

TARTU ÜLIKOOL
LOODUS- JA TÄPPISTEADUSTE VALDKOND
MOLEKULAAR- JA RAKUBIOLOOGIA INSTITUUT
GENOOMIKA INSTITUUT

Stella Eileen Päid

Eesti keskaegse arheoloogilise materjali geneetiline sugulusanalüüs Siksälä kalme näitel

Bakalaureusetöö

12 EAP

Juhendajad

PhD Kristiina Tambets

MSc Stefania Sasso

TARTU 2021

Eesti keskaegse arheoloogilise materjali geneetiline sugulusanalüüs Siksälä kalme näitel

Lühikokkuvõte: Inimeste minevikusündmuste uurimiseks kõrvutatakse erinevate valdkondade, nt ajaloo, arheoloogia ja geneetika andmeid. Käesolev töö on näide sellest, kuidas neid andmeid võrreldes saadakse terviklikum pilt meie esivanemate elust. Töös uuriti keskaegse Eesti elanikkonna geneetilist struktuuri ja sotsiaalseid sidemeid Siksälä kalme näitel. Selleks eraldati ja sekveneeriti DNA 20 indiviidi säilmetest. Suurimad sarnasused leiti Kagu-Eesti ja Läti keskaegsete indiviididega, tänapäevaste eestlastega ning ka balti ja slaavi rahvastega. Sugulusanalüüsi tulemuste abil koostati indiviididevaheliste suhete alusel üks suurem kolme põlvkonnaga sugupuu ja kolm väiksemat gruppi. Töö kõigi tulemuste põhjal võis järeldada, et Siksälä kalmesse maetud indiviidide ühiskonnakorraldust iseloomustas patrilokaalsus.

Märksõnad: vana DNA, populatsioonigeneetika, sugulusanalüüs, Eesti keskaja

CERCS: B220 Geneetika, tsütogeneetika, H220 Keskaja ajalugu, H340 Arheoloogia

Genetic kinship analysis of medieval Estonian archaeological material using the example of Siksälä burial site

Abstract: To study the human past, data from history, archaeology, and genetics among other disciplines, are compared to each other. This thesis is an example of how transdisciplinary approach helps to get a more comprehensive picture of the lives of our ancestors. The thesis studies the genetic variability and social connections of the medieval Estonian population using the example of Siksälä burial site. For that, DNA was extracted and sequenced from the remains of 20 individuals. The greatest similarities were found with other medieval individuals from South-East Estonia and Latvia as well as with modern Estonians and the Baltic and Slavic populations. Using the results of the kinship analysis the relationships between individuals of one larger three-generation family tree and of three smaller groups were reconstructed. Based on all of the results it could be concluded that the social organisation of the individuals buried at Siksälä burial site was characterised by patrilocality.

Keywords: ancient DNA, population genetics, kinship analysis, medieval Estonia

CERCS: B220 Genetics, cytogenetics, H220 Medieval history, H340 Archaeology

SISUKORD

KASUTATUD LÜHENDID	6
SISSEJUHATUS	7
1. KIRJANDUSE ÜLEVAADE.....	8
1.1 Vana DNA olemus	8
1.1.1 Säilivus	8
1.1.2 Keemilised modifikatsioonid.....	10
1.2 Uurimismeetodid	10
1.2.1 Mitokondriaalne ja Y-kromosomaalne DNA analüüs	11
1.2.2 Autosomaalne DNA analüüs	11
1.2.3 Sugulusanalüüs	12
1.3 Eesti demograafiline ajalugu Euroopa kontekstis	13
1.4 Eesti keskaeg	15
1.4.1 Siksälä kalme	15
2. EKSPERIMENTAALOSA	18
2.1 Töö eesmärgid	18
2.2 Materjal ja meetodika.....	18
2.2.1 Bioarheoloogiline materjal vana DNA analüüsiks	18
2.2.2 Materjali kataloogimine.....	18
2.2.3 DNA eraldamine.....	19
2.2.3.1 Saastuse vältimine	19
2.2.3.1 Proovi võtmine DNA analüüsiks.....	19
2.2.3.3 Pinnasaastuse eemaldamine ja DNA eemaldamine hambajuurtest	20
2.2.3.4 DNA ekstrakti puhastamine.....	21
2.2.4 DNA sekveneerimisraamatukogude koostamine.....	21
2.2.4.1 DNA fragmentide üheaahelaliste otste täitmine ja reaktsioonisegust puhastamine	22
2.2.4.2 Adapterite lisamine ja reaktsioonisegust puhastamine	22
2.2.4.3 Adapterite otste täitmine.....	23

2.2.4.4	Indeksite lisamine, DNA fragmentide amplifitseerimine ja reaktsioonisegust puhastamine	23
2.2.5	DNA sekveneerimine	23
2.2.6	Sekveneeritud järjestuse analüüs	24
2.2.6.1	Järjestuste ettevalmistamine populatsioonigeneetiliseks analüüsiks	24
2.2.6.1.1	Järjestuste joondamine.....	24
2.2.6.1.2	Järjestuste <i>in silico</i> kvaliteedikontroll	25
2.2.6.1.3	Genotüüpide määramine	25
2.2.6.1.4	Järjestuste liitmine võrdluspaneeli andmetega	26
2.2.6.2	Populatsioonigeneetiline analüüs	26
2.2.6.2.1	MtDNA ja Y-kromosoomi haplogruppide määramine.....	26
2.2.6.2.2	Peakomponentanalüüs	27
2.2.6.2.3	Sugulusanalüüs	27
2.3	Tulemused ja arutelu	28
2.3.1	Uuritavate proovide kvaliteet ja sekveneerimise tulemused	28
2.3.2	Soo määramine	30
2.3.2	Populatsioonigeneetilised analüüsid.....	31
2.3.2.1	MtDNA ja Y-kromosoomi haplogrupid	31
2.3.2.2	Peakomponentanalüüs	31
2.3.2.3	Sugulusanalüüs	32
	KOKKUVÕTE.....	36
	TÄNUSÕNAD	39
	KASUTATUD KIRJANDUSE LOETELU	40
	KASUTATUD VEEBIAADRESSID	46
	LISAD	47
	LISA 1.	47
	LISA 2.	49
	LISA 3.	50

LIHTLITSENTS.....51

KASUTATUD LÜHENDID

AMH – anatoomiliselt moodne inimene (ingl *anatomically modern human*)

ap – aluspaar (ingl *base pair*)

mtDNA – mitokondriaalne DNA

NaOCl – naatrium hüpoklorit

NGS – teise põlvkonna sekveneerimine (ingl *next generation sequencing*)

PCA – peakomponentanalüüs (ingl *principle component analysis*)

PCR – polümeraasi ahelreaktsioon (ingl *polymerase chain reaction*)

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

SRY – sugu määrav Y-piirkond (ingl *sex-determining region Y*)

TÜ GI – Tartu Ülikooli genoomika instituut

UDG – uratsiil-DNA-glükosülaas (ingl *uracil-DNA-glycosylase*)

SISSEJUHATUS

Inimese minevikusündmuste uurimisel tuginetakse erinevate valdkondade, nt ajaloo, arheoloogia ja geneetika tõenditele. Ajalugu põhineb nii kirjalikel tekstidel kui ka suulistel lugudel ja rahvaluulel. Varaseid tekste on aga vähe, kõiki sündmuseid ei kirjutatud ülesse ja paljud neist võivad olla ka aja jooksul hävinenud. Põlvkondadevaheliselt edasi antud lugusid on raske faktiliselt kontrollida. Arheoloogia allikmaterjal hõlmab näiteks tööriistade, ehete ja keraamika uurimist, aga ka nt pinnase, maastiku ja hoonete jäänuste uurimist, mille alusel tehakse oletusi populatsioonide eluolu ja etnilise kuuluvuse kohta. Geneetika aitab ajaloolaste ja arheoloogide poolt kogutud tõendeid ja nende alusel sõnastatud hüpoteese kontrollida ja neid omavahel sobitada, lisades spetsiifilisi detaile nt suguluse kohta.

Käesolev bakalaureusetöö on näide sellest, kuidas populatsiooniuringutes kõrvutatakse erinevate valdkondade andmeid, mille kaudu saadakse terviklikum pilt meie eelkäijate elust. Töös käsitletakse väikest osa Eesti kagunurka Siksälasse keskajal maetud indiviididest. Siksälä kalmet peetakse erakordseks, kuna see on väga rikas hauapanuste ja muude leidude poolest. Antropoloogiline ja arheoloogiline materjal lubab oletada ka uute elanike lisandumist Ida-Lätist ja liivlaste aladelt.

See töö teostati projekti „Keskaegse populatsiooni struktuur, tervis ja haigused Eesti territooriumil vana DNA vaatenurgast” (PRG1027) raames. Eesmärgiks oli kirjeldada Eesti keskaegse elanikkonna geneetilist varieeruvust ja sotsiaalseid sidemeid Siksälä kalme näitel. Selleks eraldati ja sekveneeriti DNA Siksälä kalmesse keskajal maetud indiviidide säilmetest ning analüüsiti nende vahelisi sugulussuhteid. Saadud andmeid võrreldi teiste kivi-, pronksi-, raua- ja keskajal Eesti aladel elanud indiviididega ning Eesti ja Euroopa tänapäevaste populatsioonidega.

1. KIRJANDUSE ÜLEVAADE

1.1 Vana DNA olemus

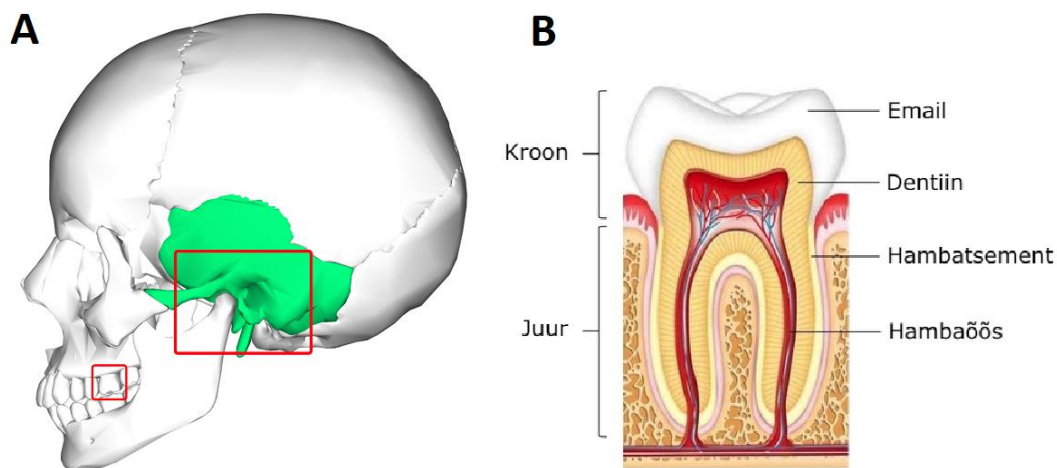
Vana DNA-ks nimetatakse surnud organismist pärit DNA-d, milles on alanud lagunemise protsessid (Herrmann ja Hummel, 1994). Organismiks võib olla inimene, loom või ka taim (Jaenicke-Despres *et al.*, 2003), kuid DNA-d saab eraldada ka näiteks sulgedest (Rawlence *et al.*, 2009), munakoortest (Oskam *et al.*, 2010), puidust (Wagner *et al.*, 2018), molluskikarpidest (Der Sarkissian *et al.*, 2017) ja setetest (Slon *et al.*, 2017). 1984. aastal suudeti esimesena sekveneerida kaks lühikest DNA fragmenti ühest väljasurnud sebra liigi esindajast (lad. k *Equus quagga*), keda säilitati muuseumieksponaadina. Sellest alates on tehnoloogia läbinud tohutu arengu, tänu millele on nüüd võimalik sekveneerida kuni miljoni aasta vanuste organismide kogu genoom (van der Valk *et al.*, 2021). Vana DNA uurimine võimaldab leida mingites kindlates aja- ja ruumipunktides elanud organismide geneetilist varieeruvust ja võrrelda seda tänapäevasega (Vicente ja Schlebusch, 2020). Selliste uurimuste tulemusena on see, kuidas inimesed hetkel minevikku mõistavad, oluliselt muutunud.

1.1.1 Säilivus

2021. aasta veebruaris avaldasid van der Valk jt artikli nende poolt eraldatud vana DNA-st, mis pärineb rohkem kui miljon aasta vanustelt mammutitelt, mis omakorda teeb sellest hetkel vanima edukalt analüüsitud vana DNA proovi (van der Valk *et al.*, 2021). Täpsemalt eraldati DNA kolme mammuti purihammastest, mis leiti Kirde-Siberist (van der Valk *et al.*, 2021). Siiani oli vanimaks vana DNA prooviks Orlando jt hobuse jäsemeluu fossiilist eraldatud DNA, mille vanuseks on 560 000 – 780 000 aastat (Orlando *et al.*, 2013). Nii mammutite purihambad kui ka hobuse jäsemeluu pärinesid igikeltsast, tänu millele seal DNA nii kaua säilis (van der Valk *et al.*, 2021; Orlando *et al.*, 2013).

DNA pika säilivuse puhul mängivad rolli mitmed tegurid. Esmalt hakkavad peale surma organismi lagundama tema enda ensüümid, seejärel erinevad mikroorganismid ja seened ning lõpuks ka putukad ja selgroogsed (Herrmann ja Hummel, 1994). Sellega saastavad need organismid surnud uuritava organismi DNA-d (endogeenset DNA-d) ka oma DNA-ga (Poinar *et al.*, 2006). Pehme koe lagunemise tagajärjel jääb alles anorgaaniline endo- või eksoskelett ning olenevalt organismist juuksed või suled. Erinevalt pehmetest kudedest on luud ja hambad lagunemise eest rohkem kaitstud tänu vähesele vee- ja ensüümisisaldusele. (Herrmann ja Hummel, 1994) Hammastel on rohkem kaitset saastuse eest tänu oma asukohale lõualuus, mille tõttu on need vana DNA uurimisel eelistatud (Higgins ja Austin, 2013). DNA eraldamiseks võetakse proov enamasti juure osast, kuna hammas on pealtpoolt kaitstud emailiga ja mujalt

hambatsemendiga (joonis 1). Täpsemalt eelistatakse dentiini, kuna selle kaudu on võimalik uurida ka vere kaudu levivaid patogeene (Margaryan *et al.*, 2018). Lisaks hammastele kasutatakse ka kaljuluud (*pars petrosa*), kuna sellest saab enamasti isegi rohkem endogeenset DNA-d kui hammastest – see on võimalik tänu kaljuluu väga tihedale struktuurile (Gamba *et al.*, 2014; Lam *et al.*, 1999).



Joonis 1. Vana DNA uurimisel kasutatavad proovid. (A) Hamba ja kaljuluu asukoht koljus (Kohandatud: <http://lifesciencedb.jp/bp3d/>). (B) Hamba ehitus (Kohandatud: Sekijima, 2016).

Kuigi endogeenne DNA on hammastes rohkem kaitstud, lagunevad aja jooksul ka sealsed DNA molekulid. Fragmentide pikkus varieerub olenevalt proovist, kuid keskmine langeb enamasti vahemikku 60–150 aluspaari (ap) (Prüfer *et al.*, 2010). Vanemate leidude puhul võivad fragmendid olla veelgi väiksemad. Näiteks Hispaaniast leitud umbes 400 000 aasta vanuse hominiini säilmetest eraldatud DNA fragmendid olid enamus alla 45 bp (Meyer *et al.*, 2014).

Endogeenset DNA-d lagundavaid protsesse on võimalik aeglustada, külmutades või kuivatades koed kiiresti peale organismi surma, mistõttu on külmad ja kuivad tingimused ka parimad DNA säilitamiseks (Dabney *et al.*, 2013). Vanimad eraldatud DNA proovid ongi säilinud just igikeltsas (van der Valk *et al.*, 2021; Orlando *et al.*, 2013). Soojades tingimustes, näiteks Egiptuses, säilib DNA ainult mõni tuhat aastat (Poinar *et al.*, 1996). Esimene sellistel tingimustel säilinud edukalt eraldatud vana DNA pärineb 4500 aasta vanuselt etiooplaselt (Gallego Llorente *et al.*, 2015), tänaseks on ka mujalt Aafrikast kogunenud andmeid taolistest genoomidest (Vicente ja Schlebusch, 2020). Konstantsetel tingimustel on suhe aja ja fragmentatsiooni vahel eksponentsiaalne, mis tähendab, et DNA laguneb püsival kiirusel. On välja arvatud, et 25°C juures laguneks DNA 1 bp pikkusteks fragmentideks 22 000 aastaga ning –5°C juures 6,83 miljoni aastaga. (Allentoft *et al.*, 2012)

1.1.2 Keemilised modifikatsioonid

Kuna peale organismi surma hakkavad rakkudes nukleasid ja mikroorganismid DNA-d lagundama (Herrmann ja Hummel, 1994) ning DNA-d parandavad ja stabiliseerivad mehhanismid enam ei toimi, viib see mitmete keemiliste modifikatsioonideni (Dabney *et al.*, 2013). DNA fragmentatsioon tekib põhiliselt depuriinimise kaudu, milleks nimetatakse protsessi, kus läbi hüdroolüüsi lõhutakse ära puriini ja desoksüriboosi vaheline glükosiidside. Selle tulemusena tekib apuriinne sait ja DNA ahel võib sealt kergelt katkeda. (Dabney *et al.*, 2013)

Vana DNA-d iseloomustavad fragmendite otstes toimuvad nukleotiidide transitsioonid (tsütosiinist tümiiniks (C→T)) ja guaniinist adeniiniks (G→A)), mis esinevad vastavalt fragmentide 5' ja 3' otstes. Enamasti leidub sekveneeritud fragmentide järjestustes just C→T transitsiooni, mille põhjustab tsütosiini muutumine uratsiiliks läbi tsütosiini deamineerimise. Sekveneerimisel loetakse uratsiil tümiiniks, kuna DNA molekulis pole uratsiili. Arvatakse, et G→A transitsioon tekib vastasahela C→T transitsiooni tulemusena. (Hofreiter *et al.*, 2001) Neid transitsioone kasutatakse vana DNA autentsuse määramiseks (Briggs *et al.*, 2007). Mõnel juhul, kui soovitakse parema kvaliteediga sekveneerimistulemusi, kasutatakse DNA ekstraktidel enne genoomsete raamatukogude koostamist uratsiil-DNA-glükosülaasi ehk UDG- (ingl *uracil-DNA-glycosylase*) töötlust, millega eemaldatakse DNA järjestuses tekkinud uratsiilid ja lõigatakse välja selle tulemusena tekkinud apürimidiinne sait. Kuigi antud töötluste järel saadud järjestused on kvaliteetsemad, siis kaob sellega võimalus määrata vana DNA autentsust, seetõttu kasutatakse täna osalist UDG-töötlust. (Rohland *et al.*, 2015)

1.2 Uurimismeetodid

Vana DNA uurimine on aastatega läbinud tohutu arengu, iseäranis seoses teise põlvkonna sekveneerimise (NGS, ingl *next generation sequencing*) väljatöötamisega. Esimesed uuringud viidi läbi kasutades lühikest osa mitokondriaalsest DNA-st (mtDNA), kuna seda on ühes rakus tuhat korda rohkem koopiaid kui tuumset DNA-d. Peale polümeraasi ahelreaktsiooni (PCR, ingl *polymerase chain reaction*) leiutamist hakati vana DNA-d järjest rohkem uurima, kuna nüüd oli võimalus vana DNA molekule amplifitseerida. Sellega suurenes aga probleem saastusega, kuna PCR-i kasutamine eeldab pikki järjestusi ning vana DNA lühikesed fragmendid jäävad tahaplaanile ja esile tuleb hoopis saastus. (Rizzi *et al.*, 2012) Tänapäevaks on PCR-i välja vahetanud NGS, mis võimaldab uurida terveid genome olenemata fragmentide pikkusest (Xu *et al.*, 2020). Juba mõned kuud peale NGS meetodi esmatutvustust 2005. aastal (Besser *et al.*, 2018) sekveneerisid Poinar jt 13 miljoni bp pikkuse osa mammuti genoomist (Poinar *et al.*, 2006) ning esimene täisgenoom, mis kuulus samuti mammutile, sekveneeriti juba

2008. aastal (Miller *et al.*, 2008). Kaks aastat hiljem suudeti sekveneerida ka esimene inimese täisgenoom, mis pärineb paleo-eskimolt Gröönimaalt (Rasmussen *et al.*, 2010).

Kuna käesolevas töös on rõhk eelkõige sugulusanalüüsil, siis kirjeldatakse järgnevalt lähemalt neid aspekte, mis genoomide uurimisel sugulussuhete väljaselgitamiseks infot annavad ja räägitakse täpsemalt sugulusanalüüsi meetodist, mis on sobilik vana DNA-le.

1.2.1 Mitokondriaalne ja Y-kromosomaalne DNA analüüs

MtDNA on rõngakujuline kaheaheelaline DNA molekul rakku energiaga varustavas mitokondris, mille suurus inimese rakus on ~16,5 tuhat ap. Arvatakse, et mitokondri oli algsest endosümbioosne bakter, mis seletaks asjaolu, miks on sellel eraldi oma DNA. Evolutsiooni käigus on aga enamus mtDNA geenidest liikunud tuumagenoomi ning mitokondrisse on alles jäänud ainult 37 geeni, mis kodeerivad kas oksüdatiivse fosforüülimise või mitokondriaalse valgusünteesi jaoks vajaminevat infot. Erinevalt tuumsest DNA-st pärandub mtDNA ainult mööda emaliini, mis tähendab, et see ei rekombineeru, muutused selle järjestuses on evolutsioonilises plaanis olulised ainult naistes ja see võimaldab uurida ajalugu konkreetselt nende kohta. (Jobling *et al.*, 2014) Samuti aitab mtDNA analüüs määrata põlvkondade vahelisi naisliinis päranduvaid põlvnemissuhteid (Merheb *et al.*, 2019).

Y-kromosoom on ~60 miljoni ap pikkune teine mitterekombineeruv osa genoomist, mis on spetsiifiline meestele ehk see pärandub ainult mööda meesliini ning sarnaselt mtDNA-le saab selle kaudu uurida ühes vanemliinis päranduvaid sugulussuhteid. Y-kromosoomis on sugu määrav Y-piirkond (SRY, ingl *sex-determining region Y*), kus paikneb SRY geen, mis põhjustab munasarja asemel testise arenemist. Kuna homoloogi puudumise tõttu ei toimu Y-kromosoomiga, sarnaselt mtDNA-le, rekombineerumist, muutub see ainult mutatsioonide, tavaliselt üksiku nukleotiidi polümorfismide (SNP, ingl *single nucleotide polymorphism*) tekke tõttu. (Jobling *et al.*, 2014)

SNP-de alusel jagatakse nii ema- kui isaliine nn haplogruppideks, mis on vastavalt mtDNA ja Y-kromosoomi fülogeneesipuu monofüleetilised harud. Nende haplogruppide alusel saab uurida ema- ja isaliinide geograafilist levikumustrit. (Jobling *et al.*, 2014)

1.2.2 Autosomaalne DNA analüüs

Kuigi mtDNA ja Y-kromosoom on kasulikud soospetsiifiliste uuringute tegemiseks, katavad nad genoomist väga väikese osa. Seetõttu kasutatakse täna täiendavalt autosoomides ehk teistes kromosoomides sisalduva geneetilise info analüüsi. Autosoomid erinevad mtDNA-st ja Y-kromosoomist selle poolest, et nende vahel toimub rekombinatsioon, tänu millele sisaldavad

autosoomid infot nii ema- kui ka isaliini ajaloo kohta. MtDNA-d ja Y-kromosoomi on varem laialdaselt kasutatud põlvnemissuhete määramiseks, kuid nende kaudu pole võimalik määrata detailseid sugulussuhteid, mille tõttu kasutatakse tänapäeval sugulusanalüüsiks autosoome. (Jobling et al., 2014) Vana DNA puhul on lood seoses järjestuste halva kvaliteediga pisut keerulisemad, kuid täna saab sugulusanalüüsi teha juba vaid 0,1x katvusega genoomidega, nt tänu READ meetodile (vt 1.2.3) (Kuhn *et al.*, 2018).

Autosomaalse DNA, mtDNA ja Y-kromosoomi analüüsid täiendavad üksteist, nende kaudu saab tervikliku pildi indiviidi genoomist ja populatsioonide geneetilisest mitmekesisusest. MtDNA ja Y-kromosoomi liinid on alati juhuslike sündmuste tõttu populatsioonist kaduma, mistõttu ei leia tänapäeva genoomidest kogu informatsiooni varasemate liinide kohta. Autosomaalsete uuringute lisamine sugukromosoomide analüüsile võimaldab vaadata veel kaugemas minevikku kui varem, aga uurida ka lähedases suguluses olevate indiviidide põlvnemissuhteid. (Colonna *et al.*, 2011) Vana DNA andmestiku lisamine tänapäevasele võimaldab omakorda paremini mõista nii tekkinud mutatsioonide kogunemise kiirust kui ka aja jooksul kaotsiläänud osa kunagisest varieeruvusest (Kivisild, 2017).

1.2.3 Sugulusanalüüs

Lisaks populatsiooni struktuuri üldisele iseloomustamisele on viimasel ajal geneetilisi meetodeid kasutatud nii kuulsate perekondade sugulussuhete analüüsimiseks (Alterauge *et al.*, 2021; O'Sullivan *et al.*, 2018) kui ka tervete ühiskondade sotsiaalse elukorralduse uurimiseks sugulussuhete kaudu (Haak *et al.*, 2008; Amorim *et al.*, 2018; Mittnik *et al.*, 2019).

Sugulusanalüüsi läbiviimiseks vana DNA alusel töötati hiljaaegu välja READ (ingl *relationship estimation from ancient DNA*) tarkvara/meetod, millega on võimalik leida kas individid on identsed (kaksikud või sama isik), esimese astme sugulased (lapsevanem/laps, õed/vennad) või teise astme sugulased (vennapoeg/-tütar või õepoeg/-tütar - tädi/onu, vanavanem/lapselaps, poolõed/-vennad). Meetodit testiti andmetega, milles leiduvad sugulussuhted olid eelnevalt teada. (Kuhn *et al.*, 2018)

READ tarkvara võrdleb omavahel kahe indiviidi genoomi 1 miljoni ap suuruste kattumatute akendena ning leiab, kui suur osa alleelidest on erinevad. Selleks arvutatakse mittekattuvate alleelide suhe indiviidide paari vahel igas aknas (P0). Protsess tehakse läbi kogu genoomiga ja kõikide indiviidipaaridega. Järgnevalt normaliseeritakse see arv kasutades tarkvara poolt antud populatsiooni järgi välja arvutatud oodatud P0 väärtust. Normaliseerimisega tagatakse see, et suguluse arvutamisel on arvestatud ka populatsiooni mitmekesisusega. Normaliseeritud arvu

võrreldakse tarkvara poolt arvatud lävenditega ning paar klassifitseeritakse vastava astme sugulaseks või suguluses mitteolevaks. (Kuhn *et al.*, 2018)

Tulemustes võib esineda ka valenegatiivseid ja valepositiivsed klassifitseerimisi. Nende kogus sõltub suuresti andmete mahust, seega on piiripealsete tulemuste korral soovitatav lisada analüüsi rohkem indiviide. Valepositiivsete tulemuste arv (2,09% ka ainult 1000 kattuva SNP korral) on võrreldes valenegatiivsetega (9,95% 1000 kattuva SNP korral) palju väiksem, mis autorite hinnangul tähendab, et programm ei pruugi küll leida kõiki teise astme sugulasi, kuid leitud sugulussuhted on enamjaolt alati tõesed. (Kuhn *et al.*, 2018)

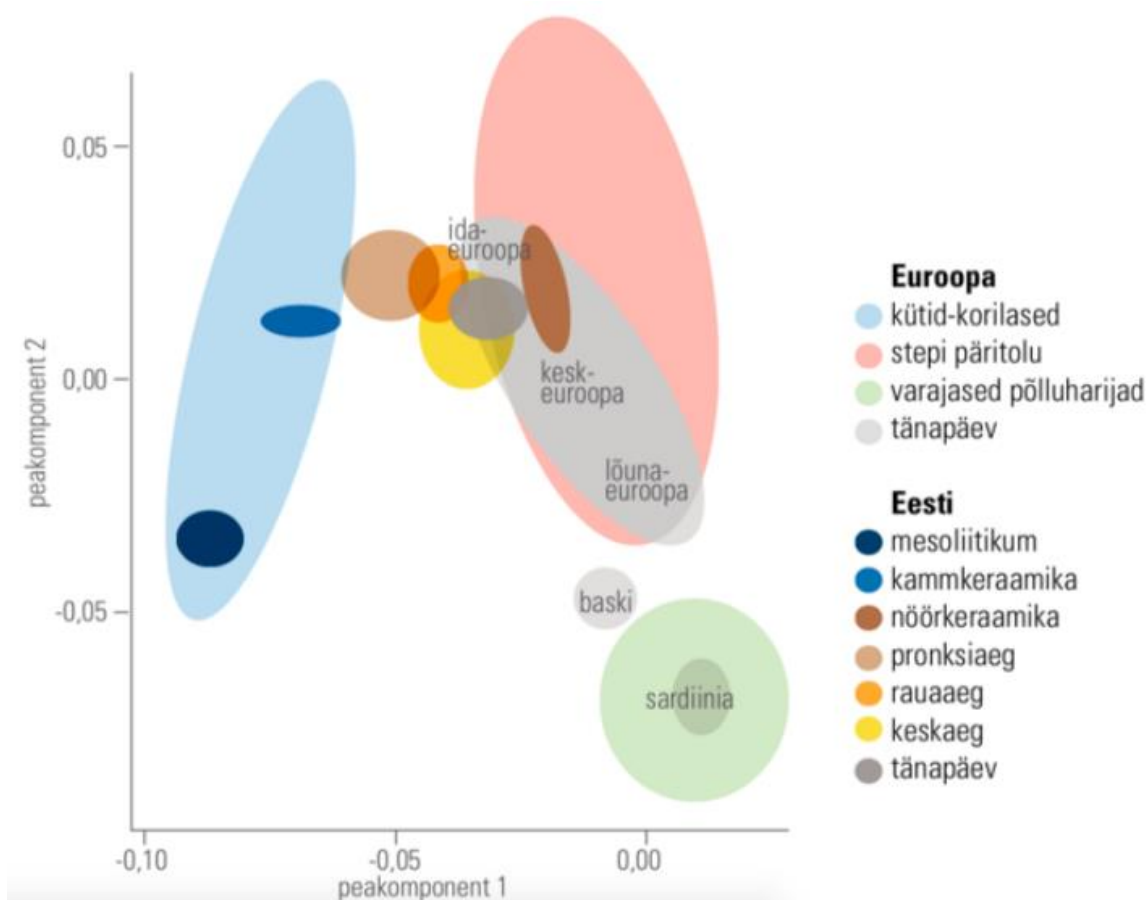
1.3 Eesti demograafiline ajalugu Euroopa kontekstis

Euroopa demograafilist ajalugu on kujundanud kolm põhilist suurt asustuslainet: esmalt jõudsid Euroopasse kütt-korilased, seejärel varajased põlluharijad ning viimaks stepipäritolu rahvad. Esimesed *Homo sapiens*'i liigi esindajad ehk anatoomiliselt moodsad inimesed (AMH, ingl *anatomically modern human*) jõudsid Euroopasse arvatavasti ~45 000 aastat tagasi (Benazzi *et al.*, 2011), kuid pidid mandrijää laienemise tõttu viimase jääaja maksimumil ~20 000 aastat tagasi Euroopa lõunapoolsematele aladele taanduma. Euroopa põhjaosad, sh Eesti alad, taasasustati kütt-korilaste poolt peale viimast jääaega. (Kriiska *et al.*, 2020) Aja jooksul toimunud nihked Euroopa ja Eesti alade rahvastiku geenipärandis on skemaatiliselt esitatud joonisel 2.

Baltikumi lõunaossa jõudsid kütt-korilased ~9000 aastat eKr, keskmisel kiviajal ehk mesoliitikumis, asustades algselt siseveekogude äärsed alad (Kriiska *et al.*, 2020) ning ~2000 aastat hiljem hilises mesoliitikumis ka rannikuäärsed alad (Kriiska, 2009). Eesti alade vanimad leiud kuuluvad Kunda kultuuri (Kriiska, 2009). Neoliitikumi ehk noorema kiviaja alguses, ~4000 aastat eKr, tulid Baltikumi aladele ida poolt kütt-korilased, kes erinesid geneetiliselt mesoliitikumi kütt-korilastest ja tõid kaasa kammkeraamika kultuuri (Kriiska, 2009; Saag jt 2017; Jones jt. 2017; Mittnik *et al.*, 2018; Kriiska *et al.*, 2020). Täna on kütt-korilaste geneetiline komponent Euroopa rahvastest suurim (20-30%) eestlaste ja leedulaste seas (Haak *et al.*, 2015).

Põlluharimine jõudis Eesti aladele teistmoodi kui mujale Euroopas. Põlluharimine ja karjakasvatus hakkas Euroopas levima ~7000 aastat eKr, liikudes Lähis-Idast algselt Balkani poolsaarele. Kagu-Euroopast rändasid põlluharijad mööda Vahemere rannikut Lõuna-Euroopasse ning sealt lõpuks Põhja-Euroopasse. (Douglas Price, 2000) Eestisse jõudis põlluharimine aga hoopis hiljem, ~2800 aastat eKr koos stepipäritolu rahvaste migratsiooniga,

mis tõi Euroopasse ka nõorkeraamika kultuuri (Kriiska, 2000; Haak *et al.*, 2015). Stepirahvaste geneetiline komponent moodustab eestlaste geenipärandist täna 60% (Tambets *et al.*, 2018).



Joonis 2. Euroopa ja Eesti rahvastiku geneetiline ajalugu vana DNA põhjal. Helehalli ja tumehalliga on märgitud vastavalt kaasaegsete eurooplaste ja eestlaste gruppide ning värvilisega vana DNA alusel uuritud populatsioonide paiknemine genoomiandmete peakomponentanalüüsi graafikul. (Kriiska *et al.*, 2018)

Pronksiajal (1200–400 a eKr) suurenes Baltikumis jälle kütt-korilaste geneetilise komponendi osakaal ning siinsed elanikud sarnanesid sel perioodil geneetiliselt rohkem lääne kütt-korilaste kui nõorkeraamikutega (Saag *et al.*, 2019). Erinevalt muust Euroopast on Eesti alade elanikkonda (ning ka nt Soome, Karjala jt uurali keeli rääkivate rahvaste geneetilist pärandit) mõjutanud ka umbes 2500 aastat tagasi siia jõudnud asustuslaine, mis tõi kaasa nn Siberi komponendi, mida on ka tänapäeva eestlaste geenipärandis 5% (Tambets *et al.*, 2018). Alates varajasest rauaajast ei ole Eesti alade rahvastiku geenipärandis väga suuri muutusi toimunud (vt ka joonis 2).

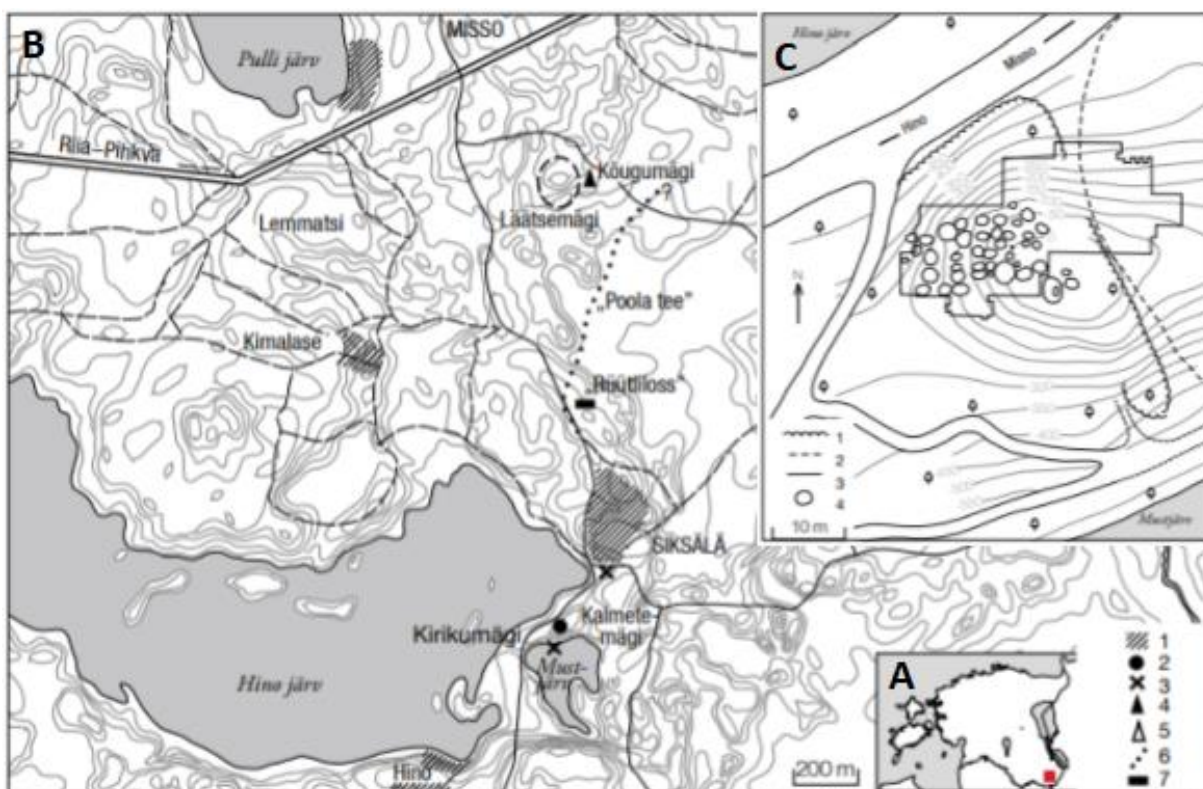
1.4 Eesti keskaeg

Eesti aladel algas keskaeg hiljem kui mujal Euroopas ehk kui Euroopas algas keskaeg juba ~500 a pKr, siis Eestis alles koos ristsõdadega 13. sajandil. Eesti keskaja lõpuks peetakse 16. sajandil toimunud Liivi sõda (1558–1583). 13. sajandi alguses kuulus Liivimaa koosseisu ainult Eesti lõunaosa, kuid 14. sajandi keskpaiku müüs Taani kuningas Põhja-Eesti Saksa ordule. (Kriiska *et al.*, 2020)

Keskaja alguses oli Eesti rahvaarv hinnanguliselt ~100 000–200 000, ent keskaja lõpuks ulatus see, tänu sisserändele ja loomulikule iibele, juba ~300 000-ni (Palli, 1996). Kogu Eesti keskaja vältel liikus siin mitmeid erinevaid rahvaid. Lisaks eestlastele elasid Põhja- ja Lääne-Eestis ka soomlased, rootslased ja taanlased ning mujal Eestis ka sakslased ja venelased. Eestlased kui talupojad elasid enamasti maapiirkondades ning linnas elava ülemkihi moodustasid hoopis sakslased, kes olid valdavalt kaupmehed ja käsitöölised, kuna enamik eestlastest ülikuid kaotas peale ristsõdu oma staatuse. (Kriiska *et al.*, 2020; Selart *et al.*, 2012) Keskaegse elanikkonna sotsiaalset kihistumist lubavad oletada ka seni tehtud vana DNA-põhised pilootuuringud, kus maa- ja jõukate linnaelanike vahel ilmnisid geneetilised erinevused, mis viitavad nende erinevale päritolule (Irtdt, 2020).

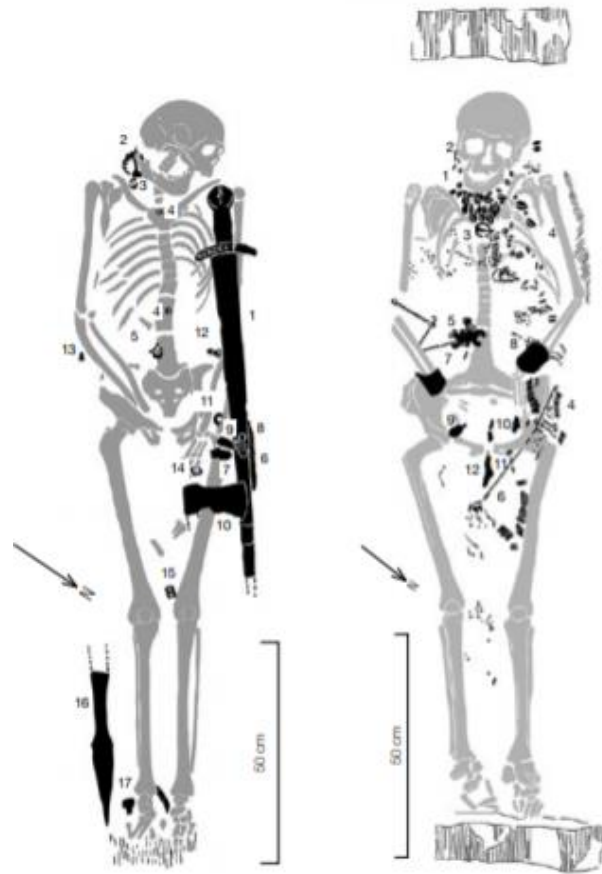
1.4.1 Siksälä kalme

Siksälä ehk Siksali kalme asub Kagu-Eestis (joonis 3A) 2 km kaugusel Misso alevikust kahel Hino järve ja Mustjärve äärsel kõrvuti oleval mäekünkal (joonis 3B) – Kalmetemäel (joonis 3C) ja Kirikumäel. Nendevahelisse nõokohta pole kedagi maetud, seega võib neid pidada kaheks eraldi matmispaigaks, aga riikliku kaitse all on need ühe muistisena. Siksälä on suurimas ulatuses uuritud keskaegne külakalme Eestis, mida iseloomustab pikk kasutusaeg. Lisaks keskaegsetele matustele on sealt leitud ka rauaaegseid ja varauusaegseid säilmeid ning see on erakordselt rikas hauapanuste ja leidude poolest. Joonisel 4 on kujutatud siinses töös analüüsitud Kalmetemäe ülikupaari, nn mõõgamehe ja tema naise matused koos unikaalsete hauapanustega. Matmiste algusaeg Siksäläs on ebaselge, kuna hulgaliseks muutusid leiud alles alates 11. sajandist. Siksälä kalmistult on leitud nii põletus- kui ka laibamatuseid, aga alates 13. sajandist maeti põhiliselt põletamata surnuid, kuigi põletusmatuseid tuli veel ette kuni 15. sajandini. Leitud põletusmatused jagunesid omakorda kaheks: hajusad puistematused, mille tõttu segunesid erinevate inimeste luustikud, ja selgepiirilised luukogumid. Siksälä kalmest on läbi uuritud 1057 m² suurune ala, mis moodustab umbes 75–80% selle kogupindalast. (Valk ja Laul, 2014)



Joonis 3. Töös uuritud arheoloogiliste kohtade kaart. (A) Siksälä asukoht Eesti kaardil (punase ruuduga). (B) Arheoloogilised leiukohad Hino ja Pulli järve ümbruses. 1 – asulakoht; 2 – kalme; 3 – õietolmuproovi võtmise koht; 4 – linnamägi; 5 – pühapuude asukoht; 6, 7 – pärimuslikud paigad. (C) Kalmetemäe skeem. 1 – 1980. a võsapiir; 2 – ühismajandi aegne haritava maa piir; 3 – kaevandi piir; 4 – kääpad. (Kohandatud: Valk ja Laul, 2014)

Maetute põhjal võib oletada, et keskaja esimesel sajandil toimus uue rahva sisserännet kaks korda. Esimesele sisserändele viitab põletusmatuste järsk asendumine laibamatustega, millest leitud rõiva- ja ehtekultuur on iseloomulik Ida-Läti elanikele. (Valk ja Laul, 2014) Seda kinnitas ka sel ajal maetud meeste pikk pea ja keskmiselt laiema nägu, mille sarnast pole seni mujalt Eestist leitud (Heapost, 2007). Teist sisserännet lubab arvata aga matuste arvu suurenemine 14. sajandi alguses, mida seostatakse selliste antropoloogiliste tunnuste muutustega, mis viitavad maetute võimalikele sidemetele liivlaste asualadega (Valk ja Laul, 2014; Heapost, 2007). Käesolevas töös analüüsitud individid pärinevad Kalmetemäelt (joonis 3C).



Joonis 4. Ülikupaari matused 14. sajandi algusest. Indiviidid SIK002 (matus 248) ja SIK003 (matus 249) (Valk ja Laul, 2014).

2. EKSPERIMENTAALOSA

2.1 Töö eesmärgid

Uurimisöö eesmärgiks oli iseloomustada Eesti keskaegse elanikkonna geneetilist varieeruvust ja sotsiaalseid sidemeid Siksälä kalme bioarheoloogilise materjali näitel. Selleks:

- 1) Eraldati ja sekveneeriti DNA Siksälä kalmesse keskajal maetud indiviidide säilmetest;
- 2) Analüüsiti Siksälä kalmesse maetud indiviidide vahelisi sugulussuhteid;
- 3) Võrreldi saadud genoomiandmeid teiste Eesti aladel elanud kivi-, pronksi-, raua- ja keskaegsete indiviidide ning Eesti ja teiste Euroopa kaasaegsete populatsioonide andmetega, et:
 - a) iseloomustada Siksälä kalmesse maetud inimeste geneetilist varieeruvust Eesti alade keskaegse ja varasema rahvastiku kontekstis;
 - b) iseloomustada Siksälä kalmesse maetud inimeste geneetilist varieeruvust Euroopa kaasaegses piirkondlikus kontekstis.

2.2 Materjal ja metoodika

2.2.1 Bioarheoloogiline materjal vana DNA analüüsiks

Antud töös analüüsiti Tartu Ülikooli ajaloo ja arheoloogia instituudi kogude arheoloogilist luumaterjali, mis on pärit Võru maakonnas Rõuge vallas asuvast Siksälä kalmest. Luumaterjal on kogutud 1980.–1991. a. läbiviidud arheoloogiliste välitööde käigus. Meie koostööpartnerid selle projekti raames olid arheoloogia prof. Heiki Valk ja osteoloog Martin Malve Tartu Ülikooli ajaloo ja arheoloogia instituudist. DNA eraldati 20 indiviidilt, täpsemalt 19 hambast ja ühest kaljuluust, mis kuulusid Siksälä kalmesse 13.–15. saj. maetud inimestele. Lisaks kaasati analüüsidesse kaks varasemalt pilootuuringus sekveneeritud indiviidi SIK002 ja SIK003 (tabel 1).

2.2.2 Materjali kataloogimine

Selleks, et luumaterjali analüüsimisega alustada, tuli see esmalt üles pildistada ja koos olemasolevate andmetega Tartu Ülikooli genoomika instituudi (TÜ GI) Eesti Biokeskuse vana DNA labori andmebaasi salvestada. Hammaste puhul pildistati neid pealtvaatest ja kahest külgvaatest, kaljuluu puhul igast võimalikust külgvaatest, et kogu luu saaks jäädvustatud (joonis 5).

2.2.3 DNA eraldamine

DNA eraldamine viidi läbi 2020. ja 2021. aastal Tartu Ülikooli nooremteaduri Stefania Sasso juhendamisel TÜ GI Eesti Biokeskuse vana DNA laboris.

Tabel 1. Vana DNA uuringus kasutatud iniviidid, nende tüpokronoloogilised dateeringud ja vanused (Valk ja Laul, 2014).

#	Indiviid	Matuse number	Dateering	Vanus	#	Indiviid	Matuse number	Dateering	Vanus
1	SIK002	248	1300-1350	40-45a	12	SIK013	135	1300-1350	40-50a
2	SIK003	249	1300-1325	>50a	13	SIK014	157	1400-1475	8-13a
3	SIK004	17	1200-1255	25-35a	14	SIK015	173	1300-1475	9-14a
4	SIK005	47	1300-1350	14-18a	15	SIK016	177	1300-1475	12-18a
5	SIK006	83	1300-1350	35-40a	16	SIK017	187	1250-1350	30-40a
6	SIK007	108	1400-1475	12-18a	17	SIK018	200	1300-1350	18-21a
7	SIK008	110A	1400-1475	35-45a	18	SIK019	246	1300-1400	25-35a
8	SIK009	110B	1400-1475	35-45a	19	SIK020	245A	1225-1300	7-11a
9	SIK010	111	1300-1475	25-35a	20	SIK021	245C	1225-1300	>50a
10	SIK011	113	1400-1475	3-5a	21	SIK022	245D	1225-1300	3-5a
11	SIK012	131	1250-1400	45-50a	22	SIK023	247B	1225-1250	30-40a

2.2.3.1 Saastuse vältimine

Puhtus on vana DNA laboris väga olulisel kohal, kuna on vaja vältida arheoloogilise materjali saastamist. Esimese asjana vahetati labori eesruumis ära igapäevased riided ja jalanõud, et hoida võimalikult palju ära mustuse sissetoomist väliskeskonnast. Järgmises ruumis pandi selga spetsiaalne kaitseülirikond, juuksevärk, kaitseprillid, näomask, kaks paari kummikindaid ja kummikud. Laboriruumides puhastati iga kord kõik tööpinnad ja kasutatavad vahendid, sh proovide hoiustamise kotid ja tuubid, enne ja pärast tööd DNA-ExitusPlus™ (ITW Reagents) lahusega. Enne tööle asumist vahetati välja ka kõik kotid, milles proovid laborisse saadeti, puhaste vastu.

2.2.3.1 Proovi võtmine DNA analüüsiks

Vana DNA analüüsiks võeti ketaspuuriga hambast kas juuretipp või mitme juurega hamba korral terve juur, kaljuluu puhul tükike kõige tihedama luustruktuuriga kohast. Puurimine toimus laminaar-tõmbekapi all eraldi ruumis. Kui muud tööpinnad ja -vahendid puhastati DNA-ExitusPlusiga, siis puurimiseks vajalikke pintsette, puuri löikekettaid ja hambasondi leotati puhastamiseks mõned minutid 6% naatrium hüpokloriti (NaOCl) lahuses, loputati Milli-Q

(Merck Millipore) veega ja lõpuks 70% etanooliga. Seda protsessi korrati iga proovi vahel, et vältida ristsaastumist.



Joonis 5. Indiviidi SIK005 (matus 47) hammas ja indiviidi SIK022 (matus 245D) kaljuluu.

Tõmbekapi alune pind kaeti fooliumiga ning volditi ka väike „sein”, et vähendada tolmu üleliigset lendumist puurimisel. Lisaks asetati tasapinnale ka väiksem tükk fooliumit, mida vahetati iga proovi järel. Puurimiseks pandi kätte ka lisakäised ja puhtad kummikindad, mida samuti vahetati peale iga puurimist. Kui hamba peal esines üleliigset mustust, näiteks mulda, siis see pühiti ühekordselt kasutatava hambaharjaga enne puurima asumist maha.

Puurimiseks hoiti hammast või luud kõvasti hambatangide vahel kaalukausi kohal ning võeti sobiv tükk proovi võimalikult vähe kahjustamata. Ühest hambast (SIK011) tuli võtta pulber, kuna mõlemad juured oli seest tühjad ning sealt poleks piisavalt DNA-d kätte saanud. Proov pandi puhtasse andmetega märgistatud kotti ning ülejäänud asetati tagasi vanasse kotti.

2.2.3.3 Pinnasaastuse eemaldamine ja DNA eemaldamine hambajuurtest

Saastuse eemaldamiseks valmistati esimese asjana ette pabersalvrätikutega kaetud alusproovide kuivatamiseks, mis asetati tõmbekappi UV valguse alla. Seejärel võeti iga proovi jaoks kaks 50 ml tuubi, kuhu pooltesse valati 5 ml 6% NaOCl lahust ja teistesse 5 ml 70% etanooli. Kõigepealt asetati proovid NaOCl lahusega tuubidesse 5 minutiks, mille jooksul neid loksutati, et võimalikult palju mustust lahti tuleks. Seejärel valati NaOCl jooksva vee all kraanikaussi, loputati iga proovi 3 korda Milli-Q veega. Järgmisena asetati proovid 2 minutiks etanooliga tuubidesse, mille ajal neid samuti loksutati. Viimaks valati etanool proovidelt ära ning need paigutati ettevaatlikult varem ettevalmistatud ja proovinumbri märgistatud alusele.

Proovid jäeti tõmbekappi UV valguse alla kuivama ~1–2 tunniks. Seda protsessi ei tehtud prooviga, mis oli pulbri kujul.

Kuivanud proovid kaaluti, et välja arvutada DNA eraldamiseks vajalike puhvrite kogused. Antud töö proovide kaalud jäid vahemikku 31–240 mg (lisa 1). Olenevalt kaalust võeti proovide jaoks kas 5 või 15 ml tuub ning 100mg luumaterjali kohta pipeteeriti tuubi 2 ml EDTA puhvrit (0,5 M, pH 8) ja 50 µl proteinaas K lahust (18,182 mg/ml) [lõppkontsentratsioonid ($C_{lõpp}$) vastavalt 0,487 M ja 0,443 mg/ml]. Proovid pandi vastavatesse tuubidesse, mis kaeti parafilmiga ja asetati aeglaselt roteeruvale loksutile (*Nutating Mixer, VWR International*) 72 tunniks. Pulbriga proov oli loksutis ainult 24 tundi.

2.2.3.4 DNA ekstrakti puhastamine

Esimese asjana pandi proovid 5 minutiks tsentrifuugi (*Centrifuge 5810 R, Eppendorf*) kiirusel 4000 pööret minutis (rpm). Tsentrifuugimise ajal valmistati ette elueerimispuhver EB (*QIAquick PCR Purification Kit, QIAGEN*), pipeteerides 1,6 ml puhvrit 5 ml tuubi, ja märgistati kontsentratori tuubid proovide nimedega. Tsentrifuugitud proovidest pipeteeriti DNA-d sisaldav ekstrakt kontsentratori tuubidesse ja asetati tagasi tsentrifuugi 20–40 minutiks (4000 rpm), kuni ekstrakti maht oli 250 µl. Ootamise ajal valmistati ette 1,5 ml tuubid, kuhu kirjutati peale vajalikud andmed, ja suure mahutiga puhastuskolonnid (*High Pure Vital Nucleic Acid Large Volume Kit, Roche*), kuhu valati 2,5 ml sidumispuhvrit PB (*QIAquick PCR Purification Kit, QIAGEN*). Kontsentreeritud ekstrakt lisati see pipetiga suspenderides sidumispuhvrile, et DNA seonduks maatriksile. Üleliigse vedeliku kättesaamiseks tsentrifuugiti proove 1 minut kiirusel 4000 rpm. Peale seda valati vedelik ära ning tuubidesse lisati 1 ml PE puhvrit (*QIAquick PCR Purification Kit, QIAGEN*), mis on vajalik DNA puhastamiseks ning proove tsentrifuugiti jälle 1 minut kiirusel 4000 rpm. Järgmisena visati vedelikuga kogumistuub ja puhvermahuti ära ning kolonn asetati 1,5 ml kogumistuubidesse, mille järel tsentrifuugiti neid kuivatamiseks 1 minut kiirusel 13 000 rpm. Peale tsentrifuugimist visati kogumistuubid jälle minema ning kolonnid asetati varem ettevalmistatud 1,5 ml tuubidesse. Kolonnidele pipeteeriti 100 µl elueerimispuhvrit EB, et DNA vabaneks maatriksilt. Proovid asetati 10 minutiks termoblokki 37°C juurde ning seejärel tsentrifuugiti proove 2 minutit (13 000 rpm). Viimaks visati kolonnid ära, tuubid ekstraktiga suleti ja pandi sekveneerimisraamatukogude tegemiseni -20°C juurde sügavkülma.

2.2.4 DNA sekveneerimisraamatukogude koostamine

DNA sekveneerimisraamatukogude koostamine koosneb DNA fragmentidele adapterite ja indeksite lisamisest. Selleks kasutati *NEBNext DNA Sample Prep Master Mix Set for 454*

(E6070, *New England BioLabs*) komplekti, mis sisaldas fragmentide otste täitmiseks vajaminevat ensüümisegu ja puhvrit, adapterite lisamiseks ensüümi, ligeerimispuhvrit ning otste täitmiseks vajalikku ensüümi ja puhvrit. Esimesed 3 etappi teostati TÜ GI Eesti Biokeskuse vana DNA laboris, 4. etapi PCR-i segu valmistati samuti vana DNA laboris, kuid DNA fragmentide amplifikatsioon toimus TÜ GI tuumiklaboris, et vältida vana DNA labori saastamist PCR-i produktidega. Kogu protsess viidi läbi TÜ nooremteaduri Stefania Sasso juhendamisel.

2.2.4.1 DNA fragmentide üheaheelaliste otste täitmine ja reaktsioonisegust puhastamine

Kõigepealt segati kokku reaktsioonisegu (1,5 ml tuubi), millega täidetakse DNA fragmentide üheaheelalised otsad. Selleks lisati 12,5 µl steriilset vett, 5 µl 10x otste täitmise puhvrit ja 2,5 µl otste täitmise ensüümisegu ning valmis segu pandi iga proovi jaoks 20 µl 0,2 ml tuubidesse. Seejärel lisati reaktsioonisegule 30 µl DNA proovi ning tuubid pandi PCR-i masinasse (*K640 Thermal Cycler, Heal Force*) 20°C juurde 30 minutiks inkubeerima.

Järgmisena puhastati proovid kasutatud reaktsioonisegust, milleks kasutati *MinElute PCR Purification Kit* (QIAGEN) komplekti. Puhastamiseks pipeteeriti esmalt kolonnidesse 550 µl PB puhvrit, mille abil saab DNA maatriksile seonduda. Puhvrile lisati 50 µl täidetud otstega DNA-d ning tuube tsentrifugeeriti kiirusel 13 000 rpm 1 minut. Kolonnist läbi läinud vedelik valati ära ning reaktsioonisegu eemaldamiseks lisati 650 µl PE puhvrit. Tuube tsentrifugeeriti veelkord kiirusel 13 000 rpm 1 minut, läbijooksnud vedelik valati välja ning asetati uuesti 1 minutiks (13 000 rpm) kuivatamiseks tsentrifugeeriti. Seejärel tõsteti kolonnid puhastesse 1,5 ml tuubidesse ümber ning kolonnile pipeteeriti DNA vabastamiseks 30 µl EB puhvrit. Tuubid tõsteti 10 minutiks 37°C juurde termoblokki ning peale seda tuli proove taaskord tsentrifugeerida kiirusel 13 000 rpm, kuid 1 minuti asemel 2 minutit. Lõpuks visati kolonnid minema ning tuubid suleti.

2.2.4.2 Adapterite lisamine ja reaktsioonisegust puhastamine

DNA fragmentide otstesse tuleb lisada spetsiifilised identsed adapterid. Selleks valmistati 1,5 ml tuubi reaktsioonisegu, mis sisaldas 1 proovi kohta 10 µl 5x ligeerimispuhvrit, 5 µl 2,5 µM adapterite segu ($C_{lõpp} = 0,25 \mu\text{M}$) ja 5 µl 5 U/µl T4 ligaasi ($C_{lõpp} = 0,5 \text{ U}/\mu\text{l}$). Valmistatud segu pandi 0,2 ml tuubidesse 20 µl iga proovi jaoks ning sellele lisati 30 µl DNA proovi. Tuubid asetati PCR-i masinasse 20°C juurde 15 minutiks inkubeerima.

Seejärel tuli DNA proovid uuesti adapterite lisamise reaktsioonisegust puhastada ning see viidi läbi täpselt samamoodi nagu eelnevalt kirjeldatud proovide puhastamine DNA fragmentide otste täitmise reaktsioonisegust punktis 2.2.4.1.

2.2.4.3 Adapterite otste täitmine

Adapterite otste täitmiseks tuli 1,5 ml tuubi kokku segada reaktsioonisegu, mis sisaldas 13 µl steriilset vett, 5 µl 10x adapteri lünkade täitmise puhvrit ja 2 µl Bst DNA polümeraasi. Reaktsioonisegu jagati jälle 20 µl kaupa 0,2 ml tuubidesse ning sellele lisati 30 µl adapteritega DNA proovi. Tuubid pandi PCR-i masinasse inkubeerima, esmalt 30 minutiks 37°C juurde ning peale seda ensüümi inaktiveerimiseks 20 minutiks 80°C juurde.

2.2.4.4 Indeksite lisamine, DNA fragmentide amplifitseerimine ja reaktsioonisegust puhastamine

Järgneva etapiks tuli valmistada PCR-i segu, milles oli 1 proovi kohta 17 µl steriilset vett, 10 µl 10x PCR puhvrit (Eurogentec), 10 µl 25 mM MgCl₂ (C_{lõpp} = 2,5 mM), 5 µl 20 mg/ml BSA-d (C_{lõpp} = 1 mg/ml), 2 µl 4x10 mM dNTP-sid (iga dNTP C_{lõpp} = 0,2 mM), 2 µl 5 U/µl *HGS Diamond Taq* polümeraasi ja 2 µl 10 µM universaalset praimerit (*NEBNext Multiplex Oligos for Illumina, New England BioLabs*) (C_{lõpp} = 0,2 µM). Valmistatud segu pipeteeriti 0,2 ml tuubidesse (48 µl) ning igasse tuubi lisati 2 µl indekseeritud praimerit (*NEBNext Multiplex Oligos for Illumina, New England BioLabs*), millest igauks oli erinev, et hiljem proove ära tunda. Peale seda lisati igasse tuubi 50 µl DNA proovi ning edasist tööd jätkati TÜ GI tuumiklaboris. Tuubid asetati PCR-i masinasse 5 minutiks eelkuumutamisse 94°C juurde. Edasine programm sisaldas 18 tsüklit, mis koosnes denaturatsioonist (30 sekundit 94°C juures), praimerite seondumisest (30 sekundit 60°C juures) ja DNA sünteesist (30 sekundit 68°C juures). Peale tsükleid hoiti proove veel 7 minutit 72°C juures, et DNA süntees kindlasti lõppeks.

Seejärel tuli DNA proovid PCR-i komponentidest puhastada, mida tehti taaskord samamoodi nagu punktis 2.2.4.1 kirjeldatud.

2.2.5 DNA sekveneerimine

Enne järgnevaid etappe määras teise põlvkonna sekveneerimise spetsialist Anu Solnik fluoromeetriliselt (Qubit, *Thermo Fisher Scientific*) sekveneerimisraamatukogude kontsentratsioonid, paralleel-kapillaarelektroforeesi meetodil (Fragment analyzer, *Agilent Technologies*) nende pikkused ja peale lahendamist kvantitatiivse PCR-iga (qPCR, *Applied Biosystems 7900HT Fast Real-Time PCR System*) täpsed kontsentratsioonid. DNA sekveneerimine viidi läbi teise põlvkonna sekveneerimisega *single-end* meetodil Illumina NextSeq 500 platvormil TÜ GI tuumiklaboris TÜ nooremteaduri Stefania Sasso juhendamisel.

2.2.6 Sekveneritud järjestuse analüüs

Sekveneritud järjestuste analüüs viidi läbi TÜ nooremteaduri Stefania Sasso juhendamisel TÜ Teadusarvutuste keskuse serveris, milles töötati programmi *Putty* (<https://www.putty.org/>) vahendusel. Järjestuste joondamine, *in silico* kvaliteedikontroll, genotüüpide määramine, andmete liitmine võrdluspaneeliga ning põhikomponentanalüüs teostati Stefania Sasso poolt autori osalusel, mtDNA ja Y haplogruppide määramine ning sugulusanalüüs töö autori poolt.

2.2.6.1 Järjestuste ettevalmistamine populatsioonigeneetiliseks analüüsiks

2.2.6.1.1 Järjestuste joondamine

Järjestuste joondamine ja kvaliteedi kontroll viidi läbi TÜ nooremteaduri Stefania Sasso poolt töö autori osalusel. Serverisse sisestatud sekvenerimise väljundfailid kopeeriti **.fastq.gz* formaadis keskaja projekti kausta. Lisaks koostati **.txt* fail, mis sisaldas sekvenerimisel saadud failide asukohti ja proovide ID-sid väljundfailide nimede jaoks. Järjestuste joondamine inimese referentsgenoomi järjestusele koosnes kolmest etapist, kus iga järgnev etapp kasutas eelmise väljundfaile lähteandmetena:

1) *Adapterjärjestuste eemaldamine ja joondamine*. Adapterid eemaldati programmiga *cutadapt* (Martin, 2011). DNA fragmendid, mis olid lühemad kui 28 ap, jäeti edasisest analüüsist välja ning saadud faili **_trimR1.fastq.gz* alusel viidi läbi järjestuste joondamine. Joondamisel kasutati referentsgenoomina inimese genoomi GRCh37 (*Genome Reference Consortium Human Build 37*, hs37d5) järjestust ja selleks kasutati programme *bwa mem* (Li ja Durbin, 2009) ja *samtools* (Li *et al.*, 2009). Tulemusena saadi **.bam* formaadis fail, mis sisaldas ainult inimese genoomile joondunud järjestusi. Illumina NextSeq 500 platvormil sekveneritakse DNA raamatukogud paralleelselt 4 rajal, iga indiviidi jaoks väljastatakse 4 faili, mis joondati eraldi.

2) *Joondatud failide liitmine, duplikaatide eraldamine ja kvaliteedikontroll*. Iga proovi joondatud failid liideti ja programmiga *picard* (<http://broadinstitute.github.io/picard/>) eemaldati PCR-i käigus tekkinud duplikaadid. Seejärel reastati uuesti indelid programmi *GATK* (McKenna *et al.*, 2010) abil. Järjestused, mille *bwa mem* (Li ja Durbin, 2009) programmi poolt antud joondamiskvaliteedi hinnang jäi alla 10, filtreeriti välja kasutades *samtoolsi* (Li *et al.*, 2009).

3) *Põhistatistikuid koondava tabeli koostamine*. Programmi *samtools* ja TÜ GI kaasprofessor Dr. Lauri Saagi loodud skripti abil koostati koondfail sekveneritud raamatukogusid iseloomustavatest statistikutest, mis esitati tabeli kujul. Väljastatud tabelis on informatsiooni

näiteks endogeense DNA sisalduse, mtDNA ja Y-kromosoomi katvuse, fragmentide pikkuse kohta jms (lisa 1.2).

2.2.6.1.2 Järjestuste *in silico* kvaliteedikontroll

MtDNA saastatuse uurimiseks kasutati programmi *contamMix* (Fu *et al.*, 2013), mis kasutab sisendina mtDNA järjestuste *.bam faile. Vastava mtDNA konsensusjärjestus liideti võrdluspaneeliga, mis sisaldas 311 erinevat mtDNA järjestust, ning seejärel joondati järjestused omavahel. Saastatus määrati tulenevalt sellest, kui suur osa mtDNA järjestustest sobitus paremini mõnele teisele võrdluspaneeli mtDNA genoomile kui uuritava proovi konsensusjärjestusele. Tulemusi loeti usaldusväärseks, kui mtDNA katvus oli >5x ja saastatus on väiksem kui 3%.

Genoomis sisalduvate surmajärgsete kahjustuste määramiseks kasutati programmi *mapDamage* (Jónsson *et al.*, 2013), mis tuvastas kui palju on fragmentide otstes toimunud nukleotiidide asendusi (5' otstes C→T, 3' otstes G→A). Tavaliselt on selliseid asendusi enam kui 5%, keskaegses valimis võib neid olla ka pisut vähem.

Tuginedes sellele, kui suur osa mõlemale sugukromosoomile kokku joondunud järjestustest on joondunud Y-kromosoomile, saab kasutades programmi *sexing.py* (Skoglund *et al.*, 2013) teostada soomäärangu. Tulemusi loeti usaldusväärseks, kui genoomi keskmine katvus oli >0,005x. Meessoost indiviidide X-kromosoomi saastatuse uurimiseks kasutati meetodit (Rasmussen *et al.*, 2011) programmis ANGSD (Korneliussen *et al.*, 2014). Tulemusi loeti usaldusväärseks, kui X-kromosoomi katvus oli >0,1x.

Grep käsuga valiti väljundfailidest indiviidide sugu, mtDNA ja meessoost indiviidide X-kromosoomi saastatust ning fragmentide kahjustusi iseloomustavad read ja saadud info salvestati *.txt failidena, mille alusel koostati lisa 1.2.

2.2.6.1.3 Genotüüpide määramine

Andmete edasiseks analüüsiks tuli määrata uuritavate proovide genotüübid. Programmi ANGSD käsuga *doHaploCall* valiti etteantud nukleotiidsete positsioonide põhjal juhuslikkuse alusel sekveneeritud DNA fragmendis olnud nukleotiid, mille tulemusena saadi pseudohaploidne järjestusinfo. Need nukleotiidsed positsioonid pärinesid TÜ GI Eesti Biokeskuse ülemaailmset kaasaegset populatsioonivalimit hõlmavast genotüpiseerimiskiibi andmestikust. Seejärel konverteeriti failid edasistes analüüsides kasutatavasse PLINK formaati kasutades käsku *haplotoPlink*. Järgmisena kasutati programmi PLINK, et muuta mittebinaarsed

failid (*.tfam, *.tped) binaarseteks (*.bed, *.bim, *.fam), mis võtsid andmemahult vähem ruumi ning mida kasutati edasisteks analüüsideks.

TÜ GI Eesti Geenivaramu täisgenoomiandmete informatiivsete positsioonidega viidi eraldi Y-kromosoomi järjestuste jaoks läbi samad etapid, et määrata Y-kromosoomide haplogrupid (vt 2.2.6.2.1) ning teha sugulusanalüüs (vt 2.2.6.2.3).

2.2.6.1.4 Järjestuste liitmine võrdluspaneeli andmetega

Peakomponentanalüüsi tegemiseks (vt 2.2.6.2.2) tuli uuritavate indiviidide andmetele võrdleva analüüsi tegemiseks juurde liita varasemalt avaldatud andmed. Tänapäevased võrdlusandmed pärinesid publikatsioonidest Behar *et al.*, 2010, 2013; Kushniarevich *et al.*, 2015; Li *et al.*, 2008; Raghavan *et al.*, 2014; Tambets *et al.*, 2018; Yunusbayev *et al.*, 2012, 2015 ja vana DNA võrdlusandmed publikatsioonidest Saag *et al.*, 2017, 2019; Mittnik *et al.*, 2018; (lisa 2). Eesti keskaja proovide info pärineb varasemalt teostatud pilootuuringust, mille tulemused on veel avaldamata, kuid mis leidsid kasutamist Kadri Irtdi magistritöös (Irtd, 2020) (lisa 2).

Andmete liitmine toimus PLINK 1.9 (Purcell *et al.*, 2007) programmi abil käsuga *bmerge*. Selle tulemusena saadud *.bed, *.bim, *.fam failid olid aluseks edasistele analüüsidele. Proovid, milles infoga kaetud SNP-de arv oli <10 000, jäeti autosoomide analüüsist välja.

2.2.6.2 Populatsioonigeneetiline analüüs

2.2.6.2.1 MtDNA ja Y-kromosoomi haplogruppide määramine

MtDNA haplogruppide määramiseks kasutati esmalt programmide paketi *samtools* (Li *et al.*, 2009) programmi *BCFtools 1.9*. Käsuga *mpileup* leiti iga indiviidi iga mitokondri positsiooni genotüüpide tõenäosused, käsuga *call* määrati nende alusel genotüübid ja varieeruvad positsioonid, mis salvestati *.vcf formaadis. Haplogrupid määrati saadud faili alusel veebiprogrammi *HaploGrep 2.0* (van Oven ja Kayser, 2009; Weissensteiner *et al.*, 2016) abil. Tulemusi loeti usaldusväärseks, kui mtDNA katvus oli >1x. Haplogrupid kontrolliti üle *PhyloTree* (van Oven ja Kayser, 2009) alusel.

Y-kromosoomi haplogruppide uurimiseks muudeti ANGSD-i väljundfailid PLINK failideks, (vt 2.2.6.1.3). Käsuga *--recode vcf-iid* väljastatati *.vcf tüüpi fail. Programmi *Bedtools 2.19* (Quinlan ja Hall, 2010) kasutades jäeti alles ainult need Y-kromosoomi regioonid, mida on võimalik usaldusväärselt joondada. Tulemusi loeti usaldusväärseks, kui Y-kromosoomi katvus oli >0,005x. Haplogruppide määrangute kontrolli teostas populatsioonigeneetika kaasprofessor Dr. Siiri Rootsi.

2.2.6.2.2 Peakomponentanalüüs

Peakomponentanalüüsi (PCA, ingl *principal component analysis*) tegemiseks kasutati esmalt PLINK 1.9 (Purcell *et al.*, 2007) programmi, kus *keep* käsuga jäeti punkti 2.2.6.1.4 väljundfailidest alles ainult meid huvitavad indiviidid. Fail sisaldas informatsiooni tänapäevaste Euroopa populatsioonide, meie regiooni kivi-, pronksi- ja rauaaegsete ning Eesti alade keskaegsete indiviidide kohta. Kasutatud populatsioonide loetelu on toodud lisa (lisa 2). Peakomponentanalüüsist jäeti kõrvale lähisugulased, sugulaste paarist valiti analüüsi proovid, mille SNP-de arv oli suurem.

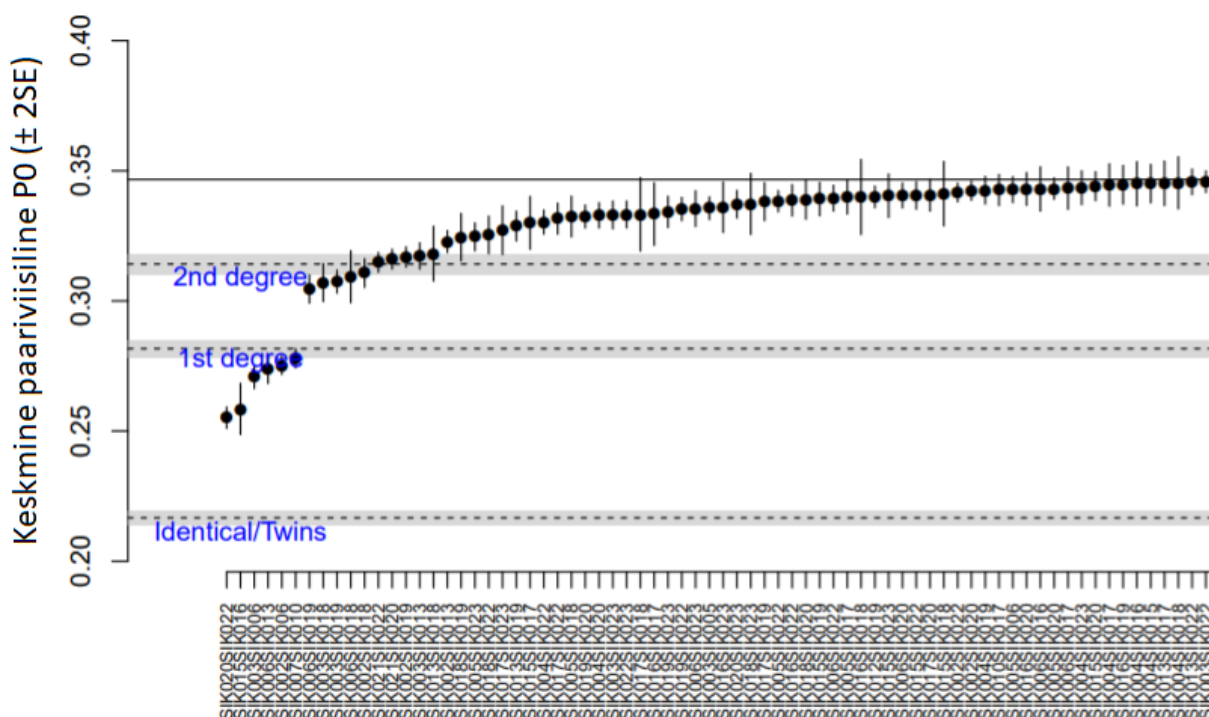
Saadud failid konverteeriti PCA analüüsiks vajalikku *eigenstrat* formaati (*.geno, *.snp, *.ind). Selleks kasutati paketi EIGENSOFT 7.2.0 (Patterson *et al.*, 2006) programmi *convertf* ning PCA viidi läbi sama paketi programmiga *smartpca*.

Graafikute tegemiseks kasutati programmi R 3.6.0 (R Core Team, 2019). Analüüsitulemuste (*eigenvectors*) failist maatriksina vektorisse lugemiseks kasutati käske *as.matrix* ja *read.table* ning graafiku tegemiseks käsku *plot*. Seejärel kuvati *text* käsuga graafikule huvipakkuvad grupid. Vana DNA proovide genotüübiandmed projitseeriti kaasaegsete indiviidide genotüüpide alusel arvutatud peakomponentidele käsuga *lsqproject*. Valminud PCA graafikud väljastas programm *.pdf formaadis.

2.2.6.2.3 Sugulusanalüüs

Sugulussidemete uurimiseks Siksälä kalme indiviidide seas kasutati READ meetodit (Kuhn *et al.*, 2018). Analüüsis kasutati ainult neid proove, mille keskmine genoomne katvus oli >0,03x ning Seega tuli antud uurimistöö 22 indiviidist sugulusanalüüsiks välja jätta 4 indiviidi: SIK008, SIK009, SIK011 ja SIK014 (tabel 1).

Sugulusanalüüsi tegemiseks kasutati eelnevalt punktis 2.2.6.1.3 ANGSD (Korneliussen *et al.*, 2014) ja PLINK (Purcell *et al.*, 2007) programmidega tehtud mittebinaarseid *.tped ja *.tfam faile. Edasise analüüsi alamskriptide käivitamiseks kasutati programme Python 2.7.11 (Van Rossum ja Drake Jr, 1995) ja R 3.1.0 (R Core Team, 2019). READ meetodiga võrreldi, kui sarnased kõik uuritavad indiviidid omavahel keskmiselt olid ja seejärel võrreldi iga indiviidide paari saadud keskmisega. Saadud tulemuste alusel koostas programm graafiku (joonis 6, lisa 3).



Joonis 6. Väljalõige uuritavate Siksälä indiviidide sugulusanalüüsist READ (lisa 3). Pidev joon tähistab mediaanväärtust ning katkendjooned tähistavad identse inimese/kaksikute (ingl *identical/twins*), 1. sugulusastme (ingl *1st degree*) ja 2. sugulusastme (ingl *2nd degree*) lävendeid. Lävendeid ümbritsev hall ala tähistab 95% usaldusvahemikku. P0 – mittekattuvate alleelide suhe indiviidide paari vahel igas 1 miljoni ap suuruses aknas; SE – standardviga (± 2).

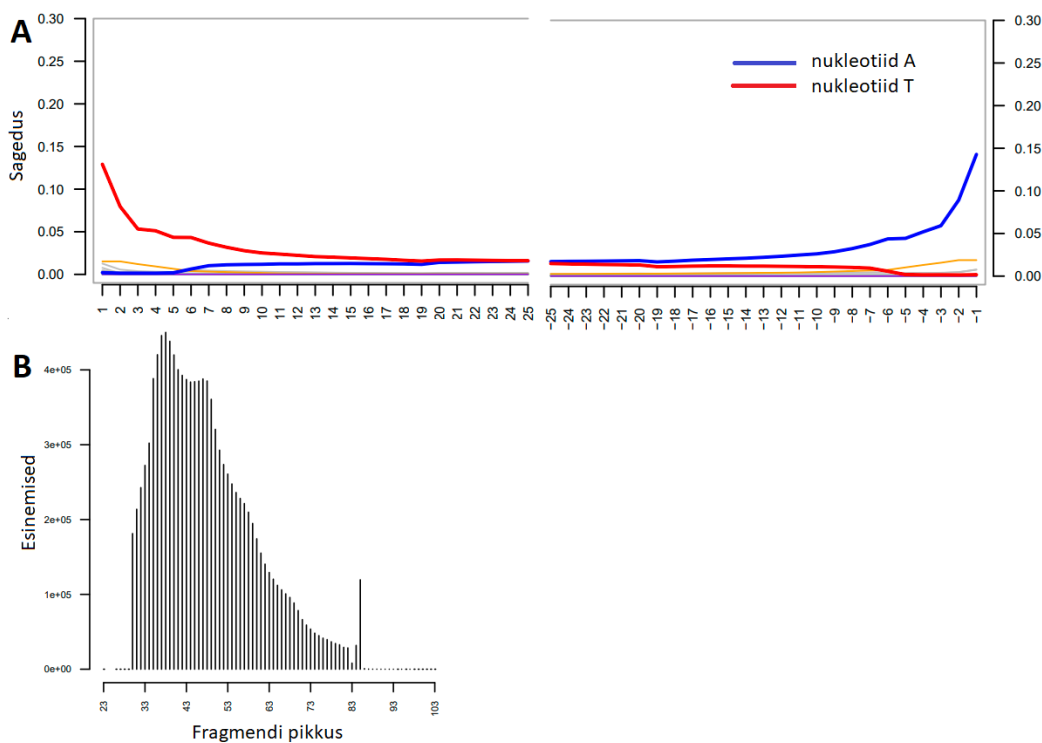
2.3 Tulemused ja arutelu

2.3.1 Uuritavate proovide kvaliteet ja sekveneerimise tulemused

Siksälä kalmesse maetud keskaegsete indiviidide vana DNA proove võib lugeda üldiselt kvaliteetseks, sest 20 proovis 22-st oli DNA säilinud piisavalt hästi, et seda edasistes analüüsides kasutada. Kuna analüüsivad proovid ei sisalda ainult uuritava indiviidi DNA-d, siis tuleb leida endogeense DNA ehk uuritava liigi DNA sisaldus (Poinar *et al.*, 2006). Antud töös analüüsitud proovide endogeense DNA sisaldus jäi vahemikku 1–77% (keskmine 34,6%, mediaanväärtus 31%) (lisa 1.2). 20-st proovist 15 oli endogeenset DNA-d >10%, mida loetakse väga heaks. Lisaks määrati ka sekveneerimisraamatukogu efektiivsus, mis näitab, kui kvaliteetne oli valmistatud raamatukogu sekveneerimiseks peale üleliigsete fragmentide (liiga lühikesed, madala kvaliteediga, genoomile mitte joondunud, duplikaadid) eemaldamist. Efektiivsuse väärtused jäid vahemikku 1–65%, keskmine väärtus oli 26,9% ning mediaanväärtus 23,7% (lisa 1.2). Selle parameetri väärtust loetakse heaks, kui see on >5% ning siinses valimis jäi 3 proovi alla selle lävendi. Tulemus oli oodatav, kuna nende 3 proovi

endogeense DNA sisaldus oli madal. Ülejäänud 2 proovi, olid endogeense DNA sisalduse puhul piiripealsed (7% ja 9%), mis selgitab asjaolu, et efektiivsust määrares loeti nende väärtused (vastavalt 6% ja 6,5%) heaks (lisa 1.2). Minimaalne keskmine fragmendi pikkus oli 42 ap, maksimaalne 64 ap ning üldine mediaanväärtus 53 ap, mille järgi võib öelda, et uuritud proovide fragmentide pikkused vastavad üldistele vana DNA fragmentide pikkustele (joonis 7). Ülegenoomsed katvused jäi vahemikku 0,0025–0,32 (keskmine 0,1169, mediaanväärtus 0,0995) (lisa 1.2). Nende väärtuste puhul vaadati sugulusanalüüsiks sobivat lävendit, milleks on $>0,03x$. Lävendit ei ületanud 4 proovi, mis kattusid varem välja toodud proovidega, mille tulemused olid halvad.

Vana DNA autentimiseks on vajalik vaadata DNA fragmentide otstes asuvate C→T asenduste hulka. Kõik peale ühe proovi, mille tulemus oli piiripealne (4%), olid üle 5%, mis on C→T asenduste hulga usaldusväärsuse lävendiks. Suurimaks leitud asenduste hulgaks oli 16%, kuid rohkem kui pooltel proovidel oli selle väärtus $>10\%$. Keskajast pärinevate proovide puhul on suhteliselt madal asenduste osakaal aktsepteeritav, kuna võrreldes nt pronksi- või rauaaja proovidega on need üldiselt palju paremini säilinud. Samuti on vaadates antud töö proovide dateeringuid näha trendi, kus proovid, mille C→T asenduste hulk oli $<10\%$, on nooremad (14.–15. sajandist), kui need, mille väärtus oli $>10\%$ (13. sajandist). (Lisa 1.2, joonis 7)



Joonis 7. Vana DNA proovide autentsuse määramise analüüs SIK020 näitel. (A) C→T asenduste hulk. (B) Fragmentide pikkuse (ap) jaotumine.

2.3.2 Soo määramine

Uuritud 20-st proovist olid 6 geneetiliselt mehed ja 12 geneetiliselt naised. Kahe indiviidi puhul luges programm nad kooskõlas olevaks mehega, mitte naisega, kuid ei andnud kindlat väljundit. See oli oodatav tulemus, kuna nende proovide kvaliteet oli madal ning ka arheoloogide poolt oli nende sugu määratlemata. Kõigi teiste indiviidide andmed peale ühe kattusid arheoloogide poolt morfoloogiliselt määratud andmetega. Ainult indiviid SIK009 (kellest oli säilinud vaid kolju) oli morfoloogiliselt määratud naiseks, kuid geneetiliselt saadi tulemuseks mees. Kinnituse leidis 3 morfoloogilise soomääratluse osas küsimärgi all olnud indiviidi, kelle puhul oli tegemist lastega (arvatavalt naissoost), mille tõttu on nende soo morfoloogiline määramine raskem. (Tabel 2)

Tabel 2. Uuritavate indiviidide morfoloogiline sugu (morfol. sugu), geneetiline sugu (gen. sugu), mtDNA haplogrupp (MT hg) ja Y-kromosoomi haplogrupp (Y hg). N/A - pole saadaval.

#	Indiviid	Matuse number	Morfol. sugu	Gen. sugu	MT hg	Y hg
1	SIK004	17	Naine	XX	J1c5a	-
2	SIK005	47	Mees	XY	H1b2	N3a3'5
3	SIK006	83	Mees	XY	H1b2	N3a3'5
4	SIK007	108	Mees	XY	U2e1f1	N3a3'5
5	SIK008	110A	Naine	XX	U4d	-
6	SIK009	110B	Naine	XY	U3a2	N/A
7	SIK010	111	Naine	XX	U2e1f1	-
8	SIK011	113	?	kooskõlas XY, aga mitte XX	U2e1f1	N/A
9	SIK012	131	Naine	XX	H13a1a1c	-
10	SIK013	135	Mees	XY	J2a1a1a3	N3a
11	SIK014	157	?	kooskõlas XY, aga mitte XX	U	N/A
12	SIK015	173	Laps (N)	XX	U8b1b	-
13	SIK016	177	Naine	XX	U8b1b	-
14	SIK017	187	Naine	XX	H1b1	-
15	SIK018	200	Naine	XX	J2a1a1a3	-
16	SIK019	246	Mees	XY	H5a1a	N3a3
17	SIK020	245A	Laps (N)	XX	U5a1a1h	-
18	SIK021	245C	Naine	XX	U5a1a1h	-
19	SIK022	245D	Laps (N)	XX	U5a1a1h	-
20	SIK023	247B	Naine	XX	I1a1c	-

2.3.2 Populatsioonigeneetilised analüüsid

2.3.2.1 MtDNA ja Y-kromosoomi haplogrupid

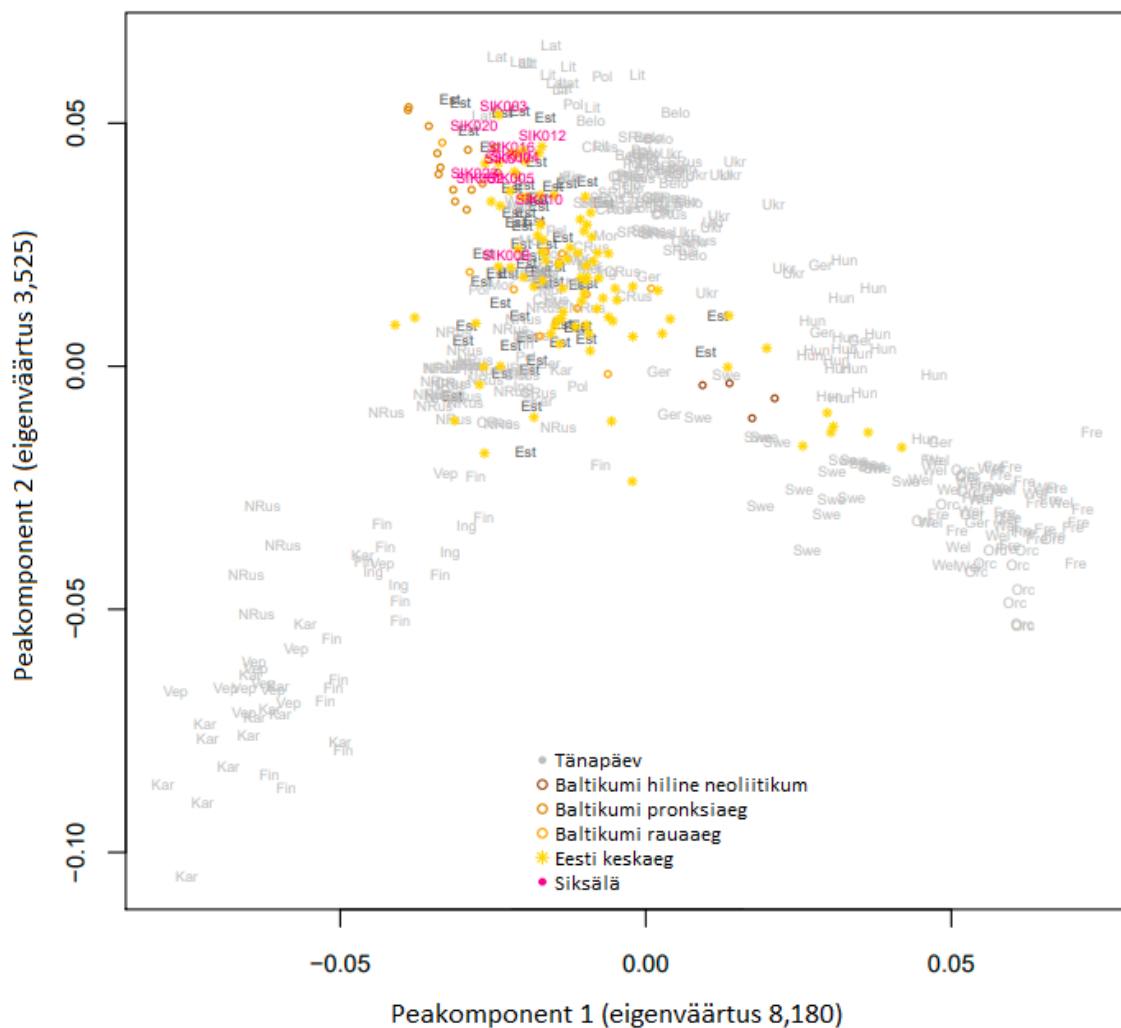
MtDNA haplogrupid õnnestus määrata kõikidel indiviididel. Tulemused näitavad, et emaliinide varieeruvus oli suhteliselt suur – täpselt pooled neist kuulusid haplogrupi U erinevatesse alamharudesse ning lisaks sellele määrati 5 indiviidi haplogruppi H, 3 haplogruppi J ning 1 haplogruppi I (tabel 2). Kõik leitud haplogrupid on väga iseloomulikud Euroopa populatsioonidele (Richards *et al.*, 1998). MtDNA-d kattuvad ainult sellistel indiviididel, kus oli tegemist vanaema/ema-laps või õdede sugulussuhtega.

Y-kromosoomi haplogrupp õnnestus määrata 5 mehel 8-st ning erinevalt uuritud valimi mitmekesistest emaliinidest kuulusid kõik need ühte ja samasse suuremasse haplogruppi nimega N3a (tabel 2). Haplogrupp N3a on Eesti aladel ja ka mujal Ida- ning Kirde-Euroopas väga levinud liin, seda just eelkõige uurali keelt kõnelevate rahvaste seas (Ilumäe *et al.*, 2016).

2.3.2.2 Peakomponentanalüüs

Peakomponentanalüüsi tegi TÜ noorenteadur Stefania Sasso autori osalusel. Analüüsis võrreldi Siksälä indiviide tänapäevaste Euroopa populatsioonidega, Baltikumi hilise neoliitikumi, pronksi- ja rauaaja indiviididega ning varasemalt vana DNA grupi poolt uuritud Eesti ja Läti alade keskaegsete indiviididega (joonis 8, lisa 2).

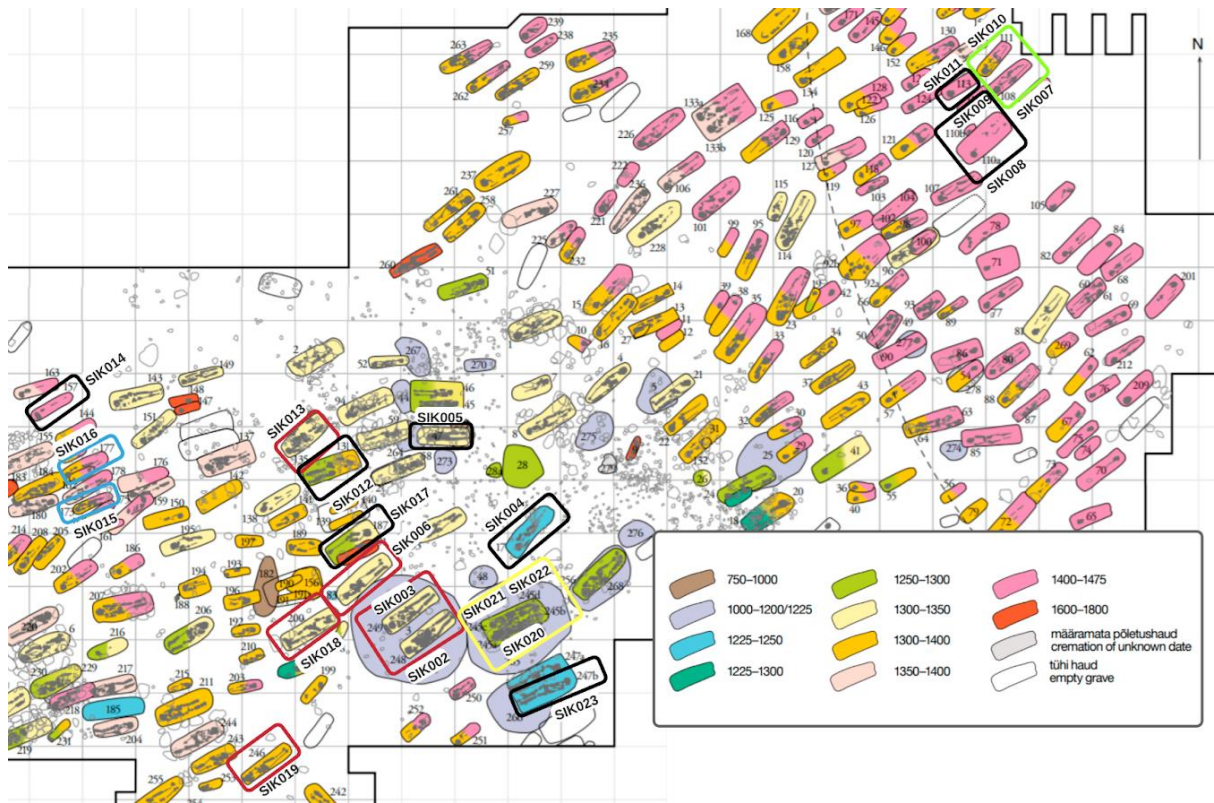
Siksälä proovid koonduvad graafikul suuresti ühte gruppi, millel on selge ühisosa keskaegsete indiviididega, kes on keskaja projekti raames tehtud pilootuuringu kohaselt valdavalt pärit Kagu-Eestist ja Lätist. Teiste keskaegsete proovidega võrreldes on Siksälä indiviidid nihkes balti ja ka slaavi rahvaste poole, mis on oodatav, kuna tegu on Kagu-Eestiga ning ka arheoloogiline taustainfo on Siksälä kalmesse maetud indiviide käsitletud lähedasena Ida-Läti elanikkonnale.



Joonis 8. Peakomponentanalüüs Siksälä indiviidide (n=11) geneetilisest varieeruvusest regionaalsel taustal. Joonisel esinevad lühendid on välja toodud lisa (lisa 2), hallid sümbolid kujutavad kaasaegseid ja värvilised sümbolid erinevate varasematel ajaperioodidel elanud indiviide. Eigenvärtused näitavad peakomponendi hajuvust, milleks võetakse võimalikult suur väärtus, mis kirjeldaks ka võimalikult suurt osa alg tunnuste varieeruvusest.

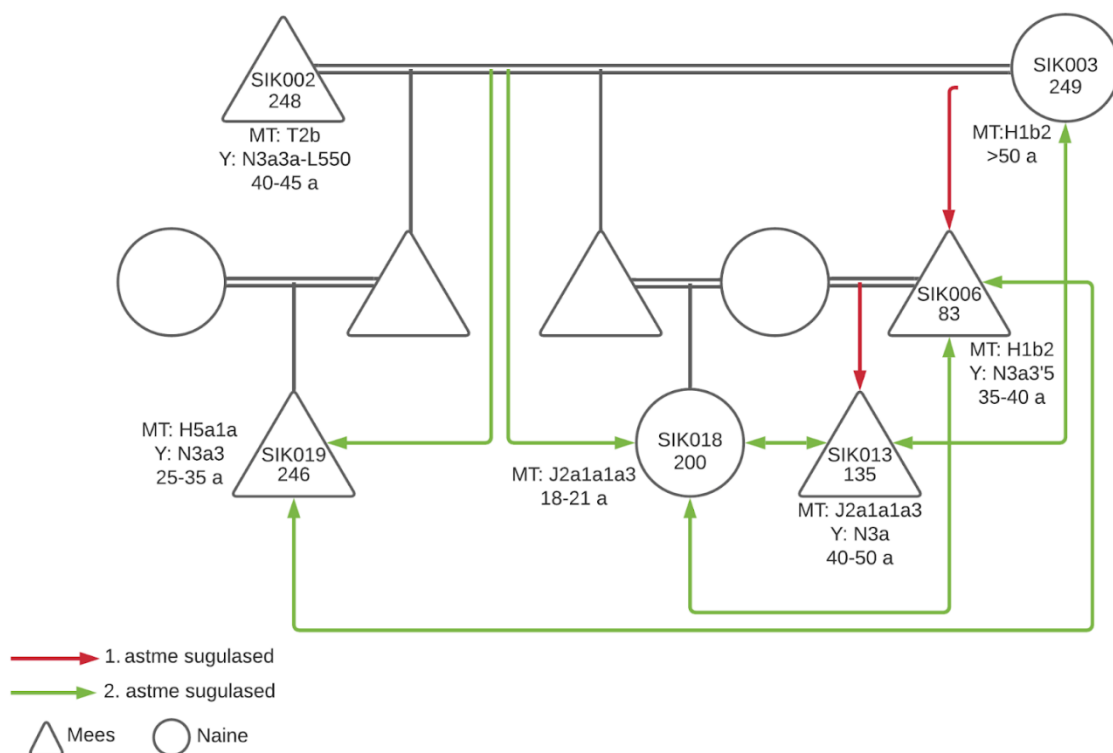
2.3.2.3 Sugulusanalüüs

Sugulusanalüüsis vaadati, kas uuritud Siksälä indiviidide seas leidub esimese ja teise astme sugulasi või identseid indiviide. Indiviidide vaheliste võimalike seoste paremaks mõistmiseks on toodud ka Siksälä kalme kaart (joonis 9).



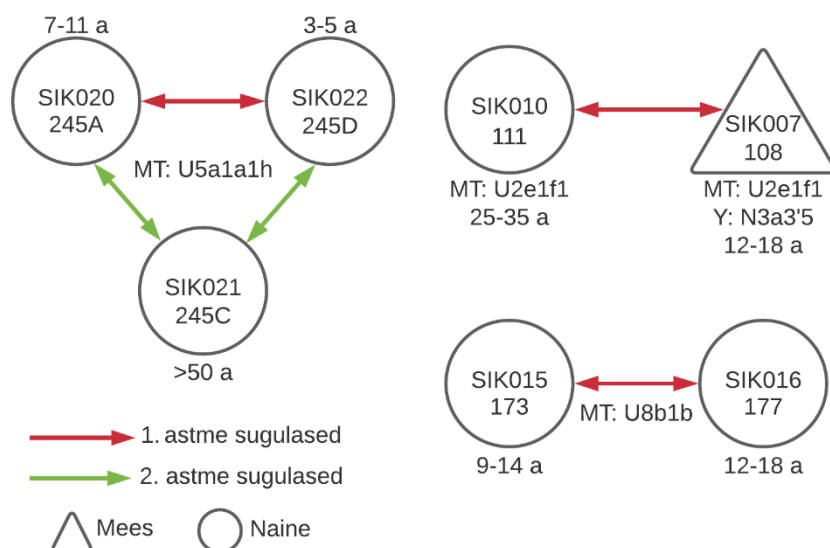
Joonis 9. Siksälä kalme kaart. Mustade kastidega märgitud individid kellel ei leitud sugulussidemeid, värvilistega märgitud erinevad leitud sugupuud ja -grupid. Legendis on toodud matuste dateeringud. (Kohandatud: Valk ja Laul, 2014)

Sugulussidemeid Siksälä individide vahel leiti mitmeid: 6 paari 1. astme sugulasi ning 10 paari 2. astme sugulasi. Leitud sugulussidemete seast õnnestus kokku panna üks suurem, kolme põlvkonda hõlmava perekonna sugupuud (joonis 10) ning kolm väiksemat gruppi (joonis 11). Suurim perekond saab alguse ülikupaarist SIK002 ja SIK003 (vt ka joonis 5) ning individist SIK006, kes on üks nende vähemalt kolmest pojast. SIK013 on 1. astme suguluses SIK006-ga ning 2. astme suguluses SIK002-ga ja SIK003-ga ehk SIK013 on SIK006 poeg. Indiviid SIK018 on 2. astme suguluses SIK002-ga, SIK003-ga (tema vanavanemad), SIK006-ga (onu) ja üllatuslikult ka SIK013-ga, mis tähendab seda, et SIK018 on ülikupaari teise poja tütar. Kuna tal on SIK013-ga sama mtDNA haplogrupp, peab see tähendama, et neil oli sama ema ning nad on poolõde ja -vend ning ühtlasi onude lapsed. Ülikupaaril leiti 2. astme sugulust ka individiga SIK019, kellel oli 2. astme sugulust ka SIK006-ga (onu), millest järeldati, et ülikupaaril on olnud ka kolmas poeg, kelle laps on SIK019. Kõigi selle perekonna meeste Y-kromosoomi haplogrupp kuulus haplogruppi N3a. (Joonis 10)



Joonis 10. Sugulusanalüüsi tulemusena rekonstrueeritud sugupuu 6 analüüsitud Siksälä indiviidi kohta. Indiviidide SIK002 ja SIK003 perekond. Kolmnurgaga on tähistatud mehi, ringiga naisi ning topeltjoonega on tähistatud paarisuhe ja ühekordse joonega antud paari last. Joonisele on lisatud leitud 1. (punane nool) ja 2. (roheline nool) astme sugulussidemed, indiviidide vanused ning mtDNA (MT) ja Y-kromosoomi (Y) haplogrupid. Tähistuseta sümbolid kujutavad sugupuu rekonstrueeritud liikmeid.

Lisaks sellele perekonnale leiti ka teisi sugulussidemeid. Indiviidid SIK020 ja SIK022 on 1. astme sugulased ning kuna neil on sama mtDNA haplogrupp ja nad on vanuselt lapsed, siis on nad tõenäoliselt õed. Neil mõlemal oli ka 2. astme sugulus ja sama mtDNA haplogrupp indiviidiga SIK021. SIK021 kõrge vanuse tõttu võib järeldada, et ta võis olla kas laste vanaema või tädi. 1. astme sugulased on ka SIK010 ja SIK007 ning vanuste ja mtDNA haplogrupi järgi järeldati, et tegemist on ema ja pojaga. Samuti leiti 1. astme sugulus indiviidide SIK015 ja SIK016 vahel ning vanuste ja ühise mtDNA haplogrupi järgi leiti, et nad on õed. (Joonis 11) Pole välistatud, et ka SIK007, kes kannab samasugust Y-kromosoomi haplogruppi kui kõik teised sugulusanalüüsis uuritud meessoost indiviidid, on joonisel 10 kujutatud suurema sugupuu liige, kuid olemasoleva valimiga seda täpsemalt määrata ei saa.



Joonis 11. Sugulusanalüüsi tulemusena rekonstrueeritud sugulussuhted kolmes väiksemas Siksälä indiviidide grupis. Kolmnurgaga on tähistatud mehed ja ringiga on tähistatud naised. Joonisele on lisatud leitud 1. (punane nool) ja 2. (roheline nool) astme sugulussidemed, indiviidide vanused ning mtDNA (MT) ja Y-kromosoomi (Y) haplogrupid.

Juba suhteliselt väikese valimi alusel Siksälä Kalmetemäelt võib oletada, et eelkõige oli tegemist meesliini kaudu suguluses olevate indiviididega. Täiskasvanud naiste vahelisi sugulussuhteid siinses töös ei leitud, ja ehkki kaugeleulatuvaid järeldusi nii väheste indiviidide põhjal teha ei saa, tunduvad saadud tulemused näitavat patrilokaalset ühiskonnakorraldust, kus mehed võtsid naised väljastpoolt kogukonda.

KOKKUVÕTE

Töö eesmärgiks oli iseloomustada Siksälä kalme näitel Eesti keskaegse elanikkonna geneetilist struktuuri ja sotsiaalseid sidemeid. Selleks eraldati DNA 20 indiviidilt ning analüüsideks lisati töösse ka kaks varasemalt sekveneeritud indiviidi, kes pärinesid samast kalmest.

Uuritavast 22-st proovist olid järgnevateks analüüsideks piisavalt kvaliteetsed 20 proovi, seega oli DNA hästi säilinud. Sugu õnnestus määrata kõikidel indiviididel ning soomäärangud kattusid vaid ühe erandiga morfoloogiliste andmetega. Uuritud indiviidide emaliinid olid mitmekesised, leiti mitmeid erinevaid haplogrupi U alamharusid ning lisaks esines ka haplogruppe H, J ja I. Erinevalt emaliinidest kuulusid kõik uuritud meesindiviidid samasse isaliini haplogruppi nimega N3a. Kõik leitud haplogrupid on iseloomulikud ka tänapäevasele Eesti geenipärandile.

Peakomponentanalüüsi abil uuriti Siksälä indiviidide geneetilist tausta võrreldes neid nii kaasaegsete populatsioonidega kui ka varasematel ajaperioodidel Eesti aladel elanud indiviididega. Suurimat sarnasust nähti oodatavalt teiste Kagu-Eesti ja Läti keskaegsete indiviididega ja tänaste eestlastega, aga ka balti ja slaavi rahvastega. Sugulusanalüüsiga leiti 6 paari 1. astme sugulasi ja 10 paari 2. astme sugulasi, mille kaudu rekonstrueeriti üks suurem kolme põlvkonnaga sugupuu ja kolm väiksemat gruppi. Kõigi tulemuste põhjal võis järeldada, et Siksälä kalmesse maetud indiviidide ühiskonnakorraldust iseloomustas patrilokaalsus.

Bakalaureusetöö oli osa TÜ GI suuremast projektist nimega „Keskaegse populatsiooni struktuur, tervis ja haigused Eesti territooriumil vana DNA vaatenurgast” (PRG1027).

Genetic kinship analysis of medieval Estonian archeological material using the example of Siksälä burial site

Stella Eileen Päid

Summary

The study of the human past is based on history, archaeology, and genetics among other disciplines. Our knowledge of history is formed on written text as well as oral stories and folklore. Archaeologists study tools, ornaments, and pottery, but also the remains of soil, landscapes, and buildings, from which they make assumptions about the life and ethnicity of populations. Genetics add specific details to help verify and reconcile the evidence gathered by historians and archaeologists. This thesis is an example of how transdisciplinary approach helps to get a more comprehensive picture of the lives of our ancestors.

Siksälä burial site is in the southeast corner of Estonia and is considered exceptional for its large number of grave goods and other finds. Anthropological and archaeological material suggest the addition of new inhabitants from eastern Latvia and Livonian territories as well.

The aim of this thesis was to describe the genetic variability and social connections of the medieval Estonian population using the example of Siksälä burial site. For that, DNA was extracted from 20 individuals who were buried at Siksälä in the Middle Ages. Two other individuals, previously sequenced in a pilot study, were added to the analyses. Out of the 22 samples, 20 had the quality necessary for the following analyses. Genetically, eight of them were male and 12 were female. The data of every individual but one overlapped with the data that had been morphologically determined by archaeologists. All found haplogroups are characteristic of modern Estonian genetic heritage. Unlike the diverse maternal lineages, where different U, H, I and J haplogroups were found, all male subjects studied belonged to the same paternal lineage having the N3a haplogroup.

The principal component analysis revealed that the studied individuals were most similar to other medieval individuals from South-East Estonia and Latvia as well as to modern Estonians and the Baltic and Slavic populations. Kinship analysis identified 6 pairs of first-degree relatives and 10 pairs of second-degree relatives, through which one larger three-generation family tree and three smaller groups were reconstructed. Based on all of the results, it could be concluded that the social organisation of the individuals buried at Siksälä burial site was characterised by patrilocality, which means that men married women from outside of the community and women moved to their husband's places of residence.

This bachelor thesis is part of the University of Tartu Institute of Genomics project called “Medieval population structure, health and diseases in Estonian territory from the perspective of ancient DNA” (PRG1027).

TÄNUSÕNAD

Esiteks tahan tänada oma juhendajaid Kristiina Tambetsit ja Stefania Sassot, kes mõlemad pakkusid suurt tuge kogu selle protsessi ajal. Tänu Kristiinale oli mul võimalus esimese bakalaureusetudengina töötada vana DNA laboris ning samuti oli ta suurimaks abiks kirjutamisel. Tänu Stefaniale sain tehtud kõik vajalikud analüüsid ja sain ka laboritööst väga tore kogemuse. Lisaks tänan ka Eesti Biokeskuse laboranti Helja Niinemäed, kes mind esimesena vana DNA laboris õpetas ja andis suurepärase aluse Stefaniaga edasi töötamiseks. Samuti soovin tänada prof. Heiki Valku ja Martin Malvet, kes usaldasid minu kätte Siksälä proovid. Tänan Heiki Valku eraldi ka selle eest, et ta mu ajaloo ja arheoloogia osa üle vaatas. Tänan ka TÜ arheoloogia, geneetika ja lingvistika sidusuuringute kolleegiumi ja selle liikmeid.

KASUTATUD KIRJANDUSE LOETELU

Allentoft, M. E., Collins, M., Harker, D., ... Bunce, M. (2012). The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proceedings. Biological sciences*, 279(1748): 4724–4733.

Alterauge, A., Lösch, S., Sulzer, A., Gysi, M. ja Haas, C. (2021). Beyond simple kinship and identification: aDNA analyses from a 17th-19th century crypt in Germany. *Forensic Science International. Genetics*. 53: 102498.

Amorim, C. E. G., Vai, S., Posth, C., ... Veeramah, K. R. (2018). Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nature Communications*. 9(1): 3547.

Barnes, I., Matheus, P., Shapiro, B., Jensen, D. ja Cooper, A. (2002). Dynamics of Pleistocene population extinctions in Beringian brown bears. *Science*. 295(5563): 2267–70.

Behar, D. M., Metspalu, M., Baran, ... Rosenberg, N. A. (2013). No evidence from genome-wide data of a Khazar origin for the Ashkenazi Jews. *Human Biology*. 85(6): 859–900.

Behar, D. M., Yunusbayev, B., Metspalu, M., ... Villems, R. (2010). The genome-wide structure of the Jewish people. *Nature*. 466(7303): 238–242.

Benazzi, S., Douka, K., Fornai, C., ... Weber, G. W. (2011). Early dispersal of modern humans in Europe and implications for Neanderthal behaviour. *Nature*. 479(7374): 525–528.

Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L. ja Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*. 24(4): 335–341.

Briggs, A. W., Stenzel, U., Johnson, P. L. F., ... Pääbo, S. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*. 104(37): 14616–14621.

Colonna, V., Pagani, L., Xue, Y. ja Tyler-Smith, C. (2011). A world in a grain of sand: human history from genetic data. *Genome Biology*. 12(11): 234.

Cooper, A., Rhymer, J.D., James, H.F., Olson, S.L., McIntosh, C.E., Sorenson, M.D. ja Fleischer, R.C. (1996). Ancient DNA and island endemics. *Nature*. 381: 484.

- Dabney, J., Meyer, M. ja Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harbor perspectives in biology*. 5(7): a012567.
- Douglas Price, T. 2000. *Europe's First Farmers*. Cambridge University Press.
- Fu, Q., Mittnik, A., Johnson, P. L. F., ... Krause, J. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology: CB*. 23(7): 553–559.
- Gallego Llorente, M., Jones, E. R., Eriksson, A., ... Manica, A. (2015). Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science*. 351(6275): aaf3945.
- Goodwin, S., McPherson, J. D. ja McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews. Genetics* 17(6): 333–351.
- Haak, W., Brandt, G., de Jong, H. N., Meyer, C., Ganslmeier, R., Heyd, V., Hawkesworth, C., Pike, A. W. G., Meller, H. ja Alt, K. W. (2008). Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. *Proceedings of the National Academy of Sciences of the United States of America*. 105(47): 18226–18231.
- Haak, W., Lazaridis, I., Patterson, N., ... Reich, D. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 522(7555): 207–211.
- Herrmann, B. ja Hummel, S. 1994. *Ancient DNA: Recovery and Analysis of Genetic Material from Paleontological, Archaeological, Museum, Medical, and Forensic Specimens*, lk 2, 60. R. C. Garber (ed.), Springer-Verlag, New York.
- Heapost, L. (2007). The cemetery of Siksälä: osteological and paleodemographical analysis. – *Siksälä. A Community at the Frontiers. Iron Age and Medieval*. Tartu Ülikool. 227–234.
- Higgins, D. ja Austin, J. J. (2013). Teeth as a source of DNA for forensic identification of human remains: A Review. *Science and Justice*. 53(4): 433.
- Ho, S. Y. W., Heupink, T. H., Rambaut, A. ja Shapiro, B. (2007). Bayesian Estimation of Sequence Damage in Ancient DNA. *Molecular Biology and Evolution*. 24 (6): 1416–1422.
- Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A. ja Pääbo, S. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic acids research*. 29(23): 4793–4799.

- Illumäe, A.-M., Reidla, M., Chukhryaeva, M., ... Rootsi, S. (2016). Human Y Chromosome Haplogroup N: A Non-trivial Time-Resolved Phylogeography that Cuts across Language Families. *American Journal of Human Genetics*. 99(1): 163–173.
- Irtd, K. (2020). Eesti keskaegse populatsiooni struktuur vana DNA vaatevinklist. Tartu: Tartu Ülikool.
- Jaenicke-Despres, V., Buckler, E. S., Smith, B. D., Gilbert, M. T. P., Cooper, A., Doebley, J. ja Pääbo, S. (2003). Early Allelic Selection in Maize as Revealed by Ancient DNA. *Science* 302(5648): 1206-8.
- Jobling, M., Hollox, E., Hurles, M., Kivisild, T. ja Tyler-Smith, C. 2014. *Human Evolutionary Genetics* (2nd ed.). New York and London: Garland Science, Taylor & Francis Group, LLC.
- Jones, E. R., Zarina, G., Moiseyev, V., Lightfoot, E., Nigst, P. R., Manica, A., Pinhasi, R. ja Bradley, D. G. (2017). The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Current Biology: CB*. 27(4): 576–582.
- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. ja Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. 29(13): 1682–1684.
- Kivisild, T. (2017). The study of human Y chromosome variation through ancient DNA. *Human Genetics*. 136(5): 529–546.
- Korneliussen, T. S., Albrechtsen, A. ja Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*. 15: 356.
- Kriiska, A. (2000). Settlements of coastal Estonia and maritime hunter-gatherer economy. *Lietuvos Archeologija*. 19: 153–166.
- Kriiska, A. (2009). The beginning of farming in the Eastern Baltic area. *The East European Plain on the Eve of Agriculture. Brit Archaeol Rep In*. 159–179.
- Kriiska, A., Lang, V., Mäesalu, A., Tvauri, A. ja Valk, H. 2020. Eesti ajalugu I. Tartu Ülikooli ajaloo ja arheoloogia instituut.
- Kriiska, A., Lang, V., Valk, H., Tambets, K., Saag, L. ja Metspalu, M. (2018). Vana DNA – uus sõnumitooja Eesti rahvastiku ajaloo kohta. *Tutulus: Eesti arheoloogia aastakiri, 2018*. 5–8.

- Kushniarevich, A., Utevska, O., Chuhryaeva, M., ... Balanovsky, O. (2015). Genetic Heritage of the Balto-Slavic Speaking Populations: A Synthesis of Autosomal, Mitochondrial and Y-Chromosomal Data. *PLoS One*. 10(9): e0135820.
- Li, H. ja Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 25(14): 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. ja 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 25(16): 2078–2079.
- Li, J. Z., Absher, D. M., Tang, H., ... Myers, R. M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*. 319(5866): 1100–1104.
- Margaryan, A., Hansen, H. B., Rasmussen, S., ... Allentoft, M. E. (2018). Ancient pathogen DNA in human teeth and petrous bones. *Ecology and Evolution*. 8(6): 3534–3542
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 17(1): 10–12.
- McKenna, A., Hanna, M., Banks, E., ... DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 20(9): 1297–1303.
- Merheb, M., Matar, R., Hodeify, R., Siddiqui, S. S., Vazhappilly, C. G., Marton, J., Azharuddin, S. ja Al Zouabi, H. (2019). Mitochondrial DNA, a Powerful Tool to Decipher Ancient Human Civilization from Domestication to Music, and to Uncover Historical Murder Cases. *Cells*. 8(5).
- Meyer, M., Fu, Q., Aximu-Petri, A., ... Pääbo, S. (2014). A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature*. 505: 403–406.
- Miller, W., Drautz, D. I., Ratan, A., ... Schuster, S. C. (2008). Sequencing the nuclear genome of the extinct woolly mammoth. *Nature*. 456(7220): 387–390.
- Mittnik, A., Massy, K., Knipper, C., ... Krause, J. (2019). Kinship-based social inequality in Bronze Age Europe. *Science*. 366(6466): 731–734.
- Mittnik, A., Wang, C.-C., Pfrengle, S., ... Krause, J. (2018). The genetic prehistory of the Baltic Sea region. *Nature Communications*. 9(1): 442.

- Monroy Kuhn, J. M., Jakobsson, M. ja Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. *PloS One*. 13(4): e0195491.
- Orlando, L., Ginolhac, A., Zhang, G., ... Willerslev, E. (2013) Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*. 499: 74–78.
- O’Sullivan, N., Posth, C., Coia, V., Schuenemann, V. J., Price, T. D., Wahl, J., Pinhasi, R., Zink, A., Krause, J. ja Maixner, F. (2018). Ancient genome-wide analyses infer kinship structure in an Early Medieval Alemannic graveyard. *Science Advances*. 4(9): eaao1262.
- Palli, H. 1996. Eesti rahvastiku ajalugu aastani 1712. *Academia* 6. Tallinn: Teaduste Akadeemia Kirjastus.
- Patterson, N., Price, A. L. ja Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*. 2(12): e190.
- Poetsch, M., Wiegand, A., Harder, M., Blöhm, R., Rakotomavo, N., Freitag-Wolf, S. ja von Wurmb-Schwark, N. (2013). Determination of population origin: a comparison of autosomal SNPs, Y-chromosomal and mtDNA haplogroups using a Malagasy population as example. *European Journal of Human Genetics: EJHG*. 21(12): 1423–1428.
- Poinar, H. N., Höss, M., Bada, J. L. ja Pääbo, S. (1996). Amino acid racemization and the preservation of ancient DNA. *Science*. 272(5263): 864–866.
- Poinar, H. N., Schwarz, C., Qi, J., ... Schuster, C. (2006). Metagenomics to Paleogenomics: Large-Scale Sequencing of Mammoth DNA. *Science* 311(5759): 392–4.
- Purcell, S., Neale, B., Todd-Brown, K., ... Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*. 81(3): 559–575.
- Quinlan, A. R. ja Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26(6): 841–842.
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.r-project.org>
- Raghavan, M., Skoglund, P., Graf, K. E., ... Willerslev, E. (2014). Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*. 505(7481): 87–91.
- Rasmussen, M., Guo, X., Wang, Y., ... Willerslev, E. (2011). An Aboriginal Australian genome reveals separate human dispersals into Asia. *Science*. 334(6052): 94–98.

- Richards, M. B., Macaulay, V. A., Bandelt, H. J. ja Sykes, B. C. (1998). Phylogeography of mitochondrial DNA in western Europe. *Annals of Human Genetics*. 62(Pt 3): 241–260.
- Rizzi, E., Lari, M., Gigli, E., De Bellis, G. ja Caramelli, D. (2012). Ancient DNA studies: new perspectives on old samples. *Genetics, Selection, Evolution: GSE*. 44:21.
- Saag, L., Laneman, M., Varul, L., ... Tambets, K. (2019). The Arrival of Siberian Ancestry Connecting the Eastern Baltic to Uralic Speakers further East. *Current Biology: CB*. 29(10): 1701–1711.e16.
- Saag, L., Varul, L., Scheib, C. L., ... Metspalu, M. (2017). Extensive Farming in Estonia Started through a Sex-Biased Migration from the Steppe. *Current Biology: CB*. 27(14): 2185–2193.e6.
- Sekijima, J. (2016). Get to know your teeth: tooth structure and support. *Filling you in blog*.
- Selart, A., Kala, T., Kaljundi, L., ... Valk, H. 2012. *Eesti ajalugu II, Eesti keskaeg*. Tartu: Tartu Ülikooli ajaloo ja arheoloogia instituut.
- Skoglund, P., Storå, J., Götherström, A. ja Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science*. 40(12): 4477–4482.
- Smith, C. I., Chamberlain, A. T., Riley, M. S., Stringer, C. ja Collins, M. J. (2003). The thermal history of human fossils and the likelihood of successful DNA amplification. *Journal of Human Evolution*. 45(3): 203–217.
- Tambets, K., Yunusbayev, B., Hudjashov, G., ... Metspalu, M. (2018). Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biology*. 19(1): 139.
- Valk, H. ja Laul, S. 2014. *Siksälä kalme I: Muistis ja ajalugu*, p. 19–137. Valk, M., Veldi, E., Valk, H. (ed.), Tartu: Tartu Ülikooli ajaloo ja arheoloogia instituut.
- van der Valk, T., Pečnerová, P., Díez-del-Molino, D., ... Dalen, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. *Nature*. 591: 265–269.
- van Oven, M. ja Kayser, M. (2009). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*. 30(2): E386–E394.
- van Rossum, G. ja Drake Jr, F. L. 1995. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Vernot, B., Zavala, E. I., Gómez-Olivencia, A., ... Meyer, M. (2021). Unearthing Neanderthal population history using nuclear and mitochondrial DNA from cave sediments. *Science*.

Vicente, M. ja Schlebusch, C. M. (2020). African population history: an ancient DNA perspective. *Current Opinion in Genetics & Development*. 62: 8–15.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A. ja Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Research*. 44(W1): W58–W63.

Xu, W., Lin, Y., Zhao, K., ... Zhou, X. (2020). An efficient pipeline for ancient DNA mapping and recovery of endogenous ancient DNA from whole-genome sequencing data. *Ecology and Evolution*. 11(1): 390–401.

Yunusbayev, B., Metspalu, M., Järve, M., ... Villems, R. (2012). The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. *Molecular Biology and Evolution*. 29(1): 359–365.

Yunusbayev, B., Metspalu, M., Metspalu, E., ... Villems, R. (2015). The genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genetics*. 11(4): e1005068.

KASUTATUD VEEBIAADRESSID

<https://www.putty.org/> (10.2020)

<http://broadinstitute.github.io/picard/> (03.2021)

<http://lifesciencedb.jp/bp3d/?lng=en> (29.05.21)

LISAD

LISA 1. Koondinfo töös kasutatud keskaegsete indiviidide (n=20) kohta.

Lisa 1.1. Uuritavate keskaegsete indiviidide arheoloogiline taustainfo ja genoomsete raamatukogude valmistamise moodsus. Morfol. – morfoloogiline; konts. – genoomse raamatukogu kontsentratsioon; N – naissoost.

#	Matuse number	Indiviid	Dateering	Morfol. sugu	Proovi kaal (mg)	Raamatukogu indeks	Sekveneeritud järjestusi	Konts. (ng/µl)
1	17	SIK004	1200-1255	Naine	194.7	NEB12	23355903	44.4
2	47	SIK005	1300-1350	Mees	240	NEB13	24298024	61.6
3	83	SIK006	1300-1350	Mees	133	NEB14	26587451	48.6
4	108	SIK007	1400-1475	Mees	190.9	NEB15	27384814	85.4
5	110A	SIK008	1400-1475	Naine	31.9	NEB16	17264852	57.4
6	110B	SIK009	1400-1475	Naine	125	NEB17	21327297	53.8
7	111	SIK010	1300-1475	Naine	100.9	NEB18	24795206	60.4
8	113	SIK011	1400-1475	?	52	NEB19	23284007	51.8
9	131	SIK012	1250-1400	Naine	215.7	NEB20	27328152	55.8
10	135	SIK013	1300-1350	Mees	101.4	NEB21	21300847	67.2
11	157	SIK014	1400-1475	?	157.7	NEB22	22440518	51.6
12	173	SIK015	1300-1475	Laps (N)	98.4	NEB1	22678415	15.50
13	177	SIK016	1300-1475	Naine	141	NEB2	20528219	10.5
14	187	SIK017	1250-1350	Naine	60.6	NEB3	17090611	0.96
15	200	SIK018	1300-1350	Naine	142.5	NEB4	28983529	2.26
16	246	SIK019	1300-1400	Mees	111.7	NEB5	22187404	2.12
17	245A	SIK020	1225-1300	Laps (N)	144.9	NEB6	22094907	7.72
18	245C	SIK021	1225-1300	Naine	85.7	NEB7	22960925	63
19	245D	SIK022	1225-1300	Laps (N)	190	NEB8	28380629	6.3
20	247B	SIK023	1225-1250	Naine	105.6	NEB9	19799492	4.88

Lisa 1. 2 Uuritavate keskaegsete indiviidide genoomsete raamatukogude sekveneermistulemuste statistikud. Endog. – endogeense; sugu – genoomne sugu; C>T – 5’ otstes esinevate tsütosiin→tümiin asenduste hulk; MT – mitokondriaalne DNA; krom. – kromosoom; hg – haplogrupp; N/A – pole saadaval.

#	Indiviid	Matuse number	Endog. DNA sisaldus	Raamatukogu efektiivsus	Keskmine fragmendi pikkus	Keskmine genoomne katvus	Sugu	C>T	Keskmine MT katvus	MT saastatus	Keskmine X-krom katvus	Meessoost indiviidi X-krom saastatus	MT hg	Keskmine Y-krom. katvus	Y hg
1	SIK002	248	84%	67%	-	0,5410	XY	7%	40,40	0,46%	0,2730	0,43%	T2b	0,162	N3a3a-L550
2	SIK003	249	34%	27%	-	0,2020	XX	4%	32,10	0,07%	0,1920	-	H1b2	-	-
3	SIK004	17	31%	29%	52	0,1070	XX	9%	28,00	0,10%	0,1060	-	J1c5a	0,00157	-
4	SIK005	47	49%	39%	54	0,1800	XY	16%	46,00	0,10%	0,0864	0,00929	H1b2	0,0522	N3a3'5
5	SIK006	83	28%	22%	50	0,1030	XY	7%	12,00	0,40%	0,0533	0,1317607	H1b2	0,0314	N3a3'5
6	SIK007	108	31%	26%	65	0,1630	XY	11%	66,10	1,50%	0,0835	0,0000661	U2e1f1	0,0503	N3a3'5
7	SIK008	110A	1%	1%	53	0,0031	XX	9%	1,07	NA	0,0030	-	U4d	0,0000569	-
8	SIK009	110B	7%	6%	53	0,0223	XY	7%	3,70	NA	0,0115	NA	U3a2	0,00718	N/A
9	SIK010	111	56%	46%	56	0,2240	XX	10%	42,40	1,70%	0,2190	-	U2e1f1	0,00298	-
10	SIK011	113	1%	1%	58	0,0025	kooskõlas XY, aga mitte XX	8%	6,53	0,40%	0,0014	NA	U2e1f1	0,000708	N/A
11	SIK012	131	72%	61%	55	0,3200	XX	4%	30,20	0,20%	0,3150	-	H13a1a1c	0,00355	-
12	SIK013	135	25%	21%	61	0,0959	XY	10%	42,70	0,10%	0,0487	0,0000661	J2a1a1a3	0,029	N3a
13	SIK014	157	1%	1%	57	0,0037	kooskõlas XY, aga mitte XX	6%	1,29	NA	0,0019	NA	U	0,00111	N/A
14	SIK015	173	22%	16%	48	0,0628	XX	5%	11,40	0,16%	0,0606	-	U8b1b	0,000925	-
15	SIK016	177	19%	13%	47	0,0450	XX	5,20%	11,90	0,15%	0,0438	-	U8b1b	0,000626	-
16	SIK017	187	34%	20%	42	0,0506	XX	9,60%	8,46	0,14%	0,0493	-	H1b1	0,000791	-
17	SIK018	200	9%	6,50%	46	0,0304	XX	7,20%	5,43	0,33%	0,0294	-	J2a1a1a3	0,000537	-
18	SIK019	246	51%	36%	46	0,1300	XY	6,30%	20,20	0,57%	0,0672	0,1568013	H5a1a	0,0374	N3a3
19	SIK020	245A	70%	53%	48	0,1970	XX	12,90%	7,06	0,39%	0,1850	-	U5a1a1h	0,00366	-
20	SIK021	245C	77%	65%	59	0,3110	XX	10%	44,40	1,50%	0,3070	-	U5a1a1h	0,00366	-
21	SIK022	245D	69%	49%	45	0,2110	XX	10,90%	23,80	0,21%	0,2110	-	U5a1a1h	0,00378	-
22	SIK023	247B	38%	25%	44	0,0763	XX	8,20%	9,74	0,16%	0,0740	-	I1a1c	0,00117	-

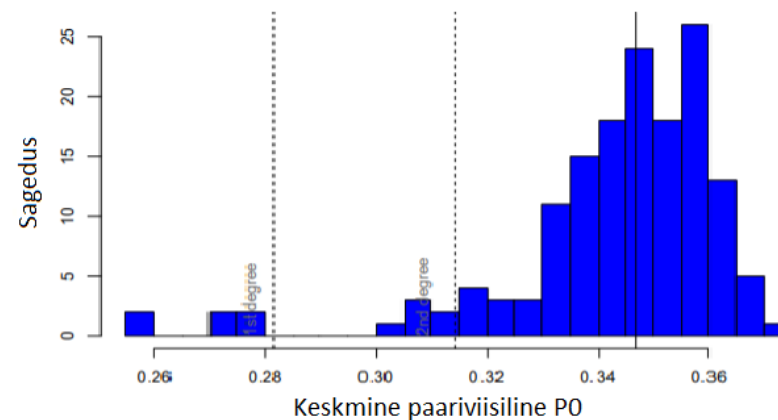
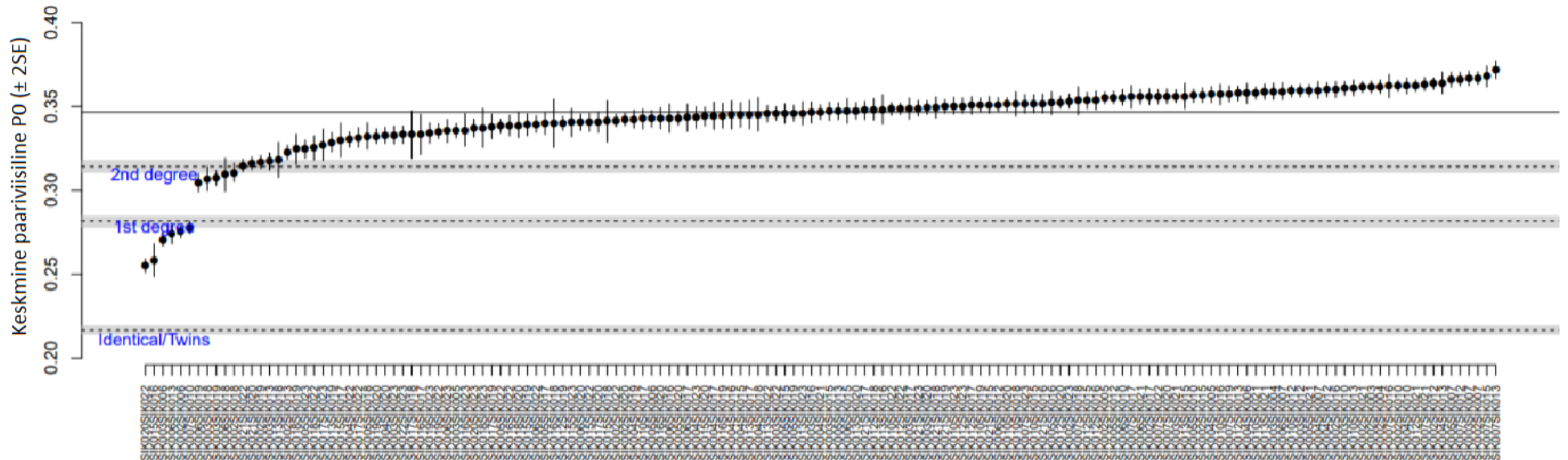
LISA 2. Peakomponentanalüüsis kasutatud tänapäevased ja muinasaja populatsioonid.

Tänapäevased populatsioonid	Lühend	Viited
Valgevenelased (n=16)	Belo	Behar et al., 2010; Kushniarevich et al., 2015
Eestlased (n=57)	Est	Kushniarevich et al., 2015; Raghavan et al., 2014; Tambets et al., 2018
Soomlased (n=19)	Fin	Tambets et al., 2018
Prantslased (n=28)	Fre	Li et al., 2008
Sakslased (n=14)	Ger	Tambets et al., 2018, Yunusbayev et al., 2012
Ungarlased (n=19)	Hun	Behar et al., 2010
Ingerisoomlased (n=6)	Ing	Tambets et al., 2018
Karjalased (n=17)	Kar	Tambets et al., 2018, Yunusbayev et al., 2012
Lätlased (n=7)	Lat	Kushniarevich et al., 2015; Tambets et al., 2018
Leedulased (n=10)	Lit	Behar et al., 2010
Mordvalased (n=15)	Mor	Yunusbayev et al., 2012
Orkadlased (n=15)	Orc	Li et al., 2008
Poolakad (n=13)	Pol	Behar et al., 2013
Keskvenelased (n=20)	CRus	Behar et al., 2010; Yunusbayev et al., 2015
Põhjavenelased (n=31)	NRus	Behar et al., 2010; Kushniarevich et al., 2015; Li et al., 2008; Yunusbayev et al., 2015
Lõunavenelased (n=15)	SRus	Yunusbayev et al., 2015
Rootslased (n=19)	Swe	Behar et al., 2013; Tambets et al., 2018
Ukrainlased (n=20)	Ukr	Yunusbayev et al., 2012
Vepslased (n=12)	Vep	Tambets et al., 2018; Yunusbayev et al., 2012
Kõmrid (n=20)	Wel	Tambets et al., 2018

Muinasaja populatsioonid	Lühend	Viited
Eesti rauaaeg (n=7)	EstIA	Saag et al., 2019
Eesti pronksiaeg (n=17)	EstBA	Saag et al., 2019
Eesti nõorkeraamikud (n=4)	EstCWC	Mittnik et al., 2018; Saag et al., 2017
Eesti keskaeg (n=59)	EstMA	Avaldamata
Läti keskaeg (n=22)	LatMA	Avaldamata

LISA 3. Sugulusanalüüsi READ tulemused.

Pidev joon tähistab mediaanväärtust ning katkendjooned tähistavad identse inimese/kaksikute (ingl *identical/twins*), 1. sugulusastme (ingl *1st degree*) ja 2. sugulusastme (ingl *2nd degree*) lävendeid. Lävendeid ümbritsev hall ala tähistab 95% usaldusvahemikku. P0 – mittekattuvate alleelide suhe indiviidide paari vahel igas 1 miljoni ap suuruses aknas; SE – standardviga (± 2). Tulpdiagrammi tulbad näitavad mittenormaliseeritud P0 keskmise väärtuse sagedusi.



LIHTLITSENTS

Lihlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, Stella Eileen Päid,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose „Eesti keskaegse arheoloogilise materjali geneetiline sugulusanalüüs Siksälä kalme näitel“, mille juhendajad on PhD Kristiina Tambets ja MSc Stefania Sasso, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.
2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commons'i litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, alates 31.12.2025 kuni autoriõiguse kehtivuse lõppemiseni.
3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.
4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Stella Eileen Päid

31.05.2021