluseline fosfataas

SNP (*single nucleotide polymorphism*) – üksiku nukleotiidi polümorfism

STR (*short tandem repeat*) – lühike kordusjärjestus ehk mikrosatelliit

## Sissejuhatus

Inimese Y-kromosoomi meessoospetsiifiline regioon (MSY) on haploidne ega rekombineeru X-kromosoomiga. Rekombinatsiooni puudumise tõttu pärandub MSY muutumatul kujul isalt pojale järgmisesse põlvkonda. Vaid aja jooksul tekkivad ja akumulereuvad mutatsioonid mõjutavad MSY-i mitmekesisust. Tänu sellele saab inimese Y-kromosoomi kasutada populatsioonigeneetilistes uurimistöodes, mis käsitlevad inimese demograafilise ajaloo erinevaid aspekte. Y-kromosoomist saame infot isaliinide leviku ja mitmekesisuse kohta.

Bakalaureusetöö käsitleb inimese Y-kromosoomi haplogrupp N3 alamhaplogruppide jaotust eestlaste ja naaberrahvaste seas. Kirjanduse osas antakse ülevaade inimese Y-kromosoomist ja selle spetsiifilistest omadustest, samuti nende rakendusvõimalustest populatsioonigeneetilistes uuringutes. Tutvustatakse Y-kromosomaalsete markerite tüüpe ja nende kasutusvõimalusi. Täpsem ülevaade antakse ka inimese Y-kromosoomi haplogruppide levikust, eelkõige haplogrupist N3 ja selle alamklaadidest.

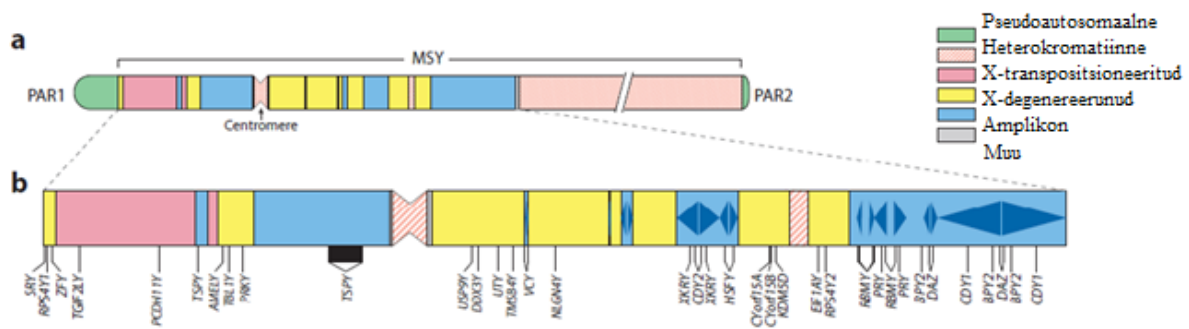
Hg N3 muudab huvipakkuvaks asjaolu, et sellel on Euraasia põhjaosas selge geograafiline struktuur ja ruumiline jaotus. Eestlaste ja eelkõige meie naabruses elavate keelesugulaste isaliinidest moodustab N3 umbes kolmandiku, samas kui Lääne-Euroopas see haplogrupp praktiliselt puudub. Hg N3 on muude Euroopas levinud isaliinide seas erakordne ka seetõttu, et on pärit Ida-Euraasiast ja seob meid täna Siberis elavate rahvastega. Käesoleva töö praktiline osa käsitleb hg N3 levikut mõnedes soomeugri, balti ja slaavi keeli rääkivates Euroopa populatsioonides. Uuritakse alamklaadi N3a3 sisestruktuuri ning N3a3 ja N3a4 geograafilist jaotust ja sagedust, saadud tulemuste alusel võrreldakse populatsioonide keelelist ja ruumilist seotust.

## 1. Kirjanduse ülevaade

### 1.1. Inimese Y-kromosoomi spetsiifilised omadused

Inimese Y-kromosoom on sugu määrava rolliga, ainult meestel esinev ja isalt pojale päranduv haploidne kromosoom. Tänu oma unikaalsele pärandumismustrile on see oluliseks informatsiooniallikaks inimese isaliini evolutsiooniuuringutes ning meditsiinis ja kohtugeneetikas [(Jobling and Tyler-Smith 2003) ja selles viidatud artiklid]. Umbes 95% Y-kromosoomi 57 Mb-sest järjestusest (Ensembl version nr. 88, avaldatud märtsis 2017, referentsjärjestus GRCh38.p10; <http://www.ensembl.org>) moodustab mitterekombineeruv ala (*non-recombining region of the Y chromosome* – NRY) ehk meessoospetsiifiline ala (*male-specific region of the Y chromosome* – MSY) (Skaletsky et al. 2003), mis ei vaheta meioosis X-kromosoomiga materjali (Joonis 1). MSY on oluline testiste arengus ja spermatogeneesis. See on ühtlasi suurim mitterekombineeruv ala inimese genoomis (Underhill and Kivisild 2007). Kaks segmenti, nn. pseudoautosomaalsed ehk X-kromosoomiga homologsed regioonid (*pseudoautosomal region* – PAR) kromosoomi otstes, rekombineeruvad X-kromosoomiga, aga need kokku katavad vähem kui 3 Mb Y-kromosoomi järjestusest. Rekombinatsiooni puudumise tõttu pärandub MSY muutumatul kujul edasi järgmistele põlvkondadele, välja arvatud vaid aja jooksul juhuslikult tekkinud ja kogunenud mutatsioonide osas (Jobling and Tyler-Smith 2003).

Kuna Y-kromosoom pärandub ainult isalt pojale, on see mutatsioonide poolt mõjutatavam kui autosomaalsed kromosoomid või X-kromosoom (Karmin et al. 2015). Kui vaadelda inimpopulatsiooni mudelina ühest mehest ja ühest naisest, arvestades sugude suhteks 1:1, siis selline paar kannab igast autosoomist nelja koopiat ja kolme X-kromosoomi, aga ainult ühte Y-kromosoomi. Populatsioonis tervikuna on seega Y-kromosoomi efektiivne populatsioonisuurus ( $N_e$ ) üks neljandik ükskõik millisest autosoomist, üks kolmandik X-kromosoomist ja sarnane haploidsele mitokondriaalsele DNA-le (mtDNA) (Hammer 1995; Thomson et al. 2000). Seetõttu on Y-kromosoom geenitriivile tundlikum kui diploidsed kromosoomid ning geenitriiv omakorda suurendab mutatsioonide akumulereerumist (Jobling and Tyler-Smith 2003). Tänapäeva inimese MSY on oluliselt degenerereerunud võrreldes X-kromosoomiga nii eukromaatilises piirkonnas (vastavalt 23Mb ja 150Mb) kui ka valke kodeerivate geenide seas (vastavalt 78 ja 800 geeni) (Skaletsky et al. 2003; Ross et al. 2005). Isaliinis akumulereerub keskmiselt  $0,74 \times 10^{-9}$  mutatsiooni aluspaari (ap) kohta aastas (Karmin et al. 2015).



**Joonis 1.** Inimese Y-kromosoomi skemaatiline ehitus; a) Y-kromosoom ja selle meessoospetsiifiline region (MSY); b) MSY eukromaatiline region. Palindroomsed järjestused on märgitud tumesiniste kolmnurkadena. Valke kodeerivad geenid on tähistatud kromosoomi MSY eukromaatilise regiooni skeemi all vertikaalsete joontega. PAR1 tähistab pseudoautosomaalse regiooni pikka ala ja PAR2 lühikest. Kohandatud (Hughes and Rozen 2012).

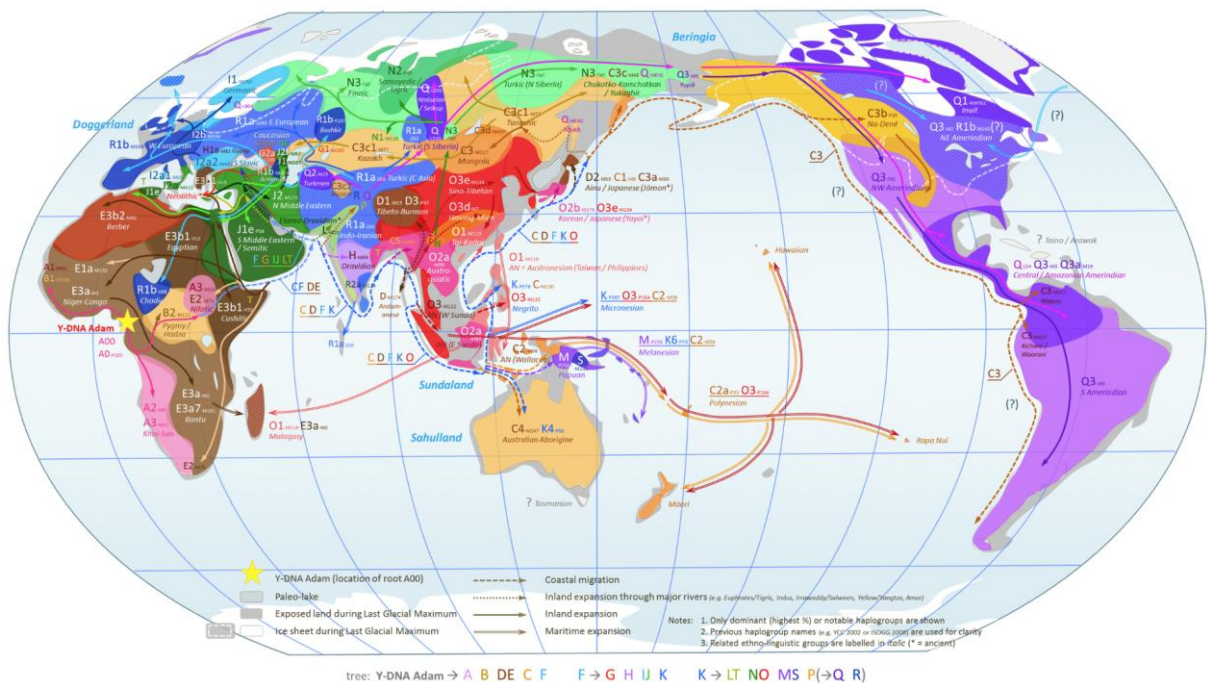
Y-kromosoomi puudumist (45, X karütüüp) tuntakse Turneri sündroomina, Y-kromosoomi geenide inaktivatsioon või kadumine põhjustab XY naise või hermafrodiidi fenotüübi kujunemist (Berta et al. 1990) ja viljatust (Sun et al. 1999; Jobling and Tyler-Smith 2003).

## 1.2. Y-kromosoom kui vahend populatsioonigeneetika uuringutes

Antropoloogias ja ajaloos, aga ka inimpopulatsioonide vaheliste geneetiliste seoste uurimisel ja soospetsiifiliste erinevuste väljaselgitamisel on Y-kromosoomi varieeruvust käsitlevatel uurimistöodel olnud suur roll (Jobling and Tyler-Smith 2003; Hughes and Rozen 2012). Paari viimase aastakümne jooksul on Y-kromosoom tänu oma unikaalsele pärandumismustrile ja suurele mutatsioonikiirusele olnud laialdaselt kasutuses informatiivse markerina inimese demograafilise ajaloo uurimisel (Rosser et al. 2000; Semino 2000; Battaglia et al. 2009; Fedorova et al. 2013; Underhill et al. 2015). Geneetilise materjali uurimisel populatsiooni demograafilise ajaloo seisukohalt tuleb lihtsustatud üldistuste tegemisel ettevaatlik olla, näiteks ei tohi geeniiline ajada segi populatsioonidega, ning tuleb meeles pidada, et Y-kromosoomi puhul on meil sisuliselt tegu ühe meessoospetsiifilise lookusega. Samas annab just Y-kromosoom (nagu ka mtDNA), mis võimaldab uurida geeniinide ajalisi-ruumilist jaotust, huvitava geneetilise andmestiku erinevate valdkondade teadlastele, sealhulgas paleontoloogidele, antropoloogidele, keeleteadlastele ja ajaloolastele (Jobling and Tyler-Smith 2003). Täna pakub võimalus sekveneerida peaaegu kogu Y-kromosoomi genoom uusi

läheneviise inimese evolutsiooni uuringuteks (Poznik et al. 2013; Batini et al. 2015; Hallast et al. 2015; Karmin et al. 2015; Ilumäe et al. 2016). Nii on suudetud väga täpselt rekonstrueerida haplogruppide fülogeneetilisi suhteid, nende võimalikku päritolu ning geograafilist jaotust ja migratsioonisuundi (Joonis 2).

Ligikaudu 70% tänapäeva ühiskondadest on patrilokaalsed (Murdock 1967; Burton et al. 1996). See tähendab, et kui eri regioonidest pärit mees ja naine abielluvad, siis pigem kolib uude kohta elama naine. Matrilokaalsuse puhul liiguvad kodukohast eemale mehed. Patrilokaalsus põhjustab Y-kromosoomi, matrilokaalsus aga emaliinis päranduvate mtDNA haplogruppide suurema lokaalse diferentseerituse (Oota et al. 2001). Patrilokaalsuse puhul peaks mtDNA seega vastupidiselt Y-kromosoomile näitama kahanevat geograafilist klasterdumist. Sellist tendentsi ongi näidatud näiteks Euroopas (Seielstad et al. 1998) ja Kagu-Aasia saartel (Kayser et al. 2001).



**Joonis 2.** Põhiliste Y-kromosoomi haplogruppide globaalne jaotus maailma rahvaste seas ja võimalikud migratsiooniteed. Kollase tähekesega on tähistatud Y-kromosomaalne Aadam – kõikide tänaste Y haplogruppide eellasvariant. Kaardil on erinevate värvidega tähistatud erinevate haplogruppide levilad ja nende nomenklatuuris kasutatavad tähised (allikas: [https://en.wikipedia.org/wiki/Human\\_Y-chromosome\\_DNA\\_haplogroup#/media/File:World\\_Map\\_of\\_Y-DNA\\_Haplogroups.png](https://en.wikipedia.org/wiki/Human_Y-chromosome_DNA_haplogroup#/media/File:World_Map_of_Y-DNA_Haplogroups.png)).

Lisaks kaasaegsete inimeste DNA uuringutele keskendutakse üha enam ka nn. „vana“ DNA uurimisele. Arheoloogilisest materjalist eraldatud proovide DNA-järjestuste analüüsimine võimaldab tänapäevaste ja iidsete populatsioonide mitmekesisust kõrvutada ja nendevahelisi seoseid uurida, lisaks saame aimu, millisel kiirusel mutatsioonid kuhjuvad (Kivisild 2017). Näiteks Y-kromosoomi hg G, mis oli varajastel farmeritel Anatoolias ja Iraanis levinuim haplogrupp (Kivisild 2017), on tänapäeval haruldane ning säilinud eelkõige Sardiinial ja Korsikal (Rootsi et al. 2012; Francalacci et al. 2013). Samas, hg J, mida varasemalt seostati põllunduse varajase levikuga Euroopasse (Rosser et al. 2000; Semino 2000), pole leitud siiani Euroopa neoliitikumi kontekstist, vaid hoopis küttidelt-korilastelt erinevatest piirkondadest Kaukaasiast kuni Karjalani (Kivisild 2017). Lisaks Y-kromosoomist saadavale infole võrreldakse nii ka mtDNA ja autosoomide varieeruvust (Rasmussen et al. 2010; Brotherton et al. 2013; Allentoft et al. 2015), et saada inimese geneetilisest mitmekesisusest eri ajaperioodidel võimalikult täpne ülevaade.

### **1.3. Y-kromosoomi põhilised markerite tüübid**

Y-kromosoomi uuritakse DNA järjestuse variatsioonide alusel. Põhilisteks neist on:

- 1) Ühenukleotiidsed asendused ehk SNP-d (*single nucleotide polymorphisms*)
- 2) Mikrosatelliidid ehk lühikesed kordusjärjestused ehk STR-id (*short tandem repeats*)

#### **1.3.1. Inimese Y-kromosoomi SNP-d**

SNP-de puhul eeldatakse, et enamus neist on kaasaegse inimese evolutsioonis tekkinud ühekordselt, fülogeneetiliste markeritena kasutatakse peamiselt just selliseid polümorfisme. Andmebaasi Ensembl (version nr. 88, avaldatud märtsis 2017) andmetel on inimese Y-kromosoomis (referentsjärjestus GRCh38.p10) kindlaks tehtud 226 603 lühikest järjestusvarianti, kuhu lisaks SNP-dele on juurde arvestatud ka insertioonid/deletsioonid ehk indelid (<http://www.ensembl.org>). Keskmise mutatsioonikiirus SNP-idel on erinevate uuringute põhjal erinev, kuid „vana“ DNA andmestiku põhjal on see väiksem,  $0,71 \times 10^{-9}$  (Trombetta et al. 2015) kuni  $0,82 \times 10^{-9}$  ap põlvkonna kohta (Poznik et al. 2013). Kaasaegsetest proovidest eraldatud DNA järjestuste põhjal on mutatsioonikiirus  $0,78 \times 10^{-9}$  (Balanovsky 2017) kuni  $1 \times 10^{-9}$  (Xue et al. 2009), kuid keskmiselt  $0,89 \times 10^{-9}$  (Helgason et al. 2015) ap põlvkonna kohta.

Viimaste aastate jooksul kasutusele võetud järgmise põlvkonna sekveneerimisega (NGS - *next generation sequencing*) on suudetud sekveneerida suur osa inimese kogu Y-kromosoomi MSY-st (Wei et al. 2012; Francalacci et al. 2013; Mendez et al. 2013; Poznik et al. 2013; Lippold et al. 2014; Scozzari et al. 2014; Yan et al. 2014; Hallast et al. 2015). NGS võimaldab tuvastada SNP-e nõ. erapooletult, ilma eelneva infota järjestuste varieeruvuse kohta, erinevalt genotüpiseerimisest. Sekveneerimisinfo alusel saab teha fülogeneesipuid, kus haru pikkused on võrdelised ajaga ja mis võimaldavad vaadeldava mitmekesisuse alusel määrata isaliini fülogeneetiliste lahknevuste täpse ajalise järjestuse ja kõikide järjestusvariantide viimase ühise eellase (MRCA - *most recent common ancestor*) (Hallast et al. 2015; Karmin et al. 2015).

### **1.3.2. Inimese Y-kromosoomi STR-id**

SNP-de varieeruvusinfo võimaldab defineerida Y-kromosoomi haplogrupid jaotust erinevates populatsioonides ja dateerida fülogeneesipuu sõlmi. Samas kasutatakse hiljuti lahknud alamklaadide sisese varieeruvuse uurimiseks teist tüüpi markereid, Y-kromosoomi-spetsiifilisi mikrosatelliite ehk lühikesi kordusjärjestusi (*short tandem repeats* – STR-id) (Hallast et al. 2015). STR-ide mutatsioonikiirus on oluliselt suurem võrreldes haplogruppe defineerivate SNP-dega. (Ballantyne et al. 2010) leidsid, et STR-ide mutatsioonikiirus jääb vahemikku  $3,8 \times 10^{-4}$  kuni  $7,2 \times 10^{-2}$  asendust lookuse kohta generatsioonis. Kõrgema mutatsioonikiiruse tõttu kasutatakse STR-ide variatsioone Y-kromosomaalsete haplogruppide sisemise struktuuri uurimiseks. Eelkõige on STR-id kasutusel SNP-de või DNA järjestuse põhjal koostatud fülogeneesipuude hiliste harude varieeruvuse hindamisel (Novelletto 2007; Balaresque et al. 2010). Samuti on STR-id väga olulised kriminalistikas indiviidide eristamiseks ja isaduse tuvastamiseks, kus kasutatakse võimalikult varieeruvaid markereid (Ballantyne et al. 2010; Purps et al. 2014). Kuna Y-STR-ide mutatsioonikiiruste hindamisel on veel väga suuri ebakõlasid, ei sobi nad fülogeneesipuul harude lahknemiste dateerimiseks ning sel puhul kasutatakse palju usaldusväärsemaid SNP-sid (Hallast et al. 2015; Karmin et al. 2015; Balanovsky 2017).

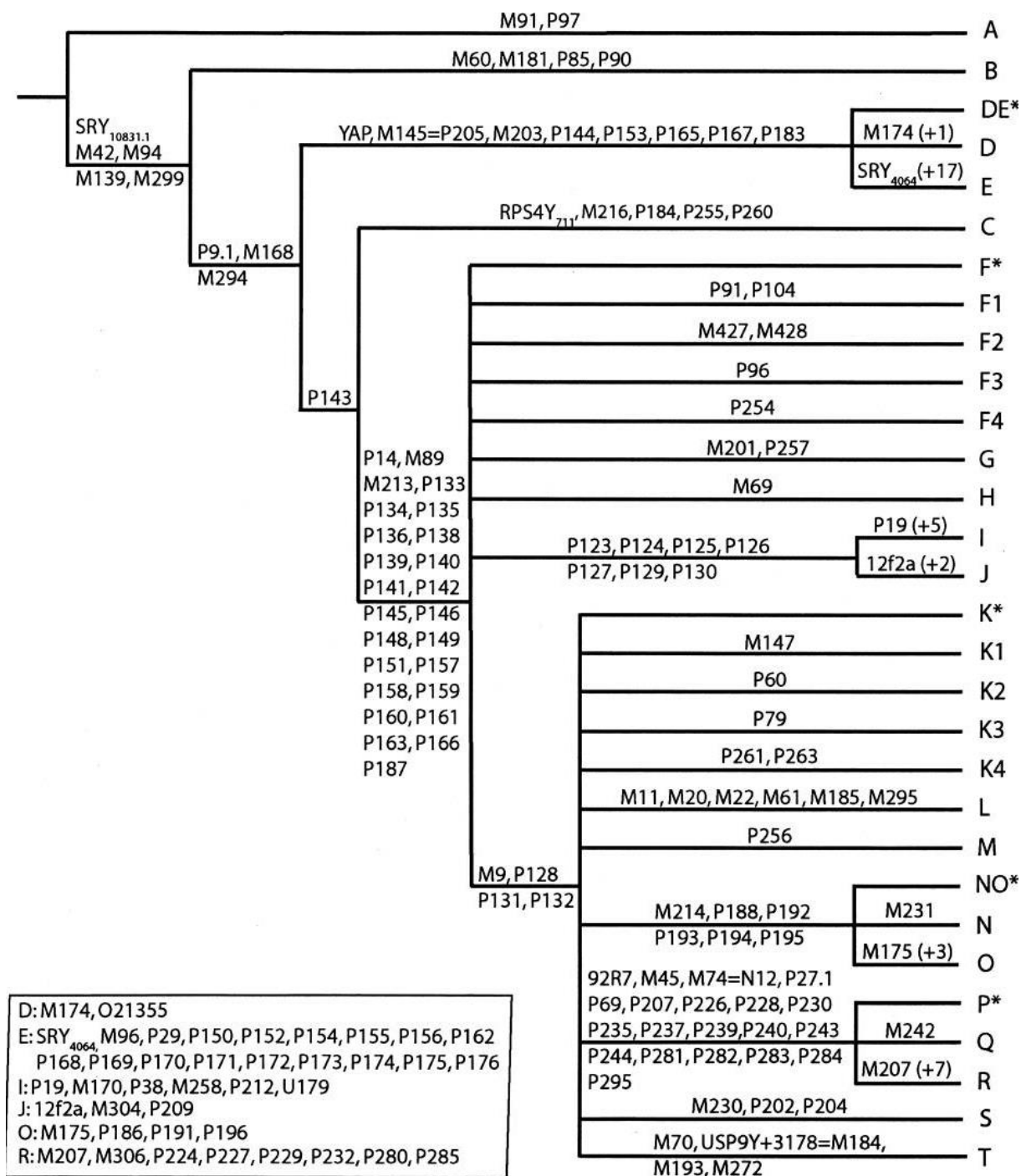
### **1.4. Inimese Y-kromosoomi fülogeneesipuu ja põhilised Euroopas levinud isaliinid**

Y-kromosoomi (ja mtDNA) populatsioonigeneetilistes uuringutes kasutatakse Y-kromosoomide kirjeldamisel ühtset nomenklatuuri, mida on pidevalt täiendatud uute

markerite ja harudega (YCC 2002; Karafet et al. 2008; van Oven et al. 2014) ning kus fülogeneetiliselt lähedased Y-kromosoomi tüübid jagatakse haplogruppidesse. Haplogrupiks nimetatakse monofüleetilist ehk ühisest eellasest pärinevat Y-kromosoomide gruppi. Ühte haplogruppi kuuluvaid Y-kromosoomide seovad ühised unikaalsed mutatsioonid, tavaliselt on nendeks SNP-d või indelid, mis on evolutsioonis eeldatavasti tekkinud ainult ühel korral ja kandunud edasi kõikidele seda järjestusvarianti kandvate indiviidide meessoost järglastele. SNP-ide järjestusvariantide alusel on koostatud globaalne Y-kromosoomi fülogeneesipuu (Joonis 3). Samas võivad ühte ja samasse haplogruppi kuuluvate Y-kromosoomide STR-id, mis määravad haplotüübi, olla varieeruvad. Haplotüüpide mitmekesisuse ja jaotuse alusel võrgustikus on võimalik teha järeldusi haplogruppide demograafilise ajaloo kohta, näiteks tähelaadne struktuur viitab hiljutisele tugevale ekspansioonile, pikad harud võrgustikus aga suuremale vanusele (Underhill et al. 2015; Ilumäe et al. 2016).

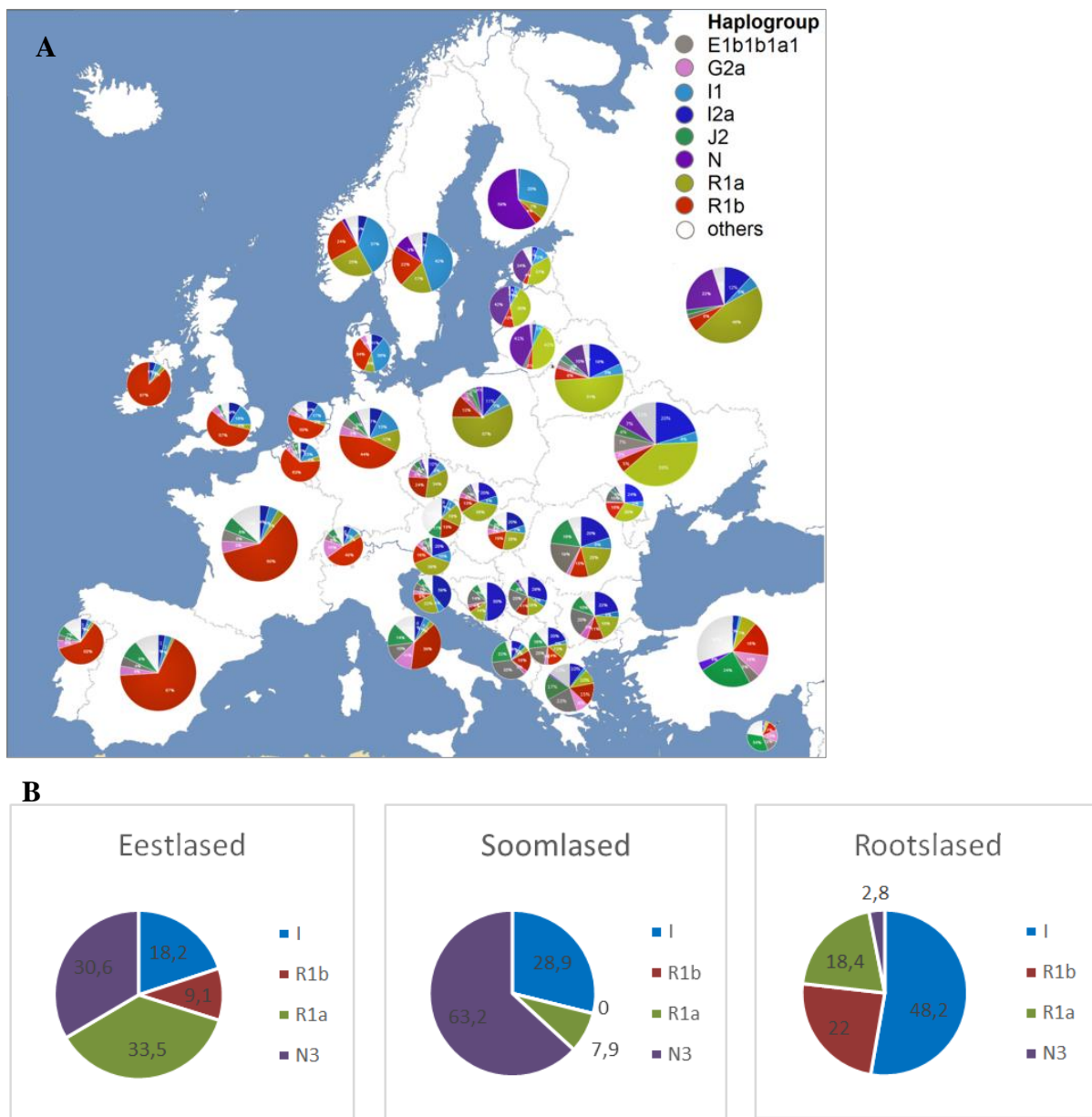
Kõikide tänaste inimese MSY-ite MRCA, vahetevahel ka Y-kromosomaalseks Aadamaks nimetatud variant (vt. Joonis 2), arvatakse olevat elanud umbes 254 000 aastat tagasi (Karmin et al. 2015). Kuigi seda meest nimetatakse Y-kromosomaalseks Aadamaks, elas temaga samal ajal veel palju teisi mehi, kes panustasid MSY-välisele geneetilisele materjalile tänapäeva inimpopulatsiooni. Seda saab vaadelda geneetilise informatsiooni isalt tütrele pärandamisega, kus Y-kromosoom läheb küll kaotsi, kuid tütar pärib pool autosomaalset genoomi ja X-kromosoomi. Y-kromosomaalset Aadamat vaadeldakse kunagi elanud mehena, kellega on MSY järjestusvariantide kaudu ühendatud kõik täna elavad mehed katkematu isa-poja liinina (Hughes and Rozen 2012).

Y-kromosomaalsetest liinidest on Euroopas täna üldiselt kõige sagedasem haplogrupp R1b (43,1%), mis on levinud eelkõige Lääne- ja Loode-Euroopas. Sellele järgnevad I1 (13,8%), I2 (9,0%), R1a (7,5%) ja J2 (7,5%) (Batini et al. 2015). Samas, enamuse Y-kromosoomi haplogruppide ja alamhaplogruppide levik on geograafiliselt struktureeritud – näiteks esineb siinses töös käsitletud, kuid mujal Euroopas haruldast hg N kõrge sagedusega eelkõige soomlastel (Joonis 4), samas kui hg-d E, G ja J on valdavalt Vahemere ümbruses (Joonis 4A). Hg N sagedus on lisaks soomlastele ka eestlastel (Joonis 4B) ja teistel Läänemere idakalda rahvastel oluliselt suurem kui näiteks Lääne-Euroopas (Joonis 4A). Eestlaste (ja soomlaste) seas on enim levinud Y-kromosomaalsed haplogrupid N, R1a ja I1 (Joonis 4B), eestlastele sarnane Y-kromosoomi haplogruppide jaotus esineb näiteks ka lätlastel, leedukatel ja põhjavenelastel. Seevastu rootslastel esineb hg N3 väga vähe, samas hg R1b ja I1a on neil oluliselt rohkem esindatud kui eestlastel ja soomlastel (Joonis 4B).



**Joonis 3.** Y-kromosoomi fülogeneesipuu lihtsustatud skeem (Karafet et al. 2008) alusel. Haplogruppe defineerivate markerite nimetused on välja toodud puu harudel ja joonise legendis. Haplogruppide nimetused on esitatud harude tipus. Siinsel skeemil ei ole harude pikkused proportsionaalsed mutatsioonide arvuga.

Üldiselt on Y-kromosoomi haplogrupid kõige mitmekesisemad Türgi ja Kreeka ning veelgi lõunapoolsemate Lähis-Ida populatsioonide seas, kõige madalam diversiteet on saamidel ja Orkney saarte elanikel Suurbritannia põhjarannikul. Mitmekesisus väheneb nii lõuna-põhja kui ka ida-lääne suunas (Joonis 4A), millest võib järeldada, et põhilised rahvasteränded Euroopasse said alguse Lähis-Idast (Batini et al. 2015).



**Joonis 4.** Euroopa põhiliste isaliinide kaart; A. Y-kromosoomi haplogruppide levik Euroopas. Sektordiagrammide värvid tähistavad joonise legendis näidatud haplogruppe; (allikas: [https://commons.wikimedia.org/wiki/File:Percentage\\_of\\_major\\_Y-DNA\\_haplogroups\\_in\\_Europe.png](https://commons.wikimedia.org/wiki/File:Percentage_of_major_Y-DNA_haplogroups_in_Europe.png)). B. Isaliinide haplogrupilise jaotuse võrdlus eestlaste (N=209), soomlaste (N=38) ja rootslaste (N=141) seas. Kohandatud Tambets et al. (2004) andmete põhjal.

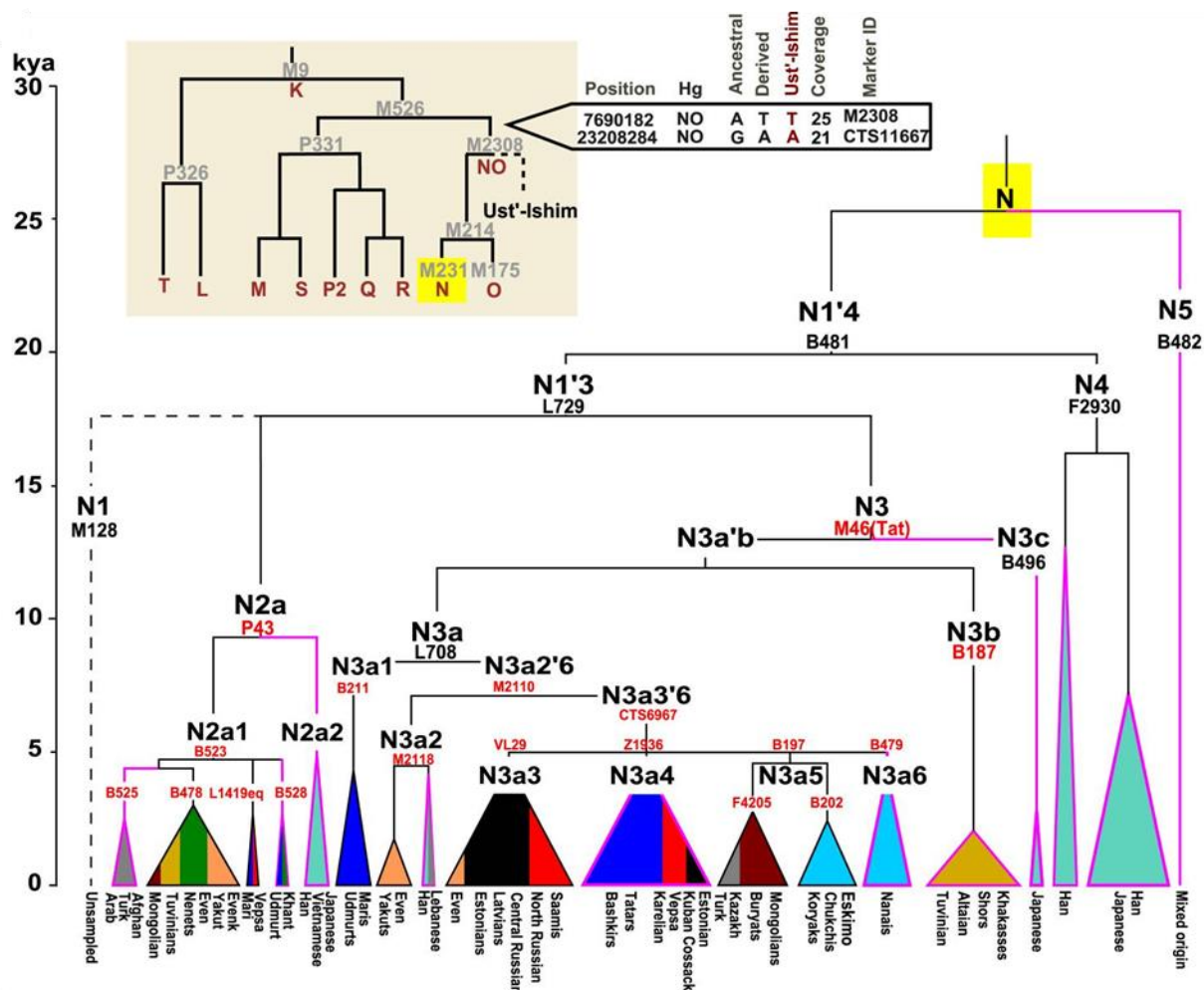
## 1.5. Haplogrupi N üldine iseloomustus ja fülogeograafiline levik

Hg N kuulub suuremasse haplogruppi NO, millel on kaks põhilist alamharu: tütarklaadid N-M231 ja O-M175 [Joonis 3; (Rootsi et al. 2007)]. Haplogruppide N ja O ruumiline sagedusjaotus erineb teineteisest suuresti, kuid nende mõlema tekkekohaks on tõenäoliselt Kagu-Aasia (Kayser et al. 2003; Hammer et al. 2006; Rootsi et al. 2007). Nende lahknemine toimus umbes 35 000 aastat tagasi (Rootsi et al. 2007). Hg O esineb täna kõrge sagedusega Ida- ja Lõuna-Aasia, aga ka suurel osal Okeania meestest, kuid ei ole levinud Euroopas. Hg N seevastu hõlmab suurt osa Ida-Euroopa, Aasia ning Kaug-Ida isaliinidest, sealjuures on selle esinemissagedus suurem kõrgematel laiuskraadidel (Zerjal et al. 1997; Karafet et al. 2002; Tambets et al. 2004). Hg N ulatuslikum ekspansioon algas umbes 5000-7000 aastat tagasi toimunud migratsioonilainega tõenäoliselt algusega Lõuna-Siberist nii ida kui lääne suunas (Ilumäe et al. 2016). Arheoloogiliste leidude põhjal on näidatud, et hg N oli valdav isaliin Põhja-Hiinas neoliitikumis 6500 aastat tagasi (Cui et al. 2013), selle sagedus seal vähenes ajapikku pronksiajal umbes 3000 aastat tagasi (Zhao et al. 2014; Gao et al. 2015). Hg N vanimad leiud Euroopas on pärit rauaaegsest Ungarist (Gamba et al. 2014), kus see hg on tänapäeval peaaegu olematu (Ilumäe et al. 2016).

Hg N põhilised alamharud hg N3 ja N2a lahkesid umbes 18 000 aastat tagasi (Joonis 5) ja nende ekspansioon algas arvatavasti vastavalt 13 000 ja 9000 aastat tagasi (Ilumäe et al. 2016).

Hg N2 on jaotunud ebaregulaarselt põhiliselt Siberi populatsioonides, kuid seda esineb ka Läänemere-äärsetel karjalastel ja vepslastel, viimastel on levikusagedus küllalt suur - 17,9%. Kõrgeimad N2 sagedused on leitud Kirde-Siberi populatsioonidel: 92% nganassaanidel, 72% eenetsitel ja 74% tundraneenetsitel (Karafet et al. 2002). Euroopas on N2 sagedus kõrgeim Volga-Uurali populatsioonides (20%) (Tambets et al. 2004). Kõige läänepoolsem N2 levikuala on Soomes, kus seda hg-d esineb juba vaid 0,4% ulatuses (Lappalainen et al. 2006).

Hilisemates uurimustöodes on defineeritud ka hg N4, mis on madala sagedusega jaotunud põhiliselt Ida-Aasia lõunapoolses mandriosas, ulatudes Kambodžast ja Vietnamist kuni Hiinani (Cui et al. 2013).



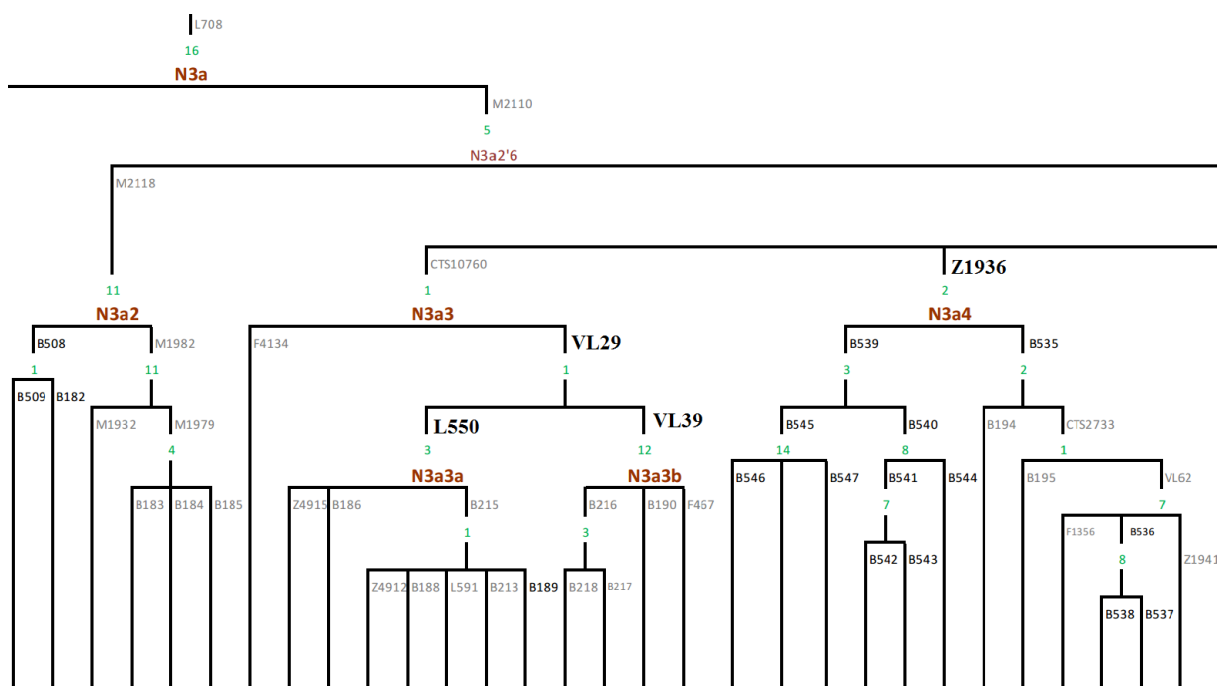
**Joonis 5.** Hg N fülogeneesipuu ja alamklaadide geograafilise jaotuse skeem (Ilumäe et al. 2016) järgi. Ülemisel skeemil on näidatud hg N teke ja positsioon suuremal fülogeneesipuul, „vana“ DNA Y-kromosoomi variant Ust' Ishim'i mehelt Lääne-Siberist ja hg NO mutatsioonid. Alumisel skeemil on välja toodud hg N alamklaadide geograafilise klasterdumise ja ajatelg (\*1000 aastat). Värvidega on tähistatud erinevad geograafilised piirkonnad. Käesolevas töös olulised alamklaadid on N3a3: must – Ida-Euroopa (eestlased, lätlased ja venelased Venemaa keskosast); punane – Põhja-Euroopa (saamid ja põhjavenelased); beež – Siber (evengid); ja N3a4: must – Ida-Euroopa (eestlased ja kubani-vene kasakad); punane – Põhja-Euroopa (vepslased, karjalased); sinine – Volga-Uurali piirkond (tatarlased ja baškiirid).

Vaatamata sellele, et hg N enim levinud alamklaadid N3 ja N2 esinevad praegu paljudes Siberi populatsioonides, sealhulgas tšuktšidel ja jupikidel Tšukotka poolsaarel Beringis, puuduvad need haplogrupid põlisameeriklaste hulgas (Bergen et al. 1999; Karafet et al. 1999; Lell et al. 2002; Bortolini et al. 2003; Seielstad et al. 2003). Selle alusel võib oletada, et hg N kromosoomid puudusid paleoliitilistel siberlastel Ameerika mandrite koloniseerimise ajal

umbes 12-17 000 aastat tagasi (Zegura et al. 2003). Kuid on ka võimalik, et hg N liinid surid asutajaefekti tõttu koloniseerimise käigus välja (Rootsi et al. 2007).

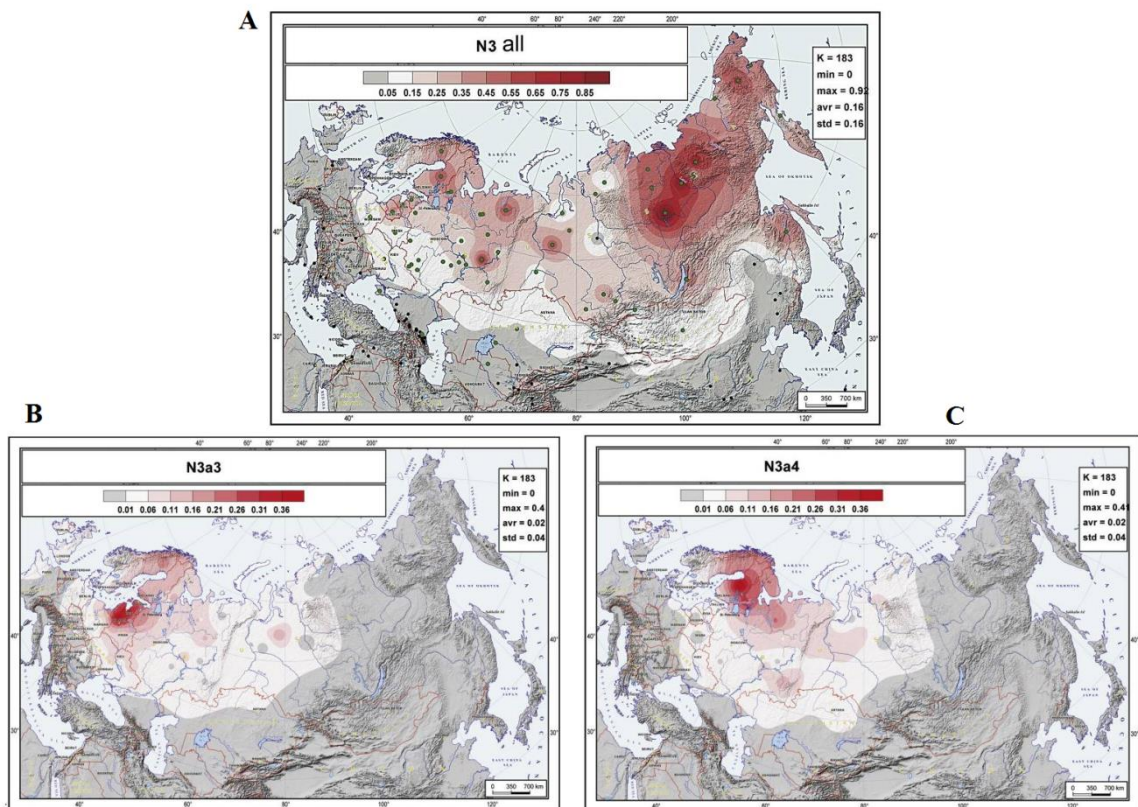
### 1.5.1. Haplogrupi N alamhaplogrupi N3 iseloomustus

Esimene suuremahuline hg N fülogeograafilist levikut käsitlev töö (Rootsi et al. 2007) näitas selle haplogrupi katkematut tsirkumarktilist levikut. Tänapäevaks on hg N suudetud struktureerida juba selge geograafilise jaotusega alamkladideks, mille lahkne misajad on teada (Joonis 5; Ilumäe et al. 2016). Hg N3 osaline sisestruktuur on esitatud joonisel 6.



**Joonis 6.** Y-kromosoomi haplogrupi N3 osaline ja lihtsustatud fülogeograafiline skeem (Ilumäe et al. 2016) alusel. Esile on tõstetud töös genotüpiseeritud N3a3 (VL29) ja selle alamkladide N3a3a-L550, N3a3b-VL39 ja N3a4-Z1936 markerid. Harudele on märgitud mutatsioonide arv ja markerite nimetused.

Hg N3 (Joonis 6) on hg N kõige sagedasem alamklad, mis on ühtlasi ka üks sagedasemaid Y-kromosoomi tüüpe Põhja-Euraasia rahvastel (Zerjal et al. 1997; Rosser et al. 2000; Semino 2000; Zerjal et al. 2001; Karafet et al. 2002; Laitinen et al. 2002; Lell et al. 2002; Villems et al. 2002; Tambets et al. 2004).



**Joonis 7.** Haplogrupi N3 (A) ning selle alamklaadide N3a3 (B) ja N3a4 (C) üldine geograafiline jaotus Euraasias. Levikusagedus on näidatud värviskaalaga (mida tumedam punane, seda suurem sagedus). Roheliste sõõridega on tähistatud (Ilumäe et al. 2016) artiklis uuritud hg N3 populatsioonid ning mustad punktid tähistavad varasemas kirjanduses käsitletud populatsioone ning on kõikidel kaartidel samad (Ilumäe et al. 2016).

Hg N3 esineb täna Hiinas, Koreas, Kalimantanil ja Jaapanis vaid vähesel määral (Kayser et al. 2003; Hammer et al. 2006; Rootsi et al. 2007), kuid just Kagu-Aasias arvatakse olevat tema tekkekoht ja algkodu. Sellele viitab hg N3 sealne suurem mitmekesisus võrreldes Lõuna-Siberiga. Siberi kaudu Ida-Euroopa suunas liikudes tegid N3 kandjad tõenäoliselt läbi tugeva „asutaja-“, või „pudelikaelaefekti“, sest vaid mõned N3 harud on jõudnud Euroopasse. Kirde-Euroopat võib vaadelda N3 teistkordse levikuallikana, kuna seda leidub kõrge sagedusega Volga-Uurali ugri rahvastel ning soomlastel, saamidel ja eestlastel (Rootsi et al. 2007). Olles valdav kogu Põhja-Aasias, on selle hg jaotus Euroopas piiratud põhja- ja idapoolsete populatsioonidega (Joonis 7). Hg N3 sagedus langeb järsult ida-lääne suunas üle Skandinaavia ning Leedu ja Poola vahel (Tambets et al. 2004).

Hg N3 alamklaadid N3c ja N3a'b lahkesid umbes 13 000 aastat tagasi. Järgmine lahkumine N3a'b sees toimus tõenäoliselt umbes 12 000 aastat tagasi (Joonis 5). Umbes 5000 aastat

tagasi toimus N3a3`6 erinevate harude kiire ekspansioon, mille tulemusena on need praeguseks jõudnud Euraasia mandri eri servadeni (Ilumäe et al. 2016). Alamklaad N3a3`6 on levinud väga laial geograafilisel alal ning esineb peaaegu pooltel uuritud proovidel nii leedukate, tšuktšide kui ka burjaadide seas. Käesolevas töös käsitletakse süvendatult hg N3 alamklaade N3a3 ja N3a4. Need alamklaadid on Põhja- Euroopas väga sagedased isaliinid ja ühtlasi kõige läänepoolsema levikuga hg N3 levialas. N3a3 alamklaadi kuulub üle kolmandiku tänapäeva meessoost eestlastest, lätlastest ja leedulastest, aga seda esineb ka saamidel, karjalastel ja soomlastel. Slaavi keeli kõnelevatel valgevenelastel, ukrainlastel ja venelastel kuuluvad peaaegu kõik hg N3 Y-kromosoomid hg-i N3a3. Ainus silmapaistev erand sellest mustrist on venelased Põhja-Euroopa poolsetest Venemaa piirkondadest, kellel on kõrgem sagedus hg N3a4, mis on teine levinum Lääne-Euraasia klaad. Hg N3a4 sagedus suureneb Loode-Euroopa suunas, kuid on ka dominantne Volga tatarlaste ja Lõuna-Uurali baškiiride hulgas (Ilumäe et al. 2016).

## 2. Töö eesmärgid

Käesoleva bakalaureusetöö eesmärgiks oli Y-kromosoomi ühe põhilise Põhja-Euraasias levinud isaliinivariandi - haplogrupi N3 - Põhja Euroopas sagedaste alamharude N3a3 ja N3a4 geograafilise leviku uurimine ja haplogrupisisese varieeruvuse täpsustamine. Selleks keskenduti nende haplogruppide sisemise struktuuri analüüsile eestlaste ja nende naaberpopulatsioonide seas. Töö on osaks suuremast hg N fülogeograafilist levikut käsitlevast projektist.

Täpsemalt uuriti:

- 1) N3a3 alamharu sisemist struktuuri uute täiendavate markerite (L550 ja VL39) abil;
- 2) N3a4 defineeriva markeri Z1936 sagedust eestlaste valimis;
- 3) N3a3 alamharudesisest varieeruvust ja N3a4 varieeruvust STR haplotüüpide alusel;
- 4) Uuritud alamharude seost uuritud populatsioonide: a) geograafilise ja; b) keelelise taustaga

### 3. Materjal ja metoodika

#### 3.1. Töös kasutatud DNA proovid

Käesolevas töös kasutatud DNA proovid on pärit Tartu Ülikooli evolutsioonilise bioloogia õppetooli ja Eesti Biokeskuse (EBK) kogust, on saadud vabatahtlikelt doonoritelt ning on kogutud kooskõlas Tartu Ülikooli eetikakomitee koodeksiga (luba nr. 252/M-17).

Töös keskenduti hg N3 Põhja-Euroopas levinud haplogruppide N3a3 ja N3a4 uurimisele osana suuremast hg N fülogeograafilist levikut käsitlevast uurimusest. Alamhaplogruppide analüüsiks oli oluline esmalt suurendada hg N kuuluvate eestlaste valimit (valim 1). Selleks genotüpiseerisin hg N markeri TAT (T→C) (Zerjal et al. 1997) osas 44 eestlast (valim 2), kelle haplogrupiline kuuluvus oli eelnevalt teadmata. Lisaks määrasin ka alamhaplogruppe defineerivad markerid antud valimis. Töö teiseks osaks oli N3a3 sisemise struktuuri täpsustamine proovidel, mille eelnev genotüpiseerimisanalüüs oli põhiliselt teostatud EBK vanemteaduri Siiri Rootsi poolt ja mis viitas nende kuuluvusele Y-kromosoomi hg N3 (Ilumäe et al., 2016). Kokku oli minu poolt sekveneeritud DNA proove 86 indiviidilt 14 erinevast populatsioonist, mille liikmed kuulusid uurali (soome-ugri) ja indo-euroopa (balti ja slaavi) keeleperekonda (Tabel 1).

**Tabel 1.** Töös kasutatud DNA-proovid ja uuritud indiviidide (N=2135) kultuuriline taust. Sulgudes on märgitud selles töös genotüpiseeritud/sekveneeritud proovide arv.

Keeleperekond	Populatsioon	Keelerühm	Uuritud proovide arv (N)	Hg N proovide arv (n)	Allikas
Soome-Ugri	Eesti (valim 1)	läänemeresoome	235	72	(Ilumäe et al., 2016)
	Eesti (valim 2)	läänemeresoome	44	(25)	Autor
	Karjala	läänemeresoome	139	52(12)	(Ilumäe et al., 2016)
	Vepsa	läänemeresoome	40	22(1)	(Ilumäe et al., 2016)
	Permikomi	permi	60	18(1)	(Ilumäe et al., 2016)
Balti	Läti	balti	86	37(1)	(Ilumäe et al., 2016)
Slaavi	Ukraina	ukraina	566	51(2)	(Ilumäe et al., 2016)
	Slovakkia	slovakkia	73	2 (2)	EBK avaldamata
	Pihkva	vene	132	30(11)	(Ilumäe et al., 2016)
	Pinega	vene	103	42(5)	(Ilumäe et al., 2016)
	Voronež	vene	96	6(1)	(Ilumäe et al., 2016)
	Kostroma	vene	51	6(1)	(Ilumäe et al., 2016)
	Belgorod	vene	143	17(1)	(Ilumäe et al., 2016)
	Smolensk	vene	100	7(2)	(Ilumäe et al., 2016)
	Valgevene	valgevene	267	26(2)	(Ilumäe et al., 2016)

Populatsioonidest olid valimis esindatud eestlased, lätlased, karjalased, vepslased, ukrainlased, slovakiid, venelased (Pinega, Voroneži, Kostroma, Belgorodi, Smolenski), valgevenelased ja permikomid.

## **3.2. DNA amplifitseerimine**

### **3.2.1. Polümeraasi ahelreaktsioon ehk PCR**

Polümeraasi ahelreaktsiooni (*polymerase chain reaction* - PCR) reaktsioonisegu tegin ühe proovi kohta 15 µl. Mõlema ahela sünteesiks vajalikke praimereid (10 pmol/µl DNA Technology A/S) võeti 0,3 µl, reaktsioonipuhvrit 3 µl (5x FIREpol® Master Mix, mis sisaldab ka 12,5 mM MgCl<sub>2</sub>, Solis BioDyne) ja milli-Q vett (Millipore) 13 µl iga DNA proovi kohta. Segu segasin kokku 1,5 ml katsetuubis ja kandsin 13 µl kaupa 200 µl-se mahuga PCR-i tuubiribadesse. Igasse tuubi lisasin 1-2 µl DNA-d, vastavalt DNA algkontsentratsioonile - slovakkide, valgevenelaste ning Voroneži, Kostroma, Belgorodi ja Smolenski venelaste proovide puhul 1 µl (5 ng/µl), ülejäänute puhul 2 µl (10 ng/µl).

PCR-i läbiviimiseks kasutati termotsüklerit UNOII (Biometra). Programm sisaldas:

- 94°C juures 3 min eelkuumutust;
- 94°C 15 sekundit kaheaahelalise DNA denaturatsiooniks;
- 15-25 sekundit 60°C juures praimerite seondumiseks DNA-le;
- 15-25 sekundit 72°C juures sünteesitava DNA ahela pikendamiseks

DNA denaturatsiooni, praimerite seondumist ja DNA ahela sünteesi korrati kogu protsessi käigus vastavalt 30-40 korda, tsükliite arv sõltus proovide DNA kontsentratsioonist, väiksema kontsentratsiooniga proovide puhul kasutati rohkem tsükliite. Tsükliite pikkus sõltus amplifitseeritud DNA lõigu pikkusest. Seejärel jäeti proovid veel 2-3 minutiks 72°C juurde, et kõik tekkinud DNA ahelad saaksid lõpuni sünteesitud, ning programmi lõpus langetati temperatuur 4°C või 10°C-ni, vältimaks DNA edasist lagunemist.

Kasutatud praimerite järjestuste tabel on Lisas 1.

### **3.2.2. Geelelektroforees**

DNA amplifitseerimise õnnestumist sekveneeritavate proovide hulgas kontrollisin horisontaalsel geelelektroforeesil 2% agarosgeelil 0,5-kordses TBE puhvril (45mM Trisboraat, 0,05 M EDTA, pH=8,3), millele lisasin 100 µl etiidiumbromiidi 100 ml geeli

kohta. Forees toimus 170 V juures elektroodide vahega 32 cm 25-30 minuti vältel. Seejärel pildistasin geeli UV-valguses (UVIPro), et hinnata saadud DNA-produkti kvaliteeti ja kogust.

### **3.2.3. DNA produkti „puhastamine“**

PCR-i segu jääkide eemaldamiseks lisisin madala kontsentratsiooniga produktidele (foreesipildil ähmane joon) kahe ensüümi segu (1 µl proovi kohta): 0,9 µl eksonukleas I (*ExoI*, Thermo Scientific, 20 U/µl), mis lagundab liigse praimerijäägid ja 0,1 µl SAP-i (*shrimp alkaline phosphatase* ehk krevetivähi aluseline fosfataas, Thermo Scientific, 1U/µl), mis lagundab üleliigsed nukleotiidid. Tuubid panin termotsüklerisse, kus neid inkubeeriti 15 minutit 37°C (ensüümide töörežiim) ja seejärel 15 minutit 85°C juures (ensüümide inaktiveerimine).

### **3.2.4. Sekvenerimine**

DNA sekvenerimine toimus Sangeri meetodil kapillaarelektroforeesil. Vajaliku PCR-i ja sadestamise teostasid mina, edasine töö tehti EBK sekvenerimise tuumiklabori sekvenaatoril (3730xl DNA Analyzer, Applied Biosystems). Sekvenerimise eesmärgiks oli tuvastada, kas uuritava proovi hg N siinses töös käsitletud alamharusid määravas positsioonis (VL29, Z1936, L550 ja VL39) on referentsjärjestusega võrreldes toimunud asendusreaktsioon või mitte.

#### **3.2.4.1. Sekvenerimisreaktsioon**

DNA sekvenerimiseks TAT (C) eestlastele ja teiste populatsioonide esindajatele vajalike markerite sekvenerimisreaktsiooni läbiviimiseks segasin 1 proovi kohta 1,5 ml katsetuubis kokku:

- 0,75 µl segu BigDye Terminator v3.1 reagent premix (BigDye Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems),
- 2 µl BigDye 5x sekvenerimispuhvrit (BigDye Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems)
- 6,1 µl milli-Q H<sub>2</sub>O
- 0,16 µl sekvenerimispraimerit (sama, mis PCR-il; 1,6 pmol)

Segu kandsin 200 µl PCR-i tuubidesse. Lisisin igasse tuubi 1 µl PCR-i produkti. Reaktsioonisegu kogumaht proovi kohta oli 10 µl. Reaktsiooni viisin läbi termotsükleris.

PCR-i programm koosnes kolmest etapist: denaturatsioon 15 sekundit 95°C juures, praimerite seondumine 10 sekundit 50°C juures ja DNA ahela süntees 1 minut 60°C juures. Kõiki etappe korrati 30 korda.

#### **3.2.4.2. DNA sadestamine**

Pärast sekveneerimisreaktsiooni lisan DNA sadestamiseks igasse tuubi:

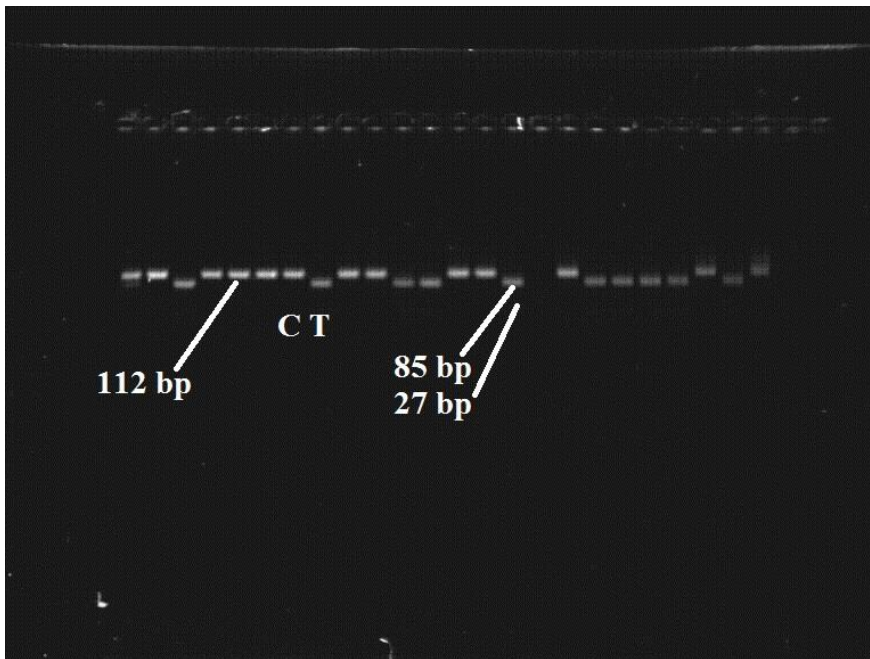
- 30 µl 96% piiritust (temperatuuril -20°C);
- 2 µl ammooniumatsetaadi ja dekstraani segu [3,5M CH<sub>3</sub>COONH<sub>4</sub> [pH>8]/EDTA (250 mM) ja 20 mg/ml punane dekstraan suhtes 1:1]

Ammooniumatsetaat on vajalik DNA sadestamiseks ja dekstraan selle seondamiseks tahkele kandjale, et teostada DNA puhastamiseks vajalikud etapid. Tuube hoidsin 10 minutit sügavkülmas -20°C juures (DNA sadeneb lahusest välja), seejärel tsentrifugisin neid 10 minuti jooksul (13 000 pöoret minutis) ning eemaldas supernatandi veejoapumbaga. DNA sade ja dekstraan jäid tuubi põhja. Seejärel lisan igasse tuubi 200 µl 70% piiritust (temperatuuril -20°C) pesemiseks. Järgnevalt tsentrifugisin veel 5 minutit (13000 pöoret minutis), eemaldas supernatandi, kandsin igasse tuubi teistkordseks pesemiseks veel 200 µl 70% piiritust (temperatuuril -20°C) ning eemaldas supernatandi ilma eelneva tsentrifugimiseta. Seejärel lasin proovidel 15 minutit 37°C termostaadis kuivada. Kõige lõpus lisan igasse tuubi 10 µl 70% formamiidi. Töö lõpus kandsin kõik proovid sekveneerimisplaadile, mille viisin edasiseks sekveneerimiseks EBK tuumiklaborisse.

#### **3.2.4.3. Restriktsioonanalüüs**

Restriktsiooni viivad läbi restriktsioonilised endonukleasid ehk restriктаasid, mis lõikavad DNA molekuli spetsiifilise nukleotiidse järjestuse alusel fragmentideks, lõhkudes ära nukleotiididevahelised fosfodietersidemed. DNA järjestuses tekkinud mutatsioonid võivad restriktsioonisaite juurde tekitada või ära kaotada. Restriktsioonanalüüsi tegin selleks, et näha, kas restriктаasid lõikavad amplifitseeritud DNA proovi või mitte.

Uurimiseks valisin 44 eestlaste proovi, kelle isaliini haplogrupiline kuuluvus ei olnud teada ja amplifitseerisin neil hg N3 defineeriva markeri TAT (T/C) järjestust sisaldava DNA lõigu (112 ap). Restriктаasina kasutasin *HinIII* (*NlaIII*) ensüümi (5 U/µl). Need proovid, mida restriктаas mutatsiooni kohalt lõikas, on lühemad ja liiguvad foreesil kiiremini, jagunedes 85 ap ja 27 ap pikkusteks fragmentideks (Joonis 8). Töö jaoks olulised hg N kuuluvad proovid olid derivaatalleeli (C) kandjad, mida restriктаas ei lõiganud.



**Joonis 8.** Markeri TAT (T/C) allelide tuvastamine *HinIII* restriksioonanalüüsil. Markeri algne (ingl. k. *ancestral allele*) alleelivariant on T, mille restriктаas lõikab 85 ja 27 ap pikkusteks fragmentideks. Derivaatalleeli (ingl. k. *derived allele*) variant on C, mis on 112 ap pikkune fragment ning mida restriктаas ei lõika.

Restriksiooni läbiviimiseks lisasin eelnevalt amplifitseeritud PCR-i produktidele 0,5 ühikut ehk 0,1 µl *HinIII* ensüümi ja 5 ühikut ehk 1 µl puhvrit ühe proovi kohta. Seejärel hoidsin proove 2 tundi 37°C termostaadis. Reaktsiooni tulemust kontrollisin geelelektrofooresil (vt. 3.2.2).

#### 3.2.4.4. STR analüüs

Hg N3 kuuluvatest populatsioonidest analüüsisin 23 erineva STR-markeri (DYS576, DYS389I/II, DYS448, DYS19, DYS391, DYS481, DYS549, DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS393, DYS458, DYS385a/b, DYS456 ja Y-GATA-H4; Promega) suhtes järgnevaid proove: 22 eestlast, 6 Pihkva venelast ja 1 lätlane (vt. Lisa 2). Need proovid kuulusid alamhaplogruppi N3a4 ja N3a4. Genotüpiseerimiseks kasutati Applied Biosystems PowerPlex® Y23 System kit-i.

Reaktsioonisegu komponendid 5 µl-s:

- 1 µl PowerPlex® Y23 5X Master Mix
- 0,5 µl PowerPlex® Y23 10X Primer Pair Mix
- 3 µl H<sub>2</sub>O
- 0,5 µl matriits-DNA-d

Reaktsiooni toimumise tingimused (PCR-i 4 esimest etappi 30 tsükli):

- 96°C 2 minutit
- 94°C 10 sekundit
- 61°C 1 minut
- 72°C 30 sekundit
- 60°C 20 minutit

Pärast PCR-i lisasin igale proovile 9,5 µl formamiidi ja 0,35 µl 5-Dye PowerPlex® STR Systems (CC5 Internal Lane Standard 500, sisemine pikkusstandard) pikkusmarkerit. Seejärel kandsin proovid sekveneerimisplaadile, lisades veel 1 µl kõigi uuritud markerite kõikide alleelide segu (ingl. k. *allelic ladder*).

STR-ide alleelide pikkused määrasin GeneMapper® Software 4.0 (Applied Biosystems) programmiga. Võrdluspopulatsioonidena kasutasin ka varasemalt Applied Biosystems AmpFISTR® Yfiler™ PCR Amplification Kit-iga genotüpiseeritud proove. Osa kasutatud proove võtsin avaldatud andmestikust: saamid (Hallast et al. 2015), eestlased (Karmin et al. 2015).

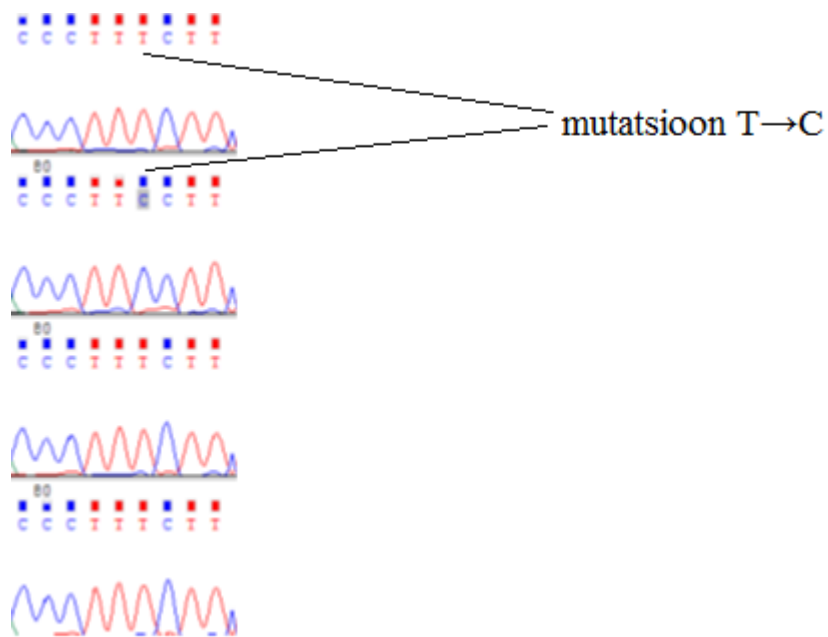
### **3.2.5. Andmeanalüüs**

#### **3.2.5.1. Fülogeneetilise võrgustiku koostamine**

Fülogeneetilise võrgustiku loomiseks kasutasin tarkvara Network 5.0.0.1. (2017) ja Network Publisheri (Fluxus Technology). Selleks koostas Microsoft Exceli tabeli põhjal sisendfaili laiendiga .ych ja kasutasin seda fülogeneetilise võrgustiku konstrueerimiseks ühendatud mediaanide (ingl. k. *median joining*) meetodil. Väljundfaili salvestasin .sto failina.

#### **3.2.5.2. Sekveneeritud järjestuste analüüs**

Sekveneeritud DNA järjestusi analüüsisin programmi Chromas Pro 1.5 (Technelysium) abil. Sellega sai tuvastada, millise alleeliga on vastava markeriga uuritavas proovis mutatsiooni määravas positsioonis tegu (Joonis 9).



**Joonis 9.** Sekveneeritud DNA lõigu järjestus markeri L550 (T/C) näitel, kus on näha mõlemad alleelivariandid; algne alleel T ja derivaatalleel C. PCR-il kasutati markerit L550. (Chromas Pro 1.5 Technelysium, ekraanitõmmis)

## 4. Tulemused ja arutelu

### 4.1. N3a3 uute markerite L550 ja VL39 ning N3a4 esinemissagedus uuritud valimis

Töö aluseks on võetud Ilumäe et al. (2016) poolt uuritud individid, kelle hg N3a3 sisestruktuuri täiendasin uute markerite L550 ja VL39 tüpiseerimise abil ning määrasin nende sageduse erinevatel populatsioonidel (Tabel 2). Samuti uurisin hg N3a4 sagedust markeri Z1936 abil. Eestlaste valimit suurendasin varasemalt genotüpiseerimata proovidega (valim 2). Minu poolt sekveneeritud hg N3 DNA proovide arv oli kokku 53 (Tabel 2).

44-st eestlasest, keda genotüpiseerisin, kuulus hg-i N3 [marker TAT (T→C)] 19 indiviidi. TAT (C) proovidest kuulus alamhg-sse N3a3 (marker VL29) 13 eestlast. Alamhg-sse N3a3a kuulus 10 indiviidi, alamhg-sse N3a3b kolm ja alamhg-sse N3a4 viis indiviidi. Üks proov, mis kuulus alamhg-i N3a3, ei olnud defineeritav kummagi uuritud markeri põhjal ja jäi N3a3\* parafüleetilisse gruppi.

**Tabel 2.** Töös kasutatud proovide arv ning haplogruppisisene jaotuvus. Sulgudes on antud töö autori poolt sekveneeritud proovide arv. Alamhaplogruppide seas on antud levikusagedused (%) iga populatsiooni kohta.

Populatsioon	N	hg N3	N3a3*	N3a3a	N3a3b	N3a4
			N3 alamharude sagedus (%)			
		TAT (T→C)	VL29	L550 (C→T)	VL39 (G→T)	Z1936 (C→T)
Eesti (valim 1)	235	72	1,2	15,8	10,2	3,4
Eesti (valim 2)	44	19 (19)	2,3	22,7	6,8	11,4
Karjala	139	51 (12)	8,6	3,6	2,2	20,1
Vepsa	40	15 (1)	2,5	0	0	32,5
Permikomi	60	14 (1)	1,7	0	0	5
Ukraina	566	48 (2)	0,2	6,7	0	1,1
Slovakkia	73	2 (2)	0	2,7	0	0
Vene (Pinega)	103	25 (5)	3,9	0	0	19,4
Vene (Pihkva)	132	29 (3)	1,5	15,9	0,8	3,8
Vene (Voronež)	96	6 (1)	0	2,1	1,0	2,1
Vene (Smolensk)	100	7 (2)	1	9	1	1
Vene (Kostroma)	51	6 (1)	0	2,0	0	9,8
Vene (Belgorod)	143	16 (1)	0,7	9,1	0	0
Valgevene	267	24 (2)	0,4	7,1	0,4	0
Läti	86	37 (1)	0	38,4	0	1,6

Hg N3 kuuluvate teiste uuritud populatsioonide esindajatele tegin PCR-i markeritega L550 ja VL39. Neist kuulusid alamhg-sse N3a3a kaks slovakk, üks karjalane ja üks ukrainlane. Alamhg-sse N3a3b kuulusid üks Pihkva, üks Voroneži ja üks Smolenski venelane ning üks valgevenelane.

Ilumäe et al. (2016) andmetel kuulus 235-st indiviidist 72 (valim 1; 30,6%) hg N3. 44-st genotüpiseeritud eestlasest (valim 2) kuulus hg N3 19 indiviidi (43,2%). Mõlema töö andmete summeerimisel saadi hg N3 koondsageduseks 32,6%. Alamhg N3a3a oli uuritud hg N3 kuuluvate populatsioonide seas enim levinud eestlaste, lätlaste, ukrainlaste, valgevenelaste ning Pihkva, Smolenski ja Belgorodi venelaste seas. Alamhg N3a3b esines siinses valimis peamiselt eestlastel (7-10%), teistel vaadeldud populatsioonidel oli seda märkimisväärselt vähem (Tabel 2), kuid kirjanduse andmetel on see alamhg 20% esindatud ka saamidel (Hallast et al. 2015). N3a4 on valdav karjalastel, vepslastel, permikomidel, Pinega ja Kostroma venelastel ning osal eestlastest. Parafüleetilisse klaadi N3a3\* kuulub osa karjalastest (8,6%) ning üksikud teiste populatsioonide esindajad. Selline parafüleetilise klaadi jaotus viitab sellele, et nende populatsioonide puhul on tõenäoliselt tegu senituvastamata hg N3 alamharuga, mis vajab edasist uurimist.

#### **4.2. N3a3 ja N3a4 varieeruvus STR-haplotüüpide alusel**

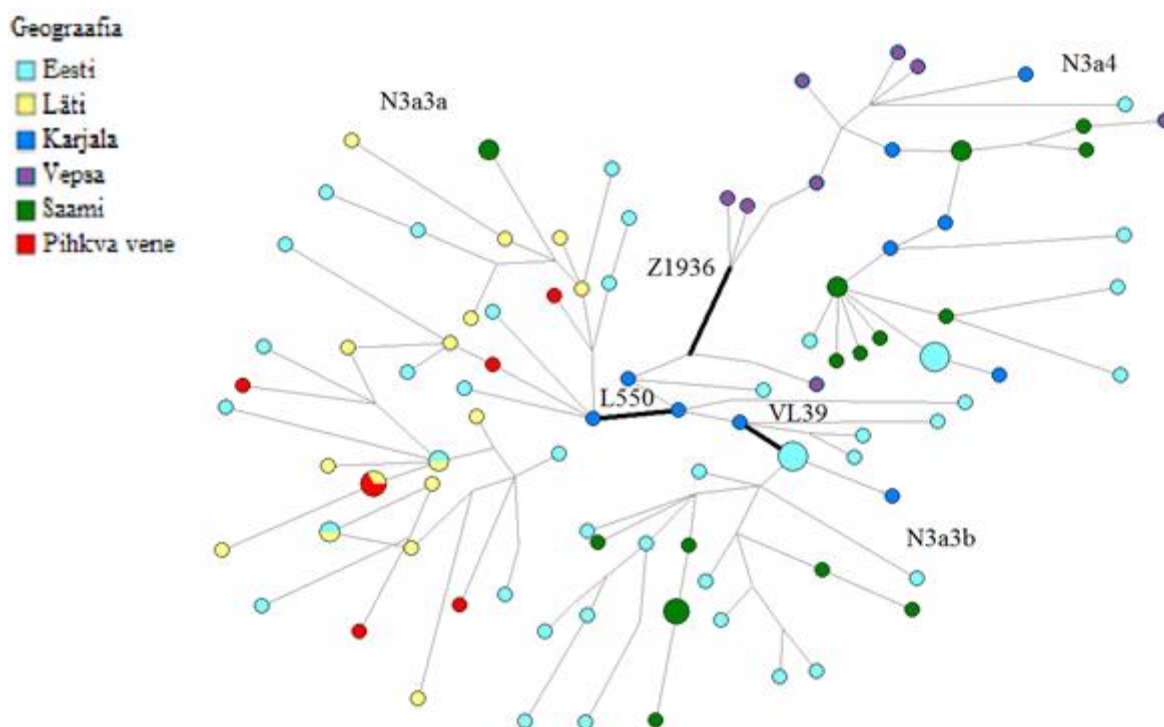
STR-analüüsi teostasid hg-i N3 kuuluvatele eestlastele. Kuna genotüpiseerimisel kasutatud proovide hulk oli väike (44 indiviidi), kaasasin analüüsi juba olemasoleva suurema eestlaste valimi ja võrdluseks ka teisi naaberpopulatsioone. Selleks lisasin üldvalimisse veel lätlaste, karjalaste, vepslaste, saamide ning Pihkva venelaste proove (Hallast et al. 2015; Ilumäe et al. 2016). Kokku kasutasin 107 proovi, millest minu poolt tüpiseeritud oli 31.

Uuritud DNA proovide seast kuulusid hg N3a3 alamharudesse eestlased, lätlased, Pihkva venelased, mõned saamid ja kaks karjalast. Hg N3a4 esines vaid saamid, vepslastel, karjalastel ja eestlastel.

STR-analüüsi andmete põhjal koostatud fülogeneesivõrgustikult (Joonis 10) on näha haplogrupisisene varieeruvus uuritud populatsioonide hulgas. SNP-de alusel jaotub võrgustik kolmeks suuremaks rühmaks. Marker L550, mis defineerib hg-i N3a3a, on eelkõige levinud lätlastel ja Pihkva venelastel, veidi vähem eestlastel. Genotüpiseeritud karjalastest ja saamidest esines seda üksikutel, vepslastel aga mitte ühelgi. Marker VL39, mis defineerib hg-

i N3a3b, esines eestlastel, saamidel ja ühel karjalasel. Hg N3a4 defineeriv marker Z1936 on levinud eestlastel, karjalastel, vepslastel ja saamidel.

Joonise 9 keskel olevad üksikud defineerimata harud tähistavad neid haplotüüpe (9 tk.), mis ei jaotunud ühtegi siin määratud alamhaplogruppi. Nad kuuluvad küll hg N3a3, kuid ei kuulu uuritud markeritega (L550, VL39 ja Z1936) defineeritud harudesse ja jäävad siinses töös parafüleetilisse gruppi N3a3\*.



**Joonis 10.** STR andmete põhjal konstrueeritud fülogeneetiline võrgustik. Võrgustik on koostatud mediaanvektorite ühendamise meetodil (vt. täpsemalt 3.2.5.1.). Sõlmed vastavad haplotüüpidele, sõlme suurus haplotüübi esinemissagedusele ning värv legendis näidatud populatsioonile. Sõlmedevaheliste lülde pikkus vastab mutatsioonide arvule kahe haplotüübi vahel. Jämedad jooned tähistavad SNP markereid Z1936, L550 ja VL39.

#### 4.3. N3a3 ja N3a4 alamharude seos uuritud populatsioonide geograafilise ja keelelise taustaga

Töö tulemustest selgub, et uuritud individid jagunevad alamhaplogruppidesse põhiliselt vastavalt nende geograafilisele paiknemisele. Kõrvuti asetsevad populatsioonid on üldiselt

sarnase haplotüübilise kuuluvusega. Näiteks on eestlaste Y-DNA uuritud markerite osas sageli sarnane lätlaste ja Pihkva venelaste omale. Omavahel on sarnased ka saamid, karjalased, vepslased ning osa eestlasi. Kuna geograafiliselt on tegu naaberrahvastega, vastab tulemus ootustele. Samas võiks keelelise kuuluvuse põhjal arvata, et soome-ugri rahvad moodustavad lätlastest ja venelastest eraldiseisva rühma. Balti ja slaavi keeled kuuluvad indoeuroopa keelerühma, soome-ugri keeled aga uurali keelkonda. Need kaks rühma on lingvistiliselt väga erinevad. Seega näiteks saamid ja eestlased on omavahel suuremas keelesuguluses, kuid Y-kromosomaalne sarnasus tundub olevat suurem näiteks eestlaste ja lätlaste vahel. Seetõttu võib arvata, et naaberpopulatsioonide vahel toimub geneetiline segunemine kiiremini kui keeleline. Teine võimalus, et on toimunud keelte segunemine, on väiksem, sest üksikud laensõnad ei muuda oluliselt keelestruktuuri, mille põhjal keelesugulust määratakse.

Uuritud valimis on aga näha ka selget keelelist jagunemist, mida „rikuvad“ üldiselt üksnes eestlased, kes kuuluvad koos balti ja slaavi keeli rääkivate populatsioonidega hg N alamhg-sse N3a3a ja koos teisi soome-ugri keeli rääkivate populatsioonidega alamhg-sse N3a3b ja N3a4 (Joonis 10). Samas, siinses töös on eestlaste heterogeensus küll teiste populatsioonide omast märksa suurem (vt. haplotüübilist jaotust joonisel 9), kuid on oluline silmas pidada, et uuritud üldvalim on ebaühtlane, siit on puudu soomlased ja uuritud eestlaste hulk teiste rahvaste esindajatega võrreldes palju suurem. Samas on valim alamhaplogruppide võrdlemiseks liiga väike, et saadud erinevustele statistikast tuge saada. Seega - siinses töös leitud huvitavaid detaile saab eelkõige kasutada edasises uurimistöös, et kultuurilisel erinevusel põhinevate eripärade paikapidavust suuremas valimis kontrollida.

## Kokkuvõte

Käesolevas uurimistöös käsitletakse inimese Y-kromosoomi hg N3 alamklaadide N3a3 ja N3a4 sagedust ja jaotust eestlaste ja naaberpopulatsioonide hulgas. Hg N3 muudab teiste Euroopas levinud isaliinide seas huvipakkuvaks asjaolu, et see on pärit Ida-Euraasiast, on levinud peaaegu üle kogu Põhja-Euraasia parasvöötme ja seob meid täna Siberis elavate rahvastega. Euroopas on hg N3 alamklaadid N3a3 ja N3a4 levinud eelkõige soome-ugri ja balti keeli rääkivate rahvaste seas.

Töö praktiline osa käsitleb hg N3 levikut mõnedes soomeugri, balti ja slaavi keeli rääkivates Euroopa populatsioonides, keda genotüpiseeriti kahe uue (L550 ja VL39) ja kolme teadaoleva (TAT; VL29 ja Z1936) hg N3 SNP-markeri osas ja 23 STR-markeri abil. Uuriti hg N3 alamklaadi N3a3 sisestruktuuri ning N3a3 ja N3a4 geograafilist jaotust ja sagedust, saadud tulemuste alusel võrreldi uuritud populatsioonide keelelist ja ruumilist seotust.

Leiti, et hg N3 kuulus kõigist genotüpiseeritud eestlastest umbes kolmandik. See on kooskõlas varem avaldatud andmetega.

Valdav osa eestlastest kuulus hg N3 alamklaadi N3a3, eelkõige selle alamharusse N3a3a. Alla poole eestlastest jaotus teiste hg N3a kladide vahel. Alamklaadil N3a3 oli eestlaste seas ligi neli korda suurem osakaal kui alamklaadil N3a4.

Teiste rahvaste esindajatest kuulusid uuritud slaavi (ukrainlased, slovakid, valgevenelased ja venelased) ja balti (lätlased) keelte rääkijad põhiliselt alamklaadi N3a3a, seevastu soome-ugri keelte rääkijad (karjalased, vepslased ja permikomid, aga ka saamid) peamiselt alamklaadidesse N3a3b ja N3a4. Seega tõstatasid uurimistöö tulemused huvitavaid küsimusi seoses populatsioonide keelelise taustaga, mida edaspidi suurema valimi abil testida.

## **The distribution and internal structure of human Y chromosomal haplogroup N3a sub-clades N3a3 and N3a4 among Estonians and their neighbouring populations**

Silja Aruhein

### Summary

This study revealed the distribution of human Y chromosomal haplogroups (hg) N3a3 and N3a4, the sub-clades of hg N3, among Estonians and their neighbouring populations. Hg N3 stands out among other European paternal lineages for it originates from eastern Europe, is distributed almost across the entire northern Eurasian temperate zone and connects us with contemporary Siberian people. In Europe the sub-clades of N3, N3a3 and N3a4, are distributed mostly between Finno-Ugric and Baltic speaking populations.

In the practical part of this study the distribution of hg N3 sub-lineages among some Finno-Ugric, Baltic and Slavic speaking European populations was studied by genotyping two new (L550 and VL39) and three previously known (TAT, VL29 and Z1936) hg N3 specific markers together with 23 STR-markers. The internal structure of the sub-lineage N3a3, the geographic distribution and frequency of the clades N3a3 and N3a4 were studied. The results were compared to the spatial-linguistic background of the populations in the study.

It was found that in accord with previous results, hg N3 comprised a third of all the genotyped Estonians. Majority of Estonians belonged to the hg N3 sub-clade N3a3, mostly to the sub-branch N3a3a. Less than a half of Estonians was distributed between other clades of N3a. The sub-clade N3a3 was almost four times more frequent among Estonians than N3a4.

The other populations studied here – the Slavic-speaking (Ukrainians, Slovaks, Belarusians and Russians) and Baltic-speaking (Latvians), belonged mostly to the sub-clade N3a3a, whereas Finno-Ugric speakers (Karels, Vepsas and Permian Komis, but also Saamis) belonged mostly to the sub-clades of N3a3b and N3a4. Thus, the results from this study have raised interesting questions about the linguistic background of the populations carrying these lineages to be tested further in the larger sample sets.

## **Tänuavaldus**

Soovin tänada oma juhendajaid Kristiina Tambetsit ja Siiri Rootsit ning EBK ja TÜ evolutsioonilise bioloogia õppetooli meeskonda igakülgse toetuse ja abivalmiduse eest.

## Kasutatud kirjandus

- Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, et al. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522:167–172.
- Balanovsky O. 2017. Toward a consensus on SNP and STR mutation rates on the human Y-chromosome. *Hum. Genet.*
- Balaresque P, Bowden GR, Adams SM, Leung HY, King TE, Rosser ZH, Goodwin J, Moisan JP, Richard C, Millward A, et al. 2010. A predominantly neolithic origin for European paternal lineages. *PLoS Biol.* 8.
- Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, Choi Y, Van Duijn K, Vermeulen M, Brauer S, et al. 2010. Mutability of Y-chromosomal microsatellites: Rates, characteristics, molecular bases, and forensic implications. *Am. J. Hum. Genet.* 87:341–353.
- Batini C, Hallast P, Zadik D, Delser PM, Benazzo A, Ghirotto S, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Dupuy BM, et al. 2015. Large-scale recent expansion of European patrilineages shown by population resequencing. *Nat. Commun.* 6:7152.
- Battaglia V, Fornarino S, Al-Zahery N, Olivieri A, Pala M, Myres NM, King RJ, Rootsi S, Marjanovic D, Primorac D, et al. 2009. Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe. *Eur. J. Hum. Genet.* 17:853–853.
- Bergen AW, Wang CY, Tsai J, Jefferson K, Dey C, Smith KD, Park SC, Tsai SJ, Goldman D. 1999. An Asian-Native American paternal lineage identified by RPS4Y resequencing and by microsatellite haplotyping. *Ann. Hum. Genet.* 63:63–80.
- Berta P, Hawkins JB, Sinclair AH, Taylor A, Griffiths BL, Goodfellow PN, Fellous M. 1990. Genetic evidence equating SRY and the testis-determining factor. *Nature* 348:448–450.
- Bortolini M-C, Salzano FM, Thomas MG, Stuart S, Nasanen SPK, Bau CHD, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, et al. 2003. Y-Chromosome Evidence for Differing Ancient Demographic Histories in the Americas. *Am. J. Hum. Genet.* 73:524–539.
- Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Jane Adler C, Richards SM, Sarkissian C Der, Ganslmeier R, Friederich S, et al. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. *Nat. Commun.* 4:1764.
- Burton ML, Moore CC, Whiting JWM, Romney AK, Aberle DF, Barcelo JA, Dow MM, Guyer JI, Kronenfeld DB, Levy JE, et al. 1996. Regions based on social structure. *Curr. Anthropol.* 37:87–123.
- Cui Y, Lindo J, Hughes CE, Johnson JW, Hernandez AG, Kemp BM, Ma J, Cunningham R, Petzelt B, Mitchell J, et al. 2013. Ancient DNA analysis of mid-holocene individuals

from the Northwest Coast of North America reveals different evolutionary paths for mitogenomes. *PLoS One* 8:e66948.

Fedorova SA, Reidla M, Metspalu E, Metspalu M, Rootsi S, Tambets K, Trofimova N, Zhadanov SI, Hooshiar Kashani B, Olivieri A, et al. 2013. Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol. Biol.* 13:127.

Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, Atzeni R, Pilu R, Busonero F, Maschio A, Zara I, et al. 2013. mtDNA coalescence times are not significantly different. This conclusion would hold whether or not an alternative approach would yield more definitive T. 341:565–569.

Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki L, Kővári I, Pap I, Anders A, et al. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* 5:5257.

Gao SZ, Zhang Y, Wei D, Li HJ, Zhao Y Bin, Cui YQ, Zhou H. 2015. Ancient DNA reveals a migration of the ancient Di-qiang populations into Xinjiang as early as the early Bronze Age. *Am. J. Phys. Anthropol.* 157:71–80.

Hallast P, Batini C, Zadik D, Maisano Delser P, Wetton JH, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Destro Bisol G, Dupuy BM, et al. 2015. The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Mol. Biol. Evol.* 32:661–673.

Hammer MF, Karafet TM, Park H, Omoto K, Harihara S, Stoneking M, Horai S. 2006. Dual origins of the Japanese: Common ground for hunter-gatherer and farmer Y chromosomes. *J. Hum. Genet.* 51:47–58.

Hammer MF. 1995. A recent common ancestry for human Y chromosomes. *Nature* 378:376–378.

Helgason A, Einarsson AW, Guðmundsdóttir VB, Sigurðsson Á, Gunnarsdóttir ED, Jagadeesan A, Ebenesersdóttir SS, Kong A, Stefánsson K. 2015. The Y-chromosome point mutation rate in humans. *Nat. Genet.* 47:453–457.

Hughes JF, Rozen S. 2012. Genomics and Genetics of Human and Primate Y Chromosomes. *Annu. Rev. Genomics Hum. Genet.* 13:83–108.

Illumäe A-M, Reidla M, Chukhryaeva M, Järve M, Post H, Karmin M, Saag L, Agdzhoyan A, Kushniarevich A, Litvinov S, et al. 2016. Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. *Am. J. Hum. Genet.* 99:163–173.

Jobling MA, Tyler-Smith C. 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–612.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830–838.

- Karafet TM, Osipova LP, Gubina MA, Posukh OL, Zegura SL, Hammer MF. 2002. High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Hum Biol* 74:761–789.
- Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, et al. 1999. Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64:817–831.
- Karmin M, Saag L, Vicente M, Sayres MAW, Järve M, Talas UG, Rootsi S, Ilumäe A-M, Mägi R, Mitt M, et al. 2015. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25:459–466.
- Kayser M, Brauer S, Weiss G, Schiefenhover W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M. 2003. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am J Hum Genet* 72:281–302.
- Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P a., Stoneking M. 2001. Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* 68:173–190.
- Kivisild T. 2017. The study of human Y chromosome variation through ancient DNA. *Hum. Genet.* 136:1–18.
- Laitinen V, Lahermo P, Sistonen P, Savontaus ML. 2002. Y-chromosomal diversity suggests that Baltic males share common Finno-Ugric-speaking forefathers. *Hum Hered* 53:68–78.
- Lappalainen T, Koivumäki S, Salmela E, Huoponen K, Sistonen P, Savontaus ML, Lahermo P. 2006. Regional differences among the finns: A Y-chromosomal perspective. *Gene* 376:207–215.
- Lell JT, Sukernik RI, Starikovskaya YB, Su B, Jin L, Schurr TG, Underhill PA, Wallace DC. 2002. The dual origin and Siberian affinities of Native American Y chromosomes. *Am J Hum Genet* 70:192–206.
- Lippold S, Xu H, Ko AA, Li M, Renaud G, Butthof A, Schröder R, Stoneking M, Jobling M, Heyer E, et al. 2014. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. *Investig. Genet.* 5:13.
- Mendez FL, Krahn T, Schrack B, Krahn AM, Veeramah KR, Woerner AE, Fomine FLM, Bradman N, Thomas MG, Karafet TM, et al. 2013. An African American paternal lineage adds an extremely ancient root to the human y chromosome phylogenetic tree. *Am. J. Hum. Genet.* 92:454–459.
- Murdock GP. 1967. *Ethnographic Atlas*. Pittsburgh: University of Pittsburgh Press
- Novelletto A. 2007. Y chromosome variation in Europe: continental and local processes in the formation of the extant gene pool. *Ann. Hum. Biol.* 34:139–172.

- Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat. Genet.* 29:20–21.
- Van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MHD. 2014. Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Hum. Mutat.* 35:187–191.
- Poznik GD, Henn BM, Yee M-C, Sliwerska E, Euskirchen GM, Lin AA, Snyder M, Quintana-Murci L, Kidd JM, Underhill PA, et al. 2013. Sequencing Y Chromosomes Resolves Discrepancy in Time to Common Ancestor of Males Versus Females. *Science* (80-. ). 341:562–565.
- Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, Angustia SMT, Santos LH, Anslinger K, Bayer B, et al. 2014. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci. Int. Genet.* 12:12–23.
- Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, et al. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757–762.
- Rootsi S, Myres NM, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Varendi K, Sahakyan H, et al. 2012. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. *Eur. J. Hum. Genet.* 20:1275–1282.
- Rootsi S, Zhivotovsky LA, Baldovic M, Kayser M, Kutuev IA, Khusainova R, Bermisheva MA, Gubina M, Fedorova SA, Ilumae AM, et al. 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur J Hum Genet* 15:204–211.
- Ross MT, Grafham D V, Coffey AJ, Scherer S, Mclay K, Muzny D, Platzer M, Howell GR, Burrows C, P C, et al. 2005. DNA sequence of the human X chromosome. *Genome* 48:325–337.
- Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, et al. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67:1526–1543.
- Scozzari R, Massaia A, Trombetta B, Bellusci G, Natalie M, Novelletto A, Cruciani F, Darwin BC, Università S, Biologia D, et al. 2014. chromosome and reveals a deep phylogenetic structure in Africa. :535–544.
- Seielstad M, Yuldasheva N, Singh N, Underhill P, Oefner P, Shen P, Wells RS. 2003. A novel Y-chromosome variant puts an upper limit on the timing of first entry into the Americas. *Am J Hum Genet* 73:700–705.
- Seielstad MT, Minch E, Cavalli-Sforza LL. 1998. Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278–280.

- Semino O. 2000. The Genetic Legacy of Paleolithic Homo sapiens sapiens in Extant Europeans: A Y Chromosome Perspective. *Science* (80-. ). 290:1155–1159.
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.
- Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, Page DC. 1999. An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. *Nat Genet* 23:429–432.
- Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF. 2003. High-Resolution SNPs and Microsatellite Haplotypes Point to a Single, Recent Entry of Native American Y Chromosomes into the Americas. *Mol. Biol. Evol.* 21:164–175.
- Zerjal T, Beckman L, Beckman G, Mikelsaar A V, Krumina A, Kucinskas V, Hurles ME, Tyler-Smith C. 2001. Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. *Mol Biol Evol* 18:1077–1087.
- Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, et al. 1997. Genetic relationships of Asians and Northern Europeans, revealed by Y- chromosomal DNA analysis. *Am J Hum Genet* 60:1174–1183.
- Zhao Y-B, Zhang Y, Li H-J, Cui Y-Q, Zhu H, Zhou H. 2014. Ancient DNA evidence reveals that the Y chromosome haplogroup Q1a1 admixed into the Han Chinese 3,000 years ago. *Am. J. Hum. Biol.* 26:813–821.
- Tambets K, Rootsi S, Kivisild T, Help H, Serk P, Loogvali EL, Tolk H V, Reidla M, Metspalu E, Pliss L, 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.
- Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. 2000. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc. Natl. Acad. Sci.* 97:7360–7365.
- Trombetta B, D’Atanasio E, Massaia A, Myres NM, Scozzari R, Cruciani F, Novelletto A. 2015. Regional differences in the accumulation of SNPs on the male-specific portion of the human y chromosome replicate autosomal patterns: Implications for genetic dating. *PLoS One* 10:1–18.
- Underhill PA, Kivisild T. 2007. Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations. *Annu Rev Genet* 41:539–564.
- Underhill PA, Poznik GD, Rootsi S, Järve M, Lin AA, Wang J, Passarelli B, Kanbar J, Myres NM, King RJ, et al. 2015. The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *Eur. J. Hum. Genet.* 23:124–131.

- Wei W, Ayub Q, Chen Y. 2012. A calibrated human Y-chromosomal phylogeny based on resequencing. *Accepted Email alerting service A calibrated human Y-chromosomal phylogeny based on resequencing.* :388–395.
- Villems R, Rootsi S, Tambets K, Adojaan M, Orekhov V, Khusnutdinova E, Yankovsky N. 2002. Archaeogenetics of Finno-Ugric speaking populations. In: Julku K, editor. *The Roots of Peoples and Languages of Northern Eurasia IV*. Oulu: Societas Historiae Fenno-Ugricae. p. 271–284.
- Xue Y, Wang Q, Long Q, Ng BL, Swerdlow H, Burton J, Skuce C, Taylor R, Abdellah Z, Zhao Y, et al. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.* 19:1453–1457.
- Yan S, Wang CC, Zheng HX, Wang W, Qin ZD, Wei LH, Wang Y, Pan XD, Fu WQ, He YG, et al. 2014. Y chromosomes of 40% Chinese descend from three neolithic super-grandfathers. *PLoS One* 9.
- YCC. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348.

## **Kasutatud veebiaadressid**

<http://www.ensembl.org>

[http://www.ensembl.org/Homo\\_sapiens/Location/Chromosome?r=Y%3A1-1000](http://www.ensembl.org/Homo_sapiens/Location/Chromosome?r=Y%3A1-1000)

<http://www.fluxus-engineering.com>

[https://commons.wikimedia.org/wiki/File:World\\_Map\\_of\\_Y-DNA\\_Haplogroups.png](https://commons.wikimedia.org/wiki/File:World_Map_of_Y-DNA_Haplogroups.png)

[https://en.wikipedia.org/wiki/Y-](https://en.wikipedia.org/wiki/Y-DNA_haplogroups_in_populations_of_Europe#/media/File:Percentage_of_major_Y-DNA_haplogroups_in_Europe.png)

[DNA\\_haplogroups\\_in\\_populations\\_of\\_Europe#/media/File:Percentage\\_of\\_major\\_Y-DNA\\_haplogroups\\_in\\_Europe.png](https://en.wikipedia.org/wiki/Y-DNA_haplogroups_in_populations_of_Europe#/media/File:Percentage_of_major_Y-DNA_haplogroups_in_Europe.png)

## Lisad

### Lisa 1. Töös kasutatud praimerid

DNA amplifitseerimiseks kasutatud praimerid, nende SNP-i positsioonid, defineeritavad klaadid, nukleotiidjärjestused, seondumistemperatuurid, saadud produkti pikkused, SNP-i positsioonid produktil ning algsed (*ancestral*) ja omandatud (*derived*) alleelivariandid.

SNP positsioon	Marker	Klaad/ haplogrupp		Forward (F) praimer 5' → 3' Reverse (R) praimer 5' → 3'	Produkti pikkus (bp)	SNP positsioon produktil	Algne	Omandatud
14570424	VL29(CTS2929)	N3a3	F	TGGACATATACACCCACACT	228	106	T	C
			R	GATTGGGGAAAAGTTGGTCA				
21285948	L550	N3a3-L550	F	CATCTCATAACCCCTTTGAGCTC	235	131	C	T
			R	TTAATGACACCAATACTTCAGTGC				
21333667	V39_eq	N3a4-V39	F	CCAATAGCTGTCTGGATATTACAAA	241	146	G	T
			R	TATAGGGTCTGGCAGTTGGTTAC				
21463326	Z1936	N3a4	F	CTAAACTCGTCCCTCAGTCA	232	150	C	T
			R	GCTTAACTCTGCCTGACTTC				

## Lisa 2. STR-markerid ja tandemkorduste arv uuritud indiviididel.

Fülogeneetilise võrgustiku (Joonis 10) koostamisel kasutati ainult STR-ide täisprofiiliga indiviide (Est – eestlased; Lat – lätlased; Prus – Pihkva venelased; nd – STR määramata).

Proovi ID	DYS19	DYS385a	DYS385b	DYS389 I	DYS389 II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS481	DYS533	DYS549	DYS570	DYS576	DYS635	DYS643	YGATAH4
Est308	14	11	13	nd	29	23	11	14	13	14	10	10	19	14	17	20	10	12	18	20	21	11	12
Est361	14	11	13	15	32	24	11	14	14	14	10	10	19	14	17	20	11	11	17	17	21	11	12
Est425	14	11	13	14	30	23	10	14	14	15	10	10	19	14	17	21	11	11	19	17	23	11	12
Est602	14	11	13	14	30	24	10	14	14	14	9	11	19	14	16	18	11	11	20	18	21	11	12
Est604	14	11	13	14	30	24	11	14	14	14	11	10	19	14	17	20	11	11	17	18	21	11	12
Est605	14	11	12	14	31	23	11	14	14	14	10	10	19	14	17	20	11	11	20	21	22	12	12
Est606	14	11	13	14	30	23	10	15	13	14	10	10	19	14	17	20	11	11	19	15	23	11	12
Est607	15	11	14	13	29	23	11	14	13	14	10	10	19	14	16	20	11	11	21	21	22	11	11
Est609	14	11	14	14	30	26	11	14	14	14	10	10	19	14	17	20	11	11	19	21	22	11	12
Est610	14	12	14	14	30	24	11	14	14	14	10	10	19	13	17	20	11	11	19	21	21	11	12
Est613	15	11	14	14	30	23	11	14	14	14	10	10	19	14	17	20	11	12	19	16	24	12	12
Est614	14	11	13	14	30	23	10	14	14	15	10	nd	19	14	18	20	11	11	19	20	22	11	12
Est619	14	11	13	13	29	nd	11	14	14	14	10	nd	19	14	17	20	10	nd	19	21	17	11	12
Est621	14	11	13	14	30	24	11	14	14	14	10	10	19	14	17	22	11	11	18	21	21	11	12
Est626	15	12	14	13	29	23	11	14	14	14	10	11	19	14	18	20	11	11	20	21	22	10	12
Est628	15	11	14	13	29	23	11	15	13	14	10	10	19	14	16	20	11	11	20	21	21	11	12
Est631	14	11	11	13	29	23	10	14	14	14	10	nd	19	14	17	20	11	13	20	20	21	11	12
Est635	15	11	13	13	29	23	10	14	13	14	10	10	19	15	18	20	11	12	19	15	22	12	12
Est638	16	11	13	14	30	23	11	14	14	14	10	10	19	14	18	20	11	11	16	15	22	11	11
Est640	14	11	13	13	28	23	11	11	14	14	10	10	19	14	17	20	12	11	17	21	22	11	12
Est74	14	11	14	14	30	22	11	11	13	14	10	9	19	14	17	20	11	11	19	16	23	11	11
Est90	14	11	13	14	30	23	11	14	13	14	10	10	19	14	18	21	11	11	21	16	21	11	12
Lat138	15	11	13	13	29	23	11	14	13	14	10	10	19	14	17	20	11	11	18	15	22	11	12
PRus129	15	11	14	13	29	23	11	14	14	14	10	10	19	14	17	20	11	11	20	21	21	11	12
PRus23	14	11	13	14	30	23	12	14	14	14	10	10	19	14	17	21	11	11	nd	22	21	11	12
PRus31	16	11	14	13	29	23	10	14	14	14	10	10	19	14	18	20	11	11	19	20	22	11	12
PRus32	15	11	14	13	29	23	11	14	14	14	10	10	19	14	17	20	12	11	20	21	21	12	12
PRus4	15	11	14	13	29	23	10	15	14	14	10	11	19	14	17	20	11	11	20	20	23	11	12
PRus9	15	11	14	13	30	24	11	14	13	14	10	10	19	12	16	20	11	11	17	21	22	11	12

## **Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks**

Mina, Silja Aruhein

(sünnikuupäev: 14.05.1993)

annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose

„Inimese Y-kromosoomi haplogrupi N3a alamharude N3a3 ja N3a3 leviku ja sisestruktuuri analüüs eestlaste ja nende naaberrahvaste seas“,

mille juhendajad on Kristiina Tambets ja Siiri Rootsi,

reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu alates **31.12.2020** kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

olen teadlik, et nimetatud õigused jäävad alles ka autorile.

kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 26.05.2017