

Aus dem Fachbereich der Biologie der Johannes Gutenberg Universität Mainz

Populationsgenetische Analyse prähistorischer Individuen aus Griechenland

**PALÆOGENETICS
GROUP**



at the Mainz Institute

Dissertation
zur Erlangung des
Dr. *rer. nat.* im Fachbereich der Biologie
Institut für Anthropologie
Arbeitsgruppe Palaeogenetik

vorgelegt von
Susanne Kreuzer
aus Biedenkopf
Mainz, 2017



JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Tag der mündlichen Prüfung: 14.12.2018

Zusammenfassung

Die vorliegende Arbeit gibt einen Einblick auf den genetischen Hintergrund der Populationen während der Neolithisierung des griechischen Festlands, eine Region welche am Anfang der Verbreitungswelle aus der neolithischen Kernzone nach Europa steht. Es wurden erstmals Skelettmaterialien prähistorischer Funden aus Griechenland untersucht, mit dem Ziel, archäologische Thesen durch die genetische Struktur zu überprüfen.

Präzise Quantifizierungen des biomolekularen Erhaltungszustand der DNA, extrahiert aus dem vorliegenden Probenmaterial (Zahn und Knochen), zeigten mäßige Qualität in Form von hohen DNA Degradierungsraten und einem niedrigen endogenen DNA Gehalt (<0,5%). Aus Felsenbeinen (*Pars petrosa ossis temporalis*) hingegen konnte ein endogener DNA Gehalt bis zu 60% isoliert werden, jedoch in Kombination mit Desaminierungsraten bis 60% und einer Fragmentlängenverteilung $\sim 60\text{bp}$. In dieser Arbeit konnte das erste Mal gezeigt werden, dass DNA Degradierungsraten nicht mit dem Alter einer Probe korrelieren, sondern das untersuchte Skelettelement und klimatische Bedingungen einen größeren Einfluss haben. Generell zeigen Skelettmaterialien aus temperierten Regionen hohe DNA Degradierungsraten und es konnte der Nachweis erbracht werden, dass der Anstieg der Desaminierungsrate mit der Abnahme der Fragmentlänge der DNA korreliert. Unter Anwendung von Hochdurchsatzmethoden konnten in dieser Arbeit 53 mitochondriale Genome sequenziert werden. Durch die Etablierung sensitiver Protokolle während der Probenbearbeitung konnten Proben bis zu einem endogenen DNA Gehalt von 0.05% hinsichtlich des mitochondrialen Genoms angereichert und authentische Sequenzen erstellt werden. Für eine populationsgenetische Analyse wurde ein mitochondrialer Datensatz erstellt, welcher zwei Individuen aus der mesolithischen Schicht der Höhle von Theopetra (7500 - 7000 kal. v. Chr.) enthält, 16 Individuen aus dem frühen Neolithikum (6700/6500 - 5800/5600 v. Chr.) und 35 Individuen aus dem fortschreitenden Neolithikum (5800/5600 - 3300/3100). Dabei liegt der Fokus dieser Arbeit auf der neolithischen Population aus Makedonien. Ausgewählte Proben aus dem frühen und späten Neolithikum wurden mit einer Abdeckung des Genoms von 1-2x sequenziert.

Mitochondriale Linien aus dem Mesolithikum liefern einen Anhaltspunkt für eine genetisch separierte Population im Vergleich zu den Jäger Sammlern aus Zentraleuropa. Diese genetische Linie konnte jedoch in präkeramischen Individuen aus Zentralanatolien gefunden werden und stimmen mit den Linien aus dem griechischen Neolithikum überein. Das griechische Neolithikum zeigt geringe genetische Distanzen zu den ersten Farmer der Starcevo und LBK Kultur. Es herrschte eine genetische Kontinuität über die gesamte Dauer des Neolithikums in Nordgriechenland, so dass eine massive Einwanderung einer genetisch differenzierten Population ausgeschlossen werden kann. In dieser Arbeit kann eine genetische Verknüpfung der frühenolithischen Population der Marmara Region mit dem Norden Griechenlands gegeben werden. Ein hohe genetische Identität auf Basis von Genomanalysen lassen die ersten Farmer der Ägäis als eine genetisch homogene Population erscheinen, in welcher keine Einflüsse von westeuropäischen Jäger und Sammler identifiziert werden können. Trans-ägäische Verbindungen wurden über die Dauer des Neolithikums aufrechterhalten, was durch eine hohe genetische Ähnlichkeit von Individuen der ägäischen Küstenregion im späten Neolithikum gezeigt werden kann. Signale von Genfluss aus der kaukasischen Jäger Sammler Gesellschaft spiegeln wider, dass vereinzelt Migrationen aus dem späten Neolithikum der Kernzone über die Ägäis in den Norden Griechenlands stattgefunden haben. Da keine genetische Kontinuität zwischen der neolithischen Population zur modernen Population in Griechenland nachgewiesen werden kann, liegt die Annahme nahe, dass ab dem Beginn der Bronzezeit in der Ägäis Prozesse stattgefunden haben, welche die Ausbildung der genetischen Struktur der heutigen Bevölkerung in Griechenland geprägt haben.

Inhaltsverzeichnis

Einleitung	1
1.1 Die Besiedlung Europas durch den Menschen	1
1.2 Die Analyse prähistorischer DNA Moleküle	6
1.3 Genetische Diversität - Entstehung und Nachweis	15
1.4 Fortschritte der Paläogenetik	21
Die Besiedlungsgeschichte Griechenlands & Probenmaterial	25
2.1 Das Jungpaläolithikum	25
2.2 Das Mesolithikum	26
2.3 Das Neolithikum	28
Fragestellung der Arbeit	37
Sequenzierung alter DNA aus archäologischen Skeletten	39
3.1 Datengrundlage	41
3.2 Methodische Ziele und Strategien	43
3.3 Arbeitsweise mit archäologischen Skelettmaterialien	44
3.4 Sanger Sequenzierung der mitochondrialen HVSI-Region	47
3.5 <i>Next-Generation Sequencing</i>	48
3.6 <i>Shotgun</i> Sequenzierung	53
3.7 Anreicherung des mitochondrialen Genoms	53
3.8 NGS Datenanalyse	58
Ergebnisse der Laborarbeit	63
4.1 Genereller Erhaltungszustand	63
4.2 Extraktion	64
4.3 Bestimmung der Probenqualität	65
4.4 Sequenzierung kompletter Genome (<i>Shotgun</i>)	76
4.5 Auswertung der <i>Capture</i> Experimente	80
4.6 Datenanalyse der mitochondrialen Genome	90
Populationsgenetische Analyse der Sequenzdaten	97
6.1 Der prähistorische Datensatz aus Griechenland	97
6.2 Mitochondriale Genomanalyse	100
6.3 Ergebnisse der mitochondrialen Genomanalyse	104
6.4 Räumlich-explicite Simulationen	116
Diskussion	119
7.1 Diskussion der Laborergebnisse	119
7.2 Die genetische Struktur im griechischen Mesolithikum	125
7.3 Die genetische Struktur des griechischen Neolithikums	126
7.4 Das trans-ägäische Netzwerk und Konnektivität	128
7.5 Das ägäische Neolithikum im überregionalen Vergleich	131
7.6 Die Rolle Griechenlands in der Verbreitung des Neolithikums	133
7.7 Das griechische Neolithikum im Kontext der heutigen europäischen Bevölkerung	137

7.8 Schlussfolgerung	139
Literaturverzeichnis	141
Anhang I - Methoden	151
A.1 Experiment zur Silika Extraktion	151
A.2 <i>Library</i> Erstellung	153
A.3 Quantifizierung	158
A.4 <i>Capture</i> des mitochondrialen Genoms	159
Anhang II - Abbildungen, Tabellen und Publikation	161
B.1 Probenvorbereitung	161
B.2 Ergebnisse PCR und Sanger Sequenzierung Phase A/B	162
B.3 DNA <i>Screening</i> (Leerkontrollen)	163
B.4 Durchführung und Parameter einzelner <i>Capture</i> Experimente	163
B.5 Durchführung und Parameter einzelner <i>Capture</i> Experimente	164
B.6 <i>Capture</i> Ergebnisse kontaminierter Proben	169
B.7 Publikation Hofmanova, Kreuzer <i>et al.</i>	172
Anhang IV - Geräte und Chemikalien	179
C.1 Geräte	179
C.2 Verbrauchsmaterialien und Reinigungsequipment	179
C.3 Chemikalien und Kits	180
Lebenslauf	183

Einleitung

1.1 Die Besiedlung Europas durch den Menschen

Hominine Formen im späten Pleistozän und frühen Holozän

Im Pleistozän, eine geologische Epoche, welche vor 2,5 Millionen Jahren begann und mit dem Übergang in das Holozän vor 11.000 Jahren endete, entwickelten sich hominine Formen des Genus *Homo* in Afrika. Das Pleistozän ist geprägt durch sich wiederholende Eiszeiten, welche mit der letzten, über etwa 6000 Jahre andauernden Maximalvergletscherung der nördlichen Hemisphäre ihren Höhepunkt erreichte (LGM, *last glacial maximum*, 24.000-18.000 vor heute). Diese Zeit ist geprägt durch einen niedrigen Meeresspiegel und einer Steppenlandschaft in den Gebieten Zentraleuropas, die im Norden und im Alpenraum den Charakter einer Eiswüste trug. Dem LGM folgte eine Warmphase, auf welche sich eine erneute Phase der Abkühlung, das Jüngere Dryas (um 10.500 v. Chr.), angeschlossen hat. In der folgenden Warmphase, dem Präboreal, erfolgte eine bis heute anhaltende Steigerung der allgemeinen Durchschnittstemperatur mit klimatischer Stabilität und leitete die geologische Epoche des Holozäns ein.

Die Klassifizierung hominier Formen erfolgt anhand ihres technischen Fortschrittes beginnend mit der Oldowan Industrie in Afrika vor 2,6 - 1,7 Millionen Jahren. Aufgrund der Nutzung primitiver Steinwerkzeuge beginnt hier die kulturelle Einordnung in das Paläolithikum (aus dem griechischen: palaios = "alt", lithos = "Stein"). Im mittleren Paläolithikum (vor 300.000 - 45.000 Jahren) besiedelten archaische Menschenformen wie *Homo heidelbergensis*, *Homo neandertalensis* und Denisova-Menschen Teile des eurasischen Kontinents, während sich im subsaharischen Afrika der anatomisch moderne Mensch (AMM, *Homo sapiens*) entwickelte. Zum Ende des mittleren Paläolithikums wanderte eine Teilpopulation des AMM über die saudische Halbinsel aus und besiedelte über die tropische Küstenroute und Melanesien den australischen Kontinent. Zeitlich verzögert startete die Kolonisation der Levante, Asiens und Europas. Anzunehmen ist, dass dieses Ereignis durch mehrere Migrationswellen geprägt ist [146, 182]. Genetische Studien belegen, dass das Aufeinandertreffen des modernen Menschen auf archaische Menschenformen zu regionalen Paarungsereignissen führte, da im Genpool des AMM archaische Genstrukturen zu finden sind. Dass das Aufeinandertreffen von Neandertaler und *Homo sapiens* nach dem *Out-of-Africa Event* stattgefunden hat, zeigen Genstrukturen der modernen nicht-afrikanischen Bevölkerung, welche etwa 2% der genetischen Information mit dem Neandertaler teilen [57, 95]. Südostasiatische und ozeanische Populationen tragen zusätzlich genetische Strukturen des Denisova, welche besonders ausgeprägt in Melanesien und Australien zu finden sind [121].

Innerhalb der Neandertal Population ist durch die Sequenzierung dreier Genome aus unterschiedlichen Teilen des eurasischen Kontinents - Europa (Vindija-Höhle), Georgien (Mesmaiskaja-Höhle), Sibirien (Okladnikow-Höhle/Altai) - eine Struktur zu erkennen. Laut Kuhlwilm und Kollegen [95], hat sich der europäische Neandertaler von der Population aus dem Altai Gebirge vor rund 100.000 Jahren getrennt. Während der Altai Neandertaler ca. 0,1 - 2,1% genetisches Material einer frühen *Homo sapiens* Population aufgenommen hat, teilt die nicht-afrikanische Bevölkerung 0,3-2,6% archaisches Genmaterial des europäischen Neandertalers. Es wird eine geringe Größe beider Neandertalpopulationen angenommen, da Heterozygotitätskoeffizienten weit unter der Variation heutiger menschlicher Populationen liegen. Analysen im Altai Neandertaler zeigen lange Abschnitte des genetischen Materials ohne Mutationen (ROHs), welche für nahe Verwandtschaften und Inzucht sprechen [144], möglicherweise ein Ergebnis von geographischer Isolation in den Altai Gebirgen [95].

Eine dritte archaische Menschenform besiedelte im mittleren Pleistozän Afrika, den Mittelmeerraum sowie weite Teile Zentraleuropas, wie Funde aus Atapuerca (Spanien), Mauer/Heidelberg (Deutschland) und Petralona (Griechenland [69]) belegen (Übersicht in [179]). Derzeit arbeitet eine Gruppe des Max Planck Instituts (Leipzig) an der Sequenzierung des ersten *Homo heidelbergensis* Genoms. Im Jahre 2014 erstellten Meyer und Kollegen das erste Mitochondrium eines 400.000 Jahre alten Heidelbergensis aus Atapuerca, jedoch konnten Verwandtschaftsbeziehungen zu den archaischen und der modernen Menschenform durch mitochondriale Daten alleine nicht gelöst werden, da errechnete Divergenzzeiten, Probenalter und anthropologische Merkmale keine schlüssigen Ergebnisse erzielten [119]. Mit der folgenden Analyse von genomischen Daten zeigte sich auf dieser Ebene eine nahe Verwandtschaft zum Vorfahren des Neandertalers, während das mitochondriale Genom Ähnlichkeiten zur Denisova Population aufweist. Da das Probenalter etwa in die Divergenzzeit von Neandertaler und Denisova fällt, rücken die Autoren von der Klassifizierung einer Heidelbergensis Population ab und sprechen von einer homininen Form, deren Beziehungen nicht genau geklärt sind [117].

Erste Funde des anatomisch modernen Menschen auf dem europäischen Kontinent werden mit dem Beginn des späten Paläolithikums ab etwa 45.000 v. Chr. sichtbar. In Süditalien befindet sich die älteste Fundstelle mit Überresten des modernen Menschen mit einem Alter von 45.000 - 43.000 Jahren (Grotta del Cavallo, [10]). Auf eine rasche Art der Verbreitung deutet ein Fund datiert auf 44.200 - 41.500 Jahre in England hin [70]. In Griechenland hingegen sind Funde des anatomisch modernen Menschen aus dem Beginn des späten Paläolithikums im europäischen Vergleich minimal (siehe Kapitel 2, Abschnitt 2.1). Es besteht die Annahme, dass die erste Kolonisation Europas auf terrestrischem Weg stattgefunden hat, so dass Halbinseln bzw. Landzungen zu dieser Zeit meist als letztes besiedelt wurden [22].

Abhängig der klimatischen Bedingungen und dem resultierenden Wandel des Lebensraumes ist die Besiedlungsstruktur Europas geprägt durch eine ständige Bewegung der menschlichen Population. Die erste Besiedlung in einer interglazialen Phase wurde durch die vor 24.000 Jahren einsetzende Eiszeit auf einen Breitengrad von 52°N zurückgedrängt. Dabei gelten die südlichen Teile Europas als Refugium für Flora, Fauna und den Menschen [195]. Mit dem Ende des LGM um 18.000 v. Chr. und der Rückbildung der Maximalvergletscherung der nördlichen Hemisphäre wurde Zentral- und Nord-Europa durch Jäger-Sammler (JS) Populationen rekolonialisiert. Die mit dem Holozän einhergehende Phase der klimatischen Erwärmung und Stabilität sorgte für die Ausbildung von bewaldeten Gebieten, was den Rekolonisten eine erweiterte Existenzgrundlage angeboten hat. Angepasst an den neuen Lebensraum kam es zu einer Änderung der Lebensweise, die sich im beginnenden Mesolithikum (mittlere Steinzeit, ab 11.700 v. Chr.) vermehrt auf das Jagen kleiner Wildtiere und die Ausbildung von Fischereien konzentrierte.

Das europäische Neolithikum - Entstehung und Verbreitung

Das Neolithikum (aus dem altgriechischem "neos" = neu, "lithos" = Stein) beschreibt den technologischen Fortschritt sowie den charakteristischen Wandel der Siedlungsstruktur in der neuen Steinzeit. Archäologisch erfasst wurden diese Entwicklungen zunächst im Fruchtbaren Halbmond, dem Entstehungsgebiet des europäischen und südwest-asiatischen Neolithikums, jedoch sind parallele Tendenzen in der Entwicklungsgeschichte des modernen Menschen ebenfalls in Asien, Afrika und Amerika zu sehen, so dass der Begriff des Neolithikums nicht für den europäischen Gebrauch generalisiert werden darf. Da sich die Merkmale des Neolithikums graduell über angrenzende Regionen ausgebreitet haben, ist von Archäologen schon früh auf eine Migrationswelle aus der Region der Levante und dem Nahen Osten in Gebiete europäischer Jäger-Sammler Gesellschaften hingewiesen worden [31]. Technologi-

scher Fortschritt in der Nahrungsproduktion, sowie mit der Sesshaftigkeit einhergehende dorfähnliche Strukturen, sind ein Schlüsselereignis in der Entwicklung moderner Gesellschaften. Allgemein wird das Konzept des "neolithischen Pakets" zur Beschreibung der Entwicklung neolithischer Kulturstufen und ihrer geographisch-chronologischen Verbreitung verwendet [130, 32]. Einzelnen Komponenten des neolithischen Paketes, z.B. eine durch den Neolithiker konstruierte Siedlungsstruktur, ihre Domestikate, Keramiken und Werkzeuge des alltäglichen Lebens, haben sich stets weiter entwickelt (siehe Abbildung 1.1). Dabei hat sich nicht ein homogenes Paket verbreitet, sondern unterschiedliche Ausprägungsformen, wobei einzelne Komponenten verstärkt in den Vordergrund neolithischer Fundsituationen rücken [129, 130]. Auf archäologischer Basis werden Merkmale, die wiederholt im neolithischen Kontext auftreten, überregional verglichen, mit dem Ziel Migrationsrouten anhand des Transfers von Technologien zu identifizieren.

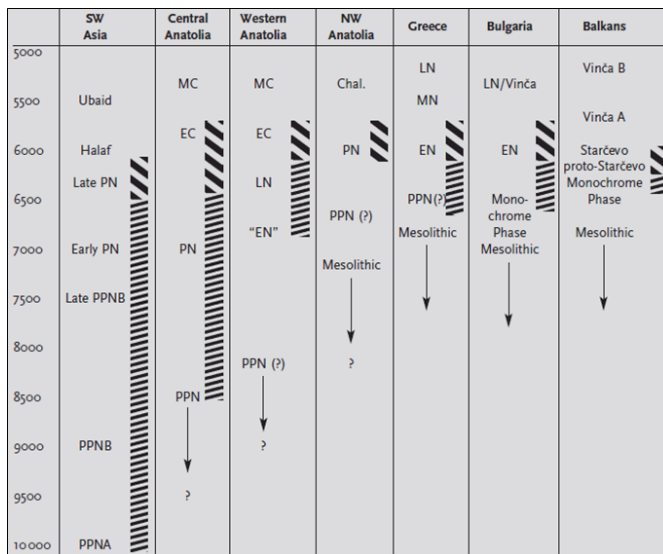


Abbildung 1.1: Geographisch-chronologischer Vergleich des Erscheinens des neolithischen Pakets ausgehend vom Punkt der Entwicklung im "Fruchtbaren Halbmond" bis auf den Balkan. Entnommen aus Çilingiroglu, 2005 [32], Tabelle 3. Dünne Linien beschreiben das erste Auftreten des neolithischen Pakets. Dickere Linien zeigen entwickelte Formen der Komponenten innerhalb einer neolithischen Kultur.

Während sich mesolithische JS Kulturen im beginnenden Holozän im Rahmen der Rekolonisation Europas weiter in nordöstliche Regionen ausgebreitet haben und ihre Populationsgröße wieder zugenommen hat, entstanden in dem Gebiet des fruchtbaren Halbmondes um 10.000 BC die ersten Vorstufen früher Farmerkulturen (PPNA, "pre pottery neolithic A" [130], 10.000 - 9.000 v. Chr. [32]). Es wird angenommen, dass Epipaläolithische Jäger und Sammler der Natufian-Kultur der Levante und Syromesopotamiens während der Kaltphase des Dryas ihre Ernährungsstrategie hinsichtlich eines kontrollierten Anbaus von Getreiden und Hülsenfrüchten änderten [8]. Während dieser Zeit entstanden Siedlungsstrukturen, bestehend aus runden Hütten mit innenliegender Kochstelle. In der zweiten Phase des vorkeramischen Neolithikums (PPNB, "pre pottery neolithic B" 9000-7.000 v. Chr. [32]) begann die Domestikation von Schaf, Rind, Schwein und Ziege in unterschiedlichen Regionen des Nahen Ostens (siehe Abbildung 1.2).

Die vollständige Ausbildung des Neolithikums dauerte etwa 3000 Jahre (PPNA/B, siehe Abbildung 1.3 - dunkelgrau), bevor es sich in angrenzende Regionen ausbreitete. Die ersten Charaktere des neolithischen Paketes sind zu diesem Zeitpunkt nicht einheitlich über das Gebiet des fruchtbaren Halbmondes verbreitet, sondern zeigen eine regionale Strukturierung, welche sich durch Siedlungsarchitektur und Domestikate (siehe Abbildung 1.2 [207]), sowie verwendeter Werkzeuge und Bestattungsrituale unterscheidet (siehe Phase A in Özdoğan (2014) [130]). Während der Entwicklungsphasen (PPN) hat durch Änderung der Nahrungsgrundlage und Ausbildung gesellschaftlicher Strukturen die Populati-

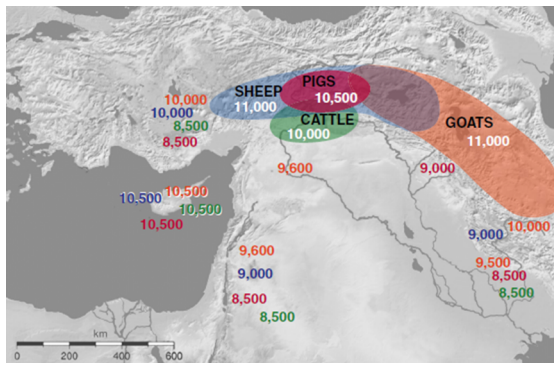


Abbildung 1.2: Herkunft und Verbreitung der Haustiere während des sich entwickelnden Neolithikums. Zeitangabe erfolgt kalibriert vor heutiger Zeitrechnung. Entnommen aus Zeder (2008), Abbildung 1 [207].

onsdichte in neolithischen Gesellschaften stark zugenommen. Dies führt ab dem späten PPNB (ab 7500 v. Chr.) zu einer schnellen Verbreitung der neolithischen Lebensweise in westliche Richtung an die Küste der Ägäis (7000 - 6500 v. Chr., siehe Abbildung 1.3 - gelb) und in die angrenzende nord-west-anatolische Marmara Region (ab 6600 v. Chr., siehe Abbildung 1.3 - hellgrau). Ab 7.000 v. Chr. sind Keramiken als Aufbewahrungsbehälter für Nahrungsprodukte eine wichtige Komponente des neolithischen Paketes. Die Art der Herstellung und künstlerische Gestaltung geben Anlass zur Verknüpfung von Fundsituationen über größere Distanzen. Mit der ersten Verbreitungswelle des Neolithikums auf den europäischen Kontinent sind zunächst sehr einfach gestaltete, monochrome Keramiken verbunden. Mit der folgenden Besiedlung des griechischen Festlandes (Thessalien, ab 6500 v. Chr.) und Kretas über die Ägäis hat erstmalig der Übertritt des Neolithikums von Klein-Asien nach Europa stattgefunden. Kenntnisse der Navigation und Seewege haben sich in der Ägäis bereits im späten Paläolithikum während der Kaltzeit des Jüngeren Dryas entwickelt [22]. Mit steigenden Temperaturen und klimatischer Stabilität im beginnenden Holozän hat sich, ausgehend von der Insel Melos, ein trans-ägäisches Netzwerk für den Transport von Obsidian ausgebildet. Diese Fähigkeit wurde möglicherweise später von Neolithikern an der Küste der Levante bzw. der südwestlichen Küste Anatoliens genutzt, um die neolithische Lebensweise zu transferieren [99, 152]. Aus dem nordwestlichen Teil Anatoliens setzte sich die westlich gerichtete Ausbreitung der neolithischen Kultur auf dem Landweg fort und erreicht ab 6400 v. Chr. türkisch Thrakien und den nördlichen Teil Griechenlands, sowie die angrenzenden Teile Bulgariens [141, 152, 130]. Die Ausbildung früh-neolithischer Gesellschaften in Süd-Ost Europa hat recht zügig stattgefunden - innerhalb von 500 Jahren waren erste Anzeichen ihrer Kultur in Bulgarien (Karanovo-Kultur ab 6000 v. Chr., [92]) und Kroatien/Ungarn angekommen (Starčevo- und Körös-Kultur [122, 14]). Als Anlass für die schnelle Verbreitung des Neolithikums zu dieser Zeit sehen Klimaforscher einen 600-jährigen Kälteeinbruch (6600 - 6000 v. Chr., Hudson-Event), welcher klimatisch bedingt zu Missernten und einer resultierenden Nahrungsknappheit führte [201].

Angekommen auf dem Balkan trafen die ersten Farmer auf rauhe klimatische Bedingungen im Vergleich zum Mittelmeerraum, was die nördliche Expansion der neolithischen Population zunächst verlangsamte bzw. einschränkte. Diese "ökologische Barriere" (CEB AEB *Central European-Balkan Agro-ecological Barrier*, [180]), befindet sich in etwa auf Höhe des Plattensees und teilt damit die heutige Region Ungarns in einen südlichen und nördlich-transdanubischen Teil. Um diese zu überwinden mussten Existenzstrategien hinsichtlich Ernährungs und Siedlungsstruktur angepasst werden. An dieser Stelle kam es zu Kontakten zwischen ersten Farmern und mesolithischen Jäger-Sammlern, was durch das Werkzeuginventar erster Farmer belegt ist und starke Ähnlichkeiten zum mesolithischen Inventar dieser Region zeigt [128]. Während sich die Körös-Kultur im Karpathenbecken weiterentwickelte und die Alföld-LBK bildete, entstand um 5600 v. Chr. die formative Phase der Linearbandkeramik (LBK) ausgehend aus der späten Starčevo Kultur. Neben der bezeichnenden Keramik, welche Gestaltungen von linearen Bandmustern enthält, entstanden zu dieser Zeit die für die LBK Phase typischen

Langhäuser. Die ersten Farmer, welche die CEB AEB überschritten, ließen sich nördlich des Plattensees nieder und bildeten die Kulturstufe der LBK-Transdanubien aus. Mit der Formation der LBK in Transdanubien war der Weg nach Zentraleuropa geebnet und erste Fundstellen in Österreich und Süddeutschland zeigten ab 5500 v. Chr. die Ankunft des Neolithikums.



Karte Isochrone Neolithic, Stand: 21.04.2016

Abbildung 1.3: Ausbreitung des Neolithikums aus der Region des fruchtbaren Halbmondes nach Europa (und Südost-Asien). Datierungen sind in kalibrierten Jahren vor unsere Zeitrechnung angegeben. Die isochrone Darstellung der neolithischen Welle erfolgt anhand erster neolithischer Funde eines geographischen Abschnittes. Quelle: Prof. Dr. Joachim Burger.

Die Gestaltung von Keramiken hat während der Weiterentwicklung der neolithischen Kultur hochkomplexe Ausarbeitungsformen angenommen. Beginnend mit der Impresso Kultur (6100 - 5700 v. Chr.) und ihrer Verbreitung in der Ägäis und dem östlichen Mittelmeer, hat sich während des balkanischen Neolithikums die Linearbandkeramik (LBK) entwickelt. Ab diesem Zeitpunkt werden neolithische Kulturstufen, unabhängig anderer Komponenten des neolithischen Paketes, nach Art der Gestaltung ihrer Keramiken benannt. Weiterentwicklungen der LBK sind in nördlichen Regionen Europas die Trichter- und Glockenbecher Kultur. Die Ausbreitung nach Zentral- und Nord-Europa entspricht keinem homogenen Prozess, sondern einem Mosaik aus ersten Farmergesellschaften, die auf europäische JS Gesellschaften treffen. Dabei sind eine Vielzahl von Szenarien sichtbar, beginnend mit einer isolierten Lebensweise bis hin erster Vermischungsereignisse beider Gesellschaften [18, 174]. Seit dem Aufkommen der ersten Farmergesellschaften im fruchtbaren Halbmond vergingen etwa 6000 - 7000 Jahre bis das Neolithikum die nördlichen Breitengrade Europas erreichte (siehe Abbildung 1.3).

Neben der transdanubischen Route zur Verbreitung des Neolithikums über den Balkan gibt es eine weitere Route, welche die Kolonisation des Mittelmeerraumes und der iberischen Halbinsel beschreibt. Maritime Kenntnisse wurden genutzt und die Netzwerke zur Besiedlung der Adria und Italiens ausgeweitet (6100 - 5500 v. Chr.). Die Ausbreitungsrichtung und folgende Besiedlung des westlichen Mittelmeeres ist gekennzeichnet durch die Impresso/Cardial Kultur, welche ihren Namen Keramiken

mit Muschelabdrücken verdankt. Die neolithische Lebensweise wird in vielen kleinen Wellen über das Mittelmeer verbreitet und erreichte die iberische Halbinsel um 5500 v. Chr. (siehe Abbildung 1.3).

Die Neolithisierung Europas wird seit der Bergung erster neolithischer Fundsituationen kontrovers diskutiert - erfolgte die Verbreitung der neu erworbenen Fortschritte des Neolithikums durch die Wanderung einzelner Farmerpopulationen (demische Verbreitung) oder wurden die Fortschritte durch die Übermittlung der Information ohne Migration einzelner Individuen weitergegeben (kulturelle Verbreitung)? Aufgrund des breiten Interpretationsspektrums ist diese Fragestellung aus archäologischer Sicht schwer lösbar, so dass genetische Daten zur Aufklärung dieser Prozesse herangezogen werden.

1.2 Die Analyse prähistorischer DNA Moleküle

Enthalten archäologische Fundstellen Bestattungen von Mensch oder Tier, dann kann das Skelettmaterial für eine genetischen Analyse herangezogen werden. Die Konservierung von DNA in archäologischen Funden über Zeiträume von mehreren Tausend Jahren ist eine Begebenheit, die sich die molekulare Evolutionsforschung zu Nutzen macht, um genetische Linien von Lebewesen zu charakterisieren und miteinander in Beziehung zu setzen. Die Matrix des Skelettmaterials bietet Raum zur Konservierung der DNA eines Individuums, jedoch ist die DNA Effekten aus der Umgebung ausgesetzt, welche spezielle Schäden an den DNA Molekülen induzieren. Diese Schäden häufen sich in Abhängigkeit des Zeitfensters von Bestattung zur Bergung eines Individuums sowie des Ausmaßes der einwirkenden Kraft an. Alte DNA Datensätze werden für eine Rekonstruktion der Populationsstruktur innerhalb diverser Zeitstufen der Entwicklungsgeschichte einer Spezies herangezogen. Durch den methodologischen Fortschritt innerhalb der DNA Sequenzierung im letzten Jahrzehnt stehen heutzutage Referenzdatensätze archaischer Menschenformen und des sich entwickelnden anatomisch modernen Menschen (AMM) zur Verfügung.

Eigenschaften alter DNA

Nach dem Tod eines Organismus setzen taphonomische Prozesse ein, die durch physikalische und chemische Wechselwirkungen einen Organismus zersetzen. Tief in der Knochenstruktur eingebettet befindet sich die DNA in Knochenzellen (Osteocyten, Osteoblasten und Osteoclasten), welche nach dem Zelltod durch den Verlust zelleigener Reparaturmechanismen zersetzt werden. Die Knochenstruktur, aufgebaut aus mineralisierten Kollagenfibrillen, schützt aus den Zellen freigesetzte DNA vor dem Abbau durch mikrobiellen Nukleasen. Die Porengröße der Kollagenfibrillen verhindert das Eindringen von Mikroben, deren Proteasen und Kollagenasen die Knochenstruktur zersetzen würden. Diagenetische Prozesse, wie z.B. ein saurer pH des umgebenden Milieus, sorgen für eine Demineralisierung des Knochens, so dass mit steigender Porosität Mikroben Zugang erhalten und der Abbau des Skelettmaterials beginnt [124]. Unter optimalen Bedingungen hingegen bleiben Skelettelemente über tausende von Jahren erhalten, so dass diese für paläowissenschaftliche Analysen zur Verfügung stehen. Für paläogenetische Studien wird die DNA aus pulverisierten Skelettelementen gewonnen. Während der Extraktion von DNA aus Skelettmaterialien wird DNA aus Bodenorganismen coextrahiert (exogene DNA) und überlagert die Analyse der humanen Ziel DNA (endogene DNA). Das Verhältnis an endogener zu exogener DNA ist abhängig von Knochenstruktur und Porosität. Die NGS Technologie erlaubt heutzutage eine direkte Quantifizierung des endogenen human-genetischen Materials (Bestimmung des endogenen DNA Gehalts siehe Kapitel 3.5), welches für die Wahl der Sequenzierungsstrategie herangezogen werden kann.

Da Wasser die Poren der Kollagenfibrillen passieren kann, ist die DNA hydrolytischen Effekten ausgesetzt. So kann die N-glykosidische Bindung zwischen einer Purinbase und der Desoxyribose hydrolytisch gespalten werden (hydrolytische Depurinierung, siehe Abbildung 1.4-A, [39]) und es entsteht eine Apurin-Stelle (AP-Stelle). Diese Positionen stellen einen Angriffspunkt für eine β -Eliminierung dar, resultierend in der Spaltung des Zucker-Phosphat Rückgrats der DNA. Folglich entsteht ein Einzelstrangbruch an abasischen Stellen der DNA, was mit der Zeit zu einer Fragmentierung des DNA Doppelstranges führt (siehe Abbildung 1.5a, [106, 39]). Durch die Fragmentierung entstehen einzelsträngige DNA-Überhänge, in welchen die Basen relativ ungeschützt vorliegen. Ein weiterer hydrolytischer Effekt betrifft die Basen Cytosin und Adenin in einzelsträngigen Überhängen, wobei durch Desaminierung die Aminogruppe durch eine Sauerstoffgruppe ersetzt wird (siehe Abbildung 1.4-B) [39]. Aus Cytosin entsteht Uracil und damit geht eine Donorgruppe für die Ausbildung der dritten Wasserstoffbrücke verloren. Polymerasen, die während enzymatischer Modifikationen im Labor verwendet werden, bauen an dieser Stelle komplementär ein Adeninmolekül ein. Analog entsteht bei der hydrolytischen Desaminierung von Adenin die DNA Base Hypoxanthin, welche eine zusätzliche Akzeptorgruppe bietet, so dass eine Polymerase komplementär Cytosin einbaut. Folglich werden in einer Sequenzierung inkorrekte Position ausgelesen, welche sich in hoher Rate am Ende eines Moleküls befinden. Durch den Abgleich der ausgelesenen Sequenz mit einer Referenz sind diese Desaminierungsfrequenzen auf bioinformatischem Weg quantifizierbar (siehe Abbildung 1.5) [78].

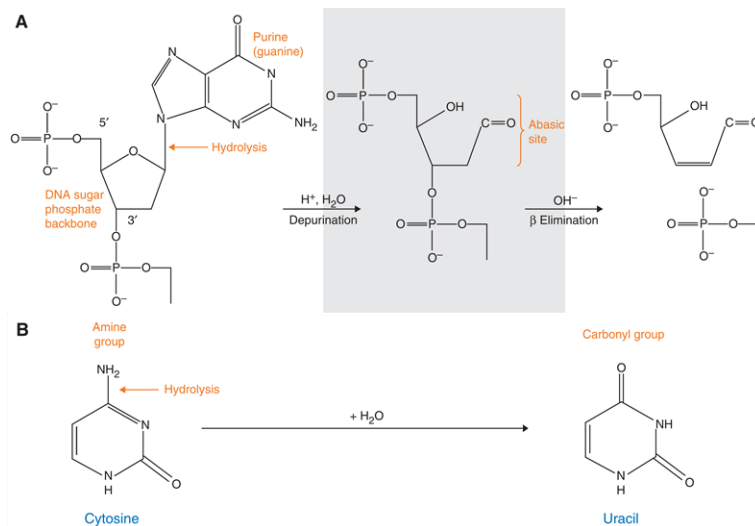
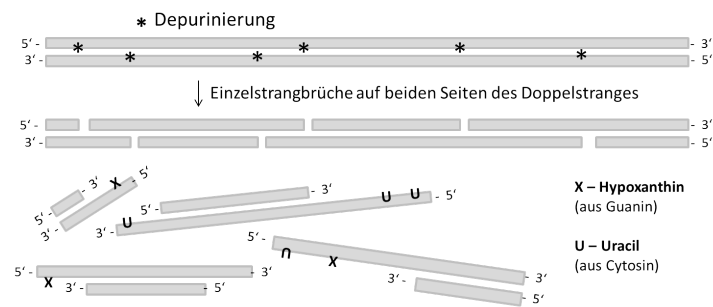
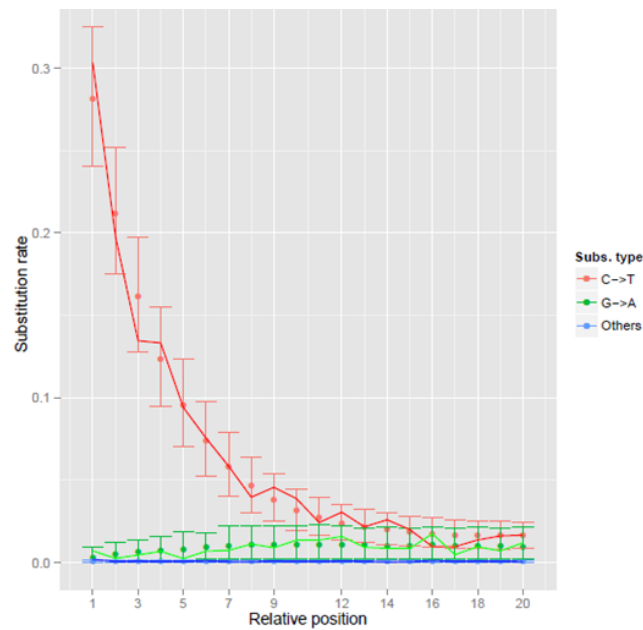


Abbildung 1.4: Chemische Modifikationen von DNA, die durch hydrolytische Veränderungen entstehen (Abbildung aus [39]). A - Chemischer Ablauf der hydrolytischen Depurinierung am Beispiel der Purinbase Guanin. B - Das Ergebnis der Desaminierung ist der Austausch einer Aminogruppe durch eine Carbonylgruppe.

Beide Effekte, die durch hydrolytische Modifikation entstehen, werden als Authentifikationskriterien während der Analyse paläogenetischer Materialien herangezogen, da mittels NGS Analysen komplette Moleküle sequenziert und charakterisiert werden können. Heutzutage ist es möglich den endogenen DNA Gehalt einer Probe, sowie Fragmentlänge und Desaminierungsrate zu bestimmen (siehe Kapitel 3.5). Alte DNA Moleküle zeigen kurze, durchschnittliche Fragmentlängen von 50 - 150bp und Desaminierungsfrequenzen an den äußeren Positionen eines Moleküls von bis zu 50%, welche exponentiell gegen die Mitte eines Moleküls abnehmen. Klimatische Bedingungen haben neben der Zeit den stärksten Einfluss auf den Erhalt von DNA in archäologischen Skelettmaterialien. Während in nördlichen Breitengraden (z.B. Sibirien, Russland, Grönland) ein DNA Erhalt von über 100.000 Jahren möglich ist, was z.B. durch die Sequenzierung von 560 - 780 Tausend Jahre alter DNA aus einem Pferdeknöchel, gelagert unter Permafrost Bedingungen gezeigt werden konnte [127],



(a) Schäden an alten DNA Molekülen ausgelöst durch hydrolytische Modifikationen.



(b) Quantifizierung der Desaminierungsraten an den Endpositionen eines DNA Moleküls durch das Programm *mapDamage 2.0* [78]. Gezeigt wird die Substitutionsrate in Abhängigkeit zur relativen Position der DNA Moleküle.

Abbildung 1.5: Auswirkung hydrolytischer Effekte auf DNA Moleküle - Entstehung und Quantifizierung.

nehmen die Erhaltungsbedingungen in Richtung des Äquators stetig ab. In einem vorangegangenen Projekt dieser Arbeit wurden die schlechten Erhaltungsbedingungen der DNA aus prähistorischen Proben Griechenlands deutlich. Aufgrund des niedrigen endogenen DNA Gehalts in Kombination mit geringer Fragmentlänge und hohen Desaminierungsraten müssen zunächst die Protokolle sensitiviert werden, um eine populationsgenetische Analyse zu ermöglichen¹.

¹Die erhobenen Daten dieses Projekts werden im Abschnitt 3.1 (Datengrundlage) dargestellt.

Next Generation Sequencing Methoden in der Paläogenetik

Während bei konventionellen PCR Analysen einzelne *Loci* aus einem Genom herauskopiert und anschließend sequenziert werden, gehen bei einer Hochdurchsatz Sequenzierung alle vorliegenden Moleküle eines Extraktes in annähernd unveränderter Form in die Sequenzierung ein (Multilokus Sequenzierung). Dadurch hatte die paläogenetische Forschung das erste Mal die Möglichkeit den Zustand dieser alten DNA Moleküle zu beurteilen. Es ist offensichtlich, dass Skelettelemente aus temperierten Regionen eine stark reduzierte Fragmentlängenverteilung der Moleküle enthalten, welche für ein Primersystem in einer PCR nur in geringem Anteil zu erfassen sind. Desweiteren ist der Anteil an alter DNA des Zielorganismus - in diesem Fall die humane, endogene DNA - im Verhältnis zur exogenen DNA quantifizierbar. Wie im vorherigen Abschnitt beschrieben, ist der endogene Anteil innerhalb eines DNA Extraktes meist weit unter 1%.

Unter der Vielzahl an *Next Generation Sequencing* (NGS) Methoden ist die Illumina Technologie derzeit Marktführer und wird auch in der vorliegenden Arbeit zur Erstellung des Datensatzes verwendet. Allgemein basiert die Illumina Sequenzierung auf der Ligation universeller Adapterstrukturen an die Gesamtheit der DNA Moleküle eines Extraktes. Diese Adapterstrukturen ermöglichen sowohl die künstliche Amplifikation der Extraktmoleküle als auch ihre Sequenzierung, da sie diverse Bindestellen für Primer bieten. Diese sequenzierfähigen Moleküle werden allgemein *Library* Moleküle genannt. Bei der direkten Sequenzierung der Gesamtheit jener Extraktmoleküle spricht man von einer *Shotgun* Sequenzierung, eine Begriffsbildung, welche analog zur Schrotschusssequenzierung im Zuge des humanen Genom Projektes [98] entstanden ist. In der paläogenetischen Analyse ist man in den meisten Fällen lediglich an der endogenen DNA interessiert, so dass im Falle einer *Shotgun* Sequenzierung der größte Anteil des Datensatzes nach der Analyse verworfen wird. Aus diesem Grund ist es notwendig einzelne Proben einer Studie im Vorfeld einer genetischen Analyse nach ihrer Qualität zu beurteilen, so dass gezielt Proben mit einem hohen endogenen DNA Gehalt gewählt werden.

Des weiteren können spezielle Anreicherungsverfahren einer Sequenzierung vorgeschaltet werden. Diese Verfahren sind weitreichend bekannt aus der personalisierten Genomanalyse in der Medizin, in welcher genomische Strukturen hinsichtlich der diagnostischen Fragestellung aus der Erbinformation des Patienten isoliert werden können. In der Paläogenetik kann nach Anreicherung der Anteil der endogenen DNA gegenüber der exogenen, meist bakteriellen DNA, stark erhöht werden (*whole genome capture*, WGC [30]). Ist man nur an bestimmten Positionen eines Genoms interessiert, so können gezielt Genregionen angereichert werden (*whole exome capture*, WEC). Momentan wird der größte Anteil an paläogenetischen Daten durch die Arbeitsgruppe des Genetikers David Reich (Boston) produziert, welche in einem Anreicherungsverfahren gezielt funktionelle und herkunftsspezifische Polymorphismen untersuchen. Dieses Anreicherungsverfahren startete in einer ersten Studie mit 350 Tausend Positionen des humanen Genoms von Interesse [61] und wurde mittlerweile auf 3,7 Millionen Positionen erweitert [48]. Auf dieser Basis stehen etwa 280 prähistorische Individuen zur Analyse zur Verfügung, jedoch ist anzumerken, dass nur etwa ein Dutzend dieser Individuen eine Abdeckung der Zielregionen zeigen, welche eine diploide Genotypisierung ermöglichen. Dieses Anreicherungsverfahren ist mit etwa 624.000 Positionen durch Affymetrix kommerziell zu erwerben (Axiom[®] *Genome-Wide Human Origins 1 Array*).

Prähistorische und rezente Referenzdaten

Um genetische Zusammenhänge einzelner prähistorischer Individuen oder Populationen zu ergründen, werden diese mit einem aus der Literatur zusammengestellten Referenzdatensatz ko-analysiert. Dafür stehen mittlerweile mitochondriale und genomische Datensätze aus einem breiten Zeit und Raum Fenster zur Verfügung. Besonders NGS Daten können durch methodische Unterschiede oder verwendete Technologie Diskrepanzen in Quantität und Qualität der Daten aufweisen. Unter Verwendung einheitlicher Parameter hinsichtlich sequenzierter Basenqualität, *Mapping* Qualität und analysierter Positionen eines Genoms müssen Datensätze aus diversen Studien zu einem standardisierten Referenzdatensatz zusammengefasst werden.

Diversitätsstudien auf Basis von DNA-Sequenzdaten starteten mit der Entwicklung der Sanger Sequenzierung im Jahre 1975 und erfuhren einen revolutionären Wandel mit dem Aufkommen der Hochdurchsatz Sequenzierung. Das Interesse der Erforschung von Genstrukturen hinter phänotypischen Ausprägungen, sowie der räumlichen Verbreitung der genetischen Variation führte zu Beginn dieses Jahrtausends zur Fertigstellung der Sequenz des menschlichen Genoms [98]. Mehrere Gruppen und Institutionen begannen im Anschluss humane Genome weltweit beprobter Populationen zu sequenzieren. Die Analyse von Populationsstrukturen erlaubt Rückschlüsse innerhalb der Entstehungsgeschichte des AMM und deckt genetische Zusammenhänge einzelner Populationen auf. Während dieses Forschungsgebiet zunächst auf die Analyse uniparentaler Marker (z.B. *Genographic Consortium*, [199]) und vereinzelte Analysen auf genomischer Basis aufbaute, kann heutzutage auf genomische Datensätze einzelner Populationen zurückgegriffen werden.

Im Oktober 2015 wurde die dritte Phase des 1000 *Genomes* Projektes unter Anwendung von *Shotgun* Sequenzierungen und Sequenzierung nach Anreicherung des Exoms veröffentlicht. Koordiniert durch das europäische Bioinformatik Institut (EMBL-EBI) besteht mittlerweile die Möglichkeit auf eine genetische Datenbank mit 2504 Individuen aus 26 Populationen zurückzugreifen (Phase 3, *1000 genomes* Projekt [35]). Eine weitere Quelle populationsspezifischer Referenzdaten bietet das *Human Genome Diversity Cell Line Panel* (HGDP-CEPH), welches durch die Verwaltung von 1050 Zelllinien von Individuen aus 52 Populationen eine Basis für Forschergruppen zur Genotypisierung und Resequenzierung darstellt. Das genetische Material kann in Form von DNA/RNA-Extrakten angefordert werden. In diesem Rahmen stehen diverse Datensätze zur Verfügung, welche abhängig von der Fragestellung genotypisierte Positionen bzw. ganze Genome enthalten. Für die Auswertung paläogenetischer Datensätze stehen auf dieser Plattform sowohl moderne Referenzdaten typisiert durch das *Genome-Wide Human Origins 1 Array* zur Verfügung, als auch tief sequenzierte Genome von 10 Individuen, welche innerhalb der Analyse des Denisova Genoms verwendet wurden [121] ¹.

Kommerziell erwerbbar sind SNP Chips zur Genotypisierung werden von Firmen wie Illumina, MyBait und Affymetrix angeboten, um auf eine kostengünstige Weise die genetische Variation einzelner Individuen zu erfassen, ohne das komplette Genom zu sequenzieren. Diese Genotypisierungen werden vorzugsweise in der Medizin eingesetzt, um in Genom-weiten Assoziierungsstudien (GWAS) schädliche Allele einem Krankheitsbild zuzuordnen. Für die Auswahl der untersuchten Positionen im Genom, die ein SNP Chip wie z.B. Illumina 550W enthält, wurden alle durch den *HapMap*² Datensatz erfasst

¹http://www.cephb.fr/en/hgdp_panel.php; *Dataset 11 Harvard Genetic Department & Dataset 12* Max Planck Institute, Leipzig : MPI-EVA-Denisova

²*HapMap* stellte im letzten Jahrzehnt eine weitere Quelle für rezente Referenzdaten dar, mit dem Ziel Haplotypen innerhalb des menschlichen Genoms zu kartieren. Mit 300 Individuen wurde HapMap in den letzten Jahren durch die Panels 1000genomes und HGDP-CEPH weit übertroffen, so dass dieses Projekt nun eingestellt wurde (Quelle: http://www.ncbi.nlm.nih.gov/variation/news/NCBI_retiring_HapMap)

Polymorphismen hinsichtlich ihrer Allelfrequenz und Genkopplung untersucht. Es wurden jene gewählt, die eine minimale Allelfrequenz (MAF) von mindestens 5% im Rahmen des Gesamtdatensatzes zeigten. Sind SNPs in einem hohen Maß miteinander gekoppelt, d.h. sie werden gemeinsam vererbt, dann wurde lediglich ein SNP dieses Haplotyps für die Analyse gewählt ("*tag*" SNP) [140]. In einer Studie durch Busby und Kollegen (2015) zur Untersuchung der Populationsdynamik in Westurasien wurde aus diversen SNP Chip Genotypisierungen ein Datensatz aus 2192 Individuen, eingegliedert in 144 Populationen, zusammengestellt. Dieser Datensatz enthält 477812 überlappende Polymorphismen je Individuum [28].

Prähistorische Referenzen

Für die Analyse der Position der prähistorischen Population aus Griechenland anhand genetischer Marker wurde ein umfangreicher Referenzdatensatz aus der Literatur zusammengestellt. Die in dieser Arbeit dargestellten Daten umfassen zum einen das Mitochondrium, sowie das gesamte Genom¹. Eine Übersicht der zugrunde liegenden demographischen Geschichte der vorgestellten Individuen bzw. Populationen der Referenzdaten ist in dieser Arbeit in dem Kapitel "Fortschritte der Paläogenetik" zu finden (Abschnitt 1.4 für das Paläolithikum, Mesolithikum und das europäische Neolithikum).

Der prähistorische, **mitochondriale Referenzdatensatz** umfasst chronologisch gruppierte Populationen aus dem späten Paläolithikum bis in die Bronzezeit. Aus geographischer Sicht umspannt er eine Breite von der Ägäis bis in den hohen Norden Europas, wobei der Fokus auf der danubischen Ausbreitungsrouten des Neolithikums liegt. Der Umfang des mitochondrialen Referenzdatensatzes ist in Tabelle 1.1 dargestellt. Die zeitliche Einordnung der Proben erfolgt entweder anhand von ¹⁴C Datierungen oder aus der archäologischen Fundsituation, da abhängig der stratigraphischen Einordnung eines Fundes Angaben zum Alter einer Probe getätigt werden können. Insgesamt konnte durch Literaturrecherchen ein Datensatz aus mehr als 1000 Individuen, eingegliedert in 28 Populationen erstellt werden. Die mitochondriale Datenbank umfasst die HVS-I Region (Position 16051 - 16400), da diese den kleinsten gemeinsamen Nenner zwischen den verfügbaren Daten aus der Literatur darstellt (siehe Kapitel 6.2). Vollständige mitochondriale Genome sind aus allen genomischen Studien verfügbar (siehe Tabelle 1.1), sowie aus Einzelstudien nach Anreicherung des mitochondrialen Genoms (Blätterhöhle [18], europäisches Paläolithikum [143] und europäisches Neolithikum, H-Linien [23]).

¹Zusammenfassung der Publikation Hofmanová, Kreutzer *et al.* (2016)

Tabelle 1.1: Der mitochondriale Referenzdatensatz (Stand März 2017). Die Einordnung der nordeuropäischen Populationen erfolgte nach dem Schema der Dissertation von Anna Schulz (Institut für Anthropologie, Universität Mainz) [170]. In der Spalte "genomische Studien" sind Individuen aufgelistet, für welche bereits Analysen von genomischen Daten bereit stehen (dargestellt in Abbildung 1.6).

Gruppe	Abkürzung	Anzahl Individuen	Probenalter (kal. v.Chr.)	Kulturstufe	Referenzen	genomische Studien
Oberes Paläolithikum	Pre-LGM	16	40.000 - 25.000	Gravettien/Aurignacien	[143, 11, 91]	Kostenki [171]
Spätes oberes Paläolithikum (post-LGM)	SOP-JS	16	17.000 - 11.000	Epigravettien/Magdalenien	[143, 47, 77, 170]	Bichon [77]
Holozän Jäger-Sammler	Holo-JS	19	10.000 - 6000	Mesolithikum	[19, 18, 47, 102]	Loschbour [102]
Präneolithisch Nord-Ost Europa	NEE-JS	23	9000 - 5800	Mesolithikum	[19, 41, 170]	-
Präneolithisch Nordeuropa	NE-JS	25	7600 - 4800	Mesolithikum	[19, 174, 102, 170]	StoraFörvar11 [174], Motala [102]
Paraneolithisch südliches Baltikum	Neo-JS	18	4000 - 3000	Paraneolith. Mesolithiker	[19, 170]	-
Paraneolithisch Skandinavien	PWC-JS	28	3000 - 2400	Pitted Ware Culture	[109]	Ajvide [174]
Frühes Neolithikum Türkei	ANA-EN	27	6500 - 5600	Fikirtepe	[114, 71]	BAR [71]; Barcın, Mentese [114]
Frühes Neolithikum Ungarn/Kroatien	STA-EN	44	6000 - 5500	Starčevo	[181]	BAM22, 25 [61]
Frühes Neolithikum Alföld (Ungarn)	KÖR-EN	18	6000 - 5600	Körös	[81, 53]	KO1/2 [53]
Frühes Neolithikum Transdanubien (Ungarn)	LBK-T-EN	39	5300 - 4900	LBK Transdanubien	[181]	SZEH4 [61]
Mittleres Neolithikum Alföld I (Ungarn)	ALBK-I-MN	62	5500 - 5200	frühe Alföld LBK / Szatmar	[81, 53]	NE1, NE4 [53]
Mittleres Neolithikum Alföld II (Ungarn)	ALBK-II-MN	119	5300 - 4900	späte Alföld LBK (Szakálhát, Tiszadob, Esztár)	[81, 53]	NE2/3/5/6 [53]
Spätes Neolithikum Alföld (Ungarn)	TIS-LN	48	~ 4750	Tisza	[81]	-
Frühes Neolithikum Mitteleuropa	LBK-EN	120	5500 - 4775	LBK	[60, 59, 23, 18, 20]	Stuttgart [102]
Mittleres Neolithikum Mitteleuropa	CE-MN	49	4600 - 4000	Rössen, Schöningen	[23, 20, 170]	Rössen_EN, Schöningen_MN [61]
Spätes Neolithikum Mitteleuropa	CE-LN	64	4000 - 2800	Baalberge, Salzmünde	[20, 23]	Baalberge_MN, Salzmünde_MN [61]
Finales Neolithikum Mitteleuropa	CE-FN	64	2800 - 2200	Corded Ware, Glockenbecher	[20, 23]	BZH, QLB, ROT [61]
Frühes Neolithikum Westeuropa	GUR	39	5000 - 4000	LBK- und Kardialelemente	[156]	-
Frühes Neolithikum Iberia	WE-EN	29	5475 - 4960	Karidialkultur	[52, 96, 160, 126]	Spain_EN [61], CB13 [126]
Spätes Neolithikum Iberia	WE-LN	61	4335 - 3635		[52, 160]	Spain_MN [61]
Finales Neolithikum Iberia	WE-FN	86	2900 - 1740	teilw. Glockenbecherkultur	[160]	
Frühes Neolithikum Nordeuropa	NE-EN	22	3600 - 2500	Trichterbecher Kultur	[109, 170]	Gökhem [174]
Spätes Neolithikum Nordeuropa	NE-FN	11	2500 - 1800	Trichterbecher Kultur	[170, 115]	
Bronze Zeit Kreta	MIN-BA	37	2450 - 1750	Minoer	[73]	-
Bronze Zeit Steppe	YAM-BA	28	3300-2300	Yamnaya	[204]	Yamnaya aus Haak (2015) [61]
Bronze Zeit Mitteleuropa	CE-BA	94	2200 - 1750	Unetice	[20]	Unetice_EBA [61]
Späte Bronzezeit Steppe	CAT-BA	23	2450-1950	Katakombengrabkultur	[204]	-

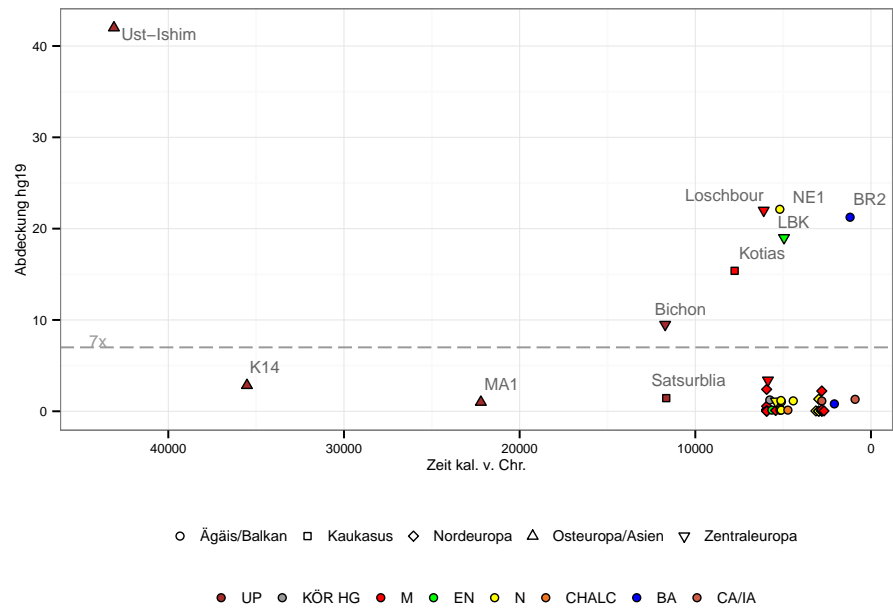
Autosomale Sequenzdaten

Mit der Sequenzierung des Neandertaler und Denisova Genoms im Jahre 2010 [57, 148] sind **paläogenomische Sequenzdaten** für den anatomisch modernen Menschen (AMM) ab dem Jahr 2012 stetig am zunehmen. Mittlerweile kann auf prähistorische Referenzen von weit über 300 Individuen zurückgegriffen werden, jedoch besitzen lediglich 7 dieser Datensätze eine durchschnittliche Abdeckung des gesamten Genoms von 7x (Stand Mai 2016).

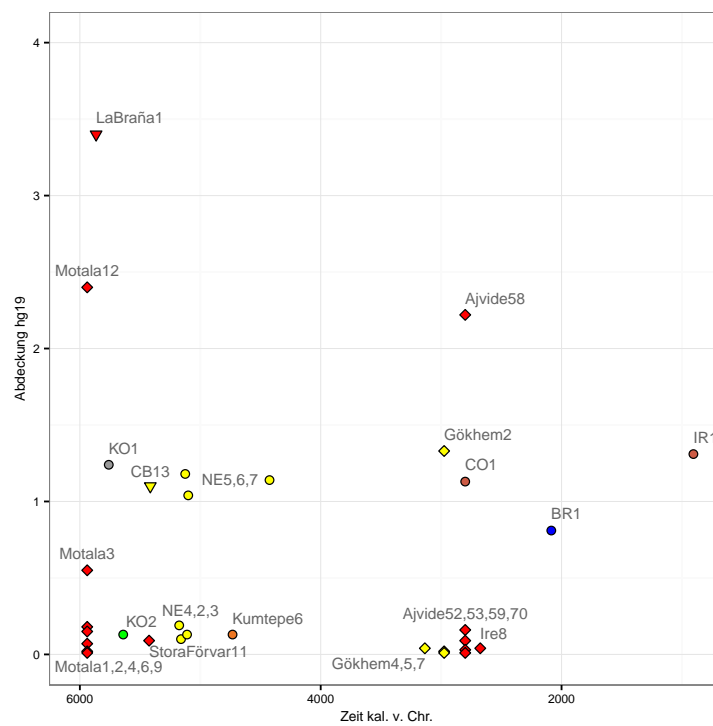
Die Abdeckung des Genoms (*coverage*, Anzahl sequenzierter Moleküle je Position eines Genoms) ist ein wichtiges Kriterium während der Bestimmung eines Genotyps. Desaminierungen alter DNA führen zu fehlerhaft ausgelesenen Basen, so dass eine Position von mindestens 7 ausgelesenen Sequenzen abgedeckt sein sollte. Die meisten paläogenetischen Genome haben eine Abdeckung unter 7x, so dass eine eindeutige diploide Genotypisierung nicht möglich ist. An dieser Stelle wird zur Ermittlung eines Genotyps eine haploide Allel Präsenz Bestimmung durchgeführt, wobei unter Einbeziehung einer Basenqualität von >30 zufällig eine vorhandene Allelvariante ausgewählt wird (siehe [57, 174]). Eine erweiterte Form ist die "Majority Typisierung", bei welcher das in Überzahl vorkommende Allel an einer Position gepickt wird [61]. Um Fehler durch Desaminierungen zu vermeiden, werden lediglich Transversionen zur Referenzsequenz für eine statistisch genomische Auswertung herangezogen. Um auch Transitionen in eine populationsgenetische Analyse der Sequenzdaten mit einzubeziehen, wurde an der Universität Fribourg (Schweiz, Wegmann Lab) eine Methode der Genotypisierung speziell für stark geschädigte DNA entwickelt, welche Desaminierungsraten erlernt und während der Genotypisierung berücksichtigt [90]. Diese Methode kam erstmals in der Publikation Hofamnová, Kreutzer *et al.* (2016) zum Einsatz, welche genomische Analysen der vorliegenden Arbeit enthält [71].

Durch *Shotgun* Sequenzierung stehen zusätzlich 32 Individuen mit einer Abdeckung von 0,01x - 4x zur Verfügung, welche durch eine Allel Präsenz Typisierung für statistische Analysen herangezogen werden können; eine graphische Übersicht bietet Abbildung 1.6. Es wird deutlich, dass sich dieses Forschungsgebiet gerade in den letzten Jahren verstärkt auf das Holozän und die Neolithisierung Europas konzentriert (siehe Abbildung 1.6b)¹. Aus dem Datensatz der Arbeitsgruppe von David Reich stehen etwa 280 prähistorische Individuen in Form von Genotypen (PLINK Format aus [61]) oder als Sequenz Alignments zur Verfügung. Innerhalb dieses Spektrums sind die Individuen des HGDP-CEPH typisiert worden und dienen als moderne Referenz (HGDP-CEPH *Dataset 11*). Dieser Referenzdatensatz eignet sich durch die Wahl der Positionen für die Aufklärung von Populationsstrukturen mittels statistischen Analysen, jedoch nicht für demographische Modellierungen. Die Positionen innerhalb dieses Anreicherungssystems wurden so gewählt, dass sie zum einen eine Abdeckung in bereits erstellten Datensätzen des Schimpansen, Denisova Menschen und Neandertaler zeigen und zum anderen populationspezifisch eine hohe Abgrenzungsmöglichkeit bieten. Dabei wurden Lokalisationen wie die Nähe zu Genstrukturen oder auch die vollständigen Erfassung einzelner Haplotypen außer Acht gelassen (siehe technische Beschreibung [107]). Einzelne Individuen der Datensätze nach Haak 2015, Mathieson 2015 und Fu 2016 besitzen eine Abdeckung der Zielregion von weit unter 1x. In diesen Studien liegt der Fokus auf der Anzahl an SNPs, welche über Allel-Präsenz Typisierung in eine Analyse einfließen. Mit variierender Anzahl untersuchter Positionen zwischen den genannten Studien ist es schwierig Angaben zur Abdeckung zu treffen, so dass diese Daten nicht in Abbildung 1.6 dargestellt werden.

¹Populationen der eurasischen Bronze Zeit [2] werden in der vorliegenden Arbeit nicht in den Referenzdatensatz einbezogen, da diese nicht im Rahmen der Fragestellung liegen.



(a) Gesamtansicht Referenzgenome (ohne archaische Menschenformen).



(b) Vergrößerung auf Proben aus dem Holozän.

Abbildung 1.6: Übersicht paläogenetischer Referenzgenome nach *Shotgun* Sequenzierung, dargestellt anhand Datierung des Individuums (kalibriert v. Chr.) und Abdeckung des Genoms (Referenzgenom GRCh37/hg19). Grundlage dieser Darstellung ist der Referenzdatensatz aus [71], *Additional Dataset S1*; weitere Informationen der Proben können dieser Tabelle entnommen werden. Die gestrichelte Linie zeigt eine Abdeckung von 7x. Abkürzungen: UP= Oberes Paläolithikum, KÖR HG= Körös Jäger Sammler, M=Mesolithikum, EN=frühes Neolithikum, CHALC= Chalkolithikum, BA=Bronzezeit, CA/IA=Kupferzeit/Eisenzeit.

1.3 Genetische Diversität - Entstehung und Nachweis

Das **Mitochondrium** entspricht einem kompakten Zellorganell, welches ein autonom replizierendes, zirkuläres Genom mit einer Länge von 16569bp enthält. Spermatozoen enthalten nur wenige Mitochondrien, welche bei der Befruchtung nicht in die Eizelle eindringen. Man spricht von einem uniparental vererbten Marker, da nur die mitochondriale DNA der Mutter an die Folgegeneration weitergegeben wird, analog zum Y-Chromosom, welches vom Vater auf die männlichen Nachkommen vererbt wird. Uniparentale Marker haben keine Möglichkeit der Rekombination des genetischen Materials und stellen damit eine genetische Linie dar, auf der sich mit zunehmender Evolutionszeit Mutationen anhäufen. Einzelne mitochondriale/Y-chromosomale Linien werden, definiert über ihre Mutationen, in separate, der Abstammungslinie entsprechende Haplogruppen eingeordnet (siehe Abbildung 1.7 zur Einteilung der mitochondrialen DNA Linien ¹).

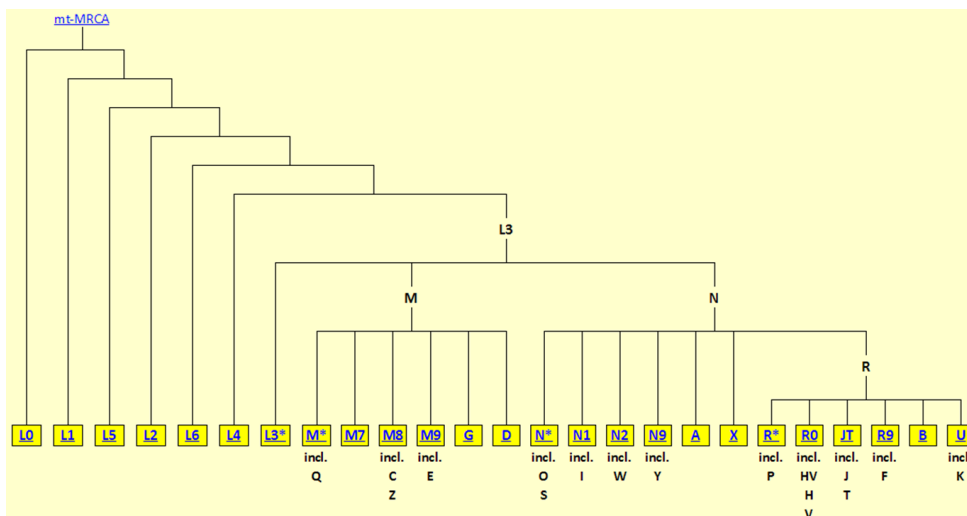


Abbildung 1.7: Darstellung weltweit erfasster mitochondrialer Linien, phylogenetisch eingegliedert in Haplogruppen (Quelle: *mtDNA Tree Build 17* (18. Feb. 2016), <http://www.phylotree.org/tree/index.htm> [193]).

Innerhalb des **autosomalen genetischen Materials** werden physikalisch auf einem Gameten lokalisierte Gene in ihrer Ausprägungsform gemeinsam an die Folgegeneration vererbt (Genkopplung). Unregelmäßigkeiten in ihrem Vererbungsschema können durch homologe, meiotische Rekombination während der Zellteilung in der Keimbahn hervorgerufen werden, welche einen Austausch von Chromosomenabschnitten zur Folge hat. Im Gegensatz zur mitochondrialen bzw. Y-chromosomalen DNA bietet diese Möglichkeit die Grundlage der Einzigartigkeit des autosomalen Datensatzes eines Individuums und eine gesteigerte genetische Diversität innerhalb von Populationen. Neue Kombinationen werden an die Folgegeneration weitergegeben und bieten der Evolution die Option zur Ausbildung neuer Merkmale, welche durch Selektion forciert bzw. eliminiert werden können. Basiert z.B. eine phänotypische Prägung auf mehreren genetischen Elementen (Polygenie), führt Rekombination zu neuartigen Merkmalsausprägungen.

Die Rekombinationsrate (r) drückt die Wahrscheinlichkeit aus, dass ein Rekombinationsereignis zwischen zwei Loci stattfindet. Innerhalb der genomischen Architektur existieren unterschiedliche Re-

¹Analog steht zur Einteilung Y-chromosomaler Linien ein Stammbaum unter <http://www.phylotree.org/Y/tree/> zur Verfügung, dessen Einteilung auf Y-spezifischen SNPs besteht, welche unter http://www.phylotree.org/Y/marker_list.htm abgerufen werden können.

kombinationsraten. Während genomische Regionen nahe der Centromere eine geringere Rekombinationswahrscheinlichkeit als Telomerregionen zeigen, gibt es Rekombinations "hot spots" die räumlich nahe von PRDM9 Loci liegen, welche für eine meiose-spezifische Histon-Methyltransferase [9] kodieren. Da gekoppelte Abschnitte den gleichen evolutionären Prozessen unterliegen, kann im Falle von Selektion eines der Allele der gleiche selektive Effekt auf eine gekoppelte, aber neutrale Position ausgeübt werden - man spricht im Falle von positiver Selektion von *Hitchhiking* und bei Verdrängung dieser Varianten aus dem Genpool von einer im Hintergrund ablaufender Selektion (*Background selection*). Das Ausmaß an Rekombination über die Zeit wird anhand der Zusammensetzung der Allele eines Haplotyps innerhalb einer Population berechnet. Maß für die Korrelation zweier Abschnitte zueinander ist der LD *linkage disequilibrium*-Wert, welcher mit dem Abstand zweier Positionen auf einem Chromosom (Centi Morgan) korreliert. Nachdem durch Mutation ein neuer Haplotyp generiert wird (maximaler LD Wert), werden durch Rekombination in Folgegenerationen Haplotypen neu kombiniert, so dass die Kopplung vermindert wird (LD Wert sinkt) und sich mit der Zeit ein Kopplungsgleichgewicht einstellt (*linkage equilibrium*, LD=0).

Mutation und Rekombination sind Kräfte der Evolution welche die genetische Diversität auf individueller Ebene formen, indem neue Allele ausgebildet und kombiniert werden. Abhängig der Zusammensetzung und Geschichte einer Population haben diese Kräfte unterschiedliche Auswirkungen auf die genetische Diversität. Die Ausprägung der genetischen Diversität unterliegt einem stochastischen Prozess, welcher zu einer Verschiebung von Allelfrequenzen in der Subpopulation im Vergleich zur Ursprungspopulation führt (genetische Drift). Die Wirkungsweise der genetischen Drift ist abhängig von der effektiven Größe einer Population (N_e). Während in Populationen mit vielen Individuen Allelfrequenzen konstant bleiben, kommt es in isolierten Populationen rasch zur Fixierung oder Eliminierung einzelner Allele [205]. Eine strukturierte Population basiert folglich darauf, dass Panmixie nur begrenzt in Subpopulationen stattfindet. Neben der geographischen Distanz können auch kulturelle Werte (z. B. Religion, Kastensystem) zur Bildung von Subpopulationen beitragen. Variierende Umwelteinflüsse können einen selektiven Druck auf die Individuen der Subpopulation ausüben, so dass bestehende Genvarianten, welche in der Ursprungspopulation keinen Einfluss haben, in einer neuen Umgebung der Subpopulationen positiv wirken. Folglich kann ein der Umwelt angepasstes Allel eine schnelle Fixierung innerhalb der Subpopulation durch positive Selektion erlangen.

Isolierte Populationen entstehen durch Expansion in neue Gebiete (Gründereffekt, *founder effect*). Des weiteren kann es beispielsweise durch Naturgewalten oder Epidemien zu Populationseinbrüchen kommen, welche ebenfalls die genetische Variabilität reduzieren oder eliminieren (Flaschenhals-Effekt, *bottleneck*). Im Gegensatz dazu kann eine Erhöhung der genetischen Diversität durch Migration von Individuen von einer Subpopulation in eine benachbarte hervorgerufen werden (Genfluss, *gene flow*).

Die Erfassung der genetischen Diversität

Die Beschreibung der genetischen Diversität basiert auf der Ausprägung polymorpher Stellen des Genoms innerhalb homologer Positionen bzw. Sequenzdaten. Innerhalb von Populationen kann für diese Positionen eine Frequenz für die vorliegenden Allele errechnet werden (Genotypenfrequenz). Bei Betrachtung mehrerer Positionen des Genoms können auf diese Weise deskriptive Parameter, wie die Gendiversität oder die Nukleotiddiversität innerhalb einer Population, erhoben werden. Beim Vergleich dieser Parameter zwischen Populationen können demographische Ereignisse angenommen werden, wie z.B. eine abnehmende genetische Vielfalt durch Kolonisation neuer Gebiete, da nur ein Anteil der genetischen Diversität durch abwandernde Individuen mitgenommen wird. Die Heterozygotenrate beschreibt allgemein den Anteil an bi- bzw. triallelischen Positionen innerhalb einer Population und ist

damit ein Abbild der genetischen Diversität. Unter der Annahme von Hardy-Weinberg Bedingungen sollte diese in der Folgegeneration konstant bleiben, wirken jedoch besondere demographische Ereignisse, ändern sich Allelfrequenzen und es entsteht ein Ungleichgewicht. Das Paarungssystem innerhalb und die effektive Populationsgröße von Populationen haben einen großen Einfluss auf die Heterozygotenrate. Der Extremfall der Inzucht ruft z.B. eine starke Reduktion der genetischen Diversität hervor (Heterozygotie sinkt), da während der Rekombination Abschnitte aufeinander treffen, welche identisch in ihrer Abstammungslinie sind (IBD, *identical by descent*), so dass die Autozygotie innerhalb einer Population steigt. Gleiche Effekte können in isolierten Populationen oder durch einen Einbruch der Populationsgröße entstehen. Durch die Messung von homozygoten Abschnitten innerhalb eines Individuums oder auch gemittelt in Populationen, kann auf diesem Weg eine Aussage zur Geschichte einer Population in Bezug auf Paarungseigenschaften und Populationsgröße getroffen werden (ROH; *runs of homozygosity* [85]).

Mitochondriale Datenanalyse

Bei der mitochondrialen Datenanalyse werden die mitochondrialen Abstammungslinien zunächst einzelnen Haplogruppen zugeordnet, eine Nomenklatur, welche die Klassifizierung der beobachteten Linien vereinfacht. Für eine deskriptive Analyse der Sequenzdaten werden die Daten einzelner Individuen zu Populationen zusammengefasst, gruppiert nach kulturellen, chronologischen und geographischen Parametern. Folglich stellt eine Population eine Stichprobe einer expliziten Kulturstufe innerhalb der Besiedlungsgeschichte des AMM über Europa dar. Der in dieser Arbeit verwendete mitochondriale Referenzdatensatz umfasst humane Populationen aus dem Paläolithikum bis in die heutige Zeit (siehe Tabelle 1.1). Zur Beschreibung der genetischen Diversität wird innerhalb einer Population die **Haplotypendiversität** bestimmt (z.B. nach Nei). Die Gendiversität nach Nei beschreibt die Wahrscheinlichkeit, dass sich zwei Haplotypen, zufällig gezogen aus einer Population, unterscheiden. Liegen Mutation und Drift in einem Gleichgewicht, sollten sich die Diversitätswerte zwischen der Ursprungs- und Folgegenerationen nicht unterscheiden. Eine Erhöhung der Diversität findet durch Migration statt, wohingegen eine Reduktion durch den Gründer- bzw. Flaschenhalseffekt hervorgerufen werden kann. Analog hierzu kann auf Sequenzebene die Nukleotiddiversität berechnet werden, welche die Unterschiede zwischen Haplotypen miteinbezieht.

Populationsstrukturen können in mitochondrialen Datensätzen mittels des Fixierungsindex (F_{ST}) dargestellt werden, wobei eine genetische Distanz zwischen allen Populationen bzw. Subpopulationen errechnet wird. Anhand der Varianz der Allelfrequenzen der einzelnen Populationen (Subpopulationen) im Vergleich zur Varianz der Metapopulation (totale Population), wird ein Wert zwischen 0 und 1 errechnet, welcher die genetische Distanz darstellt. Eine Datenmatrix genetischer Distanzen zwischen $m \times m$ Populationen kann durch eine multidimensionale Skalierung (MDS) auf 2 Vektoren reduziert werden. Gegeneinander aufgetragene Vektoren in einem Streudiagramm stellen die genetischen Distanzen der einzelnen Populationen im Kontext der Variation aller erfasster Populationen graphisch dar.

Autosomale Datenanalyse

Der individuelle Charakter eines autosomalen Datensatzes, untersucht anhand von Millionen variabler und unabhängiger Positionen, bietet eine tiefgreifende Möglichkeit der Differenzierbarkeit von Individuen und Populationen, welche durch uniparentale Marker nicht gegeben sind. Eine formelle Statistik zur Visualisierung von Populationsstrukturen bietet die Hauptkomponentenanalyse, bei

der die genetische Variation, abgebildet durch die Genotypen von Einzelindividuen, anhand der ersten zwei Hauptkomponenten dargestellt wird. Weitere quantitative Analysen bestimmen das Maß an genetischer Drift zwischen einzelnen Populationen (f_3/f_4 -Statistik).

Die Hauptkomponentenanalyse (*principal component analysis*, **PCA**) wurde 1978 von Cavalli-Sforza und Kollegen erstmals zur Analyse genetischer Strukturen genutzt. Anhand von Blutgruppenpolymorphismen (HLA Marker) wurde die detektierte Variation innerhalb der europäischen Population in die ersten fünf Hauptkomponenten zerlegt und jeweils auf einer geographischen Karte dargestellt ("synthetic map"). Die Autoren interpretierten die geographische Änderung der Variation durch historische Migrationsereignisse und sprachen sich für eine demische Migrationsbewegung der ersten Farmer in Europa aus [116]. Dass das interpretierte Muster jedoch lediglich ein mathematisches Artefakt darstellt, hervorgerufen durch "isolation by distance" der genetischen Muster zur Ursprungspopulation, zeigten Novembre & Stephens im Jahre 2008 [125]. Die Darstellung der genetischen Diversität durch einen Biplot der ersten zwei Hauptkomponenten eines Datensatzes spiegelt stets ein Muster wieder, welches einer geographischen Kartierung gleicht, jedoch kann aus diesem Abbild kein spezifisches Migrationsereignis abgelesen werden. Folglich zeigt eine PCA eine Strukturierung des vorliegenden Datensatzes an, wobei die Ursache dieser rein spekulativ ist. Während zunächst nur wenige Positionen eines Genoms zugänglich waren, wurde diese Analyse anhand von Allelfrequenzen einzelne Marker auf Populationsebene durchgeführt - im Prinzip ähnlich zur graphischen Darstellung von genetischen Distanzen per MDS (siehe oben). Mit steigender Sequenzdatenmenge und Abdeckung von mehreren 1000 Positionen eines Genoms, kann die Strukturierung eines Datensatzes auf individueller Ebene vollzogen werden [139]. Dabei bildet eine Datenmatrix aus m Individuen \times n Positionen des Genoms die Grundlage. Es werden biallelische, nicht gekoppelte SNPs verwendet und anhand ihrer Ausprägungsform codiert (vorhandener Genotyp = 0, 1, 2). Da weniger Individuen als Positionen verwendet werden ($m < n$) muss vor Anwendung einer PCA der Datensatz in eine quadratische Matrix transformiert werden (Dimension der Matrix = Anzahl Individuen). Es wird eine lineare Transformation durchgeführt, wobei die Genotypen eines Individuums innerhalb der Varianz der Allelfrequenzen aller Individuen durch m Eigenwerte (*eigenvalue*) dargestellt werden [139]. Eine folgende Hauptkomponentenanalyse reduziert die genetische Variabilität auf m Hauptkomponenten, welche als Eigenvektoren bezeichnet werden. Die ersten zwei Eigenvektoren erfassen anteilig die höchste Diversität eines Datensatzes und können durch einen Biplot visualisiert werden. Dabei stellt die relative Distanz der Individuen zueinander das Maß an gemeinsamen genetischen Varianten dar [178]. Wie in Abbildung 1.6 dargestellt, genügt die Abdeckung des Genoms in paläogenetischen Studien meist nicht für eine diploide Genotypisierung. An dieser Stelle werden die vorhandenen Allele bestimmt und durch *in silico* Duplikation in eine diploide, homozygote Form überführt. Die Hauptkomponenten eines modernen Referenzdatensatzes werden ermittelt und bilden einen Rahmen, in welchen die Genotypen eines alten Individuums mittels *Procrustes* Transformation abgebildet werden [175]. Für die Publikation der genomischen Daten dieser Arbeit wurde das Programmpaket LASER [198] verwendet, welches die Genotypenwahrscheinlichkeit der Individuen aus einem *Alignment* extrahiert und auf einen Referenzdatensatz projiziert.

Unter Annahme, dass Populationen die gleiche evolutionäre Geschichte durchlaufen haben, können zur Quantifizierung des Anteils an gemeinsamer genetischer Drift zwischen Populationen **F-Statistiken** durchgeführt werden. Eine molekulare Phylogenie basiert auf einer Distanzmatrix paarweiser Unterschiede der Sequenzdaten einzelner Individuen/Populationen. Analog werden in der F-Statistik [149] Populationen einer Phylogenie auf Basis der Kovarianz der Allelfrequenzen miteinander in Beziehung gesetzt. Das einfachste Szenario ist die Messung der Unterschiede zwischen zwei Populationen $F_2(P_1;P_2)$ anhand der Varianz ihrer Allelfrequenzen, wobei die Astlänge innerhalb des Stamm-

baums quantifiziert wird ($F_2 = \sum (p_1 - p_2)^2$). Die *Outgroup* f_3 -Statistik bestimmt unter Annahme einer Phylogenie dreier Populationen die gemeinsame genetische Variation zweier Testpopulationen in Abhängigkeit der heterozygot vorliegenden Positionen einer Außengruppe [145]. Grundlage dieser Teststatistik ist die Verwendung einer Population als Außengruppe (P_{Out}), welche nach dem Split keinen Genfluß zu den Testpopulationen aufweist. Um den Anteil an gemeinsamer genetischer Drift mehrere Testpopulationen mit einer speziellen Population in Relation zu setzen, wird neben der Außengruppe eine zweite Population fixiert (P_1) und eine unbegrenzte Anzahl an Testpopulationen (P_2, P_3, \dots, P_X) können rotierend innerhalb dieses Testszenarios untersucht werden (siehe Abbildung 1.8). Die Statistik errechnet sich anhand der Allelfrequenzen der verwendeten Populationen $F_3(P_{Out}; P_1, P_2)$. Für die Auswertung einer *Outgroup* f_3 -Statistik bedeutet dies, je positiver der bestimmte Wert zweier Testpopulationen ist, desto geringer ist die genetische Distanz - in Abbildung 1.8 entspricht dies der gemeinsamen Astlänge der zwei untersuchten Populationen (rot).

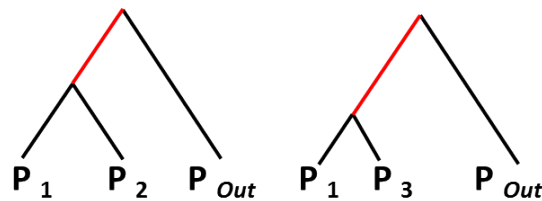


Abbildung 1.8: Graphische Darstellung der *Outgroup* f_3 -Statistik. In dieser speziellen Konstellation der F-Statistik wird eine Population fixiert (P_1) und der Wert an gemeinsamen Allelen zu den Testpopulationen P_2, P_3 in Abhängigkeit zu einer Außengruppe (P_{Out}) bestimmt.

Während die *Outgroup* f_3 -Statistik eine genetische Distanz im Rahmen einer Teststatistik quantifiziert, kann in der f_4 -Statistik eine weitere Population herangezogen und die Richtung des Genflusses innerhalb dreier Testpopulationen (A, B, C) in Relation zu einer Außengruppe (D) bestimmt werden. Unter Annahme eines Stammbaummodells mit den Kladen (A, B) & (C, D) besteht unter der Nullhypothese $f_4 = 0$ kein Genfluss zwischen beiden monophyletischen Gruppen (siehe Abbildung 1.9 (a)). Alternativ beschreiben zwei Hypothesen die Richtung des Genflusses: (1) Ein positiver f_4 -Wert zeigt Genfluss von C nach A ((A, C; B, D) - A und C bilden Klade, 1.9 (b)) und (2) ein negativer f_4 -Wert zeigt Genfluss von C nach B ((B, C; A, D) - B und C bilden Klade, 1.9 (c)).

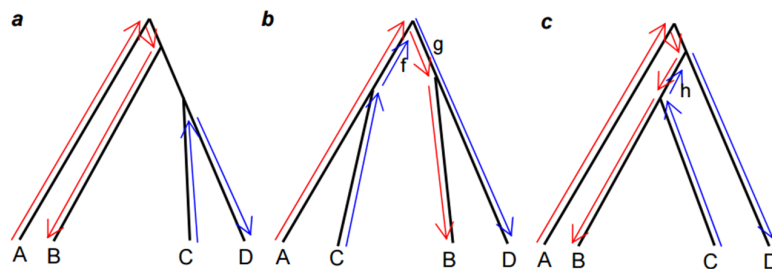


Abbildung 1.9: Quantitative Messung gemeinsamer genetischer Drift innerhalb einer Phylogenie von 4 Populationen - entnommen aus Reich *et al.*(2009) [149]

Modell-basierte *Cluster* Analysen bilden eine Population als Mischung anzestralischer Populationen ab (**ADMIXTURE** [1]). Im Spektrum eines autosomalen SNP Datensatzes mehrerer Individuen und/oder Populationen werden sequentiell Abschnitte des Genoms abgelesen und es wird nach gleichen genetischen Komponenten in Form von Allelfrequenzen nicht-gekoppelter Loci gesucht. *A priori*

werden k ancestrale Populationen angenommen, so dass die Genome des Datensatzes in Vektoren gegliedert werden, welche anteilig auf die ancestralen Populationen zurückgeführt werden können. Durch die Zusammensetzung eines individuellen Genoms aus k herkunftsspezifischer Komponenten können Rückschlüsse zu seinen Vorfahren getroffen werden.

Während die vorangegangenen Methoden auf der Analyse individueller SNPs beruht und ihre Lokalisation innerhalb des Genoms vernachlässigt wird, gibt es eine weitere, welche auf der Analyse von Haplotypen basiert. Da Mutationen auf einem Chromosom gemeinsam vererbt werden, kann innerhalb autosomaler SNP Datensätze nach herkunftsspezifischen Haplotypen innerhalb der Chromosomensegmente gesucht werden. Zur vereinfachten graphischen Darstellung werden Abschnitte gleicher Herkunft einem Farbschema zugeordnet, weshalb das Verfahren auch als *Chromosome painting* bezeichnet wird. Aufgrund von Rekombination enthält ein Chromosom Abschnitte unterschiedlicher Herkunft (z.B. nach *Admixture*), so dass diese Analyse feinste Populationsstrukturen identifiziert und anteilig die Herkunftskomponenten innerhalb einzelner Individuen/Populationen berechnet. Damit werden Haplotypen als einzelnen Marker der Evolution mit ihrer eigenen Geschichte identifiziert. Es wird ein "Empfänger" Individuum angenommen, dessen Genom aus Haplotypen der "Donor" Individuen bzw. Populationen in Form einer Kopie zusammengesetzt ist. Anschließend kann innerhalb des Empfänger Genoms die prozentuale Zusammensetzung der Donor Segmente errechnet werden (*Allel-Matching Profil*) [67, 68]. Gegenüber *ADMIXTURE* Analysen besitzt diese Methode den Vorteil, dass mehr als 100 ancestrale Populationen angenommen werden können, während bei *ADMIXTURE* Analysen dies eine zu hohe Rechenleistung darstellt (maximal $k=10$). Folglich ist durch das *Chromosome painting* eine feinere Analyse möglich.

Populationsmodelle: Kontinuität und Analyse demographischer Parameter

Grundlage genetischer Kontinuität sind gleichbleibende allelische Muster zwischen Populationen zweier Zeitstufen innerhalb einer geographischen Region. Unter Annahme von genetischer Drift als alleinig wirkende Kraft kann Kontinuität zwischen Populationen zweier Zeitstufen durch demographische Prozesse wie z.B. Migration aufgehoben werden, da neu eingebrachtes genetisches Material die Zusammensetzung innerhalb einer Population verändert. Um genetische Drift zwischen zwei Zeitstufen darzustellen werden multiple Vorwärtssimulationen genutzt. Auf Basis des genetischen Materials der alten Population (Zeitpunkt 0) wird unter Annahme grundlegender populationsgenetischer Parameter (Populationsdichte, Wachstumsrate) die Folgegeneration erstellt. Die Simulation endet nach x Generationen mit Erreichen der nächsten Zeitstufe. Durch die Erhebung deskriptiver Parameter der genetischen Diversität innerhalb alten und simulierten Population kann eine Aussage zur Kontinuität getroffen werden.

1.4 Fortschritte der Paläogenetik

Paläolithikum und Mesolithikum

Mit einem Alter von 45.000 Jahren ist das Individuum aus Ust’Ishim (Sibirien) eines der ältesten, genetisch erfassten Funde des AMM [46]. Seine mitochondriale Linie R* und autosomal-genetische Studien platzieren dieses Individuum in eine Population, welche nach der Migration aus Afrika den asiatischen Kontinent besiedelt hat. Ust’-Ishim zeigt im Vergleich zu heute lebenden AMM Introgressionen von Neandertal Genen in langen Sequenzabschnitten. Dies weist auf ein Paarungsereignis vor mehr als 45.000 Jahren hin, jedoch nur wenige Tausend Jahre vor der Lebzeit von Ust’Ishim, da die Kräfte der Rekombination die eingebrachten Sequenzabschnitte noch nicht in kleinere aufgebrochen haben.

Eine weitere genetische Studie zweier Individuen einer Jäger-Sammler (JS) Kultur aus dem Kaukasus (Satsurblia, Kotias, [77]) weist auf eine Strukturierung der JS Population während des Paläolithikums hin. Nach der Besiedlung der Levante spalteten sich die aus Afrika kommende Population vor 40 - 50.000 Jahren auf - während sich eine in nordöstliche Richtung verbreitete und die Basis für eine kaukasische JS Population darstellte (CHG, *caucasian hunter gatherer*), kolonisierten Vorfahren der westlichen Jäger-Sammler (WHG, *western hunter gatherer*) im späten Paläolithikum den europäischen Kontinent (siehe Abbildung 1.10).

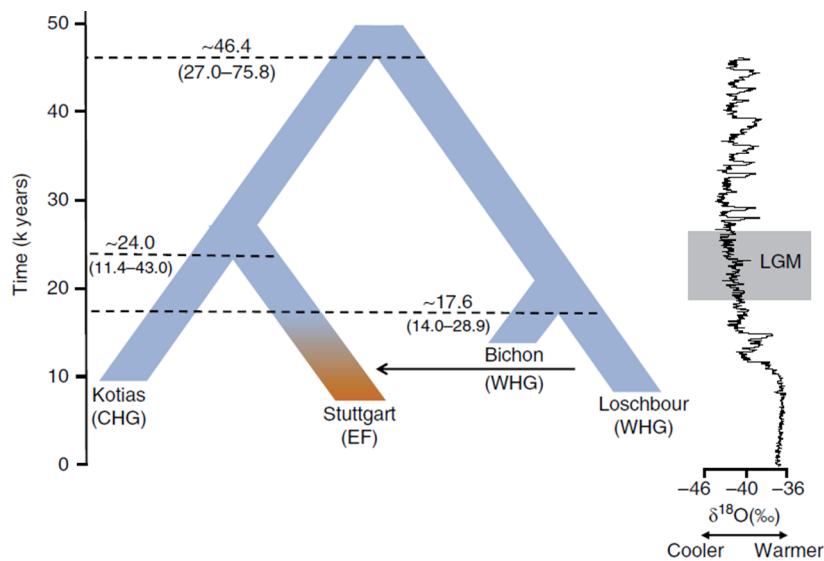


Abbildung 1.10: Populationsstrukturen im späten Paläolithikum Eurasiens (entnommen aus [77]).

Mitochondrialer Genome aus dem europäischen Pleistozän zeigen während der ersten Kolonisation des europäischen Kontinents M-, R- und U*-Linien. Eine Studie von Bramanti und Kollegen (2009) zeigte nach der Maximalvergletscherung des europäischen Kontinents ein homogenes Bild der Linie U5 innerhalb der westeuropäischen JS Populationen (WHG), welches sich bis in das Holozän fortsetzt [19]. Die Eiszeit und der damit einhergehende Rückgang der Populationsgröße führte zu einem Verlust der M- und R-Linien, folglich wird die Theorie der Verringerung der Populationsgröße durch genetisches Daten unterstützt [143]. Ein genetisches *bottleneck* kann ebenfalls auf Basis der Heterozygotenrate autosomaler Daten der nord- und zentraleuropäischen Individuen von Loschbour [102] und Ajvide [174] gezeigt werden, welche weit unter der heutigen Variation liegt und auf eine geringe Populationsgröße

nach der Eizeit hinweist. Weitere autosomal-genetische Analyse von 51 Individuen aus dem späten Pleistozän und frühen Holozän Westeuropas (43.000 - 5000 v. Chr.) durch Fu und Kollegen (2016) [48] zeigen eine Struktur innerhalb der JS Population. Eine Distanzmatrix gebildet aus der Statistik $f_3(X, Y; Mbuti)$ zeigt in einer multidimensionalen Skalierung die Ausbildung von Cluster einzelner JS Subpopulationen, welche mit geographischen (CHG), zeitlichen (preLGM/postLGM) und kulturellen (Magdalenien/Azilien, Epipaläolithikum, Mesolithikum) Kennzeichen assoziiert werden können. Die D Statistik der Form $D(Early\ European, Later\ European; Y, Mbuti)$, wobei Y eine nicht-afrikanische, moderne Population darstellt, zeigt nahöstliche Einflüsse innerhalb der Individuen nach dem LGM. Damit sind Populationsbewegungen über größere Distanzen im Zusammenhang mit der Rekolonisation Europas nach der Eiszeit nicht ausgeschlossen. Jedoch fehlt für den Beweis die genetische Verknüpfung über Individuen aus JS Populationen aus der Ägäis, der Levante und dem Nahen Osten.

Neolithikum

Durch die Analyse der genetischen Strukturen einzelner, geographisch und chronologisch verbreiteter, neolithischer Populationen konnten bereits eindeutige Beweise für einen demischen Verbreitungsprozess des Neolithikums geliefert werden. Die Pionierarbeit von Barbara Bramanti und Kollegen zeigte im Jahre 2009 keine genetische Kontinuität des mitochondrialen Markers zwischen einer JS Population und der ersten Neolithiker in Zentraleuropa. Neue genetische Strukturen mussten folglich durch Migration eingebracht worden sein, was in den folgenden Jahren durch weitere genetische Untersuchungen auf mitochondrialer Ebene innerhalb frühen Neolithikums in Zentraleuropa (LBK) durch das Auftreten neuer Linien (N1a, T2, K, J, HV, V, W, & X) bestätigt wurde [60, 20]. Dieses Bild bestätigt sich entlang der Balkan Route innerhalb der Starčevo Kultur und LBK Transdanubien [181], auf der iberischen Halbinsel [52] und auch in nördlichen Teilen Europas [109]. Durch Brandt und Kollegen wurde ein mitochondrialer Datensatz veröffentlicht, welcher einen Zeitraum von der Ankunft der ersten Farmer bis in die auf das Neolithikum folgende Bronzezeit umfasst [20]. Es zeigte sich, dass die Linienkomposition nach Ankunft des Neolithikums über einen Zeitraum von 1500 Jahren konstant geblieben ist. Ab dem späten Neolithikum ist ein Anstieg an JS Linien zu verzeichnen, was auf eine Vermischung von Farmern und JS hinweist. Interessanterweise konnten im finalen Neolithikum und dem Übergang in die Bronzezeit neue mitochondriale Linien entdeckt werden, was auf Genfluss aus einer nicht bekannten Population hinweist. Aufgrund der mitochondrialen Linien, welche im Paläolithikum und der Bronzezeit in Russland festgestellt wurden, argumentieren die Autoren für eine mögliche Migration aus der Steppe. Mitochondriale Daten unterstützen folglich die Theorie der demischen Migration nach Zentral Europa, jedoch fehlen Daten des Ursprungspunktes des europäischen Neolithikums.

Durch den Wechsel zu den Hochdurchsatz-Sequenziermethoden konnten ab dem Jahre 2012 auch autosomale Daten bezüglich dieser Fragestellung analysiert werden. Die ersten Datensätze beschreiben skandinavische [175, 174] und zentraleuropäische [102] JS im Vergleich zu lokalen Farmern. In beiden Studien ist eine verringerte genetische Diversität innerhalb der JS Genome im Vergleich zu den Neolithikern auffällig, was auf eine verringerte effektive Populationsgröße aufgrund des eiszeitlichen *bottlenecks* hinweist (siehe oben). Die Hauptkomponentenanalysen zeigen eine klare Trennung beider Kulturen, wobei sie nicht in die genetische Variation moderner Eurasischer Populationen fallen (siehe Abbildung 1.11). Eine Ausnahme stellt die heutige sardische Population dar, welche eine große genetische Ähnlichkeit zu den Individuen des Neolithikums zeigen [175]. Durch den Vergleich der Genome des frühneolithischen Farmers (Stuttgart, EEF, *early european farmer*), dem mesolithischen Genom aus Loschbour (WHG) und einem paläolithischen Individuum aus Sibirien (MA-1, ANE, *an-*

cient north eurasion), schlussfolgerten Lazaridis und Kollegen (2014) auf die Herkunft der modernen Europäer aus diesen drei ancestralen Populationen [102]. Dies zeigt sich schon in der PCA, in welcher die drei Populationen ein Dreieck bilden, welches die moderne europäische Variation umspannt. Der Einfluss des Neolithikums (EEF) wurde auf 30% für die Balkanpopulationen berechnet und nimmt Werte bis 90% in mediterranen Populationen an. Während in Westeuropäischen Populationen eine hohe Ähnlichkeit zu dem Genom des WHG besteht, welche in nördliche Richtung zunimmt, ist diese Komponente in der kaukasischen Bevölkerung nicht vorhanden. Hier zeigt sich hingegen ein Anteil von 29% an ANE Herkunft.

Durch die Erhebung weiterer Daten aus dem Holozän, konnte durch eine hohe Anzahl an Individuen und Abdeckung von Zeitstufen des Neolithikums bis in die Bronzezeit eine weitere Episode der europäischen Prähistorie beleuchtet werden. Eine Gruppe um David Reich analysierten 2015 feine Strukturen des zentraleuropäischen Neolithikums, in Kombination mit JS aus Russland und Individuen der Yamnaya Kultur der Steppenregion (siehe PCA in Abbildung 1.11). Frühneolithische Individuen aus Spanien, Ungarn und Deutschland sind genetisch sehr ähnlich, was auf einen gemeinsamen Ursprung zurückzuführen ist. Nach der Ankunft der ersten neolithischen Welle in Europa konnten die Autoren zwei weitere Populationsfluktuationen feststellen. Während des mittleren Neolithikums ist eine starke Aufnahme von genetischen Varianten der WHG Population festzustellen, was für eine Zusammenkunft beider Populationen spricht. Im weiteren Verlauf des Neolithikums wird in der Kultur der Schnurkeramik (*Corded Ware*) um 2500 v. Chr. ein starker Anstieg genetischer Elemente aus der Yamnaya Population beobachtet, was auf eine massive Migration aus der Steppe zurückzuführen ist. Anhand dieser Studie wurde nun die Herkunft der modernen Europäer auf die drei ancestralen Populationen der WHG, EEF und Yamnaya zurückgeführt. Etwas später zeigten Jones und Kollegen, dass die Herkunft der Yamnaya Population selbst in der Population der CHG und von MA-1 zu suchen ist [77].

Während genetische Daten die demische Verbreitung des Neolithikums belegen, stehen ab dem Jahr 2016 erste genetische Untersuchungen aus dem Gebiet des fruchtbaren Halbmondes zur Verfügung, welche einen Einblick in die Populationsstruktur während der Entstehungsgeschichte des Neolithikums gewähren. Ein wichtiger Aspekt, der durch Broushaki und Kollegen anhand der Untersuchung eines früh-neolithischen Genoms der heutigen Zagros Region (Iran, WC1 - 7455-7082 v. Chr.) gezeigt werden konnte, ist eine deutliche Strukturierung der früneolithischen Population. Die Probe WC1 zeigt weder genetische Ähnlichkeit zu ersten Farmern noch zur heutigen Bevölkerung Europas, jedoch zu modernen Populationen aus Pakistan und Afghanistan. Damit konnte zum ersten Mal durch genetische Daten eine parallele Entwicklungsgeschichte neolithischer Gesellschaften aus lokalen Jäger und Sammlern der Region gezeigt werden. Es wird angenommen, dass eine Trennung der vorfahrenden JS Gesellschaft vor 46.000 -77.000 Jahren stattgefunden hat [24]. Eine weitere Studie bestätigte dieses Ergebnis und lieferte durch die Untersuchung epipaläolithischer (Natufian) und frühneolithischer Individuen aus der Levante die Vorfahren europäischer Farmer [101].

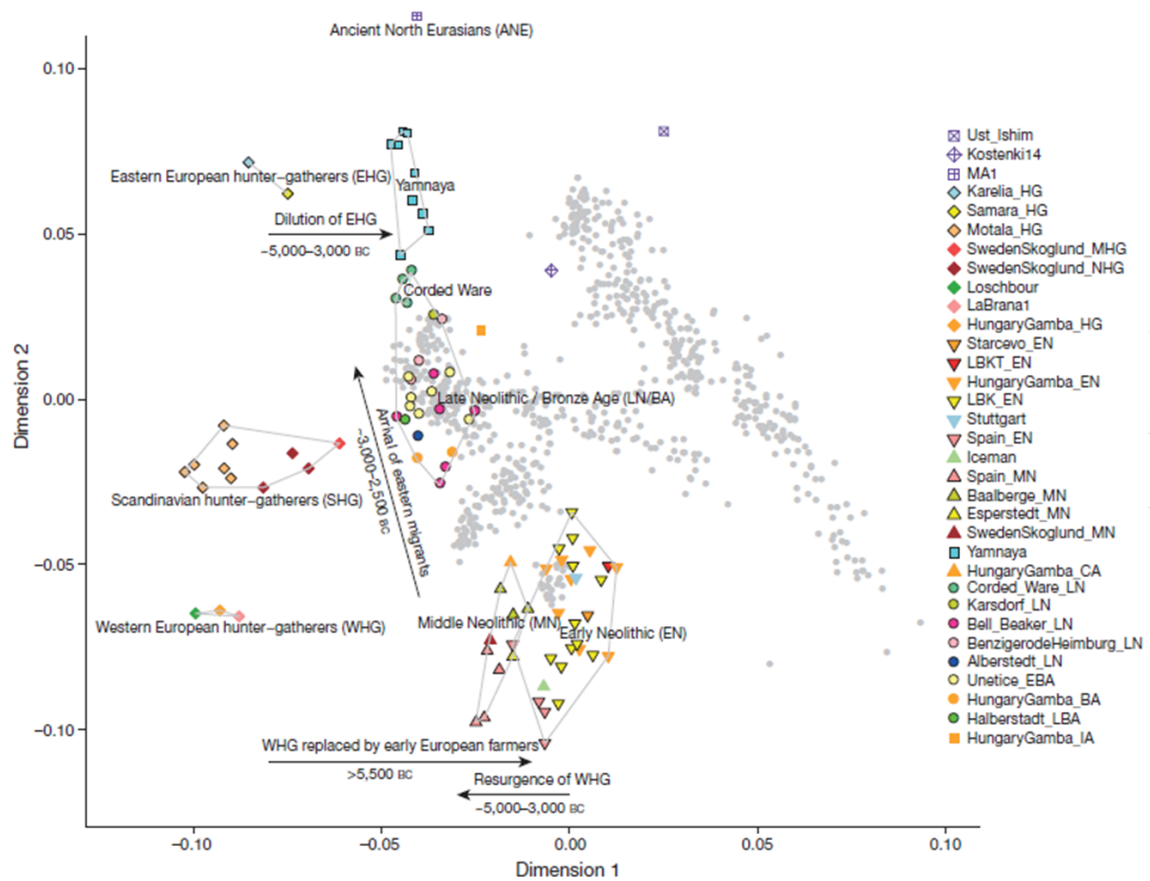


Abbildung 1.11: Die Hauptkomponentenanalyse aus Haak *et al.* 2015 [61] zeigt die Populationsstruktur prähistorischer Individuen (siehe Legende) im Vergleich zur heutigen Bevölkerung Eurasiens (hellgrau). Deutliche genetische Distanzen sind zwischen den Individuen der ersten Farmer (mitte unten) und den JS Populationen aus Zentraleuropa (WHG), Nordeuropa (SHG) und Osteuropa (EHG) zu sehen. *Admixture* Ereignisse nach Einwanderung der Steppenpopulation führen im späten Neolithikum zu einer Zusammensetzung der Bevölkerung, welche jener der heutigen Bewohner Zentraleuropas gleicht (LN/BA).

Die Besiedlungsgeschichte Griechenlands & Probenmaterial

2.1 Das Jungpaläolithikum

Das griechische Jungpaläolithikum zeigt im europäischen Vergleich wenige Funde archaischer Menschenformen, sowie eine marginale Besiedlungsdichte durch den AMM. Dies liegt zum einen in der Fokussierung der griechischen Archäologie auf die klassische Zeit, aber auch steigende Meeresspiegel können Fundstellen überschwemmt haben¹. Eine negative Fundsituation darf jedoch nicht durch die Abwesenheit menschlicher Formen interpretiert werden [161, 184]. Anhand von Klimastudien, welche über das Verhältnis der Sauerstoff-Isotope (δ^{16}/δ^{18}) Temperaturen in einzelne Epochen bestimmen (*Marine Isotopic Stage, MIS*), kann der Wechsel von Kalt- und Warmphasen der Erde angegeben werden. Klimatische Bedingungen und daraus abgeleitete Aussagen zum Stand des Meeresspiegels zeichnen ein Bild eines Lebensraumes, an welchen sich Flora und Fauna stets anzupassen hatten. Vor etwa 300.000 Jahren waren große Teile des Meeres im Eis der Arktis gespeichert und die Ägäis bildete eine Landbrücke von Klein-Asien nach Europa aus (MIS8, siehe Abbildung 2.12 [108]). Damit gilt diese Region als Korridor für die Verbreitung archaischer Menschenformen. Mit steigenden Temperaturen und Zunahme des Meeresspiegels hat Griechenland etwa 60% seiner Landmasse verloren und Zeichen der Gattung *Homo* sind heutzutage wohl auch auf dem Meeresboden zu suchen. Des weiteren zeigte diese Region eine hohe tektonische Aktivität, so dass Effekte erodierender und geomorphischer Prozesse archäologische Funde tief vergraben haben könnten [51, 184]. In den letzten Jahren ist eine zunehmende Anzahl an Projekten mit Fokus auf das Paläolithikum zu verzeichnen, was dieses Bild in Zukunft wandeln könnte. Auf dem südlichen Teil der Peloponnes, dem Pindus Gebirge und dem Ionischen Meer gibt es Anzeichen für eine längere Besiedlung durch den Neandertaler im späten Pleistozän [66]. Der Zeitraum lässt sich etwa auf 100.000 - 40.000 Jahre vor heute eingrenzen, so dass auch Griechenland als Refugium für Neandertal Populationen gilt und eine weitere mögliche Region für Paarungsereignisse mit dem AMM darstellt.

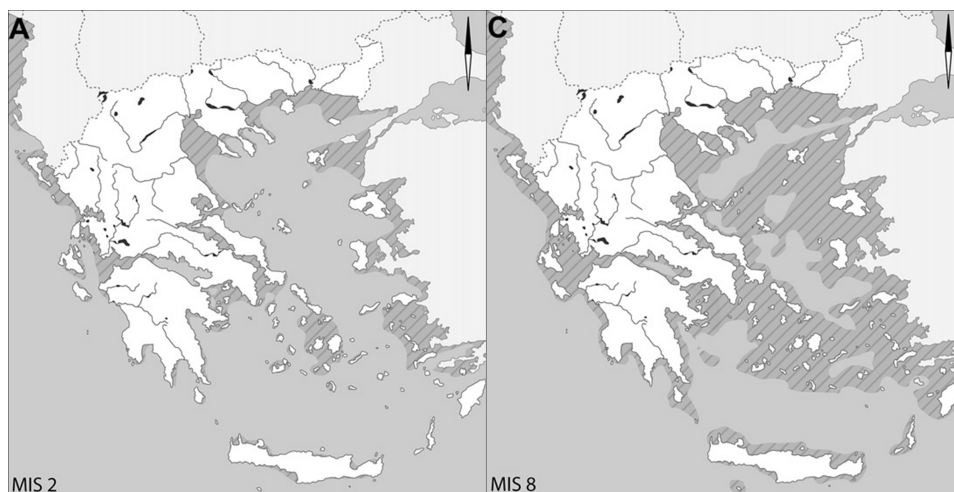


Abbildung 2.12: Stand des Meeresspiegels vor 29.000 Jahren (MIS2) und 300.000 Jahren (MIS8) [108].

¹Im Jahre 2015 wurde beispielsweise ein Siedlung aus der Bronzezeit südlich der Peloponnes in der Ägäis gefunden.

Griechenland ist seit dem späten Paläolithikum stets eine Region mit hohen Populationsbewegungen. Während die erste Besiedlung Europas auf terrestrischem Weg abgelaufen ist, wird der griechischen Halbinsel eine geringe Populationsdichte zugesprochen. Die ältesten Funde des anatomisch modernen Menschen stammen vom nördlichen Teil der Peloponnes - die Höhle von Klisoura, welche ab etwa 40.000 vor heute besiedelt wurde [89]. Dies gibt Anlass zur Annahme, dass auch auf der Peloponnes eine Möglichkeit des Aufeinandertreffens von Mensch und Neandertaler stattgefunden haben könnte [66].

Allgemein lassen sich die ersten Funde des modernen Menschen im späten Pleistozän auf die Region Epirus, Thessalien und vereinzelt auf die Peloponnes begrenzen. Gerade in den letzten Jahren haben neue Projekte Fundstellen aus dieser Zeit offen gelegt - besonders während der maximalen Eiszeit ist eine Zunahme an Funden in diesen Regionen zu beobachten. Zu Beginn des Holozäns nimmt die Anzahl der Funde wieder ab [51]. Die Höhlen von Theopetra (Thessalien) und Franchthi (Peloponnes) gehören zu den einzigen Fundstellen, auf welchen Skelettmaterialien aus dieser Zeit geborgen werden konnte. Für dieses Projekt steht eine Probe aus der Höhle von Theopetra für eine genetische Analyse zur Verfügung (Theo 3, siehe Tabelle 2.2, *Theo 3*).

2.2 Das Mesolithikum

Das Mesolithikum beginnt in Griechenland ab 10.000 v. Chr. mit dem Übergang in das Holozän und einhergehender klimatischer Stabilität. Während im späten Pleistozän eine starke Fundsituation im nord-westlichen Griechenland vorzufinden ist, geht diese anfangs des Holozäns drastisch zurück. Mit dem Beginn der Warmphase und der veränderten Ressourcengrundlage verlagert sich die beobachtete Lebensweise im Mesolithikum auf maritime Bereiche. Perles und Reingruber vermuten einen wirtschaftlichen Umschwung vom Jagen und Sammeln auf den maritimen Lebensraum, so dass die Bewohner Griechenlands dieser Zeit küstennahe Regionen bevorzugten, welche aufgrund des steigenden Meeresspiegels [97] heutzutage nicht mehr auszumachen sind [142, 151]. Mesolithische Fundstellen auf dem griechischen Festland zeigen ein verändertes Werkzeuginventar im Vergleich zu kontemporär lebenden Populationen in Zentraleuropa [151]. Auf klimatische Bedingungen zurückzuführen existieren Unterschiede in der Ausprägung des griechischen Mesolithikums - während der Norden von dicht bewaldeten Landstrichen und starkem Niederschlag geprägt gewesen ist, war der Süden Griechenlands steppenartig und trocken. Das griechische Mesolithikum war ein Mosaik diverser Lebensweisen, so dass keine Eigenschaft zwischen den mesolithischen Fundstellen identifiziert werden kann, welche als gemeinsame "kulturelle Signatur" herangezogen werden kann [51].

Während des Jüngeren Dryas, einer Zeit in der die Entwicklung der PPN Kultur in der Levante begonnen hat (Natufian Kultur), haben erste Seewege und Kenntnisse der Navigation zu einer veränderten Sichtweise der Meere geführt. Die Überwindung der Meere stellte keine Barriere mehr dar, sondern eröffnete durch seine Nutzung als Route für den Transfer und Austausch von Rohmaterialien neue Möglichkeiten. Des weiteren wird vermutet, dass eine Nahrungsknappheit in der Levante, bedingt durch Missernten während des Jüngeren Dryas, zu der Besiedlung der Insel Zypern geführt haben soll (um 9000 v. Chr.). Folglich waren zu Beginn des Holozäns erste Kenntnisse der Navigation zu See in der späten Natufian Kultur vorhanden. Ausgehend von Griechenland wurde die Ägäis ebenfalls ab dem späten Paläolithikum und zu Beginn des Holozäns durch Seefahrer bezwungen, was durch den Handel von Obsidian von der ägäischen Insel Melos zum Festland gezeigt werden kann (siehe Abbildung 2.13).



Abbildung 2.13: Handel des Rohstoffes Obsidian während des Mesolithikums (aus [152]).

In den mesolithischen Ablagerungen von Franchthi wurden Überreste von Thunfisch, einem Tiefseefisch, gefunden - ein weiterer Anhaltspunkt für die Beherrschung des Ozeans [22]. Mesolithische Fundstellen in der Ägäis, wie der "Cave of Cyclops" auf den nördlichen Sporaden (Youra), zeigen eine Jäger Sammler Gesellschaft mit ausgeprägter Fischereitechnik. Die Grundlage für das griechische Mesolithikum war ein facettenreicher Lebensraum - in der Ägäis und küstennahen Regionen hat auf Basis maritimer Kenntnisse ein kulturell-ökonomischer Wandel eingesetzt, welcher im Nordwesten (Epirus, Ionische Inseln) nicht zu verzeichnen war [51]. Dies wird auf der mesolithischen Fundstelle Maroulas auf Kythnos deutlich (Ägäis, 8086 - 7668 v. Chr. [164, 165]). Es wurde eine große Ähnlichkeit innerhalb der Architektur der Hütten und des Bestattungsrituals mit der Natufian Kultur gefunden und die Analyse von tierischen Überresten zeigten erste Anzeichen von Domestikation. Das Auftreten kultureller Elemente aus dem Epipaläolithikum der Levante, welche sich vom zentraleuropäischen Mesolithikum unterscheiden (wilde Zerealien, erste Mahltechniken, Bestattungen in Franchthi) werden als weiterer Anhaltspunkt herangezogen, dass ein kultureller Umschwung stattgefunden hat [162]. Drei Bestattung aus Maroulas stehen für eine genetische Analyse zur Verfügung (siehe Tabelle 2.2, *Mar*).

Die Höhlen von Franchthi (Peloponnes) und Theopetra (Thessalien) sind Fundplätze, welche ab dem Paläolithikum bis zum Ende des Neolithikums genutzt worden sind. Ab dem Neolithikum wurde die Höhle von Franchthi nur noch für Bestattungen genutzt, da eine Siedlung in nächster Nähe entstanden ist (*Franchthi bay*). Aus den Bestattungen in Franchthi stehen 4 mesolithische sowie 5 neolithische Proben für eine DNA Analyse zur Verfügung (mesolithisch: *Fra 1 - Fra 3, Fra 9* Tabelle 2.2; neolithisch: *Fra 4 - Fra 8*) Tabelle 2.6). Der bekannteste Fund aus Franchthi ist ein komplettes Skelett, welches seitlich liegend und in gehockter Position gefunden wurde und zu einem männlichen Individuum im Alter von 25 Jahren gehört (*Fra 1*). Die Fundplätze von Franchthi und Maroulas standen innerhalb des Obsidian Netzwerkes miteinander in Kontakt (siehe Abbildung 2.13). Die Höhle von Theopetra wurde scheinbar saisonal von paläolithischen und mesolithischen Jägern genutzt. Durch Wassereintrag und Erosion ist die Stratigrafie dieser Fundsituation recht unübersichtlich, so dass eine ^{14}C Datierung dieser Proben zwingend erforderlich ist. Aus dem griechischen Mesolithikum stehen für diese Arbeit 6 Proben zur Verfügung (siehe Tabelle 2.2; *Theo*).

Tabelle 2.2: Archäologisches Skelettmaterial aus dem Mesolithikum zur genetischen Analyse.

Die angegebenen Datierungen sind durch (+) ^{14}C Analysen im Rahmen dieser Arbeit oder (*) ^{14}C Datierung aus Literatur bzw. anhand chronologischer Einordnung des Materials innerhalb der Stratigraphie eines Fundplatzes ermittelt worden.

Labor Code	^{14}C Datierung (kal. v. Chr.)	Skelett Element	Grabnummer
<i>Theopetra - Zentral Griechenland, Trikala, Thessalien [111]</i>			
Theo 1	7605-7529*	Tibia	theopetra 2
Theo 2		Fibula	theopetra 2
Theo 3		Femur	theopetra 3
Theo 4		Ulna	theopetra 4 Δ 3 - Δ 4
Theo 5	7288-6771+	Tibia	theopetra 4 Δ 3
Theo 6		Metacarpale	
<i>Franchthi - Süd Griechenland, Nauplio, Peloponnese [36, 76, 133]</i>			
Fra 1	7320-7070*	Femur	franchti 1
Fra 2		Femur	franchti 2
Fra 3		Fibula	franchti 3
Fra 9		Knochelement	
<i>Maroulas - Ägäis [165, 123]</i>			
Mar 1	8086-7668	Tibia	T tomi13
Mar 2	8086-7668	Femur	tomi 1 burial 21-23
Mar 3	8086-7668	Humerus	tomi 1 burial 21-23

2.3 Das Neolithikum

Die chronologische Einteilung des Neolithikums in Nordgriechenland erfolgt in dieser Arbeit nach dem Schemata von Stelios Andreou, Michael Fotiadis und Kostas Kotsakis (1996) "Review of Aegean Prehistory V: The Neolithic and Bronze Age of Northern Greece" [3]:

Tabelle 2.3: Die archäologische Chronologie des Neolithikums in Nordgriechenland [3].

Archäologische Phase	Zeitraum kalibriert v. Chr.
Frühes Neolithikum	6700/6500-5800/5600
Mittleres Neolithikum	5800/5600-5400/5300
Spätes Neolithikum	5400/5300-4700/4500
Finales Neolithikum	4700/4500-3300/3100
Frühe Bronzezeit	3300/3100-2300/2200

Das frühe Neolithikum

Für die Ankunft des Neolithikums in Europa (ab 6700/6500 v. Chr.) wurde in der Ägäis zunächst das maritime Netzwerk aus dem Mesolithikum genutzt - Halbinseln wie Griechenland und weiterführend auch Süditalien wurden zeitlich vor den nördlichen Teilen (Balkan) besiedelt. Dies stellt einen bedeutenden Unterschied hinsichtlich der ersten Besiedlungswelle des AMM im Paläolithikum dar, welcher über Land abgelaufen ist [22]. Dieser maritim-orientierte Transfer des Neolithikums impliziert aber auch eine begrenzte Transportkapazität - dies gilt für Mensch, Domestikate und Technologien. So macht eine Expedition über längere Distanzen nur für kleine Gruppen Sinn und nicht für eine komplett ausgebildete Gesellschaft. Dies ist in den Zusammensetzungen der neolithischen Pakete innerhalb verschiedener thessalischer Fundsituationen ersichtlich. So vermutet Catherine Perles eine Migration mehrerer kleiner Pioniergruppen von unterschiedlichen Ursprüngen der Levante und SW Anatolien

nach Thessalien [141]. Ähnliche klimatische Bedingungen auf dem europäischen Festland vereinfachten die Adaption der neuen Fähigkeiten nach der Migration [142].

Es werden zwei Routen für die Kolonisation des griechischen Festlands durch erste Farmergesellschaften beschrieben, welche beide die Kenntnisse der Navigation zu See aus dem Mesolithikum nutzten. Einmal führte die maritime Besiedlung zur Ankunft des Neolithikums in Thessalien und ausgehend von nordwestanatolischen Küstenregionen erfolgte die Ankunft der ersten Farmer in nördlichen Teilen Griechenlands (Makedonien). Die prominenteste Fundstelle aus dem Norden, Nea Nikomedia, wurde etwa 200 Jahre nach Thessalien besiedelt (ab 6300 v. Chr.), so dass die ersten Farmer Thessaliens als Pioniere des Neolithikums auf dem europäischen Festland galten. Catherine Perles beschreibt es 2001 als Paradoxon, dass das erste Ziel der Besiedlung Griechenlands Thessalien darstellt, da es eine Region mit durchschnittlich wenig Niederschlag ist. Jedoch spiegelt dies die klimatischen Bedingungen der Levante wieder und transferierte, nahöstliche Domestikata sind an die Umgebung angepasst [142]. Neuere Grabungen und ^{14}C Datierungen haben in den letzten 15 Jahren jedoch Fundstellen zu Tage gebracht, welche mit 6600 v. Chr. zeitgleich mit jenen aus Thessalien liegen [87]. Agathe Reingruber argumentiert, dass das makedonische Neolithikum losgelöst vom thessalischen Neolithikum betrachtet werden muss. Im frühen Neolithikum in Makedonien konnten keine Obsidianartefakte nachgewiesen werden (ebenfals nicht in Theopetra), so dass zu dieser Zeit, im Gegensatz zu Thessalien, kein Anschluss an das Handelsnetzwerk ausgehend von der Insel Melos bestanden hat. Des Weiteren zeigt der "nordägäische Kulturkomplex" [150] vermehrt Übereinstimmungen mit den Artefakten der Fundplätze Illipinar (östliche Marmara Region) und Karanovo (Bulgarien), während auf dem Fundplatz Sesklo (Thessalien) östliche Eigenschaften aus der Levante wahrgenommen werden [87]. Folglich werden beide frühneolithische Gesellschaften in Griechenland über anteilige Komponenten des neolithischen Paketes (siehe Abbildung 2.14) unterschieden, welche auf differenzierte Ausgangspunkte der Migration zurückgeführt werden können.



Abbildung 2.14: Anhand bestimmter neolithischer Komponenten und ihrem Auftreten innerhalb neolithischer Siedlungen können Verbindungen geknüpft werden. Die Abbildung zeigt die Verbreitung von Stanzen, welche auf beiden Seiten der Ägäis identifiziert werden konnten und Süd-West Anatolien mit dem Norden Griechenlands verknüpfen [152].

Das Mosaik der kulturellen Ausprägungen des Mesolithikums und Arten des Transfers des Neolithikums, wie z.B. der Ausgangspunkt und die transferierten Elemente, geben Anlass zu diversen Interpretationsmöglichkeiten über Entstehung und Verbreitung des Neolithikums aus der Sicht Griechenlands. Ansichtsweisen sind von der Entstehung eines autochthonen Neolithikums in Griechenland bis zum

kompletten Transfer der Kultur aus dem fruchtbaren Halbmond in diversen Ausprägungen vorhanden. Autochthone Neolithisierungsmodelle wurden in den 1960er Jahren durch Milojevic und Theocharis angestrebt, da Schichten mit Domestikaten und ohne Keramiken (Präkeramikum) innerhalb der Stratigrafie Anlass zur Annahme einer eigenständigen Entwicklung des Domestikationsprozesses förderten (nach [150]). Dieser Ansatz würde jedoch mesolithische Schichten unterhalb neolithischer Fundstellen implizieren, da eine Kontinuität der Bevölkerung mit einhergehender Änderung der Lebensweise stattgefunden hätte. Da dies in Griechenland, mit der Ausnahme der Theopetra Höhle und Franchthi, bis heute nicht nachgewiesen werden konnte, vertritt Catherine Perles die Ansicht, dass Griechenland über die Ägäis besiedelt worden ist [142]. Durch eine Reanalyse von Radicarbonatierungen ägäischer Fundstellen hinsichtlich der Ankunftszeit des Neolithikums konnte Agathe Reingruber zeigen, dass diese einige Jahrhunderte, im Falle von Knossos gar ein Jahrtausend jünger sind als zuvor angenommen. Folglich stimmt die Ankunftszeit der neolithischen Expansionswelle eindeutig mit dem Erscheinen erster neolithischer Fundstellen in der Ägäis überein und sämtlichen Theorien einer eigenständigen Neolithisierung in Griechenland ist die Grundlage genommen [153]. Eine weitere Theorie hinsichtlich der Neolithisierung Griechenlands ist durch Runnels und van Andel gegeben, die eine Konnektivität ab dem Mesolithikum inklusive einer ersten Migrationswelle annehmen und damit Ähnlichkeiten innerhalb der Sequenz auf beiden Seiten der Ägäis erklären wollen. Auch sie vertreten die Ansicht, dass die neolithische Lebensweise, jedoch während einer zweiten Migrationswelle, transferiert worden ist [191].

Das frühe Neolithikum wurde ab Mitte des 20. Jahrhunderts in Thessalien untersucht, da prominente *Tell sites* (Erdbügel), welche über einen Besiedlungszeitraum von mehreren 1000 Jahren entstanden sind, recht einfach in der Landschaft zu identifizieren sind. Durch wiederkehrende Überbauung bildeten sich einzelne Besiedlungsschichten, welche kulturelle und technologische Elemente aus mehreren Epochen zeigen. Die größte *Tell site* stellt Sesklo dar, nach welcher die kulturellen Elemente (Stile) des thessalischen Neolithikums benannt sind. Die Fundstellen des frühen Neolithikums in Thessalien beinhalteten keine Skelettmaterialien. Erst ab 1980 wurde der Norden Griechenlands hinsichtlich dieser Fragestellung erschlossen. Fokussierung auf Makedonien und Thrakien, sowie Rettungsgrabungen, brachten im letzten Jahrzehnt frühneolithische Fundstellen zutage. Charakteristisch für das frühe Neolithikum in Nordgriechenland ist eine kurze Siedlungsdauer, welche zu dünnen Ablagerungsschichten führt und durch Überlagerungen und geologische Aktivität schwer auffindbar sind [88]. Diese eindimensionalen Fundplätze zeigen keine Kontinuität zum mittleren Neolithikum. Im frühen Neolithikum des Nordens ist lediglich die Fundstelle Nea Nikomedia (um ~ 6150 v. Chr.) eine *Tell site*. Bei der Grabung konnten zwei Phasen ausgemacht werden - eine früh-neolithische und eine weitere aufliegende Schicht aus dem späten Neolithikum. In der Stratigrafie des frühen Neolithikums waren die Domestikate Ziege und Schaf dominierend, sowie die Kultivierung von Korn, Linse und Erbse konnte nachgewiesen werden [15]. Aus den Bestattungen von Nea Nikomedia liegen Skelettelemente von 6 Individuen zur genetischen Analyse vor (siehe Tabelle 2.4, *Nea*). Auf den frühneolithischen Fundplätzen Mavropigi, Xirolimni und Pontokomi in der Kozani Region (etwa 6300 - 6000 v. Chr.) zeigte eine Isotopenanalyse von Kohlenstoff und Stickstoff aus humanen Knochenkollagen eine Änderung der Ernährungsstrategie auf Basis von Agrar- und Milchwirtschaft [134]. Für eine kurze Besiedlungsdauer der Kozani Region spricht die eindimensionale Fundplatzstruktur. Von diesen Fundstellen stehen zehn Proben für DNA Analysen zur Verfügung (siehe Tabelle 2.4, *Mau, Xir & Pon*).

Im frühneolithischen Kontext sind recht einfache Keramiken zu finden, welche kaum Anzeichen künstlerischer Gestaltung zeigen. Wenige sind im Impresso Stil dekoriert [189]. Erst zum Ende des frühen Neolithikums zeichnen sich Dekorationen im Stil weißer Punkte auf roten Keramiken ab. Diese zwei Phasen sind an der Fundstelle Revenia (6400 - 6000 v. Chr.) hervorgehoben: In der ersten Phase

sind einheitliche, monochrome Gefäße zu finden, und ab der zweiten Phase des frühen Neolithikums kommen verstärkt bemalte Keramiken hinzu (weiße Punkte auf rot gebranntem Ton, [190]). Von dem Fundplatz Revenia stehen 7 Proben für eine genetische Analyse zur Verfügung (siehe Tabelle 2.4, *Rev*).

Tabelle 2.4: Archäologisches Skelettmaterial aus dem frühen Neolithikum zur genetischen Analyse

Die angegebenen Datierungen sind durch (+) ^{14}C Analysen im Rahmen dieser Arbeit oder (*) ^{14}C Datierung aus Literatur bzw. anhand chronologischer Einordnung des Materials innerhalb der Stratigraphie eines Fundplatzes ermittelt worden.

Labor Code	^{14}C Datierung (kal. v. Chr.)	Skelett Element	Grabnummer
<i>Revenia - Nord Griechenland, Katerini, Makedonien (6500-5800 v. Chr.) [13, 185, 190, 12]</i>			
Rev 1		maxRM3	T. 2
Rev 2		maxLM1	T. 1
Rev 3		maxLM1	T. 2 Δ T
Rev 4		Pars petrosa	#422/060010 T2
Rev 5	6438-6264 ⁺	Pars petrosa	trial trench T2
Rev 6		Pars petrosa	#422/060011 T.3
Rev 7		Pars petrosa	#422/036017 T.5
<i>Mavropigi - Nord Griechenland, Kozani, Makedonien (6590/6450-6200-6010 v. Chr.) [80, 206, 134]</i>			
Mau 1	6333*	maxRM2	burial 3 tomi 801 tetragono VII
Mau 2		maxRM3	burial 5 tomi 173 tetragon XV
Mau 4	5924-5909*	rechter Femur	grave 7
Mau 5	6059-6017*	rechter Femur	grave 9
<i>Xirolimni - Nord Griechenland, Kozani, Makedonien (6200-6000 v. Chr.) [50, 49, 134]</i>			
Xir 1		maxLM2	burial 4 A.T. 45II
		linker Femur	burial 4 A.T. 45II
Xir 2		maxRM2	burial 5 A.T. 45 III
Xir4		rechter Femur	grave 7
		1. linker oberer Molar	grave 7
Xir 5	6164*	Femur	grave 10
<i>Nea Nikomedia - Nord Griechenland, Veria, Makedonien (6400-6000 v. Chr.) [159, 157, 158, 5, 186]</i>			
Nea 1		Pars petrosa	#3
Nea 2	6223 - 6075 ⁺	Pars petrosa	#7
Nea 3		Pars petrosa	T XII
Nea 4		Pars petrosa	T XVI
Nea 5		Pars petrosa	T XVII
Nea 6		Femur	T XXVIII
<i>Pontokomi - Nord Griechenland, Kozani, Makedonien [79, 134]</i>			
Pon 1		rechter Femur	grave 78
Pon 2		linker Femur	grave 112

Der mediterrane Raum hat allgemein eine geringe Biomasse zu bieten, was eine schnelle Adaption von Kultivierungstechniken förderte und einen längeren Zeitraum der Koexistenz von JS und Farmern unwahrscheinlich macht (Perles 2001). Für die Dauer des griechischen Frühneolithikums wird durch Klimaforscher und Archäologen eine Kaltphase beschrieben (RCC, *rapid climate change*, ab 6500 kal. v.Chr. [200]). Mit dem Ende dieser Phase ab 6050 v. Chr. hat sich das Neolithikum ausgehend von der ägäischen Küste über Flusstäler auf den Balkan ausgebreitet. Diese RCC Phase wird als Adaptionphase bezeichnet, da durch infrastrukturelle Änderungen eine Anpassung an ein rauhes Klima stattgefunden haben soll - dies stellt laut Krauss und Kollegen (2017) die Grundlage für die Neolithisierung des Balkans dar [93].

Das mittlere Neolithikum

Ab 5800/5600 v. Chr. ist das Neolithikum vollständig, in all seinen Facetten in Griechenland angekommen. Die im vorherigen Abschnitt beschriebenen Fundplätze zeigen keine Beständigkeit ins mittlere Neolithikum. Lediglich Nea Nikomedia ist nach einer Unterbrechung im späten Neolithikum wieder besiedelt worden. Nur wenige Fundstellen aus dem mittleren Neolithikum beherbergen Skelettmaterialien für eine genetische Analyse. Im Norden Griechenlands zählen hierzu die Fundplätze Makri, Limenaria auf Thassos & Stavroupouli. Ab dem mittleren Neolithikum kann eine Interaktion der Gesellschaften aus Makedonien und Thessalien angenommen werden, da zum einen der Norden Griechenlands an das Obsidian-Netzwerk angeschlossen wurde (siehe Abbildung 2.15) und zum anderen kulturelle Elemente aus Makedonien (Stanzen) im mittleren Neolithikum in Thessalien auftreten [152].



Abbildung 2.15: Handel des Rohstoffes Obsidian während des fortschreitenden Neolithikums in der Ägäis (aus [152]).

Der Fundplatz Stavroupouli (Thessaloniki, *Sta*) zeigt 2 Phasen der Besiedlung, eine im mittleren und die zweite im späten Neolithikum. An diesem Fundplatz konnten Weiterentwicklungen innerhalb der neolithischen Lebensweise anhand der Gebrauchsgegenstände und Siedlungsstruktur eindeutig festgestellt werden. Während in der ersten Phase runde Wohnhäuser mit außenliegender Kochstelle gebaut wurden, sind in der zweiten Phase rechteckige Häuser mit innenliegenden Herd und Ofen vorzufinden. Keramiken sind im mittleren Neolithikum aus schwarz gebranntem Ton kugelförmig gestaltet worden. Die Gestaltung der Keramiken änderte sich im späten Neolithikum in eine konische Form und es kamen vermehrt rot gebrannte Keramiken zum Einsatz [88]. Für diese Arbeit konnten 6 Proben aus der Stratigrafie des mittleren Neolithikums geborgen werden (siehe Tabelle 2.5). Die Fundplätze Makri und Limenaria sind durch ihre Lokalisation in Thrakien interessant für die Verbreitung des Neolithikums auf den Balkan und auch für den Handel von Rohmaterialien und Waren. Makri (Ost Griechenland, Alexandroupoli, *NS*) zeigt eine breit gefächerte architektonische Ausprägung mit einer zentral auf dem Fundplatz gelegenen Lagerhalle. Hier wird der gesellschaftliche Umschwung deutlich, der durch die neolithische Lebensweise ermöglicht wird. Durch die Umstellung auf eine eigene Nahrungsmittelproduktion und Arbeitsteilung ist eine Ausbildung einer komplex interagierender Gesellschaft möglich, welche ein stetiges Populationswachstum garantiert. In Makri ist eine deutliche Zunahme der Populationsdichte ab dem mittleren Neolithikum zu sehen. Nach der Besiedlung im beginnenden 6. Jahrtausend v. Chr. entwickelte sich der Fundplatz von einem kleinen Lager zu einer großen Siedlung im späten Neolithikum. Keramiken aus dem späten Neolithikum zeigen schwarze

Gefäße in der traditionellen Vinca A Ausprägung, welche ebenfalls auf den Fundplätzen Karanovo (Phase III, Bulgarien), Hoca Cesme (türkisch Thrakien) und Illipinar (NW Anatolien) zu finden sind [43]. Der Fundplatz Limenaria liegt auf einer Insel in der thrakischen Küstenregion und wurde in mehreren Abschnitten ab dem mittleren Neolithikum bis in die Bronzezeit bewohnt. Das geborgene Skelett stammt aus dem mittleren Neolithikum; aufgrund fehlender Datierung wurde diese Probe im Rahmen des Projektes datiert (5485 - 5372 kal. v. Chr.). In Sequenzen des späten Neolithikums können Keramiken im "Kum Tepe" Stil gefunden werden, welcher typischerweise im finalen Neolithikum (Chalkolithikum) Anatoliens zu finden ist. Dies spricht für eine anhaltende Konnektivität beider Seiten der Ägäis im Verlauf des Neolithikums [131].

Tabelle 2.5: Archäologisches Skelettmaterial aus dem mittleren Neolithikum zur genetischen Analyse
Die angegebenen Datierungen sind durch (+) ^{14}C Analysen im Rahmen dieser Arbeit oder (*) ^{14}C Datierung aus Literatur bzw. anhand chronologischer Einordnung des Materials innerhalb der Stratigraphie eines Fundplatzes ermittelt worden.

Labor Code	^{14}C Datierung (kal. v. Chr.)	Skelett Element	Grabnummer
<i>Makri - Ost Griechenland, Alexandroupoli, Thrakien (5000-4000 v. Chr.) [43]</i>			
NS1		molar maxilla linker Femur	
NS2		mandibularer Eckzahn mandibularer rechter Molar	
<i>Limenaria - Nord Griechenland, Thassos [131]</i>			
Thas 1	5485 - 5372 +	Pars petrosa	neolithisches Skelett
<i>Stavroupoli - Nord Griechenland, Thessaloniki, Makedonien (5800-5400 v. Chr.) [56]</i>			
St 1	5700-5600 *	Humerus	grave 3
St 2	5800-5700 *	Femur Zahn	grave 3 grave 3
St3	5650-5560*	Schädelknochen Zahn	grave 5 grave 5
St 4		Pars petrosa	Doiranis 5
St 5		Pars petrosa	Dagli 14 T.1
St 6		Pars petrosa	T.1

Das späte und finale Neolithikum

Ab dem späten Neolithikum (5400/5300 v. Chr.) entstehen im Norden Griechenlands Siedlungsstrukturen, welche permanent bis zum Ende des Neolithikums (3300/3200 v. Chr.) aufrecht erhalten wurden. Durch die wiederholte Bebauung desselben Landstriches sind über die Zeit typische *Tell sites* entstanden, wobei Grabungen chronologisch aufeinander folgende Schichten der Besiedlung freilegen. Für diese Arbeit stehen aus den *Tell sites* von Makryalos (*Mkg*, Pierra), Paliambela (*Pal*, Katerini, Makedonien) & Kremasti koilada (*Krk*, Kozani, Makedonien) Skelettmaterialien für eine genetische Analyse zur Verfügung. Die Kozani Region ist auch im fortschreitenden Neolithikum besiedelt, jedoch sind die Fundplätze Kleitos (Klei) und Toumba Kremasti Koilada (Krk) keine Fortsetzung der früh-neolithischen Siedlungen dieser Region (siehe oben, Mauvropigi, Xirolimni und Pontokomi), sondern erstrecken sich über angrenzende Gebiete, so dass an dieser Stelle keine Kontinuität anhand der Siedlungsstruktur ersichtlich ist. Siedlungsstrukturen und Keramiken des späten Neolithikums ähneln jenen aus dem mittleren Neolithikum. Auf dem Fundplatz Makryalos, welcher in aufeinander folgenden Phasen des späten und finalen Neolithikums bewohnt wurde, können Änderungen der sozialen Struktur von der Familie auf eine komplexe Gemeinschaft beobachtet werden (Anzahl, Größe der Keramiken, Verzehrpraktiken) [138]. Die Proben dieser Arbeit stammen aus der Phase des spä-

ten Neolithikums. Bestattungen, welche überdurchschnittlich viel vorhanden sind, zeigen verstreute Skelettelemente in öffentlichen Bereichen. Neben vielzähligen Funden schwarz gebrannter Keramiken sind Becher geborgen worden, mit Griffen in Tiergestalt, sowie Gefäße dekoriert mit Abbildungen des menschlichen Gesichts. Ähnliche Funde sind aus der Vinca Tradition auf dem Balkan bekannt. [88]. Aus Makryalos stehen Zahn - und Felsenbein Proben für eine genetische Analyse zur Verfügung (siehe Tabelle 2.7, *Mkg*).

Das thessalische Neolithikum hat trotz der hohen Dichte der Fundstellen nur wenig Skelettmaterial zu bieten. Hierzu zählen die Fundplätze Machrichori (*Mai*) und Palioskala (*Ska*, beide in Larissa, Thessalien). Aus den neolithischen Ablagerungen der Höhle von Franchthi stehen fünf Proben aus dem Neolithikum im Süden Griechenlands für eine Analyse zur Verfügung (siehe Tabelle 2.6, *Fra*).

Tabelle 2.6: Archäologisches Skelettmaterial aus dem südlichen Neolithikum (Thessalien, Peloponnes, Ägäis) zur genetischen Analyse

Die angegebenen Datierungen sind durch (+) ^{14}C Analysen im Rahmen dieser Arbeit oder (*) ^{14}C Datierung aus Literatur bzw. anhand chronologischer Einordnung des Materials innerhalb der Stratigraphie eines Fundplatzes ermittelt worden.

Labor Code	^{14}C Datierung (kal. v. Chr.)	Skelett Element	Grabnummer
<i>Franchthi - Süd-Griechenland, Nauplio, Peloponnes (5000-4000 v. Chr.)</i>			
Fra 4		fragil, gebrochen	franchthi 59
Fra 5		Humerus	Fr. 337=331 FFI: 38pile B: 1-2 burial
Fra 6		Femur	Fr. 63 Skelett 2
Fra 7		linker Femur	Fr. 62 05NE unit 10 NB 535 p.19
Fra 8		Femur	Fr. 339 FFI:20 (FOP 182)
<i>Tharrounia - Zentral Griechenland, Evia, Tharounia (5300-3300 v. Chr.) [163]</i>			
Tha 1			grave 5 bag A
Tha 2		Tibia	cave skoteini B7
Tha 3		Femur	grave 1
Tha 4		Femur	grave 4
<i>Alepotripa, SG - Süd-Griechenland [132, 135]</i>			
Alep 1		Ulna	$\mu\pi\alpha\alpha\text{II } 1\delta/31.8.71$ A931
Alep 2		untere 2 Molare	siehe Alep 1
Alep 3		Langknochen	A.1584 $\theta.20\alpha$ 24.7.70
Alep 4		Langknochen	A.1001 A/5a 3.12.70
Alep 5		Langknochen	A/4/47 22.8.70 (883)
<i>Kephala Kea - Ägäis [33]</i>			
Kea 1		Knochelement	Kea 23, Grave XIII
Kea 2		Knochelement	Kea 28, Grave XIII
Kea 3		Knochelement	Kea 42 , Grave XXXV
Kea 4		Knochelement	Kea 86, Grave XXXVII
<i>Machrichori - Larissa, Thessalien</i>			
Mai 1		Langknochenfragment	ATAE 135, TA43, #27
Mai 2		Langknochenfragment	TB34, #35 and #37
Mai 3		Langknochenfragment	TB34 #54
Mai 4		Langknochenfragment	B11, Inhumation
<i>Palioskala - Larissa, Thessalien</i>			
Ska 1		Langknochen	M14 #14
Ska 2		Langknochen	K12 #9 Δ
Ska 3		Clavicula	ATAE 374, Schnitt $\Lambda 14$, Inhumation

Tabelle 2.7: Archäologisches Skelettmaterial aus dem späten/finalen Neolithikum zur genetischen Analyse

Die angegebenen Datierungen sind durch (+) ^{14}C Analysen im Rahmen dieser Arbeit oder (*) ^{14}C Datierung aus Literatur bzw. anhand chronologischer Einordnung des Materials innerhalb der Stratigraphie eines Fundplatzes ermittelt worden.

Labor Code	^{14}C Datierung (kal. v. Chr.)	Skelett Element	Grabnummer
<i>Kremasti koilada - Nord Griechenland, Kozani, Makedonien (5340-4930/4700 v. Chr.)</i>			
Krk 1	5225 - 5014 ⁺	linker Femur	#1 , H27, #20 (lakkos B)
Krk 2		linker Femur	#24, N10, #45
Krk 3		linker Femur	#27, N9, #26
Krk 4		rechter Femur	#208, N23, #27
Krk 5		linker Femur	#89, Ξ 14, #2
Krk 6		linker Femur	#83, P8, #5
Krk 7		linker Femur	#77, K30, #5
Krk 8		Zahn	K30 Λ 30#13 $\gamma\rho\omega\epsilon$ $\alpha.\alpha.25$
Krk 9		Zahn	T7#7
Krk 10		Pars petrosa	TKK P20#7
Krk 11		Pars petrosa	T25
<i>Makriyalos - Nord Griechenland, Makriyalos, Pierra (5500/5400-5000 v. Chr.) [138, 186, 137, 136]</i>			
Mkg 1		mandRC	Λ 0162012
Mkg 2		maxLM1	I 0472001
Mkg 3		mandRM1	K 0412001
Mkg 4		maxLM2	Λ 0152015
Mkg 5		maxLM1	Λ 0152019
Mkg 6		mandLM3	Ξ 0681009
Mkg 7		mandRM1	I 0504004
Mkg 8		mandLM2	Ξ 0443003
Mkg 9		Pars petrosa	Λ 10152015
Mkg 10	4787 - 4615 [*]	Pars petrosa	10201012
Mkg 11		Pars petrosa	I 0504006
Mkg 12		Pars petrosa	0561043 T.35
Mkg 13		Pars petrosa	I 0493005
<i>Paliambela - Nord Griechenland, Katerini, Makedonien (4900-3200 v. Chr.) [187, 110, 62, 63]</i>			
Pal 1	4900-4500 [*]	Zahn	
Pal 2		Zahn	
Pal 3		Knochelement	#9083
Pal 4	4500-3200 [*]	Knochelement	#9092
Pal 5	4500-3200 [*]	Knochelement	#9102
Pal 6		Zahn	
Pal 7	4452-4350 ⁺	Pars petrosa	
<i>Kleitos - Northern Greece, Kozani, Makedonien (5400-4300 v. Chr.) [208, 209]</i>			
Klei 1		3M	grave 11
		Knochelement	grave 11
Klei 2		3M	grave 13
Klei 3		Knochelement	grave 34
		Pars petrosa	grave 34
Klei 5		Knochelement	grave 15
Klei 6		Knochelement	grave 11
Klei 7		Pars petrosa	grave 37
Klei 8		Pars petrosa	grave 38
Klei 10	4228-3995 ⁺	Pars petrosa	grave 9
<i>Mandra - Nord Griechenland, Thessaloniki, Makedonien</i>			
Man		2 Zähne (im Kiefer)	TM3 #24
<i>Mikri Volvi</i>			
Mikri		Langknochen (Femur, fragmentiert)	

Fragestellung der Arbeit

Die Erforschung der Populationsstruktur im Mesolithikum und Neolithikum der heutigen Region Griechenlands und ihrer Beziehungen zu kontemporär in Europa lebenden Populationen wurde bereits im Jahr 2009 durch Christina Papegeorgopoulou im Rahmen eines Humboldt Projektes bei Joachim Burger an der Universität Mainz begonnen. Erste Ergebnisse aus diesem Projekt zeigten zwar ein recht mäßigen Erhaltungszustand, jedoch konnten auf Basis erster DNA Sequenzdaten ein Projekt innerhalb der Deutschen Forschungsgemeinschaft (DFG) beantragt werden, dessen Ergebnisse in der vorliegenden Arbeit dargestellt werden.

Diese Arbeit umfasst auf methodischer Ebene die Umstellung zu neuen DNA Analysemethoden; während erste Sequenzdaten zu Beginn des Humboldt Projektes (2009) durch Sanger Sequenzierung erstellt worden sind, konnte im Laufe der Zeit auf die Methoden der Hochdurchsatz-Sequenzierung umgestellt werden. Es besteht nun die Möglichkeit parallel eine hohe Anzahl an Loci darzustellen. Dies umfasst die Sequenzierung kompletter mitochondrialer Genome, sowie die Generierung nukleärer Sequenzdaten in geringer Abdeckung (*low coverage*). Die Probenvorbereitung für das NGS ist zudem sensitiver im Vergleich zu PCR Methoden, so dass diese einige Vorteile für biomolekular stark geschädigte Proben zeigen. Auf methodischer Ebene werden in dieser Ebene eigene Ziele formuliert, welche auf den Eigenschaften des griechischen Probenmaterials basieren und in den nächsten zwei Abschnitten dieser Arbeit dargestellt werden.

Das verfügbare Probenmaterial der vorliegenden Arbeit wird in Kapitel 2 "Die Besiedlungsgeschichte Griechenlands & Probenmaterial" erläutert. Im Kontext der Besiedlung des europäischen Kontinents durch den anatomisch modernen Menschen deckt diese Arbeit eine Zeitstufe von vor 10.000 - 5.000 Jahren in der geographischen Region des heutigen Griechenlands ab. Die fossilen Proben beschreiben den Übergang vom Mesolithikum zur technologisch-kulturellen Stufe des Neolithikums in Griechenland und damit die Ankunft der ersten Farmergesellschaften auf dem europäischen Kontinent (siehe Abschnitt 2.3). Innerhalb der vorliegenden Fragestellung beschreibt eine "Population" eine Gruppierung des AMM einer technologische-kulturellen Stufe, welche geographisch die gleiche Region bewohnt. Die Individuen einer Population teilen einen ähnlichen Lebensraum, eine identische Lebensweise und rituelle Praktiken. Während für Zentraleuropa bereits populationsgenetische Datensätze zur Analyse der Transition zum Neolithikum vorliegen [53, 61, 102, 174] ist die genetische Zusammensetzung der Populationen, welche am Anfang der Migrationsroute stehen, nur durch zwei Fundplätze in NW Anaotlien abgedeckt (Barcin, Mentese [114]).

Die Fragestellung orientiert sich an den Thesen der griechischen Archäologie und an den Erkenntnissen, welche bereits durch paläogenetische Datensätze erhoben werden konnten. Mit der vorliegenden Arbeit soll die genetische Struktur der prähistorischen Populationen aus Griechenland und ihre Beziehung zu den Mesolithiker und Neolithiker Europas durch genetische Daten erörtert werden.

1. Die Archäologie beschreibt eine differenzierte Lebensweise des griechischen Mesolithikums hinsichtlich der Jäger und Sammler Populationen in Zentraleuropa. Ist dieser Unterschied auch in der genetischen Struktur zu erkennen?
2. Vor dem Hintergrund der Diskussion eines autochthonen Ursprungs des griechischen Neolithikums stellt sich die Frage nach der Art und Weise des Transfers des Neolithikums nach Griechenland. Nun soll auf genetischer Ebene ermittelt werden, ob diese Theorie belegt werden kann. Vor dem Hintergrund eines transägäischen Netzwerkes aus der Zeit des Mesolithikums soll durch

- genetische Daten erörtert werden, auf welche Weise die neolithische Kultur Griechenland erreicht hat.
3. Durch die Archäologie werden zwei Migrationsrouten beschrieben, welche zum Transfer der neolithischen Kultur auf das griechische Festland genutzt wurde. Können, unter Annahme eines demischen Verbreitungsmodells des Neolithikums, anhand der DNA Datensätze Beweise für den Ausgangspunkt der Migration geliefert werden? Stellt der Startpunkt der Migration eine genetisch homogene Population dar oder können nach der Ankunft genetische Unterschiede zwischen den neolithischen Populationen der Peloponnes, Thessaliens und dem Norden Griechenlands festgestellt werden?
 4. Bleibt Konnektivität beider Seiten der Ägäis auch im Neolithikum erhalten? Ist eine Populationsstruktur innerhalb des griechischen Neolithikums ersichtlich?
 5. Wie verhält sich das griechische Neolithikum im überregionalen Vergleich? Sind anhand der genetischen Struktur Unterschiede in den neolithischen Kulturen auszumachen? Aus Zentral-europa sind Paarungsereignisse von Jäger und Sammler mit neu angesiedelten Neolithikern im fortschreitenden Neolithikum bekannt. Ist dies auch im griechischen Neolithikum zu beobachten?
 6. Welche Rolle spielte die Population aus Griechenland in der neolithischen Transition? Ist Griechenland als weiterer Entstehungsort zu sehen, Teil der Migrationswelle oder lediglich ein Ausläufer einer aus Anatolien ausgehenden Besiedlung des europäischen Kontinents?
 7. Das griechische Neolithikum im Kontext der modernen europäischen Population. Welchen Beitrag hat das griechische Neolithikum zu dem Genpool der heutigen Bevölkerung Griechenlands und Europa geleistet? Besteht eine genetische Kontinuität zwischen dem Genpool der neolithischen Population und den heute lebenden Griechen?

Sequenzierung alter DNA aus archäologischen Skeletten

Der methodisch-experimentelle Abschnitt dieser Arbeit bedarf einer ausführlichen Darstellung, da für die Umstellung auf NGS Methoden unter Nutzung von Proben schlechter biomolekularer Erhaltung viel Zeit aufgewendet wurde. Die Palette qualitativ unterschiedlicher Proben eignet sich für die Darstellung des molekularen Verhaltens alter DNA Moleküle innerhalb des NGS Arbeitsablaufs. Neben der Charakterisierung des genetischen Materials einer Probe und der Optimierung einzelner Schritte des Versuchsablaufs können wichtige Kriterien zur Authentifizierung alter DNA Datensätze schon während der Probenverarbeitung gewonnen werden.

Die Anwendung des NGS zur Erstellung kompletter mitochondrialer Genome hat zunächst parallel zur Sequenzierung der HVS-I anhand PCR und Sanger Technologie stattgefunden. Ausgangspunkt und Grundlage des NGS Arbeitsablaufs dieser Arbeit bieten die Methoden beschrieben in Kreuzer (2011) [94] und Bollongino *et al.* (2013) [18]. Die Extraktion des Probenmaterials wurde nach Bramanti *et al.* (2009) [19] durchgeführt und anschließend wurde die DNA in eine *Library* pro Probe überführt [120]. Eine Anreicherung des mitochondrialen Genoms wurde anhand der *SureSelect Target Enrichment* Strategie (Agilent) durchgeführt (Methode siehe [18]) unter Anwendung eines selbsterstellten Designs der *RNA-baits* zur Abdeckung des kompletten mitochondrialen Genoms¹. Die Sequenzierung der *Capture* Produkte ist im Anschluss durch die Illumina Technologie erstellt worden (HiSeq 2000, 50bp SE).

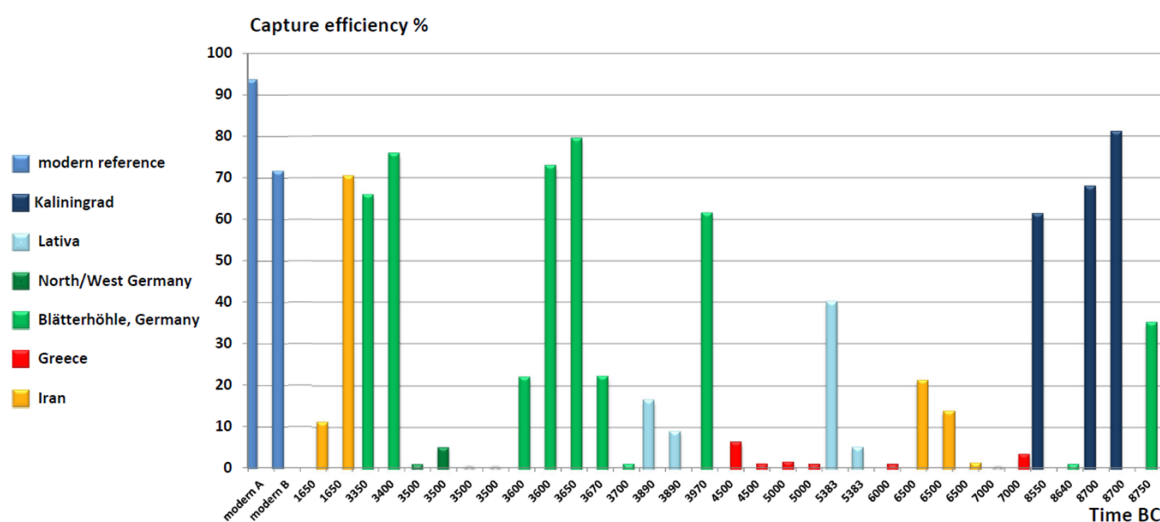


Abbildung 3.16: Darstellung der *Capture* Effizienz der beschriebenen Versuchsreihe. Abbildung entnommen aus der Posterpräsentation "Maximizing the information of DNA extracts obtained from skeletal remains", Poster, SMBE Dublin [173].

Um die Effizienz der NGS Strategie zu überprüfen wurde eine initiale Versuchsreihe durchgeführt (Jahr 2011/2012). Projekt-übergreifend wurden Proben aus geographischen Regionen vom Nahen Osten (Iran) bis Russland (Kaliningrad) und einer Zeitspanne von 8750 - 1650 v. Chr. durch ein mitochondriales *Capture* angereichert und anschließend in einem Ansatz von acht Proben auf einer HiSeq2000 *lane* sequenziert (50bp, *single-end*). Die bioinformatische Auswertung der Daten ist in [18]

¹Darstellung des *Capture* Designs in [94]

beschrieben¹.

Die Effizienz der Anreicherung, ausgedrückt in dem Verhältnis der alignierten Sequenzen gegen das mitochondriale Genom in Abhängigkeit zur sequenzierten Rohdatenmenge, spiegelt den biomolekularen Erhaltungszustand der Proben wieder - je mehr endogene DNA Moleküle in einer Probe enthalten sind, desto effektiver ist die Anreicherung. Das Ergebnis dieser Versuchsreihe ist in Abbildung 3.16 dargestellt. Es wird deutlich, dass nicht das Alter sondern die geographische Region ausschlaggebend für den Erhaltungszustand und damit dem Bearbeitungserfolg ist. So zeigen nördliche Breitengrade (Russland, Deutschland) positive Ergebnisse, welche auf einen guten Erhalt endogener DNA Moleküle zurückzuführen ist, jedoch war die verwendete Methode nicht sensitiv genug, um vollständige mitochondriale Genome in südlichen Breitengraden (Griechenland, Iran) zu erstellen.

Für eine strukturierte Darstellung wird die Projektzeit in einzelne Phasen aufgeteilt (siehe Tabelle 3.8). Bereits erhobene Daten des Humboldt Projektes (Phase A) werden als Basis dieser Arbeit zu Beginn dieses Kapitels dargestellt. Aufbauend auf den Erkenntnissen aus Phase A wurden Strategien entwickelt, um die verwendeten Methoden zu sensitivieren. Phase B-D umfassen methodische Weiterentwicklungen. Ab Phase C und D standen Felsenbeine (*Pars petrosa ossis temporalis*) als Untersuchungsmaterialien zur Verfügung.

Tabelle 3.8: Strukturierung der Projektzeit anhand Probenakquisition und methodischer Weiterentwicklung. Phase A (2009-2012) Humboldt Projekt, Phase B (04/2012-06/2014) DFG-Projekt, Phase C (07/2014-12/2014) DFG-Projekt, Phase D (01/2015-07/2015) DFG-Projekt

	Fundplatz (Anzahl Proben)	Methode
A	Franchthi (mesolithisch (4)/ neolithisch (5)), Theopetra (5), Maroulas (3), Xirolimni (2), Mavropigi (4), Revenia (3), Kleitos (3), Kremasti koilada (7), Paliambela (6), Kephala Kea (4), Tharounia (4), Makrighyalos (8)	<ul style="list-style-type: none"> • PCR-gestützte Analyse der HVS-I Region (nach [19]) • mitochondriales <i>Capture</i> und Illumina Sequenzierung (nach [94], [18]) <p><i>Ergebnisse sind im Folgenden dargestellt</i></p>
B	Xirolimi (3), Nea Makri (2), Stavropouli (3), Kleitos (3), Mandra (1), Mikri Volvi (1), Palioskala (3), Makrichiori (4)	<ul style="list-style-type: none"> • Vervollständigung PCR Analyse aus Phase A • Entwicklungen Quantifizierung (<i>Library</i>) • methodische Weiterentwicklungen in den Bereichen Extraktion ([64]), <i>Library</i> Präparation, <i>Capture</i> Sensitivierung
C	Felsenbeine - Revenia (4), Kleitos (4), Paliambela (1)	<ul style="list-style-type: none"> • Probenbearbeitung nach Optimierung in Phase B • Etablierung DNA <i>Screening</i> zur Bestimmung der Probenqualität • Erstellung genomischer Sequenzdaten aus drei neolithischen Proben (Rev 5, Pal 7, Klei 10)
D	Felsenbeine: Stavropouli (3), Kremasti koilada (2), Nea Nikomedia (5), Makrighyalos (5), Limenaria (1)	<ul style="list-style-type: none"> • DNA <i>Screening</i> inklusive Anreicherung der mitochondrialen Genome • Datenauswertung mitochondriale Genome • weitere Sequenzierungen der Probe Rev 5 zur Steigerung der durchschnittlichen Abdeckung des Genoms

¹Entfernung der Adapterstrukturen: KeyAdapterTrimFastQ_cc.py [83]; Qualitätsfilter: QualityFilterFastQ.py [83]; BWA Version 0.6.0 [104] zur Alignierung gegen rCRS (NC_012920.1, [4]); SAMtools [105]

3.1 Datengrundlage

In Phase A der Projektzeit standen 12 mesolithische und 46 neolithische Proben zur Bearbeitung zur Verfügung. Aus diesem Probenspektrum zeigten 15 Proben ein positives Ergebnis nach PCR und Sanger Sequenzierung. Sieben Proben mit positivem PCR ERgebnis wurden hinsichtlich des mitochondrialen Genoms angereichert. Diese Proben sind in Abbildung 3.16 dargestellt (rot) - eine detaillierte Auswertung erfolgt im nächsten Abschnitt.

PCR Daten

Anhand der Methode beschrieben in Bramanti *et al.* (2009) [19] wurden die Proben extrahiert, mittels PCR die HVS-I Region angereichert und nach Sanger sequenziert. Die mitochondriale HVS-I Region konnte von 15 Proben erstellt werden (siehe Tabelle 3.17), jedoch verbunden mit einem hohem Einsatzvolumen an Extrakt (8 - 12 μ l) und 42 Zyklen während der Amplifikation. Anhand der erfassten Mutationen konnten Haplotypen bestimmt werden, die in vorherigen Studien als charakteristisch genetische Linien innerhalb der zentraleuropäischen neolithischen Fundsituation eingestuft worden sind [60].

Abbildung 3.17: HVS-I Daten aus Phase A erstellt nach [19].

Range: sequenzierte Region, *Haplogroup & Quality:* Haplogruppe erstellt mit Haplogrep ([86]), SNPs: erfasste Mutationen auf der HVS-I.

SampleID	Range	Haplogroup	Quality	SNPs									
Fra1	16013-16409	X	87.3	16189C	16228T	16274A	16278T						
Fra8	16013-16409	U3	72.5	16192T	16295T	16343G							
Klei1	16013-16409	K	100.0	16224C	16311C	16327C							
Klei2	16013-16409	J	100.0	16069T	16126C								
Klei3	16013-16409	K1a10	91.0	16048A	16098C	16224C	16311C						
Krk4	16013-16409	U8b1a1	77.7	16066G	16129A	16183C	16189C	16234T					
Mar1	16013-16409	T2b2+@16304	92.3	16126C	16291T	16294T	16296T						
Mau1	16013-16409	H5	100.0	16304C									
Mau3	16013-16409	T1a4	90.6	16126C	16163G	16169T	16186T	16189C	16263C	16294T	16311C		
Tha4	16013-16409	T2b2+@16304	92.3	16126C	16291T	16294T	16296T						
Theo1	16013-16409	K	100.0	16224C	16311C								
Theo2	16013-16409	H5	71.1	16179T	16304C								
Theo4	16013-16409	T2b2+@16304	92.3	16126C	16291T	16294T	16296T						
Theo5	16013-16409	T2b2+@16304	92.3	16126C	16291T	16294T	16296T						
Xir2	16013-16409	T1a	100.0	16126C	16163G	16186T	16189C	16294T					

Next-Generation Sequencing Daten

In der zu Beginn beschriebenen Versuchsreihe dargestellt in Abbildung 3.16 wird die mäßige Verfügbarkeit endogener DNA Moleküle in den prähistorischen Proben Griechenlands deutlich. Eine detaillierte bioinformatische Auswertung dieser Proben ist in Tabelle 3.9 dargestellt. Wegen der sehr geringen Capture Effizienz (Spalte "%ALN"; entspricht Abbildung 3.16) konnte lediglich für die Proben Mau 3 & Krk 4 eine ausreichende Abdeckung des Mitochondriums von respektiv 34x und 15x erreicht werden. Problematisch war, dass nur zwei Proben den Haplotyp der HVS-I, ermittelt durch PCR/Sanger Sequenzierung, bestätigten (Mau 3, Xir 2) - ein Umstand, welcher bei allen "schwachen" Proben dieser Versuchsreihe ersichtlich war. Die Ursache liegt in einer Kombination mehrerer Fehlerquellen:

1. Biomolekular schlecht erhaltene Proben zeigen eine geringe Effizienz in der mitochondrialen Anreicherung, somit gehen zu vergleichsweise besser erhaltenen Proben wenige Moleküle in die

Tabelle 3.9: Ergebnisse des mitochondrialen Capture.

(Seq (T/Q): Sequenzen nach Trim und Qualitätsfilter, ALN rCRS: alignierte Sequenzen an rCRS, NC_012920.1 [4], % ALN: prozentuale Angabe der alignierten Sequenzen (*Capture* Effizienz), ALN Mkdup: alignierte Sequenzen nach Duplikatsentfernung, cov rCRS: Abdeckung mitochondriales Genom, mD5, mD3: Desaminierungsraten an dem jeweiligen Molekülende, HG: Haplogruppe ermittelt mit Haplofind [196].

<i>Probe</i>	<i>Rohdaten</i>	<i>Seq (T/Q)</i>	<i>ALN rCRS</i>	<i>% ALN</i>	<i>ALN Mkdup</i>	<i>cov rCRS</i>	<i>mD5</i>	<i>mD3</i>	<i>HG</i>
Fra 1	34234383	29864826	7589	0,025%	507	1,46	0,250	0,016	L2'3'4'5'6
Klei 1	19415845	17764690	993919	5,595%	1883	5,61	0,236	0,014	H32
Krk 4	41754939	36122759	87766	0,243%	11541	34,47	0,251	0,016	B
Mau 3	60398489	55018067	177771	0,323%	5340	15,94	0,244	0,013	T1a4
Theo 1	12524262	10831361	282409	2,607%	922	2,73	0,319	0,010	JT
Theo 5	34560728	31270167	74936	0,240%	576	1,7	0,272	0,007	U
Xir 2	21937452	19684598	157693	0,801%	1289	3,83	0,314	0,011	T1a5

Tabelle 3.10: *Capture* Ergebnisse nach bioinformatischer Dekontamination anhand der Strategie beschrieben in Bollongino *et al.* (2013) ([18]). Abkürzungen siehe Tabelle 3.9.

<i>Probe</i>	<i>ALN rCRS (decon)</i>	<i>cov</i>	<i>STD</i>	<i>mD5</i>	<i>mD3</i>	<i>HG</i>	<i>haploscore</i>
Fra 1	36	0,1	0,36	0,200	0,000	-	0
Klei 1	1192	3,56	2,93	0,308	0,007	K1a4a1	1
Krk 4	5288	15,81	6,48	0,245	0,006	N	0,7
Mau 3	4380	13,07	5,31	0,232	0,019	T1a4	1
Theo 1	73	0,21	0,57	0,389	0,000	L2'3'4'5'6	0,3
Theo 5	43	0,12	0,51	0,071	0,000	-	0
Xir 2	1055	3,14	2,71	0,358	0,017	T1a5	1

Sequenzierung ein

2. Proben unterschiedlichen Erhaltungszustands wurden innerhalb eines Ansatzes sequenziert. Bei der Verwendung von einem Index je Probe beträgt die Fehlerrate der Sequenzzuordnung etwa 0,3%, was auf der Ausbildung von ineinander wachsenden Cluster beruht (*mixed clusters*, siehe [84]). Falsch zugeordnete Sequenzen überlagern die endogenen Mutationen einer schwachen Probe, was eine fehlerhafte Genotypisierung zur Folge hat.

Durch einen bioinformatischen Dekontaminationsprozess (siehe [18]) anhand aller sequenzierter Proben dieser Reihe, konnten Fehlzuordnungen identifiziert und behoben werden. Dies hat jedoch einen Einfluss auf die gesamte Abdeckung des Mitochondriums innerhalb der einzelnen Proben (siehe Tabelle 3.10, Spalte *cov*). Durch die Reinigung des Datensatzes bestätigte eine weitere Probe (Klei 1) die vorliegenden PCR Ergebnisse.

3.2 Methodische Ziele und Strategien

Die Ausgangssituation zeigt einige Schwierigkeiten einen mitochondrialen DNA Datensatz aus den prähistorischen Proben Griechenlands zu erstellen. Die vorhandenen Protokolle bieten nicht genügend Erfolg um die alten DNA Moleküle anhand des NGS zu bearbeiten. Konnte das mitochondriale Genom einer Probe sequenziert werden, dann war es nur in wenigen Fällen möglich die erstellte Haplogruppe auf Basis von PCR und Sanger Sequenzierung zu replizieren. Während die Akquisition neuer Proben mit verbessertem biomolekularem Erhaltungszustand im Vordergrund steht - ebenfalls mit dem Ziel genomische Loci zu sequenzieren - soll auf methodischer Ebene der Protokollablauf beginnend mit der Extraktion und Überführung der DNA Moleküle in eine *Library* optimiert werden. Folgende Ziele werden während dieses Projektes angestrebt:

1. Charakterisierung des Probenmaterials hinsichtlich Qualität.

In paläogenetische DNA Analysen basierend auf PCR und Sanger Sequenzierung wird die DNA Qualität anhand des PCR Erfolges ausgemacht. Mit Aufkommen neuer Sequenzieretechnologien wird in dieser Arbeit eine neue Strategie gesucht, mit dem Ziel "alte" DNA Moleküle zu charakterisieren, um eine Grundlage für Folgeanalysen zu bieten. Zu Beginn wurden Proben bzw. Extrakte ohne Vorabkenntnisse für die Bearbeitung durch die NGS Technologie gewählt. Auf diesem Weg wird erst nach dem Sequenzierergebnis die Beschaffenheit des verwendeten Materials deutlich. Dies bedeutet einen hohen Materialverbrauch, verbunden mit Kosten und Zeit. Damit besteht das erste Ziel dieser Arbeit das Probenmaterial hinsichtlich endogenen Molekülen zu quantifizieren um anschließend eine geeignete Analysestrategie wählen zu können.

2. Sensitivierung von Protokollen für mäßig erhaltenens Probenmaterial.

Nach eingehender Charakterisierung des fossilen Materials und Kenntnis des Erhaltungszustandes können biochemische Bedingungen in einem Versuchsablauf auf die Qualität einer Probe angepasst werden. Nach Begutachtung des molekularen Verhaltens bestimmter DNA Strukturen kann eine biochemische Modifikation von Laborprozessen die Ausbeute der Datenmenge erhöhen.

3. Authentifizierung des Datensatzes im Rahmen der NGS Methoden.

Welche Methoden während der Probenbearbeitung und in der nachfolgenden Datenanalyse können genutzt werden, um Datensätze erhoben aus archäologischem Skelettmaterial als authentisch einzuschätzen? Wie können Kontaminationen durch rezente DNA erkannt werden und welche Methoden stehen zur Verfügung, um einen prähistorischen Datensatz von moderner Kontamination zu befreien?

4. Welche diagenetischen Prozesse wirken auf den Erhalt der DNA während der Konservierung in Skelettmaterialien?

Diese Frage stellte sich im Laufe der Arbeit, nachdem unterschiedliche Erhaltungszustände innerhalb der Skelettelemente deutlich wurden. Hier liegt der Fokus auf Desaminierungs- und Fragmentierungsraten alter DNA Moleküle, wie z.B die Abhängigkeit dieser Degradierungsprozesse vom Alter einer Probe.

3.3 Arbeitsweise mit archäologischen Skelettmaterialien

Laborstruktur und Kontaminationsvermeidung

Da das Vorkommen an DNA in archäologischen Skelettmaterialien reduziert und geschädigt ist (Kapitel 1.2), müssen bei der Bearbeitung besondere Vorsichtsmaßnahmen eingehalten werden. DNA Moleküle befinden sich überall in unserer Umwelt und eine einmalige Berührung von Proben und Labormaterialien durch ungeschützte Hände hinterlassen Hautzellen. Jeglicher Eintrag von DNA aus der Umwelt, insbesondere menschlicher DNA, führt zu Kontaminationen, die von Mixturen bis hin zur vollständigen Überlagerung eines DNA Profils führen können und Fehlinterpretationen hinsichtlich der populationsgenetischen Auswertung verursachen. Um Kontaminationsquellen vorzubeugen wurde innerhalb der paläogenetischen Forschung eine spezielle Laborstruktur zur Bearbeitung von alter DNA entwickelt. Zum einen gibt es einen Reinstraumbereich (Spurenlabor), in welchem nach der Ankunft einer Probe alle Schritte bis hin zur Amplifikation der DNA Moleküle durchgeführt werden, und einen räumlich getrennten postPCR-Bereich zur weiteren Bearbeitung der Amplifikationsprodukte. Während der ersten Bearbeitungsphase unter Reinstraumbedingungen wird der Kontakt zu freiliegenden Körperteilen vermieden [26]. Präventionsmaßnahmen sollten jedoch nicht erst mit der Bearbeitung im Labor beginnen, sondern direkt zum Zeitpunkt der Entnahme des Materials an archäologischen Fundstellen. Hier empfehlen sich spezielle Richtlinien, welche eine gute Voraussetzung für eine genetische Analyse aus archäologischen Skelettmaterialien darstellen [27]. Sei es zu Zeiten der PCR-Ära oder auch in der heutigen NGS Zeit, DNA Moleküle müssen nach der Extraktion aus Skelettmaterialien durch Amplifikation vervielfältigt werden, um diese für die weitere Analyse greifbar zu machen. Jedoch stellen diese amplifizierten Moleküle auch die größte Gefahr durch eine Rückführung in den Spurenbereich dar. Um dies zu vermeiden wird ein Einbahnstraßensystem in Bezug auf den Zutritt der Laborräumlichkeiten eingehalten – dies gilt für Personen, Chemikalien und Geräte. Zur Einhaltung verfügen beide Laborbereiche über getrennte Eingänge und Personen dürfen den Reinstraumbereich nur frisch geduscht und mit gewaschener Kleidung betreten. Straßenkleidung wird während des Durchtrittes zweier Schleusen, die den Reinstraumbereich von dem Eingang trennen, durch spezielle Laborkleidung ersetzt (Reinstraumanzug, Mundschutz, Haube, Handschuhe, Visier). Laborgeräte und Chemikalien werden ausschließlich an den Eingang des Spurenlabors geliefert und von dort aus ihrem Bestimmungsort zugewiesen.

Als Dekontaminationsmaßnahme wird jeglicher Gegenstand mit Eintritt in den Spurenbereich durch spezielle Laborseife gesäubert (Alconox[®]– anionisches Detergenz, Sigma Aldrich), in Natriumhypochlorid (z.B. DanKlorix) Lösung inkubiert und anschließend für eine Stunde UV-bestrahlt (Wellenlänge: 254 nm). Hierbei wird DNA aus der Umwelt zum größten Teil mechanisch entfernt – Reste werden chemisch und durch Strahlung degradiert bzw. stark geschädigt. Alle Arbeitsschritte, von der Einfuhr einer Probe in das Labor bis zur Erstellung einer genomischen Bibliothek (unamplifiziert), werden räumlich getrennt und in Arbeitsboxen durchgeführt. Arbeitsbereiche und Geräte werden nach Bearbeitung einer Probe und/oder am Ende des Arbeitstages mit Seife und DNA ExitusPlus[™] (AppliChem) gereinigt und anschließend für mehrere Stunden UV bestrahlt. Um möglichen Kontaminationen aus der Produktion vorzubeugen, werden Verbrauchsgegenstände wie Pipettenspitzen und Reaktionsgefäße vor dem Gebrauch für acht Stunden UV-bestrahlt. Die für diverse Versuchsabläufe verwendeten Chemikalien können nicht im Vorfeld dekontaminiert werden, weshalb in jedem Experiment Leerkontrollen eingesetzt werden. Leerkontrollen werden im Versuchsablauf als Probe behandelt und durchlaufen ebenfalls die in Abschnitt 3.5 beschriebenen Quantifizierungsschritte, so dass Kontaminationen erkannt werden können. Alle drei Monate findet ein Großputz durch

alle Mitglieder der Arbeitsgruppe statt, wobei sämtliche Oberflächen und Schränke gesäubert werden. Putzinstituten und Technikern ist das Betreten untersagt und Reparaturen durch Techniker dürfen nur unter Aufsicht und Einhaltung der oben genannten Regeln durchgeführt werden. Eine Erläuterung der Laborstruktur und Arbeitsweise der AG Paläogenetik in Mainz ist dem folgenden Film zu entnehmen: <https://www.youtube.com/watch?v=Kawbj78IFV8>.

Probenvorbereitung

Nach einer Grabung sollten Proben sorgfältig verpackt und gekühlt transportiert werden. Im Labor eintreffende Proben werden in der Originalverpackung unter Einhaltung der Dekontaminationsprozedur (s.o.) eingeführt. Für eine langfristige Lagerung empfehlen sich -20°C , jedoch werden Proben während der Bearbeitungsphase auf 4°C gelagert um ständige Gefrier- und Auftauzyklen zu vermeiden. Nach photographischer Aufnahme wird das Probenmaterial zur Dekontamination für 45 Minuten von jeder Seite UV bestrahlt (UV1). Für molekulargenetische Untersuchungen eignen sich lediglich die kompakten Elemente (*Substantia compacta*) eines Knochens, da hier eine dichte Zellstruktur vorliegt. Zunächst wird die Oberfläche abgetragen, welche zum einen nach Bodenlagerung stark durch Bodenorganismen beeinflusst ist, und zum anderen nach der Grabung und anthropologischen Untersuchungen durch moderne, menschliche DNA kontaminiert sein könnte. Bei Knochenmaterialien wird die Oberfläche durch einen Dremel (Mafra) mit Diamantschleifern (Horico) abgenommen, bis eine kompakte Knochenstruktur freigelegt ist. Zähne werden durch einen Hochdruck-Sandstrahl gereinigt (Harnisch und Rieth). Bei beiden Materialarten ist darauf zu achten Bakterienspots und starke Verfärbungen abzutragen. Die Knochenkompakta wird anschließend in kleine Würfel ($0,4 \times 1\text{cm}$) geschnitten. Knochenwürfel und Zähne werden zur abschließenden Dekontamination erneut für 45 Minuten je Seite durch UV-Licht bestrahlt (UV2).

Zum jetzigen Zeitpunkt kann schon eine Einschätzung des Probenmaterials gegeben werden. Neben der äußerlichen Inspektion zur Beschaffenheit des Materials (Bruchstellen, Verfärbungen, Porosität) kann während des Sägens über den Geruch eine Aussage zum Erhaltungszustand in Bezug auf DNA getroffen werden. Da eine dichte Knochenstruktur der DNA Schutz bietet, korreliert der Kollagengehalt mit dem DNA Gehalt einer Probe. Entsteht während des Sägens einer Probe der Geruch nach verbrannten Proteinen, ist davon auszugehen, dass in der Probe DNA enthalten ist. Um die DNA aus dem Probenmaterial herauszulösen, werden die Proben zunächst mit einer Kugelschwingmühle pulverisiert (Retsch). Mahlstärke und Dauer sind abhängig von dem Härtegrad einer Probe. Während des Mahlvorganges hat die Probe direkten Kontakt mit einem Mahlbecher, welcher anschließend sorgfältig gereinigt werden muss. Zur sofortigen Dekontamination wird der Mahlbecher mechanisch mit Seife gereinigt und anschließend für 15 Minuten komplett benetzt in Klorix eingelegt. Um kleinste Partikel zu entfernen folgt ein Mahlvorgang mit Seesand und anschließender Reinigungsprozedur wie zuvor beschrieben. Am Ende dieses Arbeitsschrittes wird zur Kontrolle, die die allgemeine Sauberkeit des Mahlvorganges gewährleistet, eine Leerkontrolle bearbeitet. Hierfür erfolgt ein Mahlvorgang mit Hydroxylapatit. Dieses Pulver ist die erste Kontrollprobe des Laborprozesses (**Müllerkontrolle, MKO**) und wird in allen weiteren Arbeitsschritten als Probe behandelt, das bedeutet sie wird im Folgeschritt parallel zu den Proben extrahiert.

Extraktion alter DNA aus Skelettmaterialien

Unter Extraktion versteht man allgemein das Herauslösen und Aufreinigen einer biologischen Komponente aus der Umgebungsmatrix – hier beschreibt es die Freisetzung der DNA aus der pulverisierten

Knochen- bzw. Zahnstruktur. Die Auflösung der Knochenstruktur (Lyse) erfolgt unter Zugabe von EDTA, ein Komplexbildner, welcher 2-wertige Kationen bindet (z.B. Ca^{2+}). N-Laurylsarcosin, ein anionisches Detergenz, ermöglicht die Auflösung von Zellstrukturen und die Zugabe von Proteinase K baut frei werdende Proteine ab. Nach der Demineralisierung liegen alle Komponenten frei in Lösung vor, so dass die DNA aus der Mixtur aufgereinigt werden kann. An diesem Punkt gibt es mehrere Methoden zur Auswahl, wobei in dieser Studie eine Phenol-Chloroform Aufreinigung genutzt wurde. Mit drei Proben wurde im Vergleich eine Aufreinigung durch Bindung der DNA an eine Silika Matrix durchgeführt (siehe Anhang A.1). Um Extraktionen untereinander vergleichen zu können, da sie sich durch Einsatzmengen an Knochenpulver oder dem finalen Extraktvolumen unterscheiden, wurde ein Normalisierungsfaktor (N) erstellt, welcher die eingesetzte Menge an Material pro μl Extrakt wiedergibt (200mg Pulver, 200 μl Extrakt: $N=1$). Dieser Faktor kann in Folgerechnungen miteinbezogen werden, um die allgemeine Qualität von Proben oder verschiedene Extraktionstechniken untereinander zu vergleichen (siehe Abschnitt 3.5 (Quantifizierung) und Abschnitt 3.5 (Probenqualität)). In jeder Extraktion wird eine Leerkontrolle mitgeführt, welche neben der Müllerkontrolle alle weiteren Schritte der Probenbearbeitung bis zur Sequenzierung durchläuft (**Extraktionskontrolle, ExKO**).

Lyse der Knochenstruktur

In Phase A der Projektzeit sind die Proben nach [19] extrahiert worden – etwa 0,5g Knochenpulver wurden zur Demineralisierung mit 2,5ml Lysepuffer (0,44M EDTA (pH 8.0, Ambion), 0,49 mg/ml proteinase K (Roche), 0,44% N-Laurylsarcosin (Merck Millipore) versetzt und über Nacht unter ständiger Bewegung inkubiert. Unter diesen Versuchsbedingungen wird jedoch nicht der komplette Ansatz an Knochenpulver demineralisiert und es bleiben Knochenstrukturen im Bodensatz erhalten. Um die maximale Ausbeute mit zeitgleich stabilem DNA Erhalt zu erzielen, wurden im Rahmen einer Masterarbeit [64] diverse Protokolle bezüglich der Demineralisierung einer Probe getestet. Dazu zählen unterschiedliche Inkubationszeiten und -temperaturen (24 – 96h; 4°C & 37°C), die Vorbehandlung des Pulvers mit Natriumhypochlorid und die Erhöhung der verwendeten Menge an EDTA in Kombination mit zusätzlichen Chemikalien während der Lyse (PTB, DTT, SDS). Der maximale DNA Gehalt konnte unter Auflösung des gesamten Materials unter Anwendung von 7ml Lysepuffer und einer Inkubationszeit von 48 Stunden bei 37°C erzielt werden. Dieses Lyseprotokoll wurde ab Phase B dieses Projektes verwendet. Um eine vollständige Lyse zu gewährleisten, die Menge an möglichen Inhibitoren zu verringern und das begrenzt zur Verfügung stehende Material einzusparen, wurde die Einsatzmenge an Knochenpulver von 0,5g auf 0,2g reduziert.

Phenol-Chloroform Extraktion

Die DNA wurde anschließend mittels Phenol-Chloroform Extraktion aus dem Überstand aufgereinigt (Roti[®]-Phenol/Chloroform/Isoamylalkohol:25/24/1, Carl Roth). Das Prinzip dieser Methode beruht auf den unterschiedlichen Phasenlöslichkeiten von Proteinen und DNA. DNA liegt in der wässrigen Phase vor, während sich Proteine in der Interphase anreichern. Auch hier ist eine positive Korrelation zwischen dem Proteingehalt in der Interphase und dem DNA Gehalt zu erwarten. Die wässrige Phase kann durch Ko-extraktion von Bodensstoffen (Mineralien, Schwermetalle, Huminsäuren) verfärbt sein. Zur Reinigung des Extraktes und Reduktion des Volumens wird die wässrige Phase über 50k Amicons (vereinzelt 30k) aufkonzentriert.

Folgende Modifikationen wurden eingeführt, um den Extraktionsschritt zu optimieren:

- Erhöhung des Wasservolumens während der Waschschrte auf dem Amicon (von 8ml auf 13ml),

um dem erhöhten EDTA Gehalt entgegenzuwirken (EDTA kann in Folgereaktionen Enzyme inhibieren)

- Reduktion der Zentrifugalkraft während der Aufkonzentration des Extraktes auf dem Amicon auf 2500g (um dem Verlust kleiner Fragmente vorzubeugen – 50k Amicons haben einen Cut-Off Wert von 75bp)

3.4 Sanger Sequenzierung der mitochondrialen HVS-I Region

Mitochondriale HVS-I Sequenzen wurden anhand von PCR und Sanger Sequenzierung entsprechend des Versuchsaufbaus beschrieben in der Dissertation von M. Unterländer ermittelt (S.98/S.114-117 in [188]). Auf Grund der qualitativ schlechten Erhaltung des griechischen Probenmaterials mussten für erfolgreiche PCR Ergebnisse bis zu 12 μ l Extrakt eingesetzt und 42 Amplifikationsrunden durchgeführt werden. Die genutzten Primersysteme für die Anreicherung der mitochondrialen Fragmente stellen sich aus vorangegangenen Arbeiten zusammen ([60], [188] – Zusammenfassung in Tabelle 3.11 & 3.12). Für die HVS-I Region stehen zwei Strategien für die Wahl der Primer zur Verfügung, welche sich in der Länge ihrer PCR Produkte unterscheiden. Zum einen kann die HVS-I Region über 4 überlappende Primersysteme abgedeckt werden, welche eine PCR Produktlänge von 162 – 196 bp (G1 – G4) zeigen. Konnte mit diesem Primersystem keine erfolgreiche PCR durchgeführt werden, wurden kleinere Systeme, welche eine PCR Produktlänge von 116 – 139 bp (K1 – K7) zeigen, angewendet. Mit Müller- und Extraktionskontrollen wurden Kontroll-PCRs durchgeführt (Primer G3, 42 Zyklen). Nach erfolgreicher Sequenzierung und Analyse der Daten, wurden für ausgewählte Proben (X, H, U- und K-Linien) Polymorphismen innerhalb des kodierenden Bereiches des Mitochondriums zur genaueren Zuordnung der Haplogruppen untersucht [188]. Hierfür wurden Primersysteme abgebildet in Tabelle 3.12 verwendet.

Tabelle 3.11: Primersysteme zur Amplifizierung der mitochondrialen HVS-I Region.

Primer: U- Position Sequenzbeginn, L- Position Ende Sequenz, Sequenz in 5' - 3' Richtung, Länge Primersequenz, LProd: Länge des gesamten Amplikons

	Primer	Sequenz	Länge	LProd	Referenz
G1	16013_U	AGCACCCAAAGCTAAGATTCTAATTTAA	28-mer	196	Unterländer 2014 [188]
	16152_L	TGATGTGGATTGGGTTTTTATGTACTAC	28-mer		
G2	L16117	TACATTACTGCCAGCCACCAT	21-mer	162	Haak et al. 2005 [60]
	H16233	GCTTTGGAGTTGCAGTTGATGTGT	24-mer		
G3	L16209	CCCCATGCTTACAAGCAAGT	20-mer	179	Handt et al. 1996 [60]
	H16348	ATGGGGACGAGAAGGGATTTG	21-mer		
G4	L16287	CACTAGGATACCAACAAAACC	20-mer	162	Handt et al. 1996 [65]
	H16410	GCGGGATATTGATTTACGG	20-mer		
K1	16011_U	AGCACCCAAAGCTAAGATTCTAATTT	26-mer	130	Unterländer 2014 [188]
	16088_L	GTGGCTGGCAGTAATGTACGAAATAC	26-mer		
K2	16071_U	GGGTACCACCCAAGTATTGACTCA	24-mer	134	Unterländer 2014 [188]
	16153_L	TGATGTGGATTGGGTTTTTATGTACTA	27-mer		
K3	16119_U	GTACATTACTGCCAGCCACCATG	23-mer	138	Unterländer 2014 [188]
	16207_L	TGATAGTTGAGGGTTGATTGCTGTAC	26-mer		
K4	16185_U	TACATAAAAACCCAATCCACATCAAAAAC	28-mer	139	Unterländer 2014 [188]
	16271_L	GGTGGGTAGGTTTGTGGTATCCT	24-mer		
K5	16233_U	AGTACAGCAATCAACCCCTCAACTATC	26-mer	127	Unterländer 2014 [188]
	16305_L	TGTACGGTAAATGGCTTTATGTACTATG	28-mer		
K6	16274_U	AAAGCCACCCCTCACCCACTAG	22-mer	116	Unterländer 2014 [188]
	16345_L	TGGGGACGAGAAGGGATTTGAC	22-mer		
K7	16340_U	ACATAAAGCCATTTACCGTACATAGCAC	28-mer	127	Unterländer 2014 [188]
	16413_L	CACTCTTGTGCGGGATATTGATTTTC	25-mer		

Tabelle 3.12: Primersysteme zur Amplifizierung von Regionen, die SNPs auf der kodierenden Region enthalten.

SNP: SNP Position (Referenz Allel/ alternatives Allel), Primer: U- Position Sequenzbeginn, L- Position Ende Sequenz, Sequenz in 5' - 3' Richtung, Länge Primersequenz, LProd: Länge des gesamten Amplikons

SNP	Primer	Sequenz	Länge	LProd	Referenz
12308 (A/G)	12303 _U	GATAACAGCTATCCATTGGTCTTAGGC	27-mer	103	Unterländer 2014 [188]
	12352 _L	GGAAGTCAGGGTTAGGGTGGTTATAG	26-mer		
14798 (T/C)	14783 U	ATACGCAAAACTAACCCCTAATAAAA	27-mer	105	Unterländer 2014 [188]
	14839 L	GCCAAGGAGTGAGCCGAAGTT	21-mer		
6371 (C/T)	6336 U	CACCCTGGAGCCTCCGTAGAC	21-mer	116	Unterländer 2014 [188]
6392 (T/C)	6403 L	ATGGCAGGGGGTTTTATATTGATAATT	27-mer		

3.5 Next-Generation Sequencing

Die Erstellung von Libraries (*genomischen Bibliotheken*) inklusive aller biochemischer Prozesse, der Quantifizierung des genetischen Materials und Illumina Sequenzierung sind in Kreutzer (2011, Diplomarbeit) dargestellt ([94], S. 20 – 31). Zu Beginn dieses Kapitels ist ersichtlich, dass die aktuellen Methoden nicht zu einer effizienten Anreicherung des mitochondrialen Genoms führen. Neben der Extraktion müssen ebenfalls die Schritte der Erstellung und Anreicherung von NGS Libraries modifiziert werden. Die vorliegende Arbeit befasst sich mit (1) der Optimierung des *Library* Protokolls, (2) der Quantifizierung der Komplexität von Extrakten und (3) der Sensitivierung der Anreicherungsmethode hinsichtlich stark degradiertes DNA mit einem geringem endogenen DNA Gehalt. Parallel wurde eine *Pipeline* zur Auswertung von Hochdurchsatz Sequenzierdaten in einer weiteren Doktorarbeit durch Christian Sell etabliert, welche die gängigen, quelloffenen Softwarestrukturen unter LINUX verwendet [172].

Die Einführung eines doppelten Indexsystems zur Clusterzuordnung während der Sequenzierung durch Kircher und Kollegen (2012) [84] war ein großer Fortschritt für alle wissenschaftlichen Bereiche, welche die Illumina Sequenzierung nutzen. Entwickelt durch das MPI in Leipzig wurde dieses System im Anschluss von Illumina übernommen. Diese Studie weist auf die Auswirkungen falscher Sequenzzuordnung und resultierender Kreuzkontamination bei der Illumina Sequenzierungen hin - ein Problem, das in Phase A dieses Projektes deutlich wurde. In der gleichen Publikation wurde eine Modifikation im *Library* Protokoll vorgestellt, die den Aufreinigungsschritt nach dem *Fill-In* durch eine Hitze-Inaktivierung ersetzt [84]. Die methodische Durchführung ist im Folgenden detailliert erläutert, eine kompakte Version der Protokolle bietet die Veröffentlichung Hofmanová, Kreutzer *et al.* [71] (*Appendix* SI2-4).

Library Erstellung

Grundlage einer erfolgreichen Anreicherung von Zielregionen durch ein *Capture* und auch einer effektiven Sequenzierung des ganzen Genomes (*Shotgun*), ist die bestmögliche Ausbeute an einzigartigen Molekülen aus einer Extraktion bzw. aus der Überführung in eine *Library*. Der ausschlaggebende Schritt ist dabei die Ligation der Adapter an das DNA Molekül. Aber auch die Art der Aufreinigung und die PCR Bedingungen zeigen Auswirkungen auf das Resultat. Da mittlerweile mehrere Abwandlungen des Originalprotokolls [120] in der Literatur verfügbar sind (z.B. [18, 84]), wurden diese Modifikationen zunächst in einer Versuchsreihe gegen das Originalprotokoll getestet (Durchführung und detaillierte Ergebnisse siehe Anhang 7.8). Das Ergebnis zeigte die höchste Ligationseffizienz unter Anwendung des Protokolls nach Kircher *et al.* (2012) (siehe Abbildung 3.18), welches ab Phase

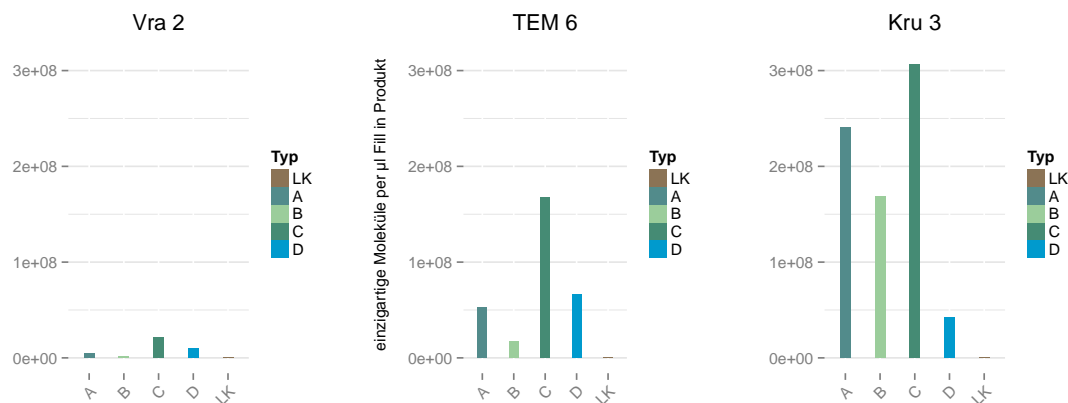


Abbildung 3.18: Ergebnis der Versuchsreihe zur *Library* Erstellung. Zum Aufzeigen der Probenunterschiede bezüglich einzigartiger Molekülzahl, sind alle y-Achsen gleich skaliert. Genaue Durchführung und detaillierte Ergebnisse finden sich im Anhang 7.8.

A: Original Protokoll nach Meyer und Kircher 2010 ([120]), B: Modifikation nach Bollongino 2013 ([18]), C: Modifikation nach Kircher 2012 ([84]), D: Kombination B & C. Durchführung und Details im Anhang A.2

B für die Bearbeitung von prähistorischen Proben aus Griechenland verwendet wurde. Aufgrund der hohen Fragmentierungsrate wird die Aufreinigung der einzelnen *Library* Schritte über Silika-Säulen durchgeführt (*MinElute Purification Kit*, Qiagen - siehe Anhang A.55). Da der PCR Ansatz nach Kircher (2012) kein Amplifikationsprodukt zeigte (siehe Anhang, Abbildung A.54a), wurde der bewährte Reaktionsansatz, beschrieben in [18], weiter verwendet (siehe Abschnitt 3.5).

Eine eigenständige Hybridisierung der Adapterstrukturen IS1/IS3 & IS2/IS3 stellt eine Gefahrenquelle dar, da aufgrund von Fehlhybridisierung die Ligationseffizienz erniedrigt werden kann. Deshalb wurde während der Projektzeit auf bereits hybridisierte Adapter umgestellt, welche in diesem Zustand von *Integrated DNA Technologies* (Leuven, Belgien) bereitgestellt wurden. Verwendete Adapter und Primer sind in Tabelle C.79 dargestellt.

PCR und Indexstrategien

PCR und Indexstrategien sind in dieser Arbeit abhängig von der Probenqualität. Das Ziel für Proben aus **Phase B** ist die Anreicherung des mitochondrialen Genoms aus qualitativ schlechten Proben. Hierfür ist es notwendig einen möglichst komplexen Probenpool mit hoher DNA-Konzentration als Grundlage für eine Anreicherung zu erstellen. Für *Capture* Versuche werden *Library* Moleküle nach dem *Fill In* in 3 PCR Parallelen (10 Zyklen) unter Anwendung der AmpliTaq Gold[®] Polymerase amplifiziert. Nach Aufreinigung (MSB Spin PCRepace, Stratec) erfolgte eine zweite PCR mit Herculase Polymerase (Agilent), durch welche die *Library* den Sättigungsbereich erreicht (3 Parallelen, 8-10 Zyklen). Diese Arbeitsweise stellt zum einen sicher, dass Uracil-tragende Moleküle erfolgreich amplifiziert werden und zum anderen wird im zweiten Ansatz durch eine effizientere Polymerase einer GC- und Längenverzerrung entgegengewirkt [38]. Dabney und Kollegen (2012) zeigten, dass die Amplifikation in den Sättigungsbereich keine Verzerrung im Sequenzierergebnis, wie z.B. eine Bevorzugung spezieller molekularer Strukturen, hervorruft. In der ersten PCR werden die *Library* Moleküle mit einem Index versehen (P7), während der P5 Adapter nicht vervollständigt wird (Primer IS7). Generell sollten PCR-Ansätze, bei welchen ein Index eingeführt oder die Adapterstruktur hinsichtlich der Bindestelle an der *flowcell* vervollständigt wird, mit mindestens 8 PCR Zyklen durchgeführt werden. Unter Annahme

einer PCR Effizienz von 80% sind 8 Zyklen notwendig, um den Hintergrund an Originalmolekülen - jene, die keinen zweiten Index bzw. die Verlängerung des Adapters tragen - unter 1% zu bringen (siehe Beispielrechnung in Tabelle 3.13). Dies gilt gleichermaßen für die erste Amplifikation von Uracil-tragenden *Libraries*. Im zweiten PCR Ansatz (Herculase Polymerase) wird lediglich eine Vermehrung der Ausgangsmoleküle unter Beibehaltung der vorliegenden Adapterstruktur durchgeführt (P5 kurz (Primer IS7)/P7 Index (Primer IS6)).

Tabelle 3.13: Kalkulation der Molekülzahl während der *Library* PCR. Angegeben ist die Zunahme der Moleküle während der Amplifikation ausgehend von 10^8 Molekülen unter Annahme einer PCR Effizienz von 80%. Die Zeile **Startmol.** beschreibt den Anteil an Originalmolekülen in der Gesamtreaktion (unvollständige Adapterstruktur/nicht sequenzierfähig, Uracil-tragend)

Start	1. Zyklus	2. Zyklus	3. Zyklus	4. Zyklus	5. Zyklus
1,00E+08	1,80E+08	3,24E+08	5,83E+08	1,05E+09	1,89E+09
Startmoleküle	55,56%	30,86%	17,15%	9,53%	5,29%
	6. Zyklus	7. Zyklus	8. Zyklus	9. Zyklus	
	3,40E+09	6,12E+09	1,10E+10	1,98E+10	
	2,94%	1,63%	0,91%	0,50%	

PCR-Effizienz 80% Erhöhung Molekülzahl um Faktor 1,8

In **Phase C/D** wurden Felsenbeine bearbeitet, bei welchen ein guter Erhaltungszustand erwartet wurde. Diese Proben eignen sich für eine Sequenzierung des gesamten Materials einer *Library* (*Shotgun*) auch ohne einen vorhergehenden Anreicherungsschritt. Bei dieser Analysestrategie ist es von Vorteil, wenn eine Probe eine hohe Komplexität an Molekülen in niedriger Konzentration besitzen, da jedes durch PCR erstellte Duplikat während der Datenauswertung nach einer Sequenzierung wieder entfernt wird. Folglich werden die Proben in 4 Parallelen und 12-16 Zyklen unter Anwendung der AmpliTaq Gold[®] Polymerase amplifiziert. Während dieser PCR wurden simultan beide Adapter, inklusive der Indices, vervollständigt.

Das **duale Index System**, bestehend aus einem 7bp langen Index (mind. 2 Unterschiede in Sequenz zu anderen Indexen des Systems) wurde im Sommer 2013 im Labor eingeführt (*Supplementary Table 1* in [84]). Im Sommer 2014 wurde das Indexsystem erweitert, um neue Kombinationsmöglichkeiten zu schaffen. Hierfür wurden die Indexsequenzen aus dem Nextera v.2 Kit (Illumina[®]) in die bestehende Adapterstruktur eingefügt. Die neuen Indexsequenzen bieten den Vorteil, dass sie aus 8bp bestehen, die untereinander einen Sequenzunterschied von mindestens 3 Basen aufweisen (siehe *Externe Tabelle XIX*).

Positiv- und Negativkontrollen während der Library Erstellung

Während der *Library* Erstellung wird neben einer Leerkontrolle standardmäßig eine Positivkontrolle mitgeführt. Die Leerkontrolle besteht aus 50µl Nuklease-freiem Wasser (**Library Leerkontrolle, LK**). An diesem Punkt werden die Müller- und Extraktionskontrollen ebenfalls in eine *Library* überführt. Die Positivkontrolle ist ein artifizielles, doppelsträngiges DNA Molekül, welches aus einer Sequenz besteht, die keinen Treffer im menschlichen Genom zeigt (Länge: 40bp, entnommen aus [166]). Dieser Sequenz wurde an beiden Enden die Roche-454-Primer Sequenzen angehängt, Sequenzen, die ebenfalls einmalig sind (zusammengesetzt: *Nonsense* Hybrid (NH), siehe Anhang, Tabelle A.58). Die beiden Einzelstränge des NH, vorwärts und revers komplementäre Sequenz, sind künstlich synthetisiert (Biosprings, Frankfurt) und werden nach Ansatz einer 100µM Stammlösung in entsprechender Verdünnung hybridisiert (siehe Anhang A.2). Während der Projektzeit wurde die Positivkontrolle durch das Einfügen von überhängenden 5'- und 3'- Enden modifiziert, um auch die Effizienz des *Blunt*

End Repairs überprüfen zu können (siehe Tabelle A.58 und Abbildung A.56). Die modifizierten Einzelstränge werden, wie auch die Adaptermoleküle, in hybridisiertem Zustand bestellt (*Integrated DNA Technologies*, Leuven, Belgien).

Die Positivkontrolle und alle Negativkontrollen (LK, ExKO, MKO) werden parallel mit den Proben nach dem *Fill In* Schritt der *Library* Erstellung durch eine quantitative Real Time PCR auf Adapterebene quantifiziert und geben auf diese Weise eine Auskunft über die Belastung durch DNA Moleküle aus der Umwelt und über die Effizienz der *Library* Reaktion (siehe folgendes Kapitel). Leerkontrollen wurden stichprobenartig sequenziert.

Quantifizierung

Quantifizierungen sind zum einen notwendig, um qualitative Aussagen über eine Proben zu treffen, aber auch um Versuchsreihen auszuwerten (siehe z.B. Test des *Library* Protokolls, Anhang A.2). Gemein haben alle genutzten Quantifizierungsmethoden die Messung eines Fluoreszenzfarbstoffes, welcher, gleichmäßig in die kleine Furche der DNA Doppelhelix eingelagert, ein messbares Fluoreszenzsignal abgibt. Für die Bestimmung der DNA Konzentration einzelner *Libraries* wird eine Qubit[®] Messung genutzt (dsDNA HS *Assay*, Invitrogen). Anhand der Qubit[®] Messung kann eine entsprechende Verdünnung für den Messbereich der *Bioanalyzer High Sensitivity* Messung (1-5 ng/ μ l, Agilent) erstellt werden, welche genutzt wird, um Fragmentlängen der Proben und das Vorkommen von Adapter- bzw. Primerdimeren zu bestimmen. Zusätzlich werden Qubit[®] Messungen genutzt, um einzelne Proben äquimolar in einen Reaktionsansatz für eine Sequenzierreaktion einzuberechnen. Je nach Erhaltungszustand einer Probe sind die Messbereiche von Qubit[®] und Bioanalyzer nicht sensitiv genug um Extrakte alter DNA zu quantifizieren, weshalb diese erst nach einer Amplifikation von *Library* Molekülen eingesetzt werden. Die quantitative Real Time PCR hingegen zeigt einen Messbereich von $10^2 - 10^9$ Molekülen pro μ l, so dass sie zur Quantifizierung während der Überführung von DNA Molekülen aus alten DNA Extrakten in eine genomische Bibliothek eingesetzt werden kann.

Quantifizierung von *Library* Molekülen mittels quantitativer Real-Time PCR

DNA Moleküle, die erfolgreich in eine *Library* eingebaut worden sind, können durch eine quantitative *Real Time* PCR nach der Auffüllung der Adapterstrukturen im *Fill In* Schritt gemessen werden [118, 120]. Unter Nutzung des Primerpaares IS7/IS8 können nur Moleküle amplifiziert werden, welche die Adapterstrukturen P5 und P7 an jeweils einem Ende des DNA Moleküls tragen und damit ein sequenzierfähiges *Library* Molekül darstellen. Da an diesem Punkt noch keine PCR zur Vervielfältigung der *Library* Moleküle stattgefunden hat, zeigt diese Messung die einzigartige Molekülzahl einer *Library* in Abhängigkeit der Molekülzahl des verwendeten Extraktes. Je höher die Molekülzahl an dieser Stelle, desto größer ist die Sequenzausbeute während einer Sequenzierreaktion nach Duplikatsentfernung – man kann folglich an dieser Stelle schon das Komplexitätslevel einer *Library* bestimmen. Nähere Erläuterungen zur Komplexität einer *Library* in Korrelation mit dem Sättigungspunkt während der Sequenzierung sind in Meyer *et al.* (2012) enthalten (siehe *Supplements* in [121]). Für die Erstellung des Reaktionsansatzes wird der KAPA Sybr Fast Universal Mastermix (PeqLab, VWR International) mit dem Primerpaar IS7/IS8 genutzt, der eine modifizierte Taq Polymerase enthält, welche 1000 Basen pro Sekunde ligiert. Durchgeführt wird die qPCR Reaktion mit dem Step One PlusTM *Real Time* PCR System (Applied Biosystems, Thermo Fisher Scientific). Anhand einer Standardreihe wird die unbekannte Molekülzahl in den Proben errechnet (einzigartige Molekülzahl pro μ l *Fill In* Produkt). Ansatz der Reaktion sowie ein Protokoll der Durchführung und Kalkulation der resultierenden Molekülzahl finden sich im Anhang (Abschnitt A.3).

Neben der absoluten Quantifikation von Molekülen innerhalb einer Probe kann auch die Belastung in Leerkontrollen und bei Mitführen einer Positivkontrolle der Erfolg der *Library* Produktion gemessen werden (siehe Abschnitt 3.5). Molekülzahlen einer *Library* werden in dieser Arbeit als "qPCR FI" gekennzeichnet. Wichtig ist die Vergleichbarkeit von Einzelreaktionen, welche durch das gleiche Ausgangsmaterial als Basis für eine Versuchsreihe gewährleistet werden kann. Abhängig von der Fragestellung können die Messungen der *Library* Molekülzahlen auf *Library* Ebene ("qPCR FI") oder Extrakt Ebene dargestellt werden. Bei dem Vergleich von Extraktionen einer Probe mit unterschiedlichen Einsatzmengen an Knochenpulver und finalen Extraktvolumina muss an dieser Stelle ein Normalisierungsfaktor mit eingerechnet werden. Der Normalisierungsfaktor 'N' beschreibt die eingesetzte Menge an Skelettmaterial (mg) während der Extraktion pro μl Extraktvolumen. In Kombination mit der eingesetzten Menge an Extrakt in eine *Library*, können auf diesem Weg unterschiedliche Extraktionen einer Probe bzw. Proben untereinander verglichen werden. Normalisierte *Fill In* Werte sind in dieser Arbeit als "qPCR FI (N)" gekennzeichnet.

MiSeq Sequenzierung und aDNA *Screening*

Bestimmung des endogenen DNA Gehalts einer Probe

Durch einen *Shotgun* Lauf mit beabsichtigt geringer Anzahl an ausgelesenen Sequenzen einer *Library* (ca. 500.000 Sequenzen) kann auf kostengünstige Weise der endogene Gehalt an DNA bestimmt werden. Dieses Verfahren ist aus der Literatur als „DNA *Screening*“ bekannt [53, 61, 71]. Ein MiSeq Lauf (50bp, *single end*) hat eine Ausbeute von ~ 20 Millionen Sequenzen, so dass ein Sequenzierungsansatz mit 20-40 Proben belegt werden kann. Die Sequenzierung erfolgte bei der auf dem Universitätscampus Mainz ansässigen Firma StarSeq. Die Datenauswertung ist in Abschnitt 3.8 beschrieben.

Als Ergebnis erhält man eine Anzahl an Sequenzen, welche gegen die Referenz des menschlichen Genoms (GRCh37/hg19) alignieren. Der endogene DNA Gehalt wird aus dem Verhältnis der am menschlichen Genom alignierten Sequenzen in Relation der Gesamtdatenmenge einer Probe bestimmt (Angabe in Prozent). Durch die Ermittlung der Desaminierungsrate [78] können weitere Aussagen über Erhaltungszustand und Authentizität der Probe getroffen werden. Bei sehr guten Proben (>100.000 Sequenzen aligniert am humanen Genom) kann zusätzlich eine Bestimmung des Geschlechtes einer Probe auf molekularer Ebene durchgeführt werden [177], welche auf dem Verhältnis von alignierten Sequenzen gegen das X- und Y-Chromosom basiert. Ab Phase C der Projektzeit wurde die Messung des endogenen DNA Gehalts für alle Proben durchgeführt, in Phase B vereinzelt. Für Proben guter Qualität - sprich einem hohen endogenen DNA Gehalt - können unterschiedliche Alignments des *Screening* Verfahrens zusammengefasst ausgewertet werden. Da die Datensätze aus unabhängigen PCR Reaktionen kommen, kann an dieser Stelle auf die Entfernung der Duplikate verzichtet werden.

Ermittlung der Komplexität einer Probe

Durch Kombination der einzigartigen Molekülzahl mit dem endogenen DNA Gehalt einer Probe kann die Komplexität einer Probe zum einen durch eine graphische Darstellung sowie rechnerisch dargestellt werden. Auch bei dieser Darstellung können die Daten auf unterschiedlichen Ebenen dargestellt werden (8Library ("qPCR FI") oder Extrakt ("qPCR FI (N)"). Einzelne Experimente dieser Arbeit verdeutlichen die Relevanz der Kombination beider Quantifizierungsmethoden, da auf diesem Weg die weitere Analysestrategie bzw. die Wahl der angewendeten Methoden gewählt werden kann. Für die graphische Darstellung werden der endogene DNA Gehalt (%) gegen den Logarithmus (\log_{10})

der einzigartigen Kopienzahl an DNA Fragmenten einer Probe in einem Streudiagramm aufgetragen [183, 202]. Die endogene Molekülzahl einer Probe je μl *Fill In* Produkt wird über die Multiplikation beider Parameter kalkuliert. Die rechnerische Darstellung ist in der Arbeit als "uniq" bzw. normalisiert auf Extraktebene als "uniq (N)" dargestellt.

Leerkontrollen

Leerkontrollen werden standardmäßig in der qPCR mitgemessen und falls diese ebenfalls sequenziert wurden, können sie in die oben beschriebenen Darstellungen miteinbezogen werden. Quantifiziert werden die Müller- und Extraktionskontrollen (MKO, ExKO) aus der Probenvorbereitung sowie die *Library* Leerkontrollen (LK). Damit können alle Schritte im Spurenbereich ab dem Sägen einer Probe erfasst und kontrolliert werden. Optimalerweise sollte die graphische Darstellung einen deutlichen Abstand zwischen den Leerkontrollen und den Proben zeigen. Ist dies nicht der Fall, sollten Proben aus dem weiteren Versuchsablauf ausgeschlossen werden. Für jede *Library* Erstellung kann ein theoretisches Kontaminationslevel berechnet werden, indem man die *Fill In* Werte der Proben ("qPCR FI") in das Verhältnis zu der entsprechenden *Library* Leerkontrolle setzt. Ist eine Leerkontrolle sequenziert, dann können auch ihre endogenen Molekülzahlen ("uniq") errechnet werden. Durch das Verhältnis der "uniq" Werte der Proben zu korrespondierenden *Library*, Extraktions- und Müllerkontrollen kann ein exaktes Kontaminationslevel für die angegebenen Schritte im Spurenbereich berechnet werden (siehe Basisrechnerung für die *Shotgun* Proben dieser Arbeit in [71], *Appendix SI2*).

3.6 Shotgun Sequenzierung

In diesem Projekt wurden für drei Proben aus zwei neolithischen Zeitstufen ein nukleärer Datensatz durch *Shotgun* Sequenzierung erstellt. Auch hier wurden desaminierte Basen als Authentifizierungskriterium erhalten. Bei einer Probe (Rev 5) wurden aufgrund der recht hohen Desaminierungsrate von >50% mit vier parallelen *Libraries* eine Behandlung durch das USER¹™ Enzym durchgeführt. Die genaue Durchführung, Sequenzierstrategie und Datenauswertung sind der zugehörigen Publikation (2015) [71] zu entnehmen (siehe *SI Appendix*, Kapitel 2 & 3).

3.7 Anreicherung des mitochondrialen Genoms

Neben methodischen Weiterentwicklungen, zu denen u.a. die zuvor besprochene Optimierung der Extraktion zählt, führten Versuchsreihen aus einer parallelen Doktorarbeit des Instituts zu der Erkenntnis, dass eine komplexe Molekülzusammensetzung, welche durch das Zusammenführen mehrerer *Libraries* einer Probe, in Kombination mit einem zusätzlichen Anreicherungsschritt zu einer gesteigerten Ausbeute der Zielregion führt [58]. Diese Erkenntnis wurde direkt für die Anreicherung mitochondrialer DNA aus den prähistorischen Proben Griechenlandes übernommen.

Grundlegendes Capture Protokoll

Zur Durchführung der Anreicherung des mitochondrialen Genoms wird das Agilent *Sure Select^{XT} target enrichment* Kit mit benutzerdefiniertem Design genutzt (nähere Erläuterung in [94], S. 34-35). Zu Beginn dieses Projektes wurde zunächst ein verbessertes Design erarbeitet (siehe Anhang, Abschnitt

¹ *Uracil-Specific Excision Reagent*

A.4). Durch Vorversuche wurde ermittelt, dass aufgrund der geringen Größe der Zielregion (16569 bp) die Herstellerangaben der Einsatzmenge an *baits* deutlich verringert werden kann. Während die Herstellerangabe 2 μl *baits* je Reaktionsansatz vorgibt, wurden für die Anreicherung eines Mitochondriums 3 μl *baits* mit 9 μl Wasser verdünnt und als Einsatz für 6 Proben genutzt (Faktor Verdünnung = 4). In Folgeversuchen wurde der Verdünnungsfaktor auf 8 erhöht (3 μl *baits* + 21 μl Wasser). Durch die Verdünnung der *baits* ist die mitgelieferte Menge an Chemikalien für die Hybridisierung und Waschstreps nicht mehr ausreichend, so dass selbsterstellte Pufferlösungen verwendet wurden (Ansatz nach [16]).

Zur Reduktion sterischer Effekte zwischen Blocker/Adapter und DNA/RNA Hybrid wurde eine neue Strategie der Blockierung der Adaptersequenzen genutzt. Der Adapter wird, den Index aussparend, lediglich auf dem reversen Strang geblockt (siehe Doktorarbeit Melanie Groß [58]). Für die zweite *Capture* Runde wurde das komplette Produkt nach einer Amplifizierung mit Herculase Polymerase (Agilent) weiterverarbeitet. Die verwendete Zyklenzahl nach Runde 1 und 2 sind in den dargestellten Versuchen variabel. Die Versuchsdurchführung wurde innerhalb der Projektlaufzeit durch neue Versionen des *MYBait* Handbuchs (MYcroarray[®]) ständig modifiziert, so dass lediglich die RNA-*baits* von Agilent stammen. Zu diesen Modifizierungen gehören (1) die *hot-binding* Strategie, (2) die Aussparung des Waschstreps 1 während der Aufreinigung und (3) die Nutzung der *on-bead* PCR für Proben aus Phase C/D.

Angepasst an die Qualität der Proben wird auch die Anreicherung des mitochondrialen Genoms auf Basis zweier Strategien durchgeführt. Schon die zuvor verwendete PCR Strategie während der *Library* Erstellung resultiert in Molekülen mit zwei Adapterstrukturen (P5 ohne Index/P5 mit Index). So gilt für die Proben aus **Phase A/B** folgende Protokollvariante: Mehrere *Libraries* einer Probe werden in einem Reaktionsansatz zusammengeführt (wobei ein geringer Anteil, etwa 30 % je *Library* ausreichend ist). Da der P5-Adapter noch nicht vollständig ausgebildet ist, muss P5-seitig nur der kurze Abschnitt und P7-seitig den Index umgebend geblockt werden. Nach der zweiten *Capture* Runde wird im Rahmen einer Index PCR das duale Indexsystem vervollständigt. Der endogene DNA Gehalt der Proben aus **Phase C/D** (Felsenbeine) ist sehr hoch, so dass für die Anreicherung des mitochondrialen Genoms nur ein Bruchteil einer *Library* verwendet wird ($\sim 15\text{-}300$ ng). Da die *On-bead* PCR eine effektive und zeitsparende Alternative darstellt, wurde diese wie in Abschnitt 3.7 (nach [103]) beschrieben angewendet. Aufgrund der geänderten PCR- und Indexstrategie (4 Parallelen, beide Indexe, 12-16 Zyklen) muss eine zusätzliche Blocksequenz während der Hybridisierung eingesetzt werden (siehe Doktorarbeit Melanie Groß [58]).

Aufstellung der durchgeführten Capture Versuche mit Sequenzierstrategie

Zur allgemeinen Übersicht der Versuchsabläufe sind in Tabelle 3.14 die durchgeführten *Capture* Reaktionen und verwendete Sequenzierstrategien angegeben. Die in der Spalte "Experiment" aufgelisteten Versuche wurden zur Sensitivierung des Protokolls hinsichtlich biomolekular mäßig erhaltenens Probenmaterial durchgeführt.

Im Verlauf des Projektes erfolgte eine Umstellung der Sequenzierstrategie mitochondrialer *Capture* Produkte. Durch Anwendung verbesserter Techniken in Extraktion, *Library* und *Capture* Durchführung konnte eine verbesserte *on-target* Rate in den *Capture* Produkten erreicht werden. Während der Projektzeit wurde deutlich, dass 500.000 Sequenzen pro Probe ausreichend sind, um ein Mitochondrium vollständig abzudecken (siehe Abschnitt 3.7). Während zu Beginn des Projektes 6-7 *Capture*produkte auf einer HiSeq2000 Lane sequenziert wurden, ist es jetzt möglich bis zu 20 Produkte auf einem MiSeq 50bp SE (30 Produkte – MiSeq 150bp SE) zu sequenzieren, was einer Preisreduktion um den

Tabelle 3.14: Übersicht der durchgeführten *Capture* Reaktionen mit Sequenzierstrategie.

PE = *paired end*, SE = *single end*, Abkürzungen der Probenamen sind den Tabellen der Probenvorbereitung zu entnehmen, siehe Kapitel 1.4 und Anhang B.1

Capture	Sequenzierstrategie	Experiment	Proben (Extraktion)
C1-2	HiSeq 100bp PE	Initialer Test der modifizierten Protokollvarianten mit qualitativ guter Probe	Krk 1
C3-4	HiSeq 100bp PE	Test Silika Extraktion und qualitativ schlechte Probe	Krk 7 / Krk 6
C6-9	MiSeq 50bp SE	Hybridisierungs-/Waschtemperatur I	Klei 4B/ Krk 2/ Krk 4/ Krk 5/ Mau 5/ Pal 5/ Pal 4/ Tha 4 (c)/ Xir 3/ Xir 5
C10-11	MiSeq 150bp SE	Hybridisierungs-/Waschtemperatur II	Klei 3/ Klei 5B/ Klei 6/ Krk 8/ Krk 9/ Mau 1/ Mau 2/ Mau 4/ Pal 1/ Pal 2/ Pal 3/ Rev 1/ Rev 2/ Rev 3/ St2B/ St 3B/ Xir 4B/ Xir 5 (reseq)
C12	MiSeq 50bp SE	Test Extrakte aus Phase A	Fra 1 (a/b)/ Fra 6/ Fra 8/ Klei 1/ Klei 2/ Tha 4 (a/b)/ Theo 1 (c)/ Theo 5 (a/b)/ Xir 1 (a/b)/ Xir 2 (a/b)
C12 reseq	HiSeq 100bp PE	Nachsequenzierung von Einzelproben aus C12	Fra 1 (a/b)/ Fra6/ Tha 4 (a/b)/ Theo 5 (a/b)
C13	MiSeq 50bp SE	Polymerasetest	Fra 3/ Klei 7/ Klei 9/ NS1A/ NS2/ Pal 2/ Pal 7/ Rev 5/ St 3B/ Xir 3
C14	HiSeq 100bp PE	Felsenbeine und Nachzügler Phase B	Fra 1(c)/ Fra 3 reseq/ Fra 8 (b)/ Klei 8/ Pal 6/ Rev 4/ Rev 6/ Rev 7/ St3B reseq/ Tha 2/ Theo 3 (a/b/c)
C15	MiSeq 150bp SE	Felsenbeine und Nachzügler Phase B	Fra 1 (c,neu Cap, gleiche Libs)/ Fra 9/ Mai 3/ Mkg 9/ Mkg 10/ Mkg 11/ Mkg 12/ Nea 1/Nea 2/ Nea 3/ Nea 4/ Nea 5/ Rev 7 (neu Cap)/ St 3A/ St 1 Thas 1/ Theo 3 (neu Cap)/ Theo 5 (c)/ Xir 4A/ Xir 4B
C16	MiSeq 150bp SE	Felsenbeine und Nachzügler Phase B	Fra 2/ Fra 3 (neu CAP)/ Krk 10/ Krk 11/ Mkg 13/ Nea 4/ St 1/ St 4/ St 5/ St 6/ Theo 2/ Xir 5

Faktor 10 entspricht. Alle Sequenzierungen wurden entweder auf dem HiSeq Gerät (2000/2500) des Instituts für Molekulare Genetik (Mainz) oder dem MiSeq der Firma Starseq durchgeführt.

Prüfung der Optimierungsstrategien aus Extraktion & Library

Anhand eines initialen Versuches zur Anreicherung des mitochondrialen Genoms aus den prähistorischen Proben Griechenlands sollen zunächst die aktuellen Protokolloptimierungen aus Extraktion und *Library* Erstellung in Kombination mit einem doppelten *Capture* Ansatzes mehrerer *Libraries* einer Probe getestet werden. Um den Rahmen des Möglichen abzustecken wurden bewusst zwei Proben gegensätzlicher Qualität ausgesucht, welche von ihrem äußeren Erscheinungsbild während der Probenvorbereitung als qualitativ hochwertig (Krk 1) bzw. niedrig (Krk 6) eingestuft werden konnten. Anhand des entwickelten Quantifizierungssystems soll das molekulare Verhalten beider Proben beurteilt werden, um einen Richtwert für weitere Versuche zu erhalten. In diesem Fall ist das theoretische Kontaminationslevel errechnet durch "qPCR FI" Daten der Proben in Korrelation zu den Leerkontrollen von Interesse. Anhand des Vergleichs dieser Daten mit dem Sequenzierergebnis sollen Erfahrungswerte für die Arbeit mit schwierigem Probenmaterial geschaffen werden. Um ausreichend Material zur Verfügung zu stellen wurden parallele Extraktionen durchgeführt, je Probe 7 *Libraries* erstellt und nach dem oben beschriebenen Schema amplifiziert (siehe 3.5, Phase B, Zyklen: 10/10). Zur Ermittlung eines oberen Schwellenwertes, wurden mit der als "gut" erachteten Probe Krk 1 zwei Einsatzmengen an *Library* Produkt in der *Capture* Reaktion getestet: (1) Pool 1 - 5µl je *Library*, (2) Pool 2 - 10µl je *Library*. Zur Ermittlung des unteren Schwellenwertes wurde von der "schlechten" Probe Krk 6 der komplette *Library* Ansatz angereichert und sequenziert.

Optimierungsstrategien des Anreicherungsverfahrens

Experimente zur Hybridisierungs- und Waschttemperatur

Zur Anreicherung des mitochondrialen Genoms aus einem 300.000 Jahre alten Höhlenbär (*Ursus deningeri*) nutzten Dabney und Kollegen (2013) [37] eine geringere Hybridisierungstemperatur von 60°C und eine Waschttemperatur von 57°C (statt 65/65°C), um auch Moleküle bis zu 30bp zurückzuhalten. Ein weiteres Protokoll, veröffentlicht von Li und Kollegen (2013) [103], beschreibt eine *Touchdown* Hybridisierung, in der die Hybridisierungstemperatur beginnend mit 65°C über 11 Stunden auf 45°C abgesenkt wird. Generell bedeutet eine geringere Hybridisierungstemperatur eine Verringerung der Spezifität, so dass auch Proben mit einer Sequenzdivergenz von bis zu 40% zwischen RNA *baits* und der Probe erfolgreich angereichert werden können (persönlich Kommunikation M. Hofreiter, Februar 2016). Um die Abdeckung des mitochondrialen Genoms nach der Sequenzierung zu steigern, wurden Protokollvarianten der oben beschriebenen Ansätze getestet. Im Vorfeld soll ein kurzer Überblick über die Ziele der einzelnen Schritte im *Capture* gegeben werden, um abzuwägen an welcher Stelle eine Verringerung der Temperatur Sinn macht (Hybridisierung und/oder Waschschrte).

Während der Hybridisierung werden einzelsträngige Moleküle der *Library* einer Probe an die komplementären Zielregionen der RNA-*baits* unter Ausbildung eines DNA/RNA Duplexes gebunden. Gleichzeitig werden die Adapterstrukturen geblockt (DNA/DNA Duplex), um eine Hybridisierung zwischen einzelnen *Library* Molekülen zu unterbinden. Der folgende erste Waschschrte dient der Aufreinigung aller nicht gebundenen *Library* Moleküle, wohingegen der temperierte zweite Waschschrte (65°C) auch unspezifisch gebundene Moleküle und die Blocksequenzen entfernt. Schmelztemperaturen sind abhängig von der Länge der Duplexstruktur eines DNA Doppelstranges, dem GC-Gehalt eines Hybrids, sowie von von der Salzkonzentration eines Reaktionsansatzes (siehe [197]). Dabei spielen einwertige Salze eine große Rolle, da diese die elektrostatische Repulsion, ausgehend von den negativ geladenen Phosphationen des DNA-Rückgrats, vermindern [168] und damit die Schmelztemperatur einer Duplexstruktur erhöhen. Weitere Faktoren, die den Schmelzpunkt beeinflussen, sind denaturierende Agenzien wie DMSO und überhängende Enden ("*dangling ends*"). Alle diese Parameter werden von MELTING ([42]), ein Programm, welches zur Berechnung der Schmelztemperatur genutzt wurde, einbezogen. Zur Berechnung der Schmelztemperaturen von DNA/DNA Duplexen erfolgte unter Annahme des *Nearest-Neighbor* Modells nach Santa-Lucia 2004 (Blocksequenzen auf dem *Library* Molekül). Für DNA/RNA Duplexe wurde das Modell nach Sugimoto 1995 genutzt (RNA-*bait* zu *Library* Molekül). Um Auswirkungen auf unterschiedliche Fragmentlängen und GC-Gehälter darzustellen, wurden mitochondriale Regionen mit jeweils 30 und 60 bp Sequenzlänge und variierendem GC-Gehalt eingesetzt (niedrig: 25%, mittel: 44%, hoch 65%; siehe Anhang, Tabelle B.66), wobei der mittlere GC-Gehalt dem durchschnittlichen Gehalt humaner, mitochondrialer DNA entspricht.

Wie in Abbildung 3.19 dargestellt, liegen die Schmelztemperaturen aller Duplexstrukturen während der Hybridisierung über 65°C, so dass die Adapterstrukturen geblockt sind und selbst kurze Fragmente mit niedrigem GC-Gehalt an die RNA-*baits* binden können. Der Waschschrte 1 wird aufgrund der Durchführung bei Raumtemperatur nicht weiter berücksichtigt. Die Anwendung eines temperierten Waschschrtes bei 65°C führt zum Verlust von Molekülen <60bp (GCn, GCm), so dass im Falle einer stark fragmentierten Probe ein großer Teil dieser Moleküle ausgewaschen wird. In dieser Abbildung wird der starke Einfluss der Salzkonzentration eines Reaktionsansatzes auf die Schmelztemperatur deutlich. Zur Gewährleistung der gleichen Ausgangssituation wurden äquimolare *Libraries* einer Probe unter unterschiedlichen Bedingungen getestet. Folgende Kombinationen für Hybridisierungs- und Waschttemperatur wurden verwendet:

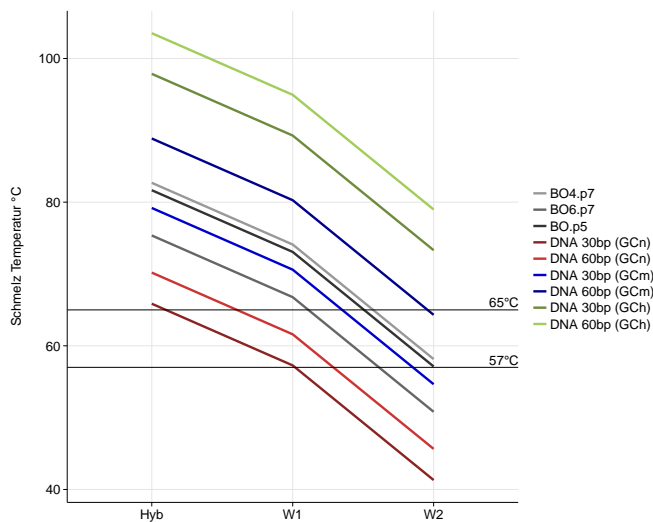


Abbildung 3.19: Berechnung der Schmelztemperatur der vorliegenden Duplexstrukturen während eines *Capture* Experimentes, unter Berücksichtigung der variierenden Salzkonzentrationen einzelner Schritte:

Hybridisierung (Hyb): $\text{Na}^+ = 650\text{mM}$,
 Waschschrift 1 (W1): $\text{Na}^+ = 150\text{mM}$,
 Waschschrift 2 (W2): $\text{Na}^+ = 15\text{mM}$

DNA-RNA Duplexe:
 GCn: niedriger GC-Gehalt (25%),
 GCm: mittlerer GC-Gehalt (44%),
 GCh: hoher GC-Gehalt (65%)

- Variante A - 65°C - 65°C (65-65) [Originalprotokoll]
- Variante B - 65°C - 57°C (65-57) [modifizierte Variante]
- Variante C - 60°C - 57°C (60-57) [nach [37]]
- Variante D - Touchdown bis 65°C - 55°C (T) [nach [103], modifiziert]

Das Originalprotokoll von Variante D beschreibt eine Touchdown Hybridisierung bis auf 45°C und gleichtemperierten Waschschrift. Abbildung 3.19 zeigt eine Schmelztemperatur von etwa 57°C für zwei der Blocksequenzen. Diese Sequenzen sollten zum größtmöglichen Teil vor der PCR entfernt werden, da sie die Primerbindungsstellen für die Amplifikation der *Capture* Produkte blockieren. Folglich wurde die finale Temperatur während der Hybridisierung und dem Waschschrift auf 55°C gesetzt. Eine Neuerung in diesem Protokoll beschreibt die *On-bead* PCR, welche direkt nach dem zweiten Waschschrift durchgeführt wird. Dadurch entfällt die Denaturierung der Ziel DNA von den *bait*s (NaOH) mit anschließender Aufreinigung des einzelsträngigen Produkts (MSB Spin PCRepace, Stratec). Die *Annealing* Temperatur während der *On-bead* PCR beträgt 65°C. Vorhergehende Versuche haben gezeigt, dass das Library Primersystem in einem Temperaturbereich von 57 - 67°C funktioniert (ermittelt durch Gradienten PCR, Daten nicht dargestellt). Die Datenauswertung der einzelnen Kombinationen einer Probe wurden auf Basis der gleichen Sequenzdatenmenge durchgeführt. Durch zufälliges Ziehen wurden die Daten auf die gleiche Anzahl an Sequenzen gebracht, so dass ermittelte Parameter innerhalb der Auswertung miteinander verglichen werden können.

Ermittlung der Sequenzdatenmenge für mitochondriale Capture Produkte

Da es sich bei dem Mitochondrium mit 16569 bp um eine vergleichsweise kleine Zielregion handelt, soll im Folgenden die Menge an Sequenzdaten ermittelt werden, die ausreichend für die Abdeckung des Mitochondriums ist. Hierfür werden durch zufälliges Ziehen mehrere Rohdatensätze mit ansteigender Anzahl an Sequenzen generiert (= *Downsampling*). Die Sättigung einer Sequenzierreaktion mitochondrialer *Capture* Produkte wird durch eine graphische Abbildung der alignierten Sequenzen (ohne Duplikate) als Funktion der steigenden Sequenzdatenmenge dargestellt. Unter Einbeziehung des Duplikationslevels werden *Capture* Produkte hinsichtlich der Komplexität einer Probe analysiert.

Polymerase Test

Schon aus PCR Zeiten sind Effekte auf die Arbeitsweise von DNA Polymerasen hinsichtlich schwieriger Matrizen, wie z.B. ein hoher GC-Gehalt oder auch repetitive Regionen, bekannt. Während der Amplifikation einer DNA-*Library* ist die Wahl der PCR Parameter ausschlaggebend für die molekulare Komplexität des Amplifikationsproduktes hinsichtlich der Repräsentation der original Moleküle aus dem Extrakt. Besonders betroffen sind genomische Bibliotheken aus Extrakten alter DNA, da diese exogene, bakterielle Moleküle enthalten, die bekanntlich einen hohen GC-Gehalt haben. Dabney und Kollegen publizierten 2012 eine Versuchsreihe zur Evaluation von 10 verschiedenen Polymerasen und Puffersystemen bezüglich GC-Gehalt und Fragmentlängen anhand moderner und alter DNA Extrakte (UDG-behandelt). Der geringste Effekt auf eine Verzerrung des GC-Gehaltes bzw. der Fragmentlänge in modernen Proben konnte nach Anwendung der Herculase Polymerase (Agilent) und der Pfu Turbo Cx Polymerase (Agilent) festgestellt werden. Für die Amplifikation aus Extrakten alter DNA zeigte die AccuPrime Pfx Polymerase die höchste Ausbeute an endogenen Molekülen. Die AmpliTaq Gold Polymerase hingegen bevorzugt Moleküle mit einem GC Gehalt >50%, was sich im Vergleich mit AccuPrime Pfx im endogenen DNA Gehalt nach Amplifikation widerspiegelt. Jedoch ist anzumerken, dass die AccuPrime Polymerase zu dem Stamm der Phusion Polymerasen gehört, denen es nicht möglich ist über Uracil zu amplifizieren. Es können folglich nur UDG behandelte Extrakte durch diese Polymerase bearbeitet werden, da Uracil Moleküle zu einem Strangabbruch führen. Da diese Polymerasen durch die paläogenetische Literatur hinweg, ohne Rücksicht auf voran gegangene UDG Behandlung, angewendet werden, soll eine Auswahl dieser nun getestet werden. Ziel ist es eine PCR Strategie zu finden, die unter Einhaltung der molekularen Komplexität eine gute Ausgangssituation für eine *Capture* Reaktion schafft.

Im Folgenden sollen die Auswirkungen der AccuPrimeTM Pfx, KAPA HiFiTM Uracil⁺, AmpliTaq Gold Polymerase und der bereits verwendeten Kombination von AmpliTaq Gold Polymerase/Herculase (siehe oben) getestet werden. Es wurden vier parallele Libraries einer Probe erstellt und nach dem *Fill In* Schritt vereint. Hieraus wurden die vier beschriebenen PCR Reaktionen in jeweils sechs PCR-Parallelen angesetzt (siehe Tabelle 3.15). Nach Aufreinigung (MSB Spin PCRepace, Stratec) wurde das gesamte PCR Produkt hinsichtlich des mitochondrialen Genoms angereichert und durch einen 50bp SE MiSeq Lauf sequenziert. Für die Auswertung wurde ein Datensatz mit der Anzahl gleicher Sequenzen erstellt.

Tabelle 3.15: PCR Bedingungen Polymerasetest. PCR Zyklus Bedingungen siehe Hersteller Protokoll

	AccuPrime (A)	KAPA Hifi (K)	AmpliTaQ (T)	Taq/Herc (T/H)
Parallelen	6	6	6	3/3
Primer Konzentration	0,14nM	0,2nM	0,2nM	0,2nM
Zyklen	15	15	15	10/10

3.8 NGS Datenanalyse

Die Datenauswertung erfolgte standardisiert anhand einer Pipeline. Diese beginnt mit den Rohdaten im FastQ Format, welche nach Indexkombinationen der einzelnen Proben sortiert sind (nähere Informationen zur bestehenden Pipeline bietet die Doktorarbeit von [172]). Die verwendeten Programme und Parameter sind in Tabelle 3.16 aufgelistet. Nach Entfernung von Adapterstrukturen aus den Sequenzdaten wurde ein Qualitätsfilter in Abhängigkeit der Sequenzierstrategie angewendet ([83]).

Dabei werden Sequenzen aussortiert, welche Basen enthalten, deren Qualitätswert 15 unterschreitet (Parameter siehe Tabelle 3.16). Es folgt ein Alignment gegen die Referenzsequenz (mitochondrial: rCRS (NC012920, [4]), humanes Genom: GRCh37/hg19). Da nach der *Library* Erstellung und besonders während der *Capture* Experimente viele PCR Zyklen angewendet wurden, müssen nun PCR Duplikate entfernt werden, um Rückschlüsse auf die ursprünglichen Moleküle und die vorhandene Abdeckung (*coverage*) ziehen zu können (Picard tools "Mark Duplicates"). "Mark Duplicates" identifiziert alle Sequenzen mit gleichen 5' Koordinaten und Orientierung und behält die Sequenz mit der höchsten Qualität (höchste Summe an Basenqualitäten >15 - siehe Picardtools Dokumentation). Bei sehr kurzen Molekülen ist die Wahrscheinlichkeit von Fehlplatzierung der Sequenz innerhalb der Referenz oder zwischen Spezies sehr hoch (man Bedenke hohe Ähnlichkeit von bakteriellen Genomen und dem humanen Mitochondrium), so dass Sequenzen unter 30bp entfernt werden (NGSutils, [21]).

Tabelle 3.16: Programme und Parameter der Datenverarbeitung

	<i>MiSeq</i>		<i>HiSeq</i>
	<i>50bp SE</i>	<i>150bp SE</i>	<i>100 bp SE</i> <i>100 bp PE</i>
Trim	KeyAdapterTrimFastQ_cc.py [83]		
Qualität	QualityFilterFastQ.py [83]		
Parameter Qualität	3 Basen <15	7 Basen <15	5 Basen <15
merge	nur PE: ea-utils Paket (6bp Überlappung mit 92% Sequenzidentität) [6]		
Alignment	BWA aln (Standardeinstellungen) [104]		
Duplikate	"Mark Duplicates" aus dem Picard tools Paket		
Sortierung/Indizierung	samtools Paket (sort, index Kommando) [105]		
Längensfilter	Filter für Sequenzen <30bp (NGSutils) [21]		

Genotypisierung der Proben

Abhängig von Insertionen oder Deletionen kann es zu Fehlalignierungen kommen, welche im nächsten Schritt fälschlicherweise als Mutationen erkannt werden können, weshalb die Sequenzdaten ohne Duplikate realigniert werden (geringerer Rechenaufwand ohne Duplikate; GATK – RealignerTargetCreator/ IndelRealigner, [40]). Da bei der Definition von Mutationen die Basenqualität, welche jeder ausgelesenen Base während der Sequenzierung zugeordnet wird, ein Hauptkriterium ist, werden die Qualitätswerte der einzelnen Basen rekaliert (GATK – BaseRecalibrator). Der Ablauf der Genotypisierung (*SNP-Call*) unterscheidet sich für haploide und diploide Abschnitte des Genoms.

Genotypisierung haploider Bereiche (Mitochondrium)

Für die Bestimmung von genetischen Varianten auf dem Mitochondrium wird der *Unified Genotyper* (GATK) verwendet (Parameter: Ploidie 1, Qualitätsfilter Basenqualität 10, Qualitätsfilter Variante 30). Anschließend werden die Varianten gefiltert nach einer Mindestabdeckung der Position von 5x und einer Qualität von >50. Zur Erstellung einer FASTA Sequenz des mitochondrialen Genoms werden die alignierten Sequenzen mit Hilfe des SAMtools Pakets [105] (mpileup/bcftools/vcfutils) über einen Zwischenschritt im FASTQ Format umgewandelt. Nicht abgedeckte Positionen wurden durch 'N' ersetzt. Haplogruppen werden auf Grundlage der Sequenz im FASTA Format mit dem Programm Haplofind [196] bestimmt. Der zusätzlich ausgegebene Wert des "*haploscore*" gibt die Wahrscheinlichkeit der Übereinstimmung mit der angegebenen Haplogruppe anhand der beobachteten Polymorphismen an.

Genotypisierung diploider Bereiche (Autosomen, X-Chromosom)

Hohe DNA-Schäden in Sequenzdaten alter DNA führen bei Nutzung von GATK zu erhöhten Heterozygotitätswerten, da die Fehlpaarung zum Referenzgenom ausgelöst durch Desaminierung von GATK als korrekt angenommen werden (siehe [172], Kapitel 2). Generell sind diploide Typisierung anhand von Genomen mit geringer Gesamtabdeckung (*low coverage*) nur begrenzt möglich bzw. fehlerhaft. Für *low coverage* Genome aus der vorliegenden Arbeit wird das Programm "ATLAS" verwendet, welches aufgrund des vorliegenden Datensatzes Sequenzierfehler und Desaminierungsraten erlernt und anhand einer Kovarianzmatrix eine Rekalibrierung der Basenqualitäten durchführt. Mit dem rekalibrierten Datensatz kann nachfolgend zum einen eine Allel Präsenz Typisierung (haploid) und eine diploide Typisierung durchgeführt werden [90]. Für genomische Daten mit einer geringer Gesamtabdeckung ist eine diploide Genotypisierung begrenzt auf wenige Abschnitte möglich (z.B. zeigt Rev 5 mit einer Abdeckung von 1.16x nur in 32% der genomischen Regionen eine 2-fache Abdeckung einer Position). Eine Erläuterung der zugrunde liegenden Algorithmen ist der Publikation Hofmanová, Kreutzer *et al.* (2016) ([71], SI Appendix SI5), sowie Kousathanas *et al.* (2016) [90] zu entnehmen.

Tabelle 3.17: Genotypisierung der Proben

	<i>mt DNA</i>	<i>Shotgun</i>
Realignierung	IndelRealigner (GATK)	
Rekalibrierung	BaseRecalibrator (GATK)	Rekalibrierung (ATLAS)
SNP Call	UnifiedGenotyper (GATK)	Allel Präsenz Typisierung (ATLAS) Diploide Typisierung (ATLAS)

Fragmentierungs- und Desaminierungsraten

Aus den Alignments können weitere Kriterien gezogen werden, die die Qualität und Authentizität der Sequenzen aus alten DNA Molekülen beschreiben. Die Bestimmung der Desaminierungsmuster an den Enden der DNA Moleküle wird durchgeführt mit *mapDamage 2.0* [78]. Standardmäßig wird die Rate der C/T Transitionen an den 5'-Enden der DNA bestimmt. Unter Verwendung von Sequenzierstrategien, deren Leseweite das gesamte alte DNA Molekül umfasst (zB. 100bp PE oder 150bp SE), kann auch die Desaminierungsrate am 3'-Ende erfasst werden. Wird das gesamte Molekül sequenziert kann zusätzlich die Fragmentlänge der Moleküle aus dem Alignment bestimmt werden, was aufschlussreich für die Größenverteilung der Moleküle innerhalb einer Probe ist.

Bestimmung der Kontaminationsraten von mitochondrialen Sequenzdaten

Für die Bestimmung von Varianten auf einem haploiden Genom wird im Fall von mehreren Allelen an einer Position jenes mit der höchsten Basenqualität gewählt (additives Verfahren) und im Ergebnis dargestellt. Zwar hat das Auftreten mehrerer Allele einen Einfluss auf den Qualitätswert des Genotyps, es ist aber lediglich durch die visuelle Betrachtung der Alignments zu erkennen (z.B. mit "samtools *view*"). Unterschiedliche Allele an einer Position können bei der Sequenzierung alter DNA durch Desaminierungen und Sequenzierfehler hervorgerufen werden. Diese Fehler treten jedoch sporadisch und zufällig verteilt über ein Genom auf. Liegen an einer Position mehrere Allele vor, dann ist dies ein starker Hinweis auf Kontamination, besonders wenn es sich um Positionen handelt, welche spezifische Änderungen einer Haplogruppe beschreiben. Da die Suche nach kontaminierenden Positionen per Hand sehr aufwändig und zeitintensiv ist, wurde ein Skript verwendet, welches einen diploiden

SNP Call über alle Positionen des mitochondrialen Alignments durchführt (GATK – HaplotypeCaller [40]). Als Ergebnis werden alle Allele mit ihrer Häufigkeit und Qualität dargestellt. Um Sequenzierfehler und Desaminierungen auszuschließen wird das Ergebnis nach Positionen gefiltert, welche eine Abdeckung von mindestens 5 Sequenzen haben, von denen 30% der Allele unterschiedlich sind. Als zusätzliche Information liefert diese Analyse Regionen, die mindestens 5-fach abgedeckt sind. Eine gängige Methode zur Darstellung der Authentizität eines alten DNA Datensatzes ist die rechnerische Bestimmung eines Kontaminationslevels bzw. eines Authentizitätswertes. Die ersten Analysen dieser Art wurden von Green und Kollegen für die Auswertung des ersten Neandertal Genoms entwickelt (MIA [57]). Anhand von fixierten Positionen des Neandertal Genoms wurde im Vergleich zu einem Datensatz von 311 modernen Individuen ein Kontaminationsspektrum errechnet. Archaische Menschenformen haben aufgrund der Divergenzzeit von >200.000 Jahren einen deutlichen Unterschied hinsichtlich fixierter Mutationen im Vergleich zum modernen Menschen – das trifft aber nicht bei der Analyse von frühen anatomisch modernen Menschen zu, auch wenn diese vor z.B. 5000 Jahren gelebt haben. Hierfür wurde die Kontaminationsbestimmung unter Einbeziehung einer Maximum-Likelihood Methode weiterentwickelt, welche gleichzeitig Sequenzierfehlern berücksichtigt und einen Authentizitätswert sowie eine Kontaminationsrate bestimmt (ContamMix [47]). Liegt der Authentizitätswert eines Datensatzes unter 95%, sprich 5% der Sequenzen können Kontaminationen zugeordnet werden, kann dieser Datensatz nicht in einer folgenden populationsgenetischen Analyse einbezogen werden. Es gibt jedoch Möglichkeiten, einen Datensatz nach authentischen, alten Molekülen zu filtern.

Dekontamination belasteter Proben

Kontaminierende Sequenzen sollten aufgrund ihres modernen Ursprungs geringere Desaminierungsraten an den Molekülen zeigen. Dieses Kriterium kann unter Anwendung des Programms *PMDtools* [176] genutzt werden, welches desaminierte Moleküle aus einem Sequenzdatensatz herauszufiltert. Für Proben mit Authentifizierungswerten unter 95% kann diese Methode angewendet werden, um potenziell kontaminierende Sequenzen zu entfernen und das authentische, alte DNA Molekül darzustellen. Der Nachteil dieser Anwendung ist, dass ca. 70% des Datensatzes verworfen wird – folglich benötigt man zu Beginn eine hohe Abdeckung des Zielgenoms. In Abhängigkeit des genutzten Schwellenwertes (*threshold*, $t_1 - t_5$) kann die Desaminierungsrate im Datensatz nach der Anwendung bis auf 80% steigen, was wiederum in Kombination mit einer geringen Abdeckung einer Position zu fehlerhaften Bestimmungen von Mutationen führen kann. *PMDtools* wird im Rahmen dieser Arbeit mit einem *threshold* von t_1 und t_3 angewendet. Die alignierten, nach Desaminierungen gefilterten Sequenzdaten wurden erneut hinsichtlich gesteigerter Desaminierungsrate, Fragmentlängenverteilung und Kontaminationsrate (bzw. Authentizitätswert) analysiert.

Seit Herbst 2015 steht ein weiteres Programm zur Verfügung, „*schmutzi*“ - entwickelt von Gabriel Renaud und Kollegen, welches über einen iterativen Vorgang aus einem Datensatz ein endogenes Mitochondrium assembliert und gleichzeitig eine Kontaminationsrate errechnet [154]. Als Parameter werden a priori eine Kontaminationsrate, Desaminierungsrate und (wenn möglich) die Fragmentlängenverteilung eines Datensatzes festgelegt, anhand derer endogene und kontaminierende Nukleotidpositionen innerhalb des Datensatzes bestimmt werden. Wie auch bei MIA/contamMix wird die Kontaminationsrate durch den Abgleich mit einer mtDNA Datenbank, welche ein Zusammenschnitt aller Haplogruppen enthält, errechnet. Vor jeder Iteration werden die Parameter neu bestimmt bis sich eine stabile Kontaminationsrate einstellt. Damit die Information der Fragmentlänge einbezogen werden kann, muss das alte DNA Molekül vollständig sequenziert sein (Option *-uselength*).

Ergebnisse der Laborarbeit

4.1 Genereller Erhaltungszustand

Allgemein zeigte das in Phase A/B untersuchte Probenmaterial makroskopisch einen schlechten Erhaltungszustand. Die Knochenelemente sind recht porös und mit Bakterienspots (schwarze Verfärbungen) versehen. Beim Sägen der Proben war nur bei zwei Proben der charakteristischer Geruch des Knochenkollagens zu erkennen (Krk 1 und Tha 2). An dieser Stelle wurden die Proben der Fundorte Alepotrypa (Alep), Mandra (Man), Mikri Volvi (Mikri, Abbildung 4.20b) und Knossos (Knoss, Abbildung 4.20a) zurückgestellt, da nur wenig kompaktes Knochenmaterial vorhanden war bzw. Zähne gebrochen waren. Zudem wurden aus den verbleibenden Fundplätzen die Einzelproben Fra 4, Pal 3 & Krk 3 aussortiert.



(a) Knossos 2



(b) Mikri volvi

Bei manchen Individuen erfolgte eine Probennahme von Zahn und Knochen. Da Zähne eine bessere DNA Konservierung nachgesagt wird, wurde von diesen Proben zunächst das Knochenmaterial untersucht, um Zähne für genomische Analysen zurückzuhalten. Von den Proben, Klei 4, Klei 5, Klei 6 & St 2 stehen jeweils die A-Proben (Zahnmaterialien) für weitere Analysen zur Verfügung. Bei St 3 und Xir 4 erbrachten die Knochenproben keine authentischen Ergebnisse, so dass in weiteren Versuchen auf die Zähne zurückgegriffen wurde.

Während der Probenvorbereitung der Felsenbeine in Phase C/D hat sich das Bild des generellen Erhaltungszustandes drastisch geändert. Schon während des Sägens war ein intensiver Geruch wahrzunehmen, so dass bei diesem Probenmaterial durch die Konservierung des Kollagens auf einen guten Erhalt der DNA geschlossen werden kann. Der innere Teil der Felsenbeine weist eine sehr dichte Knochenstruktur auf, welche durch die Kompaktheit des Materials marmoriert erscheint.



(a) Nea 5, Felsenbein, Ansicht innen



(b) Nea 5, Felsenbein, Ansicht außen

4.2 Extraktion

Die während der Probenvorbereitung getätigten Aussagen zur Probenqualität konnten in der folgenden Extraktion bestätigt werden. Allgemein zeigte sich bei Proben aus Phase B während der Phenol-Chloroform Aufreinigung der Lysate kaum eine Proteinschicht in der Interphase. Beide Parameter hinsichtlich der Qualität auf Basis der äußeren Inspektion während der Probenvorbereitung können in nachfolgenden Prozessen anhand der Bestimmung der Qualität auf molekularer Ebene bestätigt werden (siehe Abschnitt 4.3). Ein weiteres Problem dieser Probenserie ist die starke Verfärbung der Extrakte bedingt durch die Ko-Extraktion von Huminsäuren aus bodengelagerten Proben. Huminsäuren sind chemische Produkte, welche bei dem Abbau von biologischen Materialien entstehen. Ihre hochmolekulare Struktur kann eine Größe über mehrere 100kDa erreichen, so dass diese während der Reinigung/Konzentration des Extraktes anhand eines 50kDa Amikon nicht ausgewaschen werden. In Folgereaktionen können Enzyme durch Huminsäureverbindungen gehemmt werden. Die Proben der Phase C/D (Felsenbeine) zeigten während der Extraktion eine vielversprechende Proteinschicht in der Interphase. Auch hier konnten diese Beobachtungen durch die Bestimmung der Probenqualität auf molekularer Ebene bestätigt werden (siehe Abschnitt 4.3).

Die Innovationen im Ablauf der Extraktion hinsichtlich verwendeter Menge an EDTA und Inkubationszeit steigerte die allgemeine Effizienz. Es konnten mehr endogene DNA Moleküle freigesetzt werden, welche für den weiteren Ablauf zur Verfügung stehen. Eine vollständige Lyse des Skelettmaterials ist aufgrund der Knochendichte einer Probe nicht immer gewährleistet. Zu Testzwecken wurde die EDTA Menge unter Anwendung von 500mg Knochenpulver der Probe Pal 4 in einer parallelen Extraktion auf 10ml erhöht. Die Auswirkung der EDTA Erhöhung kann in der folgenden *Library* Reaktion getestet werden. In vier parallelen *Library* Reaktionen je Extrakt (Pal4 6,7/ Pal 4 10) kann anhand einer steigenden Einsatzmenge an Extraktvolumen ($5\mu\text{l}$, $20\mu\text{l}$, $35\mu\text{l}$, $50\mu\text{l}$) über die umgesetzte Molekülzahl auf *Fill In* Ebene eine Inhibition während der Umsetzung der Extraktmoleküle in *Library* Moleküle gemessen werden.

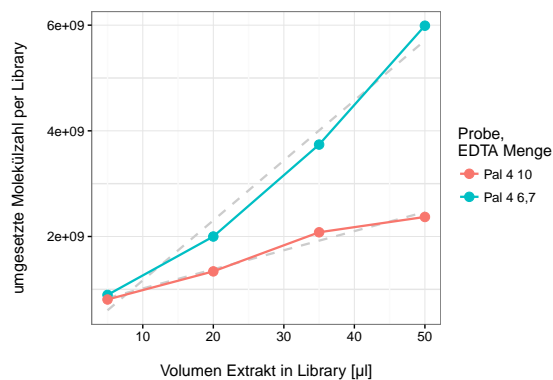


Abbildung 4.22: Inhibitionstest - unter Einsatz steigender Extraktmenge kann anhand der umgesetzten Molekülzahl die Inhibition der Enzyme im *Blunt-End* Repair gemessen werden. Es wurden 6,7 (blau) und 10 ml (rot) an EDTA zur Knochenlyse verwendet.

EDTA gilt als Inhibitor enzymatischer Reaktionen, da es Komplexe mit 2-wertigen Kationen eingeht, welche von Enzymen als Kofaktoren benötigt werden. Die Molekülzahl sollte ohne das Vorkommen von Inhibitoren mit ansteigendem Extraktvolumen linear zunehmen. In diesem Versuch zeigt die Extraktion mit der gesteigerten Menge an EDTA starke Anzeichen von Inhibition (Abbildung 4.22, Pal 4 10). Erhält man unter Einsatz von $5\mu\text{l}$ Extraktvolumen noch ähnliche Molekülzahlen, sieht man bei Verwendung von $50\mu\text{l}$ Extrakt eine 30%ig verringerte Ausbeute an einzigartigen Molekülen in der *Library* der Probe "Pal 4 10".

4.3 Bestimmung der Probenqualität

DNA Extrakte aus fossilem Probenmaterial sind aufgrund der geringen Menge an DNA Molekülen nicht quantifizierbar. Testweise wurden Extrakte alter DNA durch Qubit Messungen fluorometrisch quantifiziert. Diese Messungen zeigten meist eine Konzentration $\sim 1 \text{ ng}/\mu\text{l}$ oder kein Ergebnis (Daten werden nicht gezeigt). Folglich wird die DNA "blind" in den ersten Schritt des Versuchsablaufs, die *Library* Erstellung, eingesetzt. Nach Ligation der Adapter an die Extraktmoleküle ist die DNA durch eine quantitative PCR erfassbar - ab dieser Stelle des Versuchsablaufs können Aussagen zur Probenqualität getroffen werden.

Übersicht des allgemeinen *Library* Erfolges

Einen allgemeinen Überblick über den Erfolg einer *Library* Erstellung bietet die Darstellung der Molekülnzahlen in Proben, Leerkontrollen und Positivkontrollen nach dem *Fill In* Schritt. Zur Qualitätssicherung der technischen Durchführung der *Library* Erstellung dient die umgesetzte Molekülnzahl in der Positivkontrolle (siehe Abbildung 4.23, grün), deren Einsatzmenge mit 100ng je *Library* Ansatz stets konstant ist. Diese sollte nach der *Fill In* Reaktion als Richtwert mindestens 10^9 *Library* Moleküle enthalten. Leichte Abweichungen sind auf die Ligationseffizienz der Adapter an die Extraktmoleküle zurückzuführen¹.

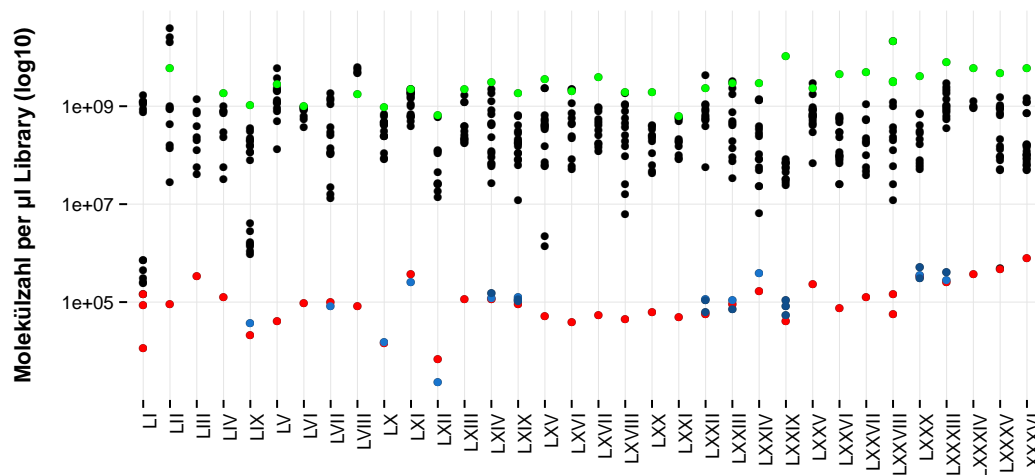


Abbildung 4.23: Darstellung des *Library* Erfolges ermittelt durch Messung der einzigartigen Molekülnzahl vor der PCR. Abgebildet sind alle *Libraries* des Projektes. Schwarz: Proben, grün: Positivkontrolle, rot: Leerkontrollen *Library*, blau: Leerkontrollen Müllern/Extraktion

Neben der Kontrolle der Arbeitsweise aller Chemikalien und Geräte durch die Positivkontrolle dient die Bestimmung der Molekülnzahl in Leerkontrollen des Müller-, Extraktions- und Libraryschrittes als wichtiges Kriterium für die Darstellung des kontaminationsfreien Arbeitens. Dabei sollten die

¹Für die Positivkontrolle wurden in den *Libraries* LI - LXXVII 100ng des selbst-hybridisierten Nonsense Moleküls eingesetzt. Die Umstellung auf eine bereits hybridisierte Positivkontrolle erfolgte in *Library* LXXVIII, in welcher die alte und neue Kontrolle parallel getestet wurden. In der neuen Positivkontrolle zeigt sich ein deutlicher Überschuss an Molekülen, was für eine ineffiziente Hybridisierung der alten Kontrolle spricht (siehe Abbildung 4.23, LXXVIII). Ab diesem Versuch wurde der Einsatz der neuen Positivkontrolle auf 25ng reduziert.

Moleküle der Proben mindestens zwei Zehnerpotenzen Abstand zu den Leerkontrollen zeigen. Eine hohe einzigartige Molekülzahl innerhalb der Proben hat einen positiven Effekt auf sämtliche Folgereaktionen. Zum einen bietet dies die Möglichkeit für eine tiefere Sequenzierung einer Probe und zum anderen wirken kontaminierende Moleküle aus der Umwelt weniger belastend auf die Proben (Darstellung in Abschnitt A.1, Tabelle A.55 & A.56). Die in Abbildung 4.23 aufgeführten Library Versuche zeigen mit einer Variation von 45% innerhalb aller gemessenen Positivkontrollen einen recht konstanten Verlauf der Ligationseffizienz. In der Literatur ist die Ligationseffizienz mit einer recht großen Abweichung von 30-70% angegeben [54]. In den Leerkontrollen finden wir eine Abweichung von 22% in der ermittelten Molekülzahl. Der größte Teil der analysierten Proben dieses Projektes zeigen eine Molekülzahl $>10^8$, jedoch ist darauf hinzuweisen, dass eine Vorauswahl durch visuelle Inspektion während der Probenvorbereitung getroffen wurde (siehe Auswertung 4.5).

Proben aus Phase A/B

Durch den Plot der einzigartigen Molekülzahl einer Probe gegen ihren Anteil an endogenen Sequenzen wird in Abbildung 4.24 die molekulare Komplexität der Proben zu Beginn der Projektzeit dargestellt. Um die Vergleichbarkeit der Proben untereinander zu wahren, wurden die einzigartigen Molekülzahlen per mg eingesetztes, pulverisiertes Skelettmaterial normalisiert (siehe Abschnitt 3.5). Allgemein ist der DNA Erhalt in den Proben recht niedrig - die meisten Proben sind unter einem Prozentsatz von 0.5% an endogener DNA vorzufinden. Für die Anreicherung des mitochondrialen Genoms wurden spezielle Versuchsreihen durchgeführt, mit dem Ziel die Ausbeute an Sequenzdaten für Proben dieser Qualität zu optimieren.

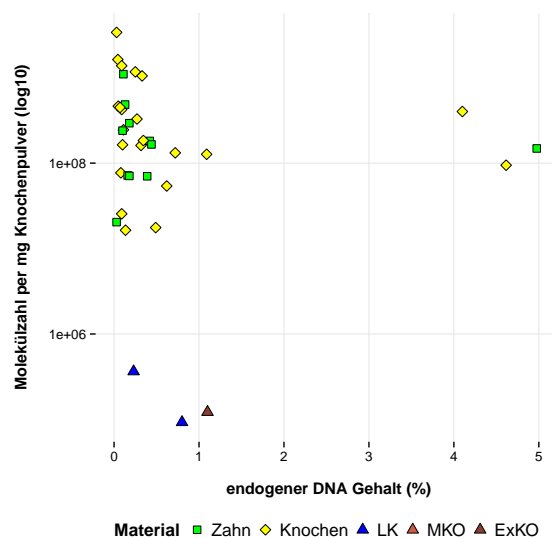


Abbildung 4.24: Darstellung der Qualität von Knochen und Zahnproben aus Phase B. x-Achse: endogener DNA Gehalt in Prozent (siehe Abschnitt 3.5), y-Achse: einzigartige Molekülzahl normalisiert per mg Pulver (Extraktion, siehe Abschnitt 3.5). Diese Graphik wurde auf Grundlage der Tabellen 4.20 - 4.26 erstellt.

LK: Leerkontrolle *Library*,
 MKO: Müller Kontrolle,
 ExKO: Extraktionskontrolle

Durch die Multiplikation des endogenen DNA Gehaltes einer Probe mit der Molekülzahl je μl *Fill In* Produkt wird die einzigartige, endogene Molekülzahl einer Probe zum Zeitpunkt des *Fill In's* berechnet (siehe Tabelle 4.18, Spalte "uniq"). Setzt man dies ins Verhältnis zur korrespondierenden Leerkontrolle, kann ein Kontaminationslevel für die Bearbeitungsphase im Spurenbereich ermittelt werden. Geräte, Chemikalien und Arbeitsweise des Müller-, Extraktion- und *Library* Schrittes können durch die zugehörige Leerkontrolle überprüft werden. In Tabelle 4.18 ist exemplarisch die Kalkulation eines theoretischen Kontaminationslevels während der *Library* Erstellung dargestellt (Proben stehen in Relation zur *Library* LK, LIII). An diesem Beispiel soll die Relevanz der einzigartigen Molekülzahl in Proben verdeutlicht werden. Je geringer die Molekülzahl - sprich Komplexität, desto höher wird

das Level an Kontamination innerhalb einer Probe. In den Schritten der Probenbearbeitung innerhalb dieses Projektes konnten auf diese Weise in Leerkontrollen keine höhere Belastung als 10^5 Moleküle pro μl festgestellt werden. Die erhobenen Parameter der Quantifizierung hinsichtlich Erhaltungszustand sind für alle Proben im Anhang dargestellt (siehe Anhang Proben *Screening* Abschnitt B.3).

Tabelle 4.18: Berechnung des theoretischen Kontaminationslevels am Beispiel von LIII.

qPCR FI (Lib): Molekülzahl per μl *Fill In* Produkt, endo [%]: endogener DNA Gehalt der Probe, uniq: Multiplikation der Parameter qPCR FI (Lib) & endo [%].

Probe	qPCR FI (Lib)	endo [%]	uniq	Kontaminationslevel (Lib)
Library LK	3,37E+05	0,78	2631	100,00%
Theo 4	3,86E+08	0,04	154400	1,70%
Fra 1	2,03E+08	0,08	162400	1,62%
Mau 1	1,29E+08	0,16	206400	1,27%
Krk 2	4,23E+08	0,1	423000	0,62%
Krk 4	7,89E+08	0,27	2128950	0,12%
Krk 5	2,88E+07	0,135	38813	6,78%
Tha 4	5,62E+07	0,09	50580	5,20%

Proben aus Phase C/D

Mit der Verfügbarkeit von Felsenbeinen als Untersuchungsgegenstand hat sich die Qualität der Proben drastisch geändert. Mehr als ein Viertel der Proben zeigen einen endogenen DNA Gehalt von $>20\%$. Nur 4 der 25 untersuchten Felsenbeine enthalten weniger als 5% humane DNA (siehe Abbildung 4.25).

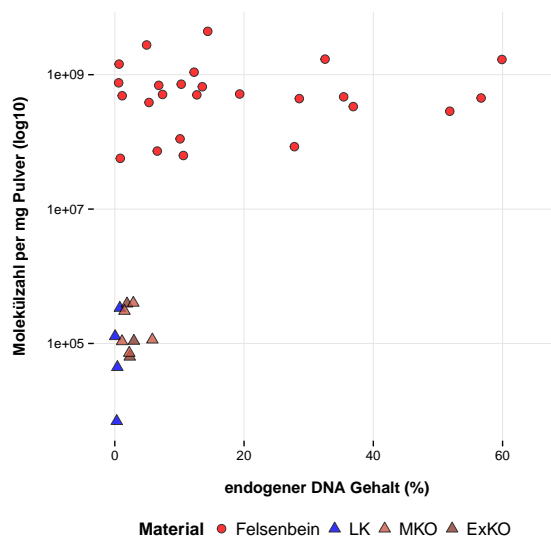


Abbildung 4.25: Darstellung der Qualität von Knochen und Zahnproben aus Phase C. x-Achse: endogener DNA Gehalt in Prozent (siehe 3.5), y-Achse: einzigartige Molekülzahl per mg Pulver (Extraktion, siehe Abschnitt 3.5).

LK: Leerkontrolle *Library*,
MKO: Müller Kontrolle,
ExKO: Extraktionskontrolle

Aus Proben mit hohem endogenen DNA Gehalt können direkt genomische Sequenzen durch eine *Shotgun* Sequenzierung erstellt werden. Um die Komplexität einer Probe zu bewahren und das Duplikationslevel niedrig zu halten, sollten diese Proben mit möglichst wenigen PCR Zyklen amplifiziert werden (siehe PCR Strategie Phase C/D, Abschnitt 3.5). Es werden direkt beide Indexe in der *Library* PCR angehängt, dass nach erfolgreicher Quantifizierung eine *Shotgun* Sequenzierung erfolgen kann. Zusätzlich werden alle Felsenbeine hinsichtlich des mitochondrialen Genoms angereichert. Die Ermittlung der mitochondrialen Kontaminationsrate (Abschnitt 3.8) kann vorab als Kriterium der

Probenwahl für genomische Analysen dienen. Ab einer Mindestanzahl von 100.000 alignierten Sequenzen (GRCh37/hg19) nach einer MiSeq Sequenzierung zur Bestimmung des endogenen DNA-Gehaltes, ist es möglich, eine molekulare Geschlechtsbestimmung durchzuführen ([177], Abbildung 4.26, Tabelle 4.19). Da einige Proben aufgrund einzelner Tests (Einführung neuer Indexsequenzen, Extraktions-tests) häufiger sequenziert worden sind, konnten diese Daten zusammengefasst werden. Das führte bei einigen Proben zu einer Gesamtabdeckung des humanen Genoms von 0.01x (siehe Tabelle 4.19).

Tabelle 4.19: Auswertung von "Screening" Daten (MiSeq 50bp SE) und molekulare Geschlechtsbestimmung von Proben mit >100.000 Sequenzen aligniert am humanen Referenzgenom (ALN hg19). Die durchschnittliche Abdeckung wurde durch QualiMap ermittelt [55]

cov: mittlerer Abdeckung hg19, STD (cov): Standardabweichung Abdeckung. Ab Spalte 5 sind die Parameter zur Geschlechtsidentifikation abgebildet ([177]); n.a.: nicht zugeordnet, (*): konsistent mit XY, nicht XX.

Probe	ALN hg19	cov	STD (cov)	Nseqs	ALN X+Y	ALN Y	R_y	SE	95% CI	Assign.
Krk1	142189	-		136395	3990	359	0,09	0,0045	0,0811-0,0989	XY
Klei9	298945	-		287837	14943	44	0,0029	0,0004	0,0021-0,0038	XX
Klei10	1856621	0.03X	0.18X	1802628	51584	4265	0,0827	0,0012	0,0803-0,0851	XY
Pla7	422532	0.01X	0.1X	577766	29755	91	0,0031	0,0003	0,0024-0,0037	XX
Rev4	534499	0.01X	0.09X	496849	16121	967	0,06	0,0019	0,0563-0,0636	n.a.
Rev5	606503	0.01X	0.1X	581623	30472	105	0,0034	0,0003	0,0028-0,0041	XX
Rev6	343222	0.01X	0.07X	316739	13684	358	0,0262	0,0014	0,0235-0,0288	n.a.
Nea2	1047237	0.02X	0.13X	999569	47415	179	0,0038	0,0003	0,0032-0,0043	XX
Nea3	639586	0.01X	0.1X	612130	28300	109	0,0039	0,0004	0,0031-0,0046	XX
Nea5	411841	0.01X	0.08X	398807	18998	64	0,0034	0,0004	0,0025-0,0042	XX
Thas1	384743	0.01X	0.08X	366510	10113	870	0,086	0,0028	0,0806-0,0915	XY
Mkg10	134381	-		119815	3385	281	0,083	0,0047	0,0737-0,0923	(*)
St5	620931	0.01X	0.1X	583799	16638	1304	0,0784	0,0021	0,0743-0,0825	(*)
St6	117787	-		108676	5498	13	0,0024	0,0007	0,0011-0,0036	XX

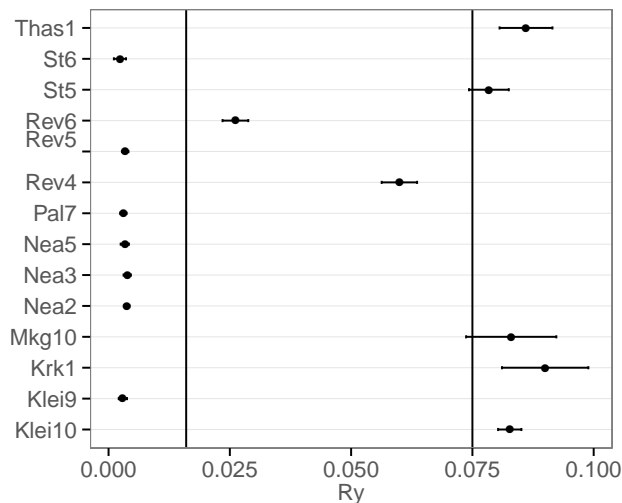


Abbildung 4.26: Molekulare Geschlechtsbestimmung anhand von Sequenzdaten.

Ry gibt das Verhältnis alignierter Sequenzen am Y Chromosom (ny) anteilig an den Sequenzen aligniert an beiden Geschlechtschromosomen (nx, ny) an: $Ry = ny / (nx + ny)$. $Ry < 0.016$ = weiblich, $Ry > 0.077$ = männlich (Abweichung anhand 95%igen Konfidenzintervall).

Molekularer Erhaltungszustand der Proben je Kulturstufe

Im Folgenden werden die die ermittelten Parameter der Quantifizierung für jede bearbeitete Probe dargestellt. Da für eine Probe parallele *Libraries* erstellt wurden, sind die Angabe der einzigartigen Molekülzahl nach *Fill In* (qPCR FI) und die ermittelten Fragmentlängen anhand der Bioanalyzer Messung (FragL (BA)) gemittelte Werte. Eine Sequenzierung zur Bestimmung des endogenen DNA Gehalts wurde hingegen mit Einzellibraries einer Probe durchgeführt. Die Charakterisierung einer Probe hinsichtlich ihrer Qualität basiert auf der molekularen Komplexität der endogenen Moleküle (uniq) und wird durch charakteristische Parameter zur Authentifizierung alter DNA Moleküle begleitet (mD5, FragL (BA)).

Proben aus dem Mesolithikum

Der molekulare Erhaltungszustand der Proben aus dem Mesolithikum ist als eher mäßig einzustufen und bestätigt das makroskopisch als porös eingestufte Skelettmaterial. Die einzigartigen Molekülzahlen zeigen einen ausreichenden Abstand zu den mitgeführten Leerkontrollen, jedoch ist nur ein geringfügiger Anteil auf einen humanen Ursprung zurückzuführen (uniq: $1,54 \cdot 10^5$ - $6,94 \cdot 10^5$). Dabei zeigt die Probe Theo 1 Potential für ein autosomal-genomische Analyse, jedoch verbunden mit einem hohen Materialverbrauch. Die Desaminierungsraten an den 5'-Positionen der DNA Moleküle sind als moderat einzustufen (mD5). Anhand des Alters der Proben wären stärkere Schäden der DNA Moleküle zu erwarten (mD5 >0.3), jedoch sind diese bei durchschnittlich 0,251 ($\pm 0,093$) vorzufinden. Geringe Desaminierungsraten in diesem Kontext können durch unsachgerechte Handhabung während der Bearbeitung erklärt werden, da exogen eingebrachte, moderne, humane DNA diese Rate senkt.

Tabelle 4.20: Molekülzahl, Fragmentlänge in bp (Bioanalyzer; BA) und MiSeq *Screening* Daten aus dem Mesolithikum

Theo = Theopetra (7600 - 6700 kal. v. Chr.), Fra = Franchthi (~ 7500 kal. v. Chr.).
qPCR FI= einzigartige Molekülzahl pro μ l vor PCR bzw. normalisiert pro g eingesetztes Skelettmaterial (N), uniq = einzigartig endogene Molekülzahl pro μ l vor PCR, mD5 = Desaminierungsrate 5'-Ende, FragL (BA) = endogene Fragmentlänge ermittelt anhand Bioanalyzer Messung in bp

Probe	Phase	Material	qPCR FI	qPCR FI (N)	endo	uniq	uniq (N)	mD5	FragL (BA)
Theo 1	A/B	Knochen	1,12E+08	5,40E+07	0,62	6,94E+05	3,35E+05	0,38	79
Theo 2	A/B	Knochen	6,21E+07	5,31E+07	-	-	-	-	90
Theo 3	A/B	Knochen	5,24E+07	1,76E+07	0,49	2,57E+05	8,62E+04	0,23	50
Theo 4	A/B	Knochen	3,86E+08	4,22E+08	0,04	1,54E+05	-	0,21	108
Theo 5	A/B	Knochen	9,27E+08	4,64E+08	0,05	4,64E+05	2,32E+05	0,25	78
Fra 1	A/B	Knochen	2,03E+08	7,70E+07	0,08	1,62E+05	6,16E+04	0,11	85
Fra 2	A/B	Knochen	8,64E+07	4,77E+07	-	-	-	-	70
Fra 3	A/B	Knochen	3,45E+08	2,41E+08	-	-	-	-	81
Fra 9	B	Knochen	3,22E+08	2,56E+08	0,09	2,90E+05	2,30E+05	0,31	73

Zur Erstellung eines mitochondrialen Datensatzes aus dem Mesolithikum wurden alle charakterisierten Proben hinsichtlich mitochondrialer DNA angereichert. Um der geringen Komplexität der Proben entgegenzuwirken wurden parallele Extraktionen einer Probe verwendet (siehe Auswertung).

Proben aus dem frühen Neolithikum

Innerhalb des früheneolithischen Probenspektrums kann eine deutliche Steigerung des molekularen Erhaltungszustandes festgestellt werden. Die Molekülzahlen der *Libraries* zeigen einen deutlichen Abstand hinsichtlich der mitgeführten Leerkontrollen ($>10^3$, Daten werden nicht gezeigt), da die *Fill In* Werte eine einzigartige Molekülzahl $>10^8$ zeigen (Ausnahme Probe Rev 7). Es sind authentische Desaminierungsraten und Fragmentierungsmuster mit einem mittleren Wert von 0,414 ($\pm 0,083$) bzw. 82 bp ($\pm 32bp$) zu erkennen. Die hohe Abweichung der Fragmentlängenverteilung ist auf die Diskrepanz des DNA Erhalts innerhalb von Knochen- und Zahnstrukturen zurückzuführen. Durchschnittlich zeigen in diesem Datensatz die Knochenelemente eine Fragmentlänge von $\sim 60-80bp$ und DNA Moleküle isoliert aus Zähnen eine Länge von $>100bp$.

Tabelle 4.21: Molekülzahl, Fragmentlänge in bp (Bioanalyzer; BA) und MiSeq *Screening* Daten aus dem Frühen Neolithikum

Mau = Mavropigi (6300 - 6000 v. Chr.), Xir = Xirolimni ($\sim 6200/6000$ v. Chr., Rev = Revenia (~ 6350 v. Chr.), Nea = Nea Nikomedia (~ 6150 v. Chr.), Pon = Pontokomi, qPCR FI= einzigartige Molekülzahl pro μl vor PCR bzw. normalisiert pro g eingesetztes Skelettmaterial (N), uniq = einzigartig endogene Molekülzahl pro μl vor PCR, mD5 = Desaminierungsrate 5'-Ende, FragL (BA) = endogene Fragmentlänge ermittelt anhand Bioanalyzer Messung in bp

Probe	Phase	Material	qPCR FI	qPCR FI (N)	endo	uniq	uniq (N)	mD5	Frag L (BA)
Mau 1	A/B	Zahn	1,29E+08	7,17E+07	0,16	2,06E+05	1,15E+05	0,42	100
Mau 2	A/B	Zahn	2,41E+08	1,82E+08	0,42	1,01E+06	7,64E+05	0,35	150
Mau 4	B	Knochen	6,04E+08	4,24E+08	-	-	-	-	69
Mau 5	B	Knochen	6,04E+08	4,24E+08	0,09	5,44E+05	3,82E+05	0,43	69
Xir 1a	A/B	Zahn	5,01E+08	1,49E+08	4,97	2,49E+07	7,39E+06	0,45	64
Xir 1b	B	Knochen	2,54E+08	1,27E+08	1,09	2,77E+06	1,38E+06	0,40	57
Xir 2	B	Zahn	8,85E+08	2,94E+08	0,18	1,59E+06	5,28E+05	0,29	62
Xir 4b	B	Knochen	4,92E+08	4,25E+08	-	-	-	-	142
Xir 4a	B	Zahn	8,13E+08	6,50E+08	-	-	-	-	67
Xir 5	A/B	Knochen	3,60E+08	1,61E+08	0,32	1,14E+06	5,11E+05	0,38	56
Rev 1	A/B	Zahn	1,85E+08	1,37E+08	-	-	-	-	158
Rev 2	A/B	Zahn	2,49E+08	1,81E+08	-	-	-	-	94
Rev 3	A/B	Zahn	1,38E+09	1,10E+09	0,11	1,52E+06	1,21E+06	0,33	136
Rev 4	C	Felsenbein	8,50E+08	6,62E+08	13,54	1,15E+08	8,96E+07	0,48	76
Rev 5	C	Felsenbein	5,47E+08	5,01E+08	12,69	6,95E+07	6,36E+07	0,54	64
Rev 6	C	Felsenbein	6,35E+08	5,08E+08	7,39	4,69E+07	3,75E+07	0,56	67
Rev 7	C	Felsenbein	6,25E+07	5,69E+07	0,85	5,31E+05	4,84E+05	0,39	60
Nea 1	D	Felsenbein	7,21E+08	1,09E+09	12,27	8,85E+07	1,34E+08	0,52	60
Nea 2	D	Felsenbein	2,86E+08	2,86E+08	51,83	1,48E+08	1,48E+08	0,34	64
Nea 3	D	Felsenbein	2,11E+08	4,50E+08	56,68	1,20E+08	2,55E+08	0,35	64
Nea 4	D	Felsenbein	6,94E+08	6,94E+08	6,80	4,72E+07	4,72E+07	0,50	63
Nea 5	D	Felsenbein	3,05E+08	4,68E+08	35,41	1,08E+08	1,66E+08	0,32	64
Pon 1	B	Knochen	1,10E+08	8,78E+07	-	-	-	-	69

In diesem Datensatz wird die Zunahme der Komplexität der endogenen DNA Moleküle isoliert aus Felsenbeinen deutlich. Die "uniq"-Werte sind mit Ausnahme der Probe Rev 7 konstant über einem Level von $4,69 \cdot 10^7$ zu finden, welche auf den hohen endogenen DNA Gehalt von durchschnittlich 21,94% zurückzuführen sind. Mit 65 bp ist die Fragmentlängenverteilung deutlich reduziert und kann konstant in allen Felsenbeinproben festgestellt werden ($\pm 5bp$). Die beschriebenen Desaminierungsraten von >0.50 gehören zu den höchsten in der Literatur beschriebenen Modifizierungen am 5'-Ende der alten DNA Moleküle.

Tabelle 4.22: Mittelwerte und Standardabweichung der Parameter zur Bestimmung der Qualität der Proben aus dem frühen Neolithikum. Zusätzlich sind diese aufgliedert in die einzelne Skelettelemente.

	n		qPCR FI	qPCR FI (N)	endo	uniq	uniq (N)	mD5	Frag L (BA)
Frühes Neolithikum	23	Durchschnitt	4.83E+08	3.97E+08	12.05	4.57E+07	5.61E+07	0.41	82
		Stdabw.	3.18E+08	2.95E+08	18.21	5.21E+07	7.69E+07	0.08	32
Knochen	6	Durchschnitt	4.04E+08	3.12E+08	0.50	1.48E+06	7.59E+05	0.40	79
		Stdabw.	1.99E+08	1.54E+08	0.52	1.15E+06	5.46E+05	0.03	36
Zahn	8	Durchschnitt	5.48E+08	3.45E+08	1.17	5.85E+06	2.00E+06	0.37	104
		Stdabw.	4.42E+08	3.54E+08	2.13	1.07E+07	3.04E+06	0.07	39
Felsenbein	9	Durchschnitt	4.79E+08	5.24E+08	21.94	8.26E+07	1.05E+08	0.44	65
		Stdabw.	2.70E+08	2.86E+08	20.67	4.59E+07	7.88E+07	0.10	5

Proben aus dem mittleren Neolithikum

Aus dem mittleren Neolithikum wurden alle verfügbaren Skelettelemente aus dem Norden Griechenlands akquiriert und im folgenden charakterisiert. Die Proben zeigen eine ausreichende einzigartige Molekülzahl um mitochondriale Genome zu erstellen, jedoch erweisen sich lediglich die Felsenbeine ausreichend komplex für eine autosomal-genetische Analyse. Für die Proben aus dem mittleren Neolithikum wurde erst ab Phase D, mit Aufkommen der Felsenbeine, der endogene DNA Gehalt konstant bestimmt. Diese zeigen mit Ausnahme von St 4 einen hohen molekularen Erhaltungszustand mit einzigartigen Molekülzahlen $\sim 10^9$ (qPCR FI). Gleichzeitig zeigen sich mit einer Desaminierungsfrequenz von ~ 0.40 authentische DNA Moleküle.

Tabelle 4.23: Molekülzahl, Fragmentlänge in bp (Bioanalyser; BA) und MiSeq *Screening* Daten aus dem Mittleren Neolithikum

Thas = Limenaria auf Thassos (5429 kal. v. Chr.), NS = Nea Makri (etwa 5500 v. Chr.), St = Stavroupouli (etwa 5750/5300 v. Chr.), qPCR FI= einzigartige Molekülzahl pro μ l vor PCR bzw. normalisiert pro g eingesetztes Skelettmaterial (N), uniq = einzigartig endogene Molekülzahl pro μ l vor PCR, mD5 = Desaminierungsrate 5'-Ende, FragL (BA) = endogene Fragmentlänge ermittelt anhand Bioanalyser Messung in bp

Probe	Phase	Material	qPCR FI	qPCR FI (N)	endo	uniq	uniq (N)	mD5	Frag L (BA)
Thas 1	C	Felsenbein	7,96E+07	8,49E+07	27,79	2,21E+07	2,36E+07	0,40	66
NS 1	B	Knochen	4,23E+08	3,62E+08	-	-	-	-	85
NS 1	B	Zahn	1,85E+08	1,48E+08	-	-	-	-	75
NS 2	B	Zahn1	9,15E+08	7,26E+08	-	-	-	-	78
NS 2	B	Zahn2	1,55E+09	1,23E+09	-	-	-	-	79
St 1	B	Knochen	6,04E+07	4,83E+07	-	-	-	-	105
St 2	B	Knochen	1,69E+09	1,05E+09	0,33	5,58E+06	3,47E+06	0,41	74
St 3	B	Knochen	3,21E+08	2,20E+08	-	-	-	-	95
St 3	B	Zahn	9,71E+08	7,77E+08	-	-	-	-	100
St 4	D	Felsenbein	8,99E+08	7,57E+08	0,6	5,39E+06	4,54E+06	0,22	66
St 5	D	Felsenbein	5,97E+08	1,70E+09	32,54	1,94E+08	5,55E+08	0,45	66
St 6	D	Felsenbein	1,52E+09	2,77E+09	4,93	7,46E+07	1,36E+08	0,50	67

Tabelle 4.24: Mittelwerte und Standardabweichung der Parameter zur Bestimmung der Qualität der Proben aus dem mittleren Neolithikum. Zusätzlich sind diese aufgliedert in die einzelne Skelettelemente.

	n		qPCR FI	qPCR FI (N)	endo	uniq	uniq (N)	mD5	Frag (BA)	L
Mittleres Neolithikum	12	Durchschnitt	7.67E+08	8.23E+08	13.24	6.04E+07	1.45E+08	0.40	80	
		Stdabw.	5.85E+08	7.97E+08	15.65	7.99E+07	2.36E+08	0.11	14	
Knochen	4	Durchschnitt	6.24E+08	4.20E+08	0.33	5.58E+06	3.47E+06	0.41	90	
		Stdabw.	7.27E+08	4.39E+08	-	-	-	-	13	
Zahn	4	Durchschnitt	9.05E+08	7.21E+08	-	-	-	-	83	
		Stdabw.	5.59E+08	4.45E+08	-	-	-	-	11	
Felsenbein	4	Durchschnitt	7.73E+08	1.33E+09	16.46	7.40E+07	1.80E+08	0.39	66	
		Stdabw.	6.00E+08	1.17E+09	16.04	8.53E+07	2.57E+08	0.12	1	

Erhaltungszustand der Proben aus dem späten/finalen Neolithikum

Das späte/finale Neolithikum setzt sich aus Proben aus dem Norden und dem Süden Griechenlands zusammen. Obwohl die Proben aus dem Süden im Schnitt 1000 Jahre jünger sind, zeigen diese einen weitaus schlechteren molekularen Erhaltungszustand. Nur die Probe Tha 2 besitzt mit einem endogenen DNA Gehalt von 4,1% Potential für eine autosomal-genetische Analyse. Des weiteren eignet sich die Probe Fra 8 für die Anreicherung des mitochondrialen Genoms.

Im Norden Griechenlands konnte ein breites Skelettinventar aus dem späten/finalen Neolithikum gesichert werden, welches in diesem Projekt nahezu vollständig charakterisiert werden konnte. Diese Materialien zeigten in Knochen- und Zahnelementen einen mässigen Erhaltungszustand, jedoch genügend, um diese Proben für eine Anreicherung des mitochondrialen Genoms zu verwenden. Hervorzuheben ist die Beprobung mehrerer Skelettelemente eines Individuums, wie es bei den Proben Krk 8 und Klei 3 der Fall ist. In beiden Proben zeigte sich eine deutliche Verbesserung des molekularen Erhaltungszustandes des Felsenbeins gegenüber dem Zahn bzw. Knochenelement des jeweiligen Individuums. Auch in dieser Zeitstufe ist eine konstante Fragmentierung der DNA Moleküle extrahiert aus den Felsenbeinen zu erkennen ($65 \pm 6\text{bp}$, siehe Tabelle 4.25).

Tabelle 4.25: Mittelwerte und Standardabweichung der Parameter zur Bestimmung der Qualität der Proben aus dem späten/finalen Neolithikum. Zusätzlich sind diese aufgliedert in die einzelne Skelettelemente.

	n		qPCR FI	qPCR FI (N)	endo	uniq	uniq (N)	mD5	Frag (BA)	L
Spätes Neolithikum	32	Durchschnitt	8.35E+08	6.69E+08	7.04	9.25E+07	7.14E+07	0.39	80	
		Stdabw.	1.13E+09	9.86E+08	13.48	3.47E+08	2.13E+08	0.10	28	
Knochen	12	Durchschnitt	1.19E+09	7.54E+08	0.53	2.23E+06	1.05E+06	0.36	72	
		Stdabw.	1.51E+09	1.00E+09	1.29	3.75E+06	1.31E+06	0.08	13	
Zahn	8	Durchschnitt	3.41E+08	2.11E+08	0.21	5.07E+05	3.34E+05	0.34	114	
		Stdabw.	3.50E+08	2.16E+08	0.17	3.73E+05	2.85E+05	0.09	37	
Felsenbein	12	Durchschnitt	8.09E+08	8.90E+08	16.98	2.29E+08	1.77E+08	0.44	65	
		Stdabw.	9.82E+08	1.22E+09	17.27	5.32E+08	3.15E+08	0.10	6	

Tabelle 4.26: Molekülzahl, Fragmentlänge in bp (Bioanalyser; BA) und MiSeq *Screening* Daten aus dem Späten/Finalen Neolithikum

Krk = Kremasti koulada (~ 5100 v. Chr.), Mkg = Makriyalos (~ 4700 v. Chr.), Pal = Paliambela (~ 4900 - 3200 v. Chr.), Klei = Kleitos (~ 5000 v. Chr.), Fra = Franchthi (~ 4000 v. Chr.), Tha = Tharrounia (~ 4300 v. Chr.), Kea = Kephala Kea, qPCR FI= einzigartige Molekülzahl pro μl vor PCR bzw. normalisiert pro g eingesetztes Skelettmaterial (N), uniq = einzigartig endogene Molekülzahl pro μl vor PCR, mD5 = Desaminierungsrate 5'-Ende, FragL (BA) = endogene Fragmentlänge ermittelt anhand Bioanalyser Messung in bp

Probe	Phase	Material	qPCR FI	qPCR FI (N)	endo	uniq	uniq (N)	mD5	Frag L (BA)
Krk 1	A/B	Knochen	2,94E+08	9,46E+07	4,62	1,35E+07	4,36E+06	0,36	80
Krk 2	A/B	Knochen	4,23E+08	1,64E+08	0,10	4,23E+05	1,64E+05	0,32	64
Krk 4	A/B	Knochen	7,89E+08	3,29E+08	0,27	2,13E+06	8,87E+05	0,37	63
Krk 5	A/B	Knochen	2,88E+07	1,64E+07	0,14	3,88E+04	2,21E+04	0,36	86
Krk 6	A/B	Knochen	4,04E+05	8,86E+04	0,25	1,01E+03	2,21E+02	0,18	50
Krk 7	A/B	Knochen	5,14E+08	2,45E+08	0,11	5,65E+05	2,70E+05	0,46	52
Krk 8	B	Zahn	3,93E+08	2,40E+08	0,10	3,93E+05	2,40E+05	0,29	109
Krk 8 (11)	D	Felsenbein	2,13E+09	4,42E+09	14,38	3,06E+08	6,36E+08	0,38	65
Krk 9	B	Zahn	9,39E+07	7,02E+07	0,39	3,66E+05	2,74E+05	0,31	153
Krk 10	D	Felsenbein	2,17E+08	7,22E+08	10,28	2,23E+07	7,42E+07	0,32	61
Mkg 1	A/B	Zahn	3,17E+07	2,04E+07	0,03	9,51E+03	6,12E+03	0,52	124
Mkg 9	D	Felsenbein	5,14E+07	6,27E+07	10,61	5,45E+06	6,65E+06	0,54	60
Mkg 10	D	Felsenbein	4,05E+08	5,16E+08	19,32	7,83E+07	9,97E+07	0,59	59
Mkg 11	D	Felsenbein	5,92E+07	7,30E+07	6,57	3,89E+06	4,80E+06	0,50	59
Mkg 12	D	Felsenbein	6,96E+07	1,11E+08	10,10	7,03E+06	1,12E+07	0,55	63
Mkg 13	D	Felsenbein	1,65E+08	3,86E+08	5,29	8,73E+06	2,04E+07	0,52	61
Pal 1	B	Zahn	6,06E+08	4,85E+08	0,13	7,88E+05	6,31E+05	0,33	138
Pal 2	B	Zahn	6,51E+07	4,76E+07	-	-	-	-	152
Pal 4	B	Knochen	2,18E+09	1,63E+09	0,05	9,80E+05	7,34E+05	0,36	86
Pal 5	B	Knochen	5,35E+09	3,39E+09	0,03	1,61E+06	1,02E+06	0,44	90
Pal 6	B	Zahn	1,07E+09	5,92E+08	-	-	-	-	116
Pal 7	C	Felsenbein	5,86E+08	3,35E+08	36,90	2,16E+08	1,24E+08	0,43	70
Klei 1a	A/B	Zahn	2,45E+08	1,66E+08	0,44	1,08E+06	7,28E+05	0,3	59
Klei 1b	B	Knochen	2,45E+08	1,84E+08	0,35	8,45E+05	6,35E+05	0,34	72
Klei 2	A/B	Zahn	2,27E+08	7,09E+07	0,18	4,08E+05	1,28E+05	0,29	62
Klei 3a	A/B	Knochen	8,04E+08	4,48E+08	0,07	5,63E+05	3,14E+05	0,48	67
Klei 3b	C	Felsenbein	6,72E+08	4,41E+08	28,55	1,92E+08	1,26E+08	0,35	73
Klei 5	B	Knochen	1,72E+09	1,17E+09	0,25	4,30E+06	2,93E+06	0,33	74
Klei 6	B	Knochen	1,94E+09	1,38E+09	0,09	1,75E+06	1,24E+06	0,35	76
Klei 7	C	Felsenbein	5,81E+08	4,86E+08	1,14	6,64E+06	5,56E+06	0,37	72
Klei 8	C	Felsenbein	1,62E+09	1,44E+09	0,67	1,08E+07	9,58E+06	0,39	66
Klei 10	C	Felsenbein	3,15E+09	1,68E+09	59,90	1,89E+09	1,01E+09	0,30	74
Fra 5	A/B	Knochen	2,56E+07	1,86E+07	-	-	-	-	56
Fra 6	A/B	Knochen	3,74E+08	2,23E+08	-	-	-	-	76
Fra 8	A/B	Knochen	1,62E+08	1,32E+08	0,72	1,17E+06	9,50E+05	0,33	84
Tha 1	A/B	Knochen	9,80E+08	6,08E+08	0,06	5,88E+05	3,65E+05	0,57	64
Tha 2	A/B	Knochen	4,02E+08	4,02E+08	4,10	1,65E+07	1,65E+07	0,24	62
Tha 4	A/B	Knochen	5,62E+07	2,55E+07	0,09	5,06E+04	2,30E+04	0,42	60
Kea 1	A/B	Knochen	4,86E+08	5,70E+08	0,01	4,86E+04	5,70E+04	-	63

Desaminierungsraten

Abbildung 4.27 zeigt alle Proben des Projektes mit zugehörigen Desaminierungsraten. Es ist auffällig, dass alte DNA Moleküle extrahiert aus Felsenbeinen erhöhte Raten im Vergleich zu Zähnen und Knochen zeigen. Manche Felsenbeine zeigen Desaminierungsraten weit über 50%, eine Eigenschaft, die in der Literatur bisher nur bei sehr alten Funden, wie z.B. den mittelpaläolithischen Funden aus Atapuerca (*Homo heidelbergensis*) beschrieben wurde. Anzumerken ist, dass Desaminierungsfrequenzen stark abhängig von Wahl der Extraktionsmethode und Polymerase sind (siehe Abschnitt 4.5).

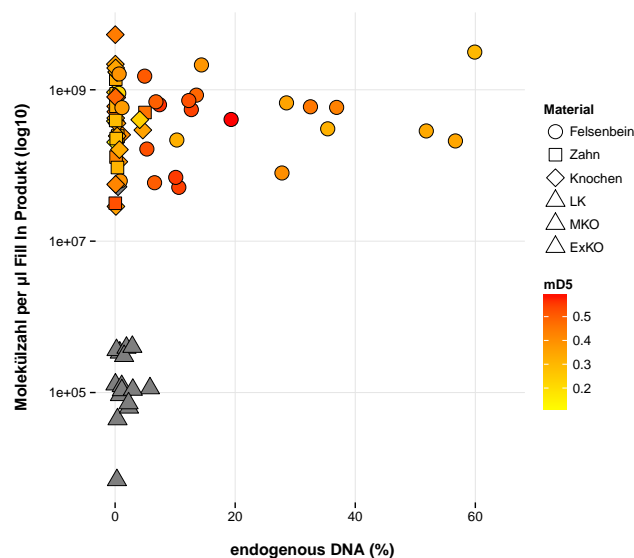


Abbildung 4.27: Alle Proben des Projektes. Die Farbgebung entspricht den Deaminierungsraten am 5'-Ende eines Moleküles.

Hingehend der Annahme, dass DNA Schäden (Fragmentierung und Desaminierung) mit dem Alter einer Probe zunehmen, werden Desaminierungsraten ausgewählter Proben in Abbildung 4.28 verglichen¹. Es kann keine Korrelation zwischen Probenalter und akkumulierten Schaden festgestellt werden, so dass Zeit als alleiniger Faktor für den DNA Erhalt ausgeschlossen werden kann (siehe im Vergleich [167, 78]). Im Gegenteil zeigt eine neolithische Probe mit 58% Desaminierung eine Erhöhung von 20% gegenüber einer 3000 Jahre älteren mesolithischen Probe (Abbildung 4.28, unten). Desweiteren können Unterschiede innerhalb gleichartiger Fundsituation und untersuchten Skelettelement festgestellt werden (Abbildung 4.28, mitte). Selbst innerhalb eines Individuums können unterschiedliche Erhaltungszustände der DNA Moleküle erfasst werden, wie es bei der Darstellung eines Zahnes und eines Felsenbeines der Probe Krk 8 abgebildet ist (Abbildung 4.28, oben).

¹Die dargestellten Desaminierungsfrequenzen an den 3'-Enden entsprechen nicht den wahren Frequenzen dieser Moleküle, da bei einem 50bp Lauf (SE) oft nicht bis zum Ende der Moleküle sequenziert wurde. Jedoch soll mit dieser Darstellung der Hinweis gegeben werden, dass über den Grad der 3'-Desaminierung im Vergleich zur 5'-Desaminierung einer Probe eine Abschätzung über die Fragmentlänge gegeben werden kann. Befinden sich beide Werte auf einem ähnlichen Level, ist mit einer Leseweite von 50bp das komplette Molekül sequenziert.

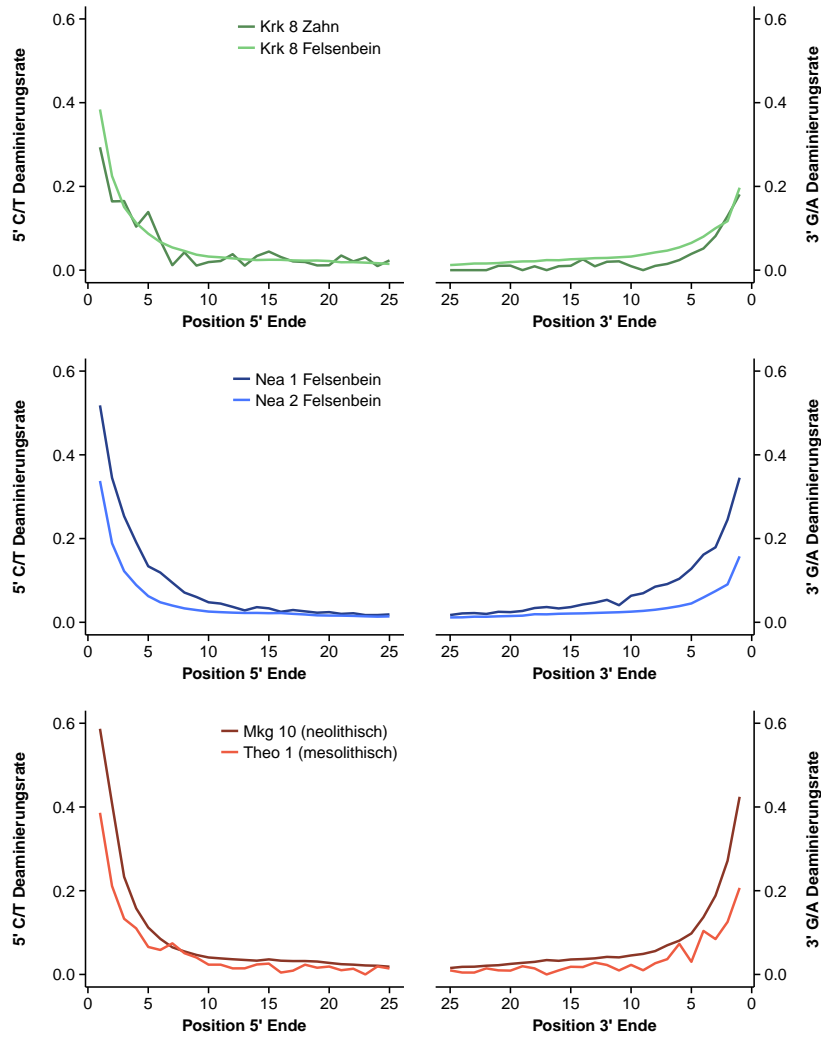


Abbildung 4.28: Darstellung der Desaminierungsraten ausgewählter Proben nach einer 50 bp Sequenzierung (MiSeq, *Screen*). Oben: Zwei Skelettelemente eines Individuums (Krk 8), Mitte: Zwei Felsenbeine unterschiedlicher Individuen eines Fundplatzes (ähnliches Alter/Fundsituation), Unten: Zwei Proben, die zeitlich etwa 3000 Jahre auseinander liegen.

4.4 Sequenzierung kompletter Genome (*Shotgun*)

Nach erfolgreichem *Screening* der Probenqualität wurden in Phase C drei neolithische Proben für eine tiefere Sequenzierung des kompletten Genoms ausgesucht. Dazu gehören die früh-neolithische Probe Rev 5 und zwei weitere Proben aus den folgenden Epochen des Neolithikums, Pal 7 (spät) und Klei 10 (final).

Für die Probe **Rev 5** waren mehrere Sequenzierung nötig um eine durchschnittlich einfache Abdeckung des Genoms zu erreichen. Da die zweite Extraktion eine verringerte Komplexität zeigte, wurden von dieser nur zwei *Libraries* sequenziert. Die erste Extraktion wurde hingegen auf einem weiteren HiSeq Lauf sequenziert. Die Probe Rev 5 zeigt im Vergleich zu den anderen Proben Griechenlands starke Fragmentierungs- und Desaminierungsraten, welche mit einer durchschnittlichen alignierten Fragmentlänge von 55 bp und einer Desaminierungsrate von 55% zu einer der extremsten in der Literatur auffindbaren Werte gehören. Bei einer dritten Extraktion dieser Probe wurde zur Aufkonzentration des Extraktes ein 30k Amikon Filter genutzt, um vermehrt kleinere Moleküle zurückzuhalten. Zur Steigerung des endogenen DNA Gehaltes wurde während dieser Extraktion eine Prelyse durchgeführt (persönliche Kommunikation Amelie Scheu). Da diese Probe in vorhergehenden Analysen während der Genotypisierung erhebliche Probleme aufgrund der hohen Desaminierungsrate zeigte, wurden die Extrakte für die letzte Sequenzierung mit dem USERTM Enzymmix behandelt. Nach einer Inkubationszeit von 3 Stunden blieben jedoch 14% der ersten Basenpositionen desaminiert (siehe Tabelle 4.27). Da durch diese Enzymbehandlung desaminierte Basen aus der DNA geschnitten werden, erfolgte eine Reduktion der Fragmentlänge um etwa 10bp im Gegensatz zum unbehandelten Extrakt, was wiederum zu einem erhöhten Verlust bei Anwendung des Längensfilters von 30bp nach der Sequenzierung führte. Zur Demonstration sind diese Schritte in Tabelle 4.29 aufgeführt: Während durch den UDG Verdau der Anteil an endogener DNA steigt (Spalte "endo [%] (Mkdup)"), kommt es durch die Anwendung des Längensfilters von 30 bp zu einem Verlust von 22-30% aller alignierten Moleküle (Spalte "seq<30bp [%]").

Die Probe **Klei 10** wurde durch einen HiSeq Lauf und zwei weitere NextSeq Läufe unter Verwendung der v.1 Chemie sequenziert. Die NextSeq Läufe bereiteten einige Schwierigkeiten während der Sequenzierung und in der folgende Analyse der Sequenzdaten. Das erste Problem lag beim Auslesen der Index-Sequenzen während der Sequenzierung, was auf die verwendete Chemie zurückzuführen ist. Laut Angabe der Illumina-Techniker, haben die Sequenzierprimer für das Auslesen der Indexsequenz nicht korrekt aligniert, so dass die Probe nicht korrekt sortiert werden konnte (persönliche Kommunikation Sven Bikar, StarSeq). An sich kein Problem, da in diesem Ansatz nur diese Probe sequenziert wurde. Während der Datenanalyse wurde jedoch das zweite Problem sichtbar, da etwa 50-60% an Daten während der Anwendung des Qualitätsfilters verloren gegangen sind (Parameter: 5% der Basen einer Sequenz haben einen Basenqualitätswert <15). Es ist anzumerken, dass exakt diese *Libraries* zuvor auf einem MiSeq Lauf quantifiziert und durch einen HiSeq Lauf tiefer sequenziert worden sind, bei welchen weder Probleme in der Indexzuordnung noch Verluste durch den Qualitätsfilter entstanden sind (durch Qualitätsfilter nach HiSeq: 10%). Nach Rücksprache der Sequenzierfirma mit Technikern der Firma Illumina wurde der gleiche Reaktionsansatz der Probe erneut auf der NextSeq Maschine sequenziert. Dieser wiederholte Lauf war nach Angabe der Sequenzierfirma weniger durch Probleme auffällig, jedoch entstand an dieser Stelle erneut ein Verlust von 30% der Daten durch den Qualitätsfilter. Da diese *Libraries* fehlerfrei in einem MiSeq Lauf zur Quantifizierung, sowie in einem HiSeq Lauf während der ersten tieferen Sequenzierung bearbeitet werden konnten, gehe ich persönlich nicht von einem Problem innerhalb der Probe, sondern von einem technischen Problem während der

NextSeq Sequenzierungen aus. Recherchiert man heute (etwa 1,5 Jahre später (Stand Februar 2016)) über NextSeq Sequenzierungen der ersten Generation (Verwendung der v.1-Chemie), ist das Ergebnis, dass von der Verwendung dieser Datensätze abgeraten wird, da das Zusammenspiel der verwendeten Chemie und der NextSeq Technologie noch nicht ausgereift war¹. Bei der NextSeq Technologie ist mit der Absicht die Sequenzierdauer zu verringern die Basenidentifizierung auf zwei Farbkanäle reduziert. Die Basen Thymin und Cytosin sind durch die Farbgebung grün und rot codiert, Adenin besteht aus einem rot-grün Mix und Guanin trägt keine Farbaddition. Die Cluster werden durch nur zwei Farbfilter gescannt worauf die Zeiteinsparung beruht. Wie andere Nutzer (siehe Fussnote) berichten, konnte zu diesem Zeitpunkt durch die NextSeq Technologie zum einen nicht die Basenqualität erreicht werden, wie sie bei einem HiSeq Lauf möglich ist, und auch das duale Indexsystem wird nur unter Nutzung der Original Illumina Indices empfohlen. Die Probe Klei 10 trägt jedoch einen selbst generierten Index (Index 26, nach [84]). Allgemein wird in diesem Forum von der Genotypisierung von *low coverage* Datensätzen erstellt durch die NextSeq v.1 Chemie abgeraten.

Die bioinformatische Auswertung zeigte, eine niedrige Basenqualitäten für die ersten Positionen der Sequenzen (FastQC Analyse, Graphiken nicht dargestellt), weshalb diese an den ersten drei Positionen manuell geschnitten und erneut analysiert wurden. Durch die Annahme, dass über den Qualitätsfilter fehlerhafte Sequenzen ausgeschlossen und damit die Genotypisierung nicht beeinflusst werden sollte, sind diese Daten populationsgenetisch analysiert worden. Durch die Anwendung von ATLAS zur Genotypisierung stellte sich jedoch eine erhöhte Heterozygotitätsrate innerhalb dieser NextSeq Läufe heraus, so dass diese Sequenzen aus folgenden Analysen ausgeschlossen wurden. Auf Ebene der *f*-Statistik war keine Änderung bezüglich der Ergebnisse feststellbar (persönliche Kommunikation Zuzana Hofmanová), jedoch war es mir wichtig mit einem sauberen Datensatz zu arbeiten, so dass diese Daten zur Veröffentlichung ausgeschlossen worden sind. (Anmerkung: Die Probe Rev 5 ist ebenfalls auf einem getrennten NextSeq Lauf sequenziert worden, jedoch waren diese Daten bezüglich Indexsortierung, Qualitätsfilter und Heterozygotitätswert unauffällig.)

Die Probe **Pal 7** war sowohl während der Probenvorbereitung, als auch der Sequenzierung und Datenanalyse unauffällig. An dieser Stelle soll noch einmal der Vorteil eines Felsenbeines für die plärogenetische Analyse hervorgehoben werden. Mit nur einer *lane* auf einem HiSeq 2500 konnte nach Sequenzierung von 4 parallelen *Libraries* einer Probe eine Abdeckung des Genoms von 1,28x erreicht werden. Eine Begebenheit, die zu Beginn dieser Projektzeit nicht denkbar gewesen wäre.

¹<http://seqanswers.com/forums/archive/index.php/t-40741.html>

Tabelle 4.27: Anzahl sequenzierter Extraktionen und *Libraries* je Probe inklusive *Screening* Ergebnisse.

Extr.= Extraktion, uniq=einzigartige Molekülzahl der *Library*, Frag BA= Fragmentlänge bestimmt durch Bioanalyser Messung, Seq= Anzahl Sequenzen, Seqs kept= Sequenzen nach Qualitätsfilter, ALN (+dupl)= alignierte Sequenzen an hg19 mit Duplikaten, ALN (filter)= alignierte Sequenzen nach Duplikatsentfernung und Anwendung des Längenfilters (30bp). endo= endogener DNA Gehalt errechnet aus dem Verhältnis der gefilterten alignierten Sequenzen zu dem Gesamtdatensatz einer Probe

Extraktion			Library						MiSeq Screening					
Probe	Extr.	Extr. Nummer	Library Nummer	Volumen Extrakt [µl]	UDG	uniq	Frag BA [bp]	PCR Zyklen (Parallelen)	Seq	Seqs kept	ALN (+ dup)	ALN (filter)	endo [%]	mD5
Rev 5	I	XXVII	LXXII.7	50	nein	5,23E+08	64	16 (4)	1201688	1176745	167643	167436	14,23%	0.541
Rev 5	I	XXVII	LXXIII.9	50	nein	4,41E+08	65	14 (4)						
Rev 5	I	XXVII	LXXV.3+4	50	nein	8,24E+08	65	14 (4)						
Rev 5	II	XXIX	LXXIII.6+7	50	nein	8,47E+07	65	14 (4)	304089	297628	26252	26218	8,81%	0.537
Rev 5	III	XXXIII	XXXIII.12	50	nein	5,51E+08	61	14 (4)	608145	599725	88745	88676	14,79%	0.510
Rev 5	III	XXXIII	LXXXIV.1-4	25	ja	1,03E+09	54	14 (4)	735259	728166	144991	101667	13,96%	0.128
Klei 10	I	XXVII	LXXII.4	50	nein	4,23E+09	74	16 (4)	2158232	2141639	1261510	1259154	58,79%	0.296
Klei 10	I	XXVII	LXXIII.8	50	nein	3,31E+09	72	14 (4)						
Klei 10	II	XXIX	LXXIII.3+4	50	nein	2,07E+09	79	14 (4)	1030749	1023482	620550	619586	60,54%	0.262
Klei 10	II	XXIX	LXXIII.3+4	50	nein	2,07E+09	79	14 (4)						
Pal 7	I	XXVII	LXXII.5	50	nein	1,03E+09	70	16 (4)	1137385	1118976	429378	428814	38,32%	0.429
Pal 7	I	XXVII	LXXV.1	50	nein	2,23E+09	68	14 (4)						
Pal 7	II	XXIX	LXXIII.5	50	nein	1,41E+08	80	14 (4)	524128	516335	183530	183267	35,49%	0.386
Pal 7	II	XXIX	LXXV.2	50	nein	7,50E+08	76	14 (4)						

Tabelle 4.28: Sequenzierstrategie und Auswertung der neolithischen Genome.

Extr.= Extraktion, seq. Lib= Anzahl sequenzierter *Libraries* je Extraktion, Seqs= Anzahl Sequenzen, Seqs kept= Sequenzen nach Qualitätsfilter, join= Anzahl überlappender Sequenzpaare nach *paired end* Sequenzierung, ALN (filter)= alignierte Sequenzen nach Duplikatsentfernung und Anwendung des Längensfilters (30bp). endo= endogener DNA Gehalt errechnet aus dem Verhältnis der gefilterten alignierten Sequenzen zu dem Gesamtdatensatz einer Probe

Probe	Extr.	seq. Lib.	Plattform	Seqs	Seqs kept	join	% join	ALN (filter)	endo	Frag BAM (bp)	cov	mD5	mD3
Rev 5	I	2	Next Seq PE	9,48E+07	3,65E+07	3,50E+07	96,02%	5,26E+06	15,03%	50,7	0,08	0,549	0,554
Rev 5	II	2	Next Seq PE	4,14E+07	1,59E+07	1,46E+07	92,11%	1,40E+06	9,59%	53,8	0,02	0,544	0,537
Rev 5	I	4	HiSeq PE	4,21E+08	1,69E+08	1,45E+08	86,04%	2,19E+07	15,07%	51,4	0,36	0,584	0,541
Rev 5	I	4	HiSeq rapid PE	2,51E+08	1,03E+08	1,01E+08	98,13%	1,47E+07	14,57%	52,0	0,25	0,525	0,541
Rev 5	III	4	HiSeq rapid SE	2,33E+08	2,33E+08	-	-	3,28E+07	14,09%	43,8	0,46	0,130	0,120
Rev 5 total					5,57E+08	2,96E+08		4,27E+07	13,67%		1,16		
Pal 7	I	2	HiSeq PE	2,74E+08	1,11E+08	9,65E+07	86,66%	4,00E+07	41,39%	64,2	0,83	0,436	0,434
Pal 7	II	2	HiSeq PE	1,46E+08	5,93E+07	5,11E+07	86,18%	2,01E+07	39,38%	72,0	0,47	0,413	0,438
Pal 7 total					1,46E+08	5,93E+07	5,11E+07	6,01E+07	40,38%		1,28		
Klei 10 I	I	2	Next Seq PE	1,76E+08	3,22E+07	3,16E+07	98,24%	1,72E+07	54,53%	60,9	0,33	0,217	0,233
Klei 10 II	II	2	Next Seq PE	7,97E+07	1,14E+07	1,12E+07	98,22%	6,41E+06	57,11%	60,7	0,12	0,258	0,264
Klei 10 I	I	2	Next Seq PE	1,75E+08	5,29E+07	5,15E+07	97,42%	2,84E+07	55,18%	64,8	0,59	0,217	0,236
Klei 10 II	II	2	Next Seq PE	6,98E+07	1,67E+07	1,62E+07	97,08%	9,32E+06	57,36%	65,5	0,19	0,258	0,264
Klei 10 I	I	2	HiSeq PE	2,31E+08	9,87E+07	8,73E+07	88,53%	5,36E+07	61,32%	69,5	1,20	0,298	0,317
Klei 10 II	II	2	HiSeq PE	1,36E+08	5,90E+07	5,23E+07	88,68%	3,36E+07	64,17%	75,4	0,82	0,263	0,279
Klei 10 total					8,67E+08	2,71E+08	2,50E+08	1,45E+08	58,28%		3,2		
Klei 10 total (ohne NextSeq)					3,67E+08	1,58E+08	1,40E+08	8,71E+07	62,75%		2,01		

Tabelle 4.29: Detaillierte Darstellung der Probe Rev 5 (dritte Extraktion: 30k Amikon, USER™)

Probe	Seqs kept	ALN (+dupl)	ALN (+dupl) [%]	ALN Mkdup	endo [%] (Mkdup)	DUP	ALN (filter)	seq<30bp	seq<30bp [%]	endo [%] (filter)	mD5	mD3
Rev-5-XXXIV-1	56601709	11587376	20,5%	11391085	20,1%	1,723%	8127709	3263376	28,65%	14,36%	0.143	0.124
Rev-5-XXXIV-2	52084597	10824181	20,8%	10651330	20,5%	1,623%	7367189	3284141	30,83%	14,14%	0.138	0.127
Rev-5-XXXIV-3	65763934	12113569	18,4%	11917354	18,1%	1,646%	8766204	3151150	26,44%	13,33%	0.125	0.121
Rev-5-XXXIV-4	58383157	11260944	19,3%	11069334	19,0%	1,731%	8554510	2514824	22,72%	14,65%	0.116	0.110

4.5 Auswertung der Capture Experimente

Prüfung der Optimierungsstrategien

Um einen ersten Eindruck über die Modifizierungen des Versuchsablaufes zu gewinnen, wurden zwei Proben des Fundplatzes Kremasti koilada bearbeitet. Anhand des äußeren Erscheinungsbilds der Proben soll das molekulare Verhalten einer "gut" erhaltenen im Vergleich zu einer "schlecht" erhaltenen Probe ermittelt werden und als Erfahrungswert für folgende Versuchsreihen genutzt werden (siehe Abbildung B.58g & B.58h; makroskopische Probeninspektion zur Probenwahl).

Schon die *Library* Erstellung zeigt einen großen Unterschied in der Anzahl der umgesetzten Moleküle (siehe Tabelle 4.30 "qPCR FI Werte"). Berechnet man ein theoretisches Kontaminationslevel anhand der Molekülzahl in den Proben gegen die Molekülzahl der Leerkontrollen, dann kann für die gute Probe Krk 1 ein maximales Level von 0,011% und für die schlechte Probe Krk 6 ein maximales Level von 56% ermittelt werden. Die Probe Krk 6 zeigte nach PCR (10/10) eine sehr geringe Konzentration, so dass diese Probe nach Reamplifizierung (8 Zyklen) komplett in das *Capture* eingesetzt wurde. Um Effekte von Einsatzmengen in einem *Capture* zu sehen, wurden von der Probe Krk 1 zwei Reaktionsansätze erstellt, wobei der erste 5 μ l (P1) und der zweite 10 μ l (P2) je *Library* enthielt. Auch wenn in P2 die doppelte Menge an Ausgangsmolekülen enthalten sind, so drückt sich dies nicht in einer gesteigerten *Capture* Effizienz aus (Tabelle 4.31, ALN mt [%]). Nach Duplikatsentfernung zeigen beide Ansätze sogar eine relativ identische Anzahl an alignierten Sequenzen.

Tabelle 4.30: *Library* Parameter der Versuchsreihe, siehe auch Abbildung 4.23.

Extr.: Extraktion, N (Lib): Normalisierungsfaktor Extraktebene, qPCR FI: Anzahl an Molekülen in der *Library*, FI Kont.: theoretisches Kontaminationslevel errechnet anhand der Moleküle der Leerkontrollen (qPCR FI), qPCR FI (N): Molekülzahl per mg eingesetzten Knochenpulver während der Extraktion

Library #	Probe	Extr.	N (Lib)	qPCR FI	FI Kont. Level	qPCR FI (N)
LI.1	LK 1			8,48E+04		
LI.2	Krk 1.1	VIII	139,5	1,68E+09	0,005%	4,81E+08
LI.3	Krk 1.2	VIII	139,5	1,22E+09	0,007%	3,50E+08
LI.4	Krk 1.3	VIII	139,5	1,22E+09	0,007%	3,49E+08
LI.5	Krk 1.4	VIII	139,5	1,28E+09	0,006%	3,67E+08
LI.6	Krk 1.5	VIII	139,5	1,09E+09	0,007%	3,14E+08
LI.7	Krk 1.6	VIII	139,5	7,51E+08	0,011%	2,15E+08
LI.8	Krk 1.7	VIII	139,5	8,50E+08	0,009%	2,44E+08
LI.9	LK 2			1,13E+04		
LI.10	Krk 6.1	X	182,5	1,42E+05	56,176%	3,12E+04
LI.11	Krk 6.2	X	182,5	2,43E+05	32,873%	5,32E+04
LI.12	Krk 6.3	X	182,5	7,34E+05	10,874%	1,61E+05
LI.13	Krk 6.4	X	182,5	7,04E+05	11,343%	1,54E+05
LI.14	Krk 6.5	X	182,5	4,39E+05	18,186%	9,62E+04
LI.15	Krk 6.6	X	182,5	3,09E+05	25,823%	6,78E+04
LI.16	Krk 6.7	X	182,5	2,57E+05	31,115%	5,63E+04
LI.17	LK 3			1,44E+05		

Für die "schlecht" erhaltene Probe Krk 6 kann eine deutliche Steigerung der *Capture* Effizienz (Tabelle 4.31, ALN mt [%]) im Vergleich zu ähnlichen Proben aus den Versuchen aus Phase A (Kapitel 3.1, Tabelle 3.9) festgestellt werden, jedoch bleiben nach Duplikatsentfernung nur 2725 Sequenzen zurück. Hier wird das Ausmaß der hohen Zahl an PCR Zyklen deutlich (10/10 + 8). Zudem zeigt die Probe Krk 6 eine hohe Kontaminationsrate (4,5 - 20,5 %), welche auf die geringe Menge an einzigartigen Molekülen zurückzuführen ist. Je näher sich eine Probe im Bereich der Leerkontrollen befindet, desto größer ist der Effekt an kontaminierenden Molekülen aus der Umgebung. Eine hohe

Molekülzahl, wie wir sie in Krk 1 vorfinden, ist hingegen weniger anfällig für Kontaminationen und bietet der *Capture* Reaktion schon bei kleinsten Einsatzmengen eine ausreichende Sättigung.

Tabelle 4.31: Datenauswertung *Capture* Produkte Krk 1 & Krk 6 nach 100bp *paired end* (PE) Sequenzierung.

(a) Parameter Qualität und Alignment

Seq: Anzahl Sequenzen (Rohdaten), Qual: Anzahl Sequenzen nach Qualitätsfilter (q15), join: Anzahl überlappende Sequenzen (PE), ALN mt: alignierte Sequenzen (Seq) und Verhältnis zu *join* [%], ALN Mkdup: nach Duplikatsentfernung

Probe	Seq	Qual	join	ALN mt	ALN mt [%]	ALN Mkdup
Krk 1 P2	20930181	17345911	7000162	3982332	56,89%	33045
Krk 1 P1	28772362	23937323	9962855	6108559	61,31%	33060
Krk 6 reamp.	19405288	18608004	8826359	3295746	37,34%	2725

(b) Fortsetzung

DUP %: Duplikationslevel, cov: mittlere Abdeckung der Zielregion (*coverage*), mD5/mD3: Deaminierungsrate am 5' - bzw. 3' Ende der Moleküle, Kont.: Kontaminationsangabe in Prozent

Probe	DUP [%]	cov	mD5	mD3	Kont. [%]
Krk 1 P2	99,17%	188	0,269	0,317	0,2 - 0,4
Krk 1 P1	99,46%	184	0,268	0,317	
Krk 6 reamp.	99,92%	17,01	0,208	0,286	4,5 - 20,5

Ermittlung der Sequenzdatenmenge für mitochondriale *Capture* Produkte

Wie in Abbildung 4.29 dargestellt, ist die mögliche Sequenziertiefe abhängig von der Komplexität einer Probe (Datengrundlage siehe Anhang, Tabelle B.65). Auch wenn ein ähnlicher endogener DNA Gehalt vorliegt, kann aus einer Probe mit einer höheren Anzahl an einzigartigen Molekülen eine größere Datenmenge herausgeholt werden (vgl. Probe Pal 4b und Pal 5 bzw. Krk 9 und Mau 2). Die qualitativ schlechteste Probe des Datensatzes (Xir 4b) zeigt, dass mit Zunahme der Sequenzdatenmenge keine adäquate Abdeckung erreicht werden kann. Proben dieser Komplexität haben schon mit 100.000 Sequenzen ein Duplikationslevel von >90% und nähern sich damit dem Sättigungspunkt, so dass eine tiefere Sequenzierung nur einen geringen Zuwachs an alignierten Sequenzen bedeutet. Für komplexe Proben dieser Darstellung (Xir 1a, Mau 2) genügen hingegen 50.000 sequenzierte Moleküle um eine 30-fache Abdeckung zu erreichen.

Anhand einer Quantifizierung des endogenen Materials einer *Library* kann folglich bereits vor der Anreicherung einer Zielregion die Qualitätsstufe einer Probe ermittelt werden und auf diesem Weg die durch eine Sequenzierung benötigte Rohdatenmenge bestimmt werden. Dies soll im folgenden dargestellt werden: Tabelle 4.32 zeigt die endogene Molekülzahl nach Quantifizierung, welche näherungsweise mit der erreichbaren Sequenziertiefe im Zusammenhang steht. Aus diesem Datensatz können Qualitätsstufen anhand der endogenen Molekülzahl pro μl *Fill In* Produkt ermittelt werden. Eine 30-fache Abdeckung der Zielregion (rCRS) ist mit etwa 10.000 alignierten Sequenzen ohne Duplikate erreicht¹ (siehe Anhang, Tabelle B.65). Während der Erstellung eines Probenpools für eine Sequenzierreaktion können nun die Proben anhand der antizipierten Rohdaten, anstatt äquimolar, zusammengeführt werden.

¹Die *coverage* wird stark von der durchschnittlichen Fragmentlänge einer Probe beeinflusst

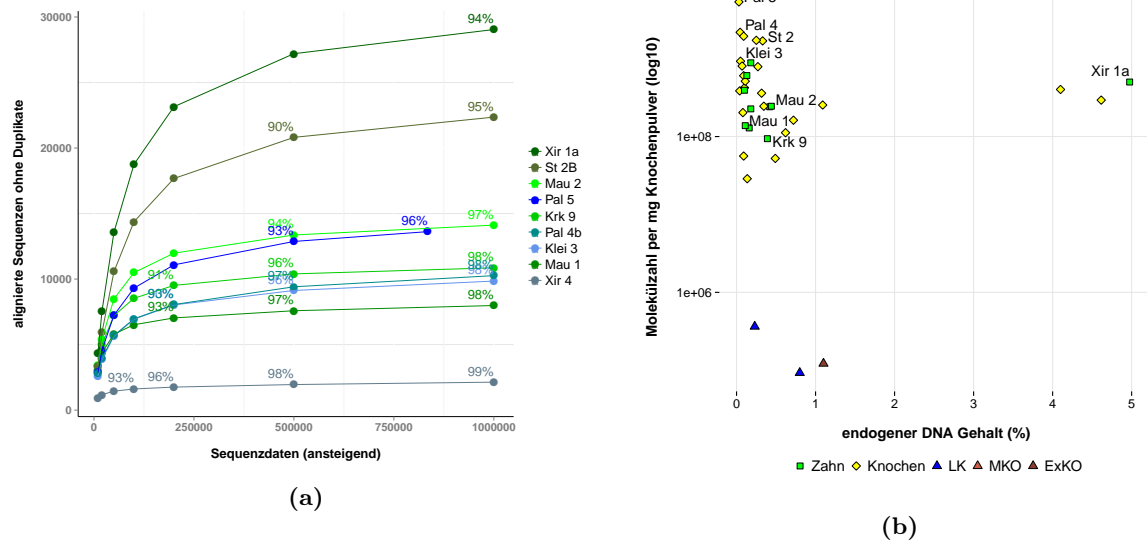


Abbildung 4.29: Aligierte Sequenzen (rCRS) ohne Duplikate für ausgewählte Proben. Farbgebung - grün: endogen >0.1 %, blau: endogen <0.1 %. Duplikationslevel sind über 90% in der Abbildung angegeben.

Tabelle 4.32: Einordnung der Proben in Qualitätsstufen und Berechnung der benötigten Rohdatenmenge nach *Capture* für eine 30x Abdeckung des Mitochondriums

Molekülzahl (qPCR FI (Lib)), endogener DNA Gehalt (endo [%]), errechnete endogene Molekülzahl pro μl *Fill In* Produkt (uniq).

Probe	qPCR FI (Lib)	endo [%]	uniq	Qualitätsstufe	antizipierte Rohdaten
Xir 1a	5,01E+08	4,97	18453900	1	50k
St 2B	1,69E+09	0,33	5577000		
Pal 5	5,35E+09	0,03	1605000	2	200k
Mau 2	2,41E+08	0,42	1012200		
Pal 4b	2,18E+09	0,04	600996	3	1000k
Klei 3a	8,04E+08	0,07	562800		
Krk 9	9,39E+07	0,39	366210		
Mau 1	1,29E+08	0,16	206400		

Experimente zur Hybridisierungs- und Waschttemperatur

Dieses Experiment beinhaltet Proben mit niedrigen (Klei 5/6, St 3B) und hohen (Mau 4, Rev 1) Fragmentierungsraten, zusätzlich wurden bewusst Proben mit mäßigem Erhaltungszustand gewählt (Rev 1, St3B). Bei allen Proben ist eine Abnahme der Spezifität mit Verringerung der Temperatur zu sehen (Tabelle 4.33, 4.34, Spalte "ALN mt [%]" und Abbildung 4.30). Bei stark fragmentierten Proben erreicht man durch die Absenkung der Temperatur eine Steigerung alignierten Sequenzen ohne Duplikate um 11 - 28% (Tabelle 4.33 4.34, Spalte "ALN Mkdup"). Diese Steigerung kann durch die Anreicherung kleinere Fragmente erklärt werden, was in Abbildung 4.30 durch eine Verschiebung der alignierten Fragmentlängen nach links ersichtlich ist. Durch die Retention kurzer Fragmente ist eine leichte Anhebung der Desaminierungsraten zu erkennen (Spalte "mD"). Aufgrund der größeren Vielfalt an Molekülen sinkt das Duplikationslevel in der Sequenzierung (Spalte "DUP %"). Für die weniger fragmentierten Proben ergibt sich eine Steigerung der Datenmenge um 3 - 7%. Bei allen Reaktionen mit gesenkter Temperatur ist eine Zunahme der Authentizität sichtbar (Ausnahme: Mau 5 - diese starke Abnahme ist nur durch eine Kontamination im Post-PCR Bereich zu erklären).

Das *bait* Design wurde anhand des humanen Genoms durchgeführt, so dass es sich in diesem Fall nicht um eine Spezies-übergreifende Reaktion handelt. Infolgedessen finden wir durch Absenkung der Hybridisierungstemperatur (60°C, T) kaum eine Steigerung in der Datenmenge erreicht, lediglich die Absenkung der Waschtemperatur ist ausschlaggebend. Die Touchdown Hybridisierung zeigt eine weitere Verbesserung, jedoch wurde hier zusätzlich eine neue Strategie zur Amplifikation des *Capture* Produktes genutzt. Die *On-bead* PCR Strategie wurde ab Phase C der Projektzeit in Kombination mit Änderung der Index Strategie angewendet.

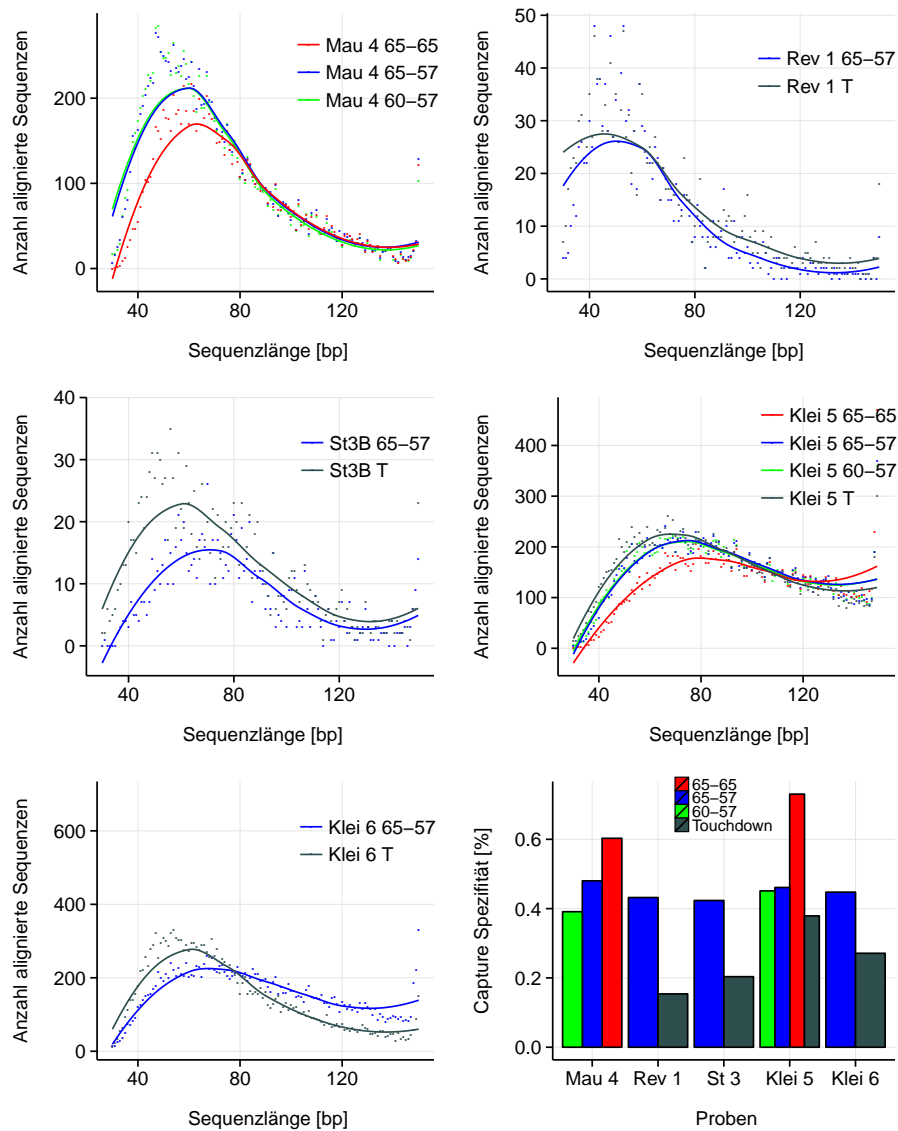


Abbildung 4.30: Fragmentlängen einer Probe nach Nutzung unterschiedlicher Hybridisierungs- und Waschtemperaturen im *Capture*. Abgebildet sind alignierte Sequenzen (ohne Duplikate) als Funktion der Fragmentlänge. Das Balkendiagramm zeigt die Spezifität der einzelnen *Capture* Reaktionen (% alignierte sequenzen mit Duplikaten in Abhängigkeit der Gesamtdatenmenge).

Tabelle 4.33: Datenauswertung nach Nutzung unterschiedlicher Hybridisierungs- und Waschttemperaturen auf Basis der gleichen Sequenzdatenmenge je Probe nach 150bp SE Sequenzierung (MISeq). Hier kann eine Angabe zur alignierten Fragmentlänge getroffen werden (Frag BAM [bp]).

Seq: Anzahl Sequenzen (Rohdaten), ALN mt: alignierte Sequenzen (Seq) und Verhältnis zu Rohdatenmenge [%], ALN Mkdup: nach Duplikatsentfernung, mD5/mD3: Deaminierungsrate am 5' - bzw. 3' Ende der Moleküle, auth. Daten: Authentizität des Datensatzes.

Probe	Seq	ALN mt	ALN mt [%]	ALN Mkdup	DUP [%]	cov	Frag BAM [bp]	mD5	mD3	auth. Daten [%]
Mau 4 65-65	552601	317824	60,3%	10407	96,7%	48,92	77,41 (27,85)	0,430	0,467	92,04
Mau 4 65-57	552601	254125	48,0%	12938	94,9%	56,13	71,65 (27,93)	0,438	0,472	93,58
Mau 4 60-57	552601	206617	39,1%	12756	93,8%	54,28	70,35 (27,17)	0,440	0,474	98,23
Rev 1 65-57	228190	94531	43,2%	1441	98,5%	5,48	63,14 (24,74)	0,422	0,451	65,83
Rev 1 T	228190	33290	15,4%	1682	94,9%	6,73	66,88 (28,88)	0,400	0,461	73,03
St 3B 65-57	268301	107843	42,3%	1653	98,5%	4,90	83,07 (29,18)	0,392	0,314	83,42
St 3B T	268301	51783	20,4%	1540	97,0%	7,05	75,90 (30,02)	0,416	0,423	93,04
Klei5B 65-65	348154	275731	83,2%	20918	92,4%	139,54	109,99 (34,88)	0,319	0,279	98,76
Klei5B 65-57	377502	167184	46,1%	22079	86,8%	135,53	101,18 (35,90)	0,323	0,315	97,25
Klei5B 60-57	377502	163651	45,1%	22096	86,5%	135,28	100,93 (36,19)	0,324	0,310	99,96
Klei 5B T	377502	137118	37,9%	22005	84,0%	127,47	95,57 (36,11)	0,328	0,329	98,64
Klei6B 65-57	812829	348688	44,8%	18688	94,6%	94,22	83,17 (32,98)	0,402	0,436	99,97
Klei 6B T	812829	210996	27,1%	19247	90,9%	92,48	79,35 (32,52)	0,405	0,452	99,98
Xir 5 65-65	1033710	442510	45,1%	2657	99,4%	11,35	79,59 (35,99)	0,367	0,321	94,19
Xir 5 60-57	1033710	517739	52,7%	2938	99,4%	14,20	72,74 (36,57)	0,399	0,383	98,24

Tabelle 4.34: Datenauswertung nach Nutzung unterschiedlicher Hybridisierungs- und Waschttemperaturen auf Basis der gleichen Sequenzdatenmenge je Probe nach 50bp SE Sequenzierung (MISeq).

Abkürzungen siehe oben

Probe	Seq	ALN mt	ALN mt [%]	ALN Mkdup	DUP [%]	cov	Frag BAM [bp]	mD5	mD3	auth. Daten [%]
Klei 4 65-65	1651985	1041456	64,4%	20202	98,1%	61,12	-	0,381	0,072	99,95
Klei 4 60-57	1651985	913834	56,4%	20807	97,7%	61,79	-	0,406	0,140	99,51
Mau 5 65-65	855893	430675	51,6%	9790	97,7%	29,24	-	0,454	0,125	96,23
Mau 5 60-57	855893	341122	41,0%	15512	95,5%	45,42	-	0,457	0,189	77,61
Xir 5 65-65	377137	178458	47,9%	2197	98,8%	5,95	-	0,380	0,168	96,67182
Xir 5 60-57	377137	204383	54,9%	2339	98,8%	6,88	-	0,408	0,145	93,48406

Polymerase Test

Nach Amplifikation durch die Test-Polymerasen zeigen alle *Library*-PCR-Produkte anhand von Bioanalyzer Messungen unterschiedliche Ausprägungen in Bezug auf Fragmentlänge und Konzentration. Die Produkte der AccuPrime Polymerase zeigen Primerdimere in hoher Konzentration, welche meist das eigentliche *Library* Produkt überragen. Hinsichtlich des Anreicherungs Erfolges des mitochondrialen Genoms zeichnet sich ein Vorteil für PCR Produkte erstellt durch die Kombination Taq/Herc ab. Bezüglich der Spezifität des Captures zeigt der Versuch kaum konsistente Ergebnisse. Deutlich ist jedoch der Unterschied in der Desaminierungsfrequenz nach Amplifikation durch AccuPrime Polymerase im Vergleich zu anderen Polymerasen, welche um 10-20% gesenkt wurde (siehe Abbildung 4.31). Ebenfalls zeigen diese Proben eine hohe Duplikationsrate, was auf einen Verlust von endogenen, einzigartigen Molekülen während der Amplifikation hinweist (siehe Tabelle B.64).

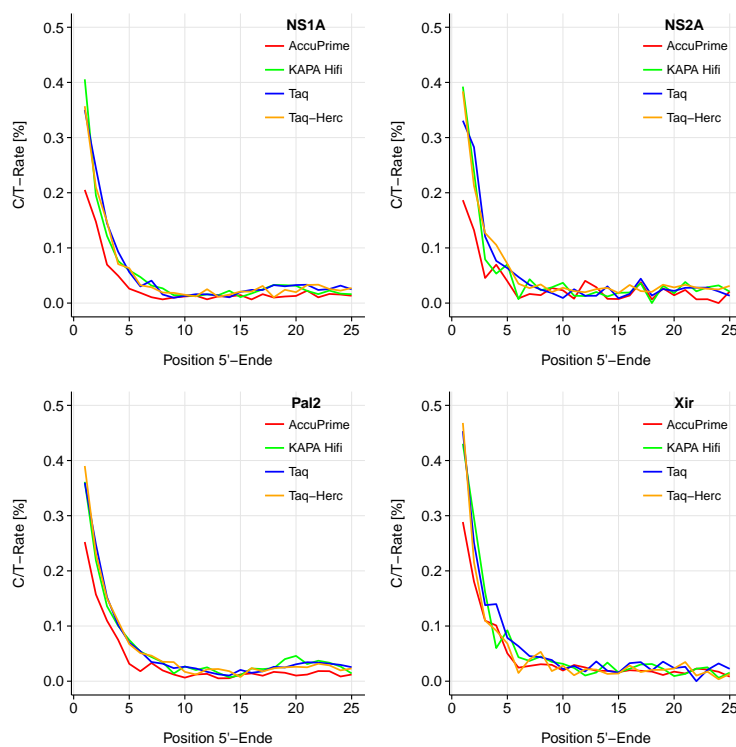


Abbildung 4.31: Desaminierungsraten vier ausgewählter Proben nach Anwendung verschiedener Amplifizierungsstrategien während der *Library* Erstellung.

Darstellung des optimierten Versuchablaufs zur Sequenzierung kompletter mitochondrialer Genome aus biomolekular schlecht erhaltenen Probenmaterialien

Nach erfolgreicher Optimierung einzelner Komponenten des Versuchablaufs soll nun anhand des modifizierten Protokolls auf die Proben (Extrakte¹) aus Phase A zurückgegriffen werden. Jene zeigten keine positiven Resultate in einer initialen Bearbeitung durch NGS Methoden (siehe Abschnitt 3.1). Von sieben Proben waren jeweils 2 Extrakte gelagert, welche nun in parallele Libraries überführt und anteilig innerhalb eines Reaktionsansatzes hinsichtlich des mt Genoms angereichert wurden (Waschtemperatur 57°C). Für drei weitere Proben wurde ein neues Extrakt erstellt (Fra 6, Theo 1, Fra 8). Die Sequenzierung der Proben erfolgte durch einen MiSeq Lauf 50 bp SE. Diese Proben wurden bereits durch Amplifikation der HVS-I Region und anschließender Sanger Sequenzierung typisiert (siehe Anhang Tabelle B.62).

Tabelle 4.35: Datenauswertung der Anreicherung von Extrakten aus Phase A (Auszug aus Anhang Tabelle B.67)

cov: mittlere Abdeckung der Zielregion (*coverage*), mD5: Deaminierungsrate am 5' Ende der Moleküle, Kont.: Kontaminationsangabe in Prozent, HG (CAP): Haplogruppe nach Capture, HG (PCR): Haplogruppe nach PCR

Probe	Extr	cov	mD5	Kont [%]	HG (CAP)	HG (PCR)
Klei 2	CPXI	73,4	0.294	0,1 - 1,7	J1c1	
Klei 2	CPI	24,0	0.297	4,0 - 10,9	J1c1	J1c
Klei 2	merge	97.44	0,295	1,19 - 3,24	J1c1	
Klei 1	CPI	11,8	0.281	1,1 - 7,2	K1a4a1	
Klei 1	CPX	93,4	0.351	0,1 - 1,2	K1a4a1	K
Klei 1	merge	105.16	0,342	0,39 - 1,55	K1a4a1	
Xir 1	CPVII	89,4	0.472	0,6 - 2,4	T2c1d1	
Xir 1	CPII	72,6	0.447	0,9 - 3,2	T2c1d1	H14b1
Xir 1	merge	161.97	0,461	1,1 - 2,5	T2c1d1	
Tha 4	CPIV	33,0	0.272	46,8 - 68,4	R	
Tha 4	CPVII	5,7	0.285	4,2 - 24,3	HV0f	T2k
Tha 4	merge	38.74	0,274	46,4 - 65,2	R1	
Fra 1	CPVI	31,5	0.217	38,3 - 56,2	U	
Fra 1	CPVIII	25,5	0.279	33,9 - 54,7	R0	X
Fra 1	merge	57.02	0,244	51,6 - 66,8	R1	
Xir 2	CPII	61,2	0.349	0,7 - 3,1	T1a5	
Xir 2	CPVII	51,9	0.350	3,2 - 7,0	T1a5	T1a
Xir 2	merge	113.08	0,350	2,2 - 4,5	T1a5	
Theo 5	CPX	25,6	0.350	7,2 - 13,4	K1c	
Theo 5	CPVIII	8,6	0.285	11,5 - 26,4	K1	T2k
Theo 5	merge	34.17	0,335	9,7 - 15,4	K1c	
Theo 4	CPVIII	11,4	0.245	44,3 - 77,8	R	
Theo 4	CPVI	6,9	0.220	15,1 - 37,9	U	T2k
Theo 4	merge	18.23	0,236	37,1 - 61,1	R1	
Fra 6	XXIV	6,1	0.228	14,9 - 42,4	R1	-
Theo 1	XXIV	21,7	0.322	1,7 - 6,8	K1c	K
Fra 8	XXIV	55,2	0.327	0,2 - 1,9	U3b3	H1aj1

Von zehn untersuchten Proben (Ausnahme Fra 6, da kein eindeutiges PCR Ergebnis) bestätigten vier Proben den Haplotyp ermittelt durch PCR (Klei 1/2, Xir 2, Theo 1). Diese Proben zeigten unter Einbeziehung beider Extraktionen (*merge*) ein maximales Kontaminationslevel von 4,5 %. Bei den Proben Xir 1, und Fra 8 ändern sich die analysierten Mutationen und daraus resultierenden

¹Anmerkung: Die Extrakte dieser Versuchsreihe waren zum Zeitpunkt der *Library* Erstellung zwischen 2-5 Jahren auf -20°C gelagert

Haplogruppen im Vergleich zu den PCR Ergebnissen, jedoch in Verbindung mit einem recht niedrigen Kontaminationslevel (max. 2,5%). Problematisch verhalten sich die Proben Tha 4, Fra 1, Theo 5 und Theo 4, bei welchen durch die NGS Strategie eine andere genetische Abstammungslinie in Kombination mit einem hohen Kontaminationslevel identifiziert werden kann. Auffällig ist, dass drei dieser Proben den Haplotyp T2k nach PCR zeigen (Theo 4, Theo 5, Tha 4).

Nachsequenzierung einzelner Proben

Kontaminierte Proben der Versuchsreihe (Tabelle 4.35) wurden auf einem HiSeq 100bp *paired end* Lauf nachsequenziert, um auf eine größere Datenmenge und die komplette Fragmentlänge der *Library* Moleküle zurückgreifen zu können. Ist das alte DNA Molekül in seiner Gesamtheit sequenziert, dann kann bei der Dekontaminationsmethode "schmutzi" mit der Option `-uselength` zusätzlich zu den Desaminierungsraten nach endogenen Fragmentlängen gefiltert werden.

Tabelle 4.36: Nachsequenzierung der kontaminierten Proben aus Phase A (Auszug aus Anhang Tabelle B.68)

cov: mittlere Abdeckung der Zielregion (*coverage*), mD5/mD3: Deaminierungsrate am 5'/3' Ende der Moleküle, Kont.: Kontaminationsangabe in Prozent, HG (CAP): Haplogruppe nach Capture

Probe	Extr	cov	mD5	mD3	Frag (BAM)	Kont [%]	HG (CAP)
Fra 1	CPVI	47.7	0,229	0,2608	107.90	43,5 - 57,4	R
Fra 1	CPVIII	48	0,268	0,3209	94.45	29.3 - 46.8	R
Fra 1	merge	93,57	0.242	0.275	106.35	53,9 - 68,7	R
Tha 4	CPIV	69,6	0,269	0,3115	99.91	65.5 - 86.1	R
Tha 4	CPVII	11	0,288	0,3202	90.80	5.4 - 23.9	HV0f
Tha 4	merge	74.96	0.270	0.311	100.17	59,4 - 78,0	R
Theo 4	CPVIII	13,9	0,224	0,2591	97.64	34.4 - 62.8	K1c
Theo 4	CPVI	22,9	0,242	0,2881	101.09	44.9 - 73.0	R12'21
Theo 4	merge	32.51	0.228	0.268	100.24	43,9 - 65,3	R1
Theo 5	CPX	15	0,279	0,3114	75.08	5.9 - 18.3	K1c
Theo 5	CPVIII	40,7	0,346	0,3774	85.88	5.1 - 10.8	K2
Theo 5	merge	48,74	0,332	0,365	77.79	0,7 - 8,7	K1c
Fra 6	XXIV	12	0,226	0,2576	104.68	21.9 - 50.3	R1

Positive Auswirkungen auf das Resultat hatte dies jedoch nur bei der Probe Theo 5, bei welcher das Kontaminationslevel gesenkt werden konnte (Tabelle 4.36). Die Proben Fra 1, Tha 4, Theo 4 und Fra 6 zeigen erneut sehr hohe Level an Kontamination. Zusätzlich weisen diese Proben recht lange durchschnittliche Fragmentlängen innerhalb der alignierten Sequenzen auf, was auf Kontaminationen durch moderne DNA zurückgeführt werden kann. Außerdem zeigen diese Proben durch die Mischung von DNA Profilen aufgrund humaner, exogener DNA zum Großteil HG R (gemeinsamer interner Knotenpunkt innerhalb der mitochondrialen Phylogenie). Die Proben Theo 5 und Fra 1 wurden in einem späteren Versuch durch eine dritte Extraktion typisiert (siehe Ergebnisteil mitochondriale Genome aus dem Mesolithikum, Abschnitt 4.6).

Dekontamination einzelner Proben

Um das Kontaminationslevel in den Proben Fra 1, Fra 6, Tha 4, Theo 4 und Theo 5 zu senken wurden zwei verschiedene Programme im Vergleich angewendet; "schmutzi" [154] und "PMDtools" in Kombination mit ContamMix [176, 47] (siehe Abschnitt 3.8). Daten aus Läufen von MiSeq bzw. HiSeq müssen getrennt dekontaminiert werden, da diese Programme die Fehlerrate der Sequenzierer und auch die Leseweite berücksichtigen. Die Dekontamination durch das Programm "schmutzi" ist ein iterativer Prozess, bei welchem nach jeder Runde ein endogenes Mitochondrium assembliert wird

(Seq endo/ endo hg) und getrennt von den konatminierenden Sequenzen einer Probe (Seq cont/cont hg) ausgewertet wird. Die Probe Klei 1 diente während des Dekontaminationsprozesses des **50bpSE MiSeq Laufes** als Positivkontrolle zur Beobachtung des Verhaltens einer Probe, da Klei 1 kein auffälliges Kontaminationslevel zeigt (siehe Tabelle 4.35). Die Positivkontrolle verhält sich unauffällig. Ihr Kontaminationslevel zu Beginn (Start Kont Est in %) stimmt in etwa mit der Bestimmung durch ContamMix überein (Tabelle 4.35, 0,39 - 1,55 % und Tabelle 4.37 0-2%) und die ermittelte Haplogruppe ist beständig¹. Für die Probe Theo 5 ergibt sich ebenfalls die Haplogruppe K1c, begleitet durch ein Kontaminationslevel von 9 (8-10) %. An dieser Stelle muss die Haplogruppe der kontaminierenden Sequenzen hervorgehoben werden (T2), welche mit dem PCR Ergebnis übereinstimmt. Für die Probe Theo 4 kann mit einem akzeptablen Kontaminations Level (5-7%) die Haplogruppe U8b ermittelt werden, welche aus phylogenetischer Sicht ein Vorläufer der Haplogruppe K1 darstellt. Die Auswertung der Probe Fra 6 musste nach 100 Iterationen abgebrochen werden, ohne dass ein adäquates Mitochondrium mit geringem Kontaminationslevel gefunden werden konnte. Die Ergebnisse der Dekontamination durch das Programm "schmutzi" sind in Tabelle 4.37 dargestellt.

Das Programm *PMDtools* sortiert einen Datensatz nach desaminierten Sequenzen und folgende Analysen wie die Genotypisierung es gefilterten Datensatzes, die Bestimmung der Desaminierungsrate, des Kontaminationslevels sowie der Haplogruppe werden per Hand durchgeführt. Anhand der Stringenz des Filterprozesses können Schwellenwerte von t1 - t5 angenommen werden (siehe Tabelle 4.38, Schwellenwert t1 und t3). Die Positivkontrolle Klei 1 verhält sich wiederholt unauffällig und es können an ihrem Ergebnis die Effekte der gewählten Stringenz verdeutlicht werden. Auf Grundlage von 34567 alignierten Sequenzdaten (Tabelle B.67, ALNMkdup) wurden durch die Anwendung von *PMDtools* (t1) etwa 75% der Sequenzen entfernt und gleichzeitig die Desaminierungsrate des Datensatzes auf 0,798 (mD5) angehoben (Tabelle 4.38a). Während die Haplogruppe gleich geblieben ist, konnte das Kontaminationslevel auf ein Maximum von 1,76% gesenkt werden. Ähnliche Effekte zeigte die Probe Theo 5 - die Haplogruppe K1c blieb konstant während das Kontaminationslevel gesenkt werden konnte. Bei den restlichen Proben konnte auch durch *PMDtools* kein authentisches Ergebnis erzielt werden².

Für den Dekontaminationsprozess der Daten generiert durch einen **HiSeq 100bp PE Lauf** wird nur mit Sequenzpaaren weitergearbeitet (*joined reads*), so dass die komplette Sequenzierung des alten DNA Moleküls gewährleistet ist. Unter diesen Voraussetzungen kann während der Dekontamination durch "schmutzi" die Fragmentlänge der Moleküle mit einbezogen werden (*-uslength*). Auch die Dekontamination auf Basis einer größeren Datenmenge brachte keine eindeutigen Resultate für die Proben Fra 1/6, Tha 4 und Theo 4. Während der getrennten Analyse beider Extraktionen zeigten sich jeweils unterschiedliche Haplogruppen. Einzig bei der Probe Theo 5 kann ein Erfolg erzielt werden. Auch hier kann durch "schmutzi" wieder das Ergebnis der PCR und Sangersequenzierung als exogene Kontamination, die Haplogruppe T2, identifiziert werden. Innerhalb der Probe Theo 4 zeigen die Resultate der Dekontamination verstärkt das Ergebnis U*/U8 und innerhalb der Probe Fra 6 H*. Für weitere Analysen ist dieses Ergebnis jedoch nicht authentisch, so dass diese Proben aus den Folgeanalysen ausgeschlossen werden.

¹Anmerkung: Die Bestimmung der Haplogruppe ist Teil des Programms "schmutzi".

²Für die Verwendung von *PMDtools* unter einer höheren Stringenz (t3) reicht die Datengrundlage nicht aus. Dies zeigt sich bereits bei der Positivkontrolle, bei welcher die gefilterten, endogenen Sequenzen auf 2498 gesenkt werden und ein hohes Kontaminationslevel errechnet wird. Dies ist auf die geringe *coverage* (7,54) und hohen Desaminierungsrate (0,905) zurückzuführen. Die ermittelte Haplogruppe bleibt gleich

Tabelle 4.37: Dekontamination durch "schmutzi" (MiSeq 50bp SE). Kontaminationsbestimmung erfolgt durch das Programm .

Start KontEst: Kontaminationsbestimmung zu Beginn, Seq endo: Anzahl endogene Sequenzen ermittelt durch "schmutzi", endo hg: endogene Haplogruppe, Seq cont: Anzahl kontaminierender Sequenzen, cont hg: kontaminierende Haplogruppe, Final KontEst: Finale Kontaminationsbestimmung

Probe	Start KontEst (min/max)	Seq endo	endo hg	Seq cont	cont hg	Final KontEst (min/max)
Klei 1 endo	2 (0,5-3,5)	19843	K1a4a1	1236	K1a4a1a	1 (0-2)
Fra 1 endo	28 (26-30)	11906	U	979	K1a4a1c1	10 (9-11)
Tha 4 endo	30 (28-32)	10509	JT	858	K1a4c	11 (10-12)
Theo 4 endo	13,5 (5,5-21,5)	5392	U8b	455	M5a4	6 (5-7)
Theo 5 endo	0 (0-1)	10078	K1c	770	T2c1	9(8-10)
Fra 6 endo	86 (84-88)	267	H1b1g	1547	HV0f	99 (98)

Tabelle 4.38: Anwendung von *PMDtools* und resultierende Kontaminationsbestimmung anhand ContamMix (MiSeq 50bp SE).

Seq: alignierte Sequenzen nach Anwendung von *PMDtools*, cov: Abdeckung rCRS, mD5: Desaminierungsrate am 5'-Ende, Kont: durch ContamMix ermittelte Kontaminationsrate, HG: Haplogruppe

(a) *PMDtools* (t1).

Probe	Seq	cov	mD5	Kont [%]	HG
Klei 1	8593	26.10	0,798	0,01-1,76	K1a4a1
Fra 1	3813	11.57	0,753	12,42-38,27	R
Tha 4	3001	8.96	0,769	12,70-39,70	R
Theo 4	1209	3.62	0,744	9,01-25,68	U
Theo 5	2959	8.63	0,787	0,35-8,27	K1c
Fra 6	400	1.20	0,763	0,14-8,44	H2a2

(b) *PMDtools* (t3).

Probe	Seq	cov	mD5	Kont [%]	HG
Klei 1	2498	7.54	0,905	0,19 - 8,87	K1a4a1
Fra 1	1345	4.07	0,887	4,94 - 23,73	U
Tha 4	1181	3.46	0,903	7,61 - 30,68	HV9c
Theo 4	600	1.79	0,899	5,25 - 19,17	R11'B6
Theo 5	1248	3.58	0,877	4,87 - 19,71	K1
Fra 6	199	0.60	0,925	0,28 - 7,94	H27

4.6 Datenanalyse der mitochondrialen Genome

Die mitochondriale Datengenerierung war in Phase B in die Prozessoptimierungen eingegliedert, so dass von einer Probe meist mehrere Datensätze zur Verfügung stehen (vgl. Tabelle 3.14). Zudem sind Proben mit vermindertem Erfolg oder Kontaminationen erneut in Phase C/D bearbeitet worden. Für eine zusammenfassende Darstellung der mitochondrialen Genomsequenzen werden im Folgenden das Endergebnis je Probe in ihrem archäologisch-kulturellen Zeitabschnitt dargestellt. Ausnahmen stellen Proben dar, bei welchen unterschiedliche Skelettelemente bearbeitet worden sind; es werden beide Bearbeitungsschritte aufgezeigt. Während dieses Projektes wurden 74 Proben bzw. Skelettelemente durch ein mitochondriales *Capture*-Verfahren bearbeitet.

Authentizitätskriterien des Datensatzes

Der mässige biomolekulare Erhaltungszustand alter DNA Moleküle aus Knochenfunden erfordert die Einhaltung strikter Kriterien für die Gewährleistung der Authentizität eines Datensatzes. Die Anwendung von Hochdurchsatzsequenzierverfahren in der Paläogenetik eröffnen durch die parallele Sequenzierung multipler Loci ein neues Spektrum an Authentifizierungskriterien, wobei das alte DNA Molekül annähernd in seiner Originalform dargestellt wird². Hingegen muss bei Anwendung von PCR und Sanger Sequenzierung die Fragmentlänge des genutzten Primersystems (Amplikonlänge) vorhanden sein, um ein positives Ergebnis zu erzielen. Speziell für stark fragmentierte Proben dieses Projektes konnte eine Diskrepanz zwischen dem erstellten Datensatz durch PCR/Sanger Sequenzierung und NGS nach Anreicherung des mitochondrialen Genoms gezeigt werden (siehe Tabelle 4.35). Während PCR Methoden keine weiterführende Analyse eines Sequenzdatensatzes ermöglichen, können innerhalb einer Probe eines NGS Datensatzes 3 weitere Kriterien zur Darstellung der Authentizität herangezogen werden:

1. Auf labortechnischer Ebene stellt die Arbeit unter Reinstraumbedingungen und die Analyse der mitgeführten Leerkontrollen das erste Kriterium für die Arbeit mit alter DNA dar. Die Müller-, Extraktions- und *Library*kontrollen wurden durch eine quantitative PCR hinsichtlich ihrer Molekülzahl untersucht worden. Diese befanden sich in einem Rahmen von 10^3 - 10^5 Molekülen je μl *Fill In* Produkt und zeigen damit einen Abstand von 10^2 - 10^5 Molekülen zu den einzigartigen Molekülzahlen innerhalb des Probenspektrums (siehe Abbildung 4.23 und Tabelle B.3). Zu jeder erstellten *Library* wurde ein theoretisches Kontaminationslevel in Abhängigkeit der Leerkontrolle des *Library* Ansatzes errechnet, um die Hintergrundkontamination einer Probe darzustellen. Die Abhängigkeit der Kontaminationsrate einer Probe von der Komplexität einer *Library* ist in Abschnitt 4.5 am Beispiel der Probe Krk 6 dargestellt. Proben, welche nicht über einen Abstand von $>10^2$ zu den Leerkontrollen verfügen, sind aus dem Versuchsablauf ausgeschlossen worden.
2. Desaminierungsraten der Moleküle während des Proben *Screenings* und nach *Capture* bzw. Shotgun-Sequenzierung.
3. Fragmentlängenverteilung innerhalb einer Probe (*Bioanalyzer* Messung, bei Sequenzierweite >100 extrahiert aus Alignment).

²lediglich während der Erstellung der glatten Enden während der *Library* Erstellung werden einzelsträngig überhängende 5'-Enden auf die Position des 3'-Endes zurückgeschnitten

4. Die Kalkulation der Kontaminationsrate, ermittelt durch ContamMix bzw. "schmutzi", anhand mitochondrialer Sequenzendaten einer Probe ist das aussagekräftigste Kriterium. In diesem Projekt werden Proben mit einem Authentizitätswert >95% als authentisch beachtet.
5. Mitochondriale *Capture* Produkte wurden nach desaminierten Molekülen gefiltert (*PMD tools t1*) und erneut Genotypisiert. Bei Übereinstimmung der Haplogruppe vor und nach der Anwendung des Filters spricht dies für die Authentizität einer Probe.

Mitochondriale Genome aus dem Mesolithikum

Die mesolithischen Proben boten zum größten Teil schlechte Bedingungen zur Anreicherung. Niedrige endogene Molekülzahlen versprachen keinen großen Erfolg (<2*10⁵ endogene Moleküle pro µl Fill In Produkt, siehe Tabelle 4.20), weshalb mehrere Extraktionen einer Probe bearbeitet wurden. Von acht analysierten Proben konnte aus zwei Proben ein authentisches Ergebnis erzielt werden (Erfolgsrate 25%, siehe Tabelle 4.39).

Tabelle 4.39: Auswertung des mitochondrialen Datensatzes aus dem Mesolithikum (nach Anreicherung).

ALN Mkdup: alignierte Sequenzen ohne Duplikate, cov: Abdeckung rCRS, mD5: Desaminierungsrate am 5'-Ende, mD3: Desaminierungsrate am 3'-Ende, Kont: Kontaminationsrate (ContamMix), auth. Daten: Authentizitätswert (ContamMix), FragL: durchschnittliche Fragmentlänge der alignierten Moleküle, HG/score: Haplogruppe und Qualitätswert

Probe	Skelett-element	ALN Mkdup	cov	mD5	mD3	Kont [%]	auth. Daten [%]	FragL	HG	score
Theo 1	Knochen	7359	21.7	0.322	-	1,84 - 6,71	96,4	-	K1c	0,4
Theo 2	Knochen	<i>unzureichende Datengrundlage zur Auswertung</i>								
Theo 3	Knochen	<i>siehe Anhang B.71 - kein authentisches Ergebnis darstellbar</i>								
Theo 4	Knochen	<i>siehe Anhang B.71 - kein authentisches Ergebnis darstellbar</i>								
Theo 5	Knochen	12600	62.74	0.344	0.365	0,05 - 3,8	99,3	81.99	K1c	0,4
Fra 1	Knochen	<i>siehe Anhang B.71 - kein authentisches Ergebnis darstellbar</i>								
Fra 2	Knochen	<i>unzureichende Datengrundlage zur Auswertung</i>								
Fra 3	Knochen	<i>siehe Anhang B.71 - kein authentisches Ergebnis darstellbar</i>								
Fra 9	Knochen	<i>unzureichende Datengrundlage zur Auswertung</i>								

Von 2 mesolithischen Proben des Fundplatzes Theopetra konnten authentische Mitochondrien gewonnen werden (Theo 1, Theo 5). Das komplette Mitochondrium von Theo 1 bestätigt nach Anreicherung einer dritten Extraktion die PCR/Sanger Ergebnisse der Extraktionen aus Phase A/B. Bei Theo 5 hingegen wurde nach Anreicherung des mitochondrialen Genoms eine anderen mitochondrialen Linie als nach PCR und Sanger Sequenzierung gefunden (siehe Tabelle B.67). Da die Kontaminationsrate erhöht war, wurde eine dritte Extraktion dieser Probe analysiert, welche den Haplotyp nach *Capture* kombiniert mit einem niedrigen Kontaminationslevels (0,4 - 9,0) bestätigte (siehe Tabelle 4.40, Theo 5 XXIX). Für die Auswertung wurden die Extraktionen XXIX und CPVIII (t1) genutzt, welche zusammengeführt in einer Analyse ein Kontaminationslevel von maximal 3,8% zeigten.

Die Probe Theo 2 zeigte nach der Datenanalyse lediglich 300 mitochondriale Sequenzen, so dass diese Probe für weitere Analysen ausgeschlossen wurde. Die Probe Theo 3 ist im Anhang dargestellt. Es wurden 3 Extraktionen, versehen mit unterschiedlichen P7-Indexen, in einem Ansatz angereichert und anschließend getrennt ausgewertet. Die drei Extraktionen zeigen unterschiedliche genetische Linien sowie sehr hohe Kontaminationslevel (siehe Anhang Tabelle B.71). Dekontaminationsprozesse zeigten kein eindeutiges Resultat, so dass keine weitere Auswertung erfolgte (siehe Anhang Tabelle B.72). Die Probe Theo 4 steht aufgrund der stratigrafischen Unsicherheit der Fundstelle Theopetra unter

Tabelle 4.40: Auswertung der Sequenzdaten der Probe Theo 5.

ALN Mkdup: alignierte Sequenzen ohne Duplikate, cov: Abdeckung rCRS, mD5: Desaminierungsrate am 5'-Ende, mD3: Desaminierungsrate am 3'-Ende, Kont: Kontaminationsrate (ContamMix), auth. Daten: Authentizitätswert (ContamMix), FragL: durchschnittliche Fragmentlänge der alignierten Moleküle, HG/score: Haplogruppe und Qualitätswert

Probe	ALN mt	cov	mD5	mD3	Kont [%]	FragL	HG	score
Theo 5 XXIX	5738	29,99	0.352	0.365	0,4 - 9,0	85,87	K1c	0,4
Theo 5 CPVIII t1	8973	40,7	0.346	0.377	5,1 - 10,8	85,88	K1c	0,4
Theo 5 merge	12600	62,74	0.344	0.365	0,05 - 3,8	81,99	K1c	0,4
Theo 5 merge t1	4717	23,26	0.678	0.719	0,08 - 3,4	81,25	K1c	0,4

Verdacht ein weiteres Skelettelement der Probe Theo 5 zu sein. Beide Proben zeigten nach PCR die Linie T2k, welche sich jedoch als fehlerhaft im Falle von Theo 5 herausstellte. Nach Anreicherung zeigt die Probe Theo 4 ein hohes Kontaminationslevel, welches in folgenden Dekontaminationsanalysen die genetische Linie U*/U8 zeigt. Diese steht phylogenetisch eine Ebene über K1, jedoch kann das Kontaminationslevel nicht gesenkt werden. Diese Probe kann nicht für Folgeanalysen verwendet werden und auch die Verwandtschaft zur Probe Theo 5 kann nicht bestätigt bzw. ausgeschlossen werden.

Die Analyse der *Capture* Produkte der mesolithischer Proben aus Franchthi zeigten keine authentischen Ergebnisse (siehe Abschnitt 4.5 für Probe Fra 1). Auch eine dritte Extraktion der Probe Fra 1, welche in zwei unterschiedlichen *Capture* Experimenten angereichert wurde, zeigte hohe Kontaminationslevel. In dieser Extraktion wurde die Haplogruppe X2 festgestellt, welche mit den PCR Ergebnissen übereinstimmt. Kontaminationsraten zeigen, dass in diesem Fall lediglich eine Kontamination sequenziert worden ist (siehe Anhang Tabelle B.71). Die Proben Fra 2 und Fra 9 zeigten nur wenige Sequenzdaten nach Anreicherung des mitochondrialen Genoms und konnten nicht weiter ausgewertet werden. Die Probe Fra 3 wurde hinsichtlich 4 separater Extraktionen angereichert, jedoch ohne Übereinstimmungen der mitochondrialen Abstammungslinie (H5, T, R, HV; siehe Anhang Tabelle B.71) und auch die Dekontamination durch "schmutzi" lieferte keine eindeutigen Ergebnisse (siehe Anhang Tabelle B.72). Bei den Proben des Fundplatzes Franchthi wird besonders deutlich, dass mögliche Kontaminationen desaminiert vorliegen und Dekontaminationsprozeduren nicht erfolgreich durchgeführt werden können. Bei Proben dieses Fundplatzes ist anzumerken, dass die Grabung schon vor mehr als 40 Jahren stattgefunden hat und die Proben im Anschluss nicht sachgerecht, in Bezug auf paläogenetische Analysen, gelagert worden sind.

Die Proben aus Maroulas sind nicht durch ein mitochondriales *Capture* angereichert worden, da diese in der Quantifizierung eine zu geringe Anzahl an einzigartigen Molekülen in Kombination mit sehr geringen endogenen Werten zeigten.

Mitochondriale Genome aus dem frühen Neolithikum

Unter den **frühneolithischen Proben** konnte im Vergleich zu dem Probenmaterial aus dem Mesolithikum eine verbesserte Probenqualität festgestellt werden - aus 22 Skelettelementen konnte aus 16 erfolgreich ein authentisches, mitochondriales Genom erstellt werden (siehe Tabelle 4.41). Die verbleibenden 6 Proben zeigten eine Authentizität der Daten <95%, wobei die Probe Mau 1 erfolgreich durch Anwendung der Dekontaminationsverfahren bereinigt werden konnte (siehe Tabelle 4.42 & 4.43). Das Zahnmaterial des Fundplatzes Revenia enthält durchgehend nicht genug DNA um Mitochondrien zu erstellen, so dass auf die Felsenbeine zurückgegriffen werden musste. Die Probe Rev 7 ist in ihren Einzelextraktionen im Anhang dargestellt (Tabelle B.71 & B.72). Diese Probe stellt eins von zwei Felsenbeinen dieser Arbeit dar, deren DNA Qualität als mäßig einzustufen ist und zwar in einem Maße,

dass kein Mitochondrium isoliert werden konnten.

Bei der Probe Xir 1 wurden erfolgreich zwei Genome aus unterschiedlichen Skelettelementen angereichert - für weiterführende Analysen werden diese Daten zusammengefasst. Die Probe Xir 4 wurde ebenfalls aus jeweils Zahn und Knochen angereichert, jedoch lieferte das Knochenelement keine verwertbaren Sequenzdaten.

Tabelle 4.41: Auswertung des früh-neolithischen Datensatzes (nach Anreicherung).

ALN Mkdup: alignierte Sequenzen ohne Duplikate, cov: Abdeckung rCRS, mD5: Desaminierungsrate am 5'-Ende, mD3: Desaminierungsrate am 3'-Ende, Kont: Kontaminationsrate (ContamMix), auth. Daten: Authentizitätswert (ContamMix), FragL: durchschnittliche Fragmentlänge der alignierten Moleküle, HG/score: Haplogruppe und Qualitätswert

Probe	Skelett-element	ALN Mkdup	cov	mD5	mD3	Kont [%]	auth. Daten [%]	Frag BAM [bp]	HG	score
Mau 1	Zahn	8422	47	0.311	0.328	9.4 - 19.7	86,6	94.33	H5	1
Mau 2	Zahn	14390	73	0.373	0.402	2.7 - 5.8	96	83.83	T2b23	1
Mau 4	Knochen	12891	54	0.438	0.476	0,6 - 3,9	98,1	70.22	H46	0,6
Mau 5	Knochen	9790	29,2	0.454	-	1,7 - 7,2	96,5	-	T2f	0,8
Xir 1	Zahn	31237	94.30	0.454	-	0.8 - 2.5	98,5	-	T2c1d1	1
Xir 1	Knochen	17367	47.58	0.416	-	0.5 - 3.5	98,3	-	T2c1d1	1
Xir 2	Zahn	20249	61.2	0.349	-	0,7 - 3,1	98,4	-	T1a5	1
Xir 4A	Zahn	15768	69.41	0.428	0.489	0,01 - 0,81	99,9	72.73	K1a1	1
Xir 4B	Knochen	2342	12.3	0.197	0.198	7.9 - 28.3	91,6	86.37	K1a24	1
Xir 5	Knochen	6693	22.45	0.463	0.536	0,02 - 2,7	99,9	55.86	K1a24	1
Rev 1	Zahn	2517	10.84	0.390	0.440	31.5 - 79.7	52,7	66.88	U	0,4
Rev 2	Zahn	1206	4.1	0.453	0.496	14.4 - 49.9	74,0	58.59	H2a5	1,0
Rev 3	Zahn	2283	9.3	0.412	0.450	19.3 - 54.4	66,8	67.59	R2'JT	1,0
Rev 4	Felsenbein	27051	122.03	0.502	0.525	0,4 - 2,9	98,8	74.55	H	1
Rev 5	Felsenbein	28654	85.81	0.550	-	0,3 - 2,4	99,01	-	X2b	1
Rev 6	Felsenbein	22775	93.41	0.561	0.588	0,1 - 3,9	99,3	67.77	H46	0,6
Rev 7	Felsenbein	5222	21.97	0.328	0.343	4,00 - 20,4	86,6	73.7	R9	0,4
Nea 1	Felsenbein	19111	85.57	0.441	0.459	0,00 - 0,53	99,9	73.90	K1a4b	1
Nea 2	Felsenbein	32214	206.27	0.328	0.339	0,51 - 1,52	99,05	105.92	K1a	1
Nea 3	Felsenbein	31822	206.13	0.322	0.322	0,22 - 1,03	99,5	107.03	K1a2	1
Nea 4	Felsenbein	13045	52.85	0.454	0.494	1,7 - 6,8	96,1	66.84	H	1
Nea 5	Felsenbein	32798	229.42	0.319	0.316	1,19 - 2,85	98,1	115.72	H	1

Tabelle 4.42: Dekontamination der Probe Mau 1 durch *PMDtools* ([176], *PMD t1*)

Probe	ALN mt	cov	mD5	mD3	Kont [%]	auth Daten [%]	FragL	HG	haplofind score
Mau 1	8422	47	0.311	0.328	9,7 - 19,7	86,6	92.13	H5	1.00
Mau 1 t1	3493	19.96	0.643	0.650	0,08 - 5,0	99,1	94.33	H5	1.00

Tabelle 4.43: Dekontamination der Probe Mau 1 durch "schmutzi" [154]

erstelt: Daten zu Beginn der Analyse, finaleIT: Daten nach finaler Iteration, ALN mt: alignierte Sequenzen an rCRS, cov: Abdeckung, contEst: Kontaminationsbestimmung durch "schmutzi", HG/score: Haplogruppe und Qualität

Probe	ALN mt	cov	contEst	min	max	HG	score
Mau 1 endo	7748	43.58	0.1	0.08	0.12	H5	1
Mau 1 cont	524	3.06				T2c1a	0,8
finale IT	ALN mt	cov	contEst	min	max	HG	score
Mau 1 endo	7794	43.65	0.07	0.06	0.08	H5	1
Mau 1 cont	538	3.16				T2c1a	0,8

Mitochondriale Genome aus dem fortschreitenden Neolithikum

Aus der Epoche des **mittleren bis späten Neolithikums** wurden 41 Skelettelemente bearbeitet wovon aus 38 erfolgreich ein mitochondriales Genom sequenziert werden konnte. Es wurden drei Individuen doppelt beprobt (Krk 8 (Zahn) = Krk 11 (Felsenbein); Klei 1 (Zahn/Knochen); Klei 3 (Knochen/Felsenbein)), so dass 35 Mitochondrien für Folgeanalysen zur Verfügung stehen.

Tabelle 4.44: Auswertung des neolithischen Datensatzes (nach Anreicherung).

ALN Mkdup: alignierte Sequenzen ohne Duplikate, cov: Abdeckung rCRS, mD5: Desaminierungsrate am 5'-Ende, mD3: Desaminierungsrate am 3'-Ende, Kont: Kontaminationsrate (ContamMix), auth. Daten: Authentizitätswert (ContamMix), FragL: durchschnittliche Fragmentlänge der alignierten Moleküle, HG/score: Haplogruppe und Qualitätswert

Probe	Skelettelement	ALN Mkdup	cov	mD5	mD3	Kont [%]	auth. Daten [%]	FragL	HG	score
NS1	Zahn	24007	72.16	0.356	-	0,4 - 2,7	98,6	-	H1	1
NS2	Zahn	21270	63.82	0.366	-	1,5 - 4,7	97,3	-	J1c7	0,4
St 2B	Femur	23250	115	0.405	0.424	1.2 - 3.1	97,9	81.61	N1a1a1a3	1
St3A	Zahn	5449	19.42	0.386	0.459	0,31 - 4,44	98,6	58.78	K1a	1
St 4	Felsenbein	749	4.20	0.294	0.323	7,2 - 26,7		92.59	-	-
St5	Felsenbein	31520	162.25	0.422	0.442	0,86 - 2,7	98,4	84.80	H	1
St 6	Felsenbein	31053	136.87	0.494	0.536	4,0 - 6,6	94,7	72.85	N1a1a1a3	1
Krk 1	Femur	33062	300.05	0.265	0.313	0,0 - 0,4	99,9	148.86	K1a1	1
Krk 2	Femur	11215	33.4	0.420	-	2,7 - 11,9	94,4	-	H	1
Krk 4	Femur	19498	58.9	0.376	-	0,1 - 1,7	99,4	-	U8b1a1	1
Krk 5	Femur	6876	20.8	0.191	-	25 - 48	65,6	-	-	-
Krk 6	Femur	2725	17.01	0.208	0.286	4.5 - 20.5	89,9	103.83	R	1
Krk 7	Knochen	16084	80.32	0.406	0.458	0,1 - 1,5	99,5	82.53	K1a3a3	1
Krk 8	Zahn	31417	250.99	0.323	0.211	0,0 - 0,4	99,9	129.97	N1a1a1a	1
Krk 9	Zahn	11012	49	0.392	0.443	0,0 - 2,0	99,7	73.89	H	1
Krk 10	Felsenbein	28202	154.02	0.310	0.325	1,1 - 2,4	98,4	89.72	K1a4a1	1
Krk 11	Felsenbein	32572	214.58	0.356	0.371	0,2 - 1,0	99,5	108.78	N1a1a1a	1
Mkg 9	Felsenbein	10040	36.87	0.531	0.558	0,9 - 4,5	97,9	60.47	T2b	1
Mkg 10	Felsenbein	9506	31.37	0.557	0.579	0,0 - 2,3	99,9	54.57	J1c3j	0,5
Mkg 11	Felsenbein	22406	89.30	0.499	0.533	0,0 - 0,1	99,7	65.85	K1a	1
Mkg 12	Felsenbein	8923	32.67	0.533	0.569	0,3 - 5,1	98,5	60.50	U7b	1
Mkg 13	Felsenbein	12336	43.73	0.477	0.498	0,1 - 1,9	99,4	59.44	K1a	1
Pal 1	Zahn	31624	243	0.341	0.275	0,1 - 0,6	99,8	127.06	U4b3	1
Pal 2	Zahn	15978	53.76	0.374	-	1,8 - 5,0	96,9	-	K1a7	1
Pal 4	Knochen	15488	46.79	0,281	-	3,5 - 6,5	95,2	-	U2e2a	1
Pal 5	Knochen	14051	41.89	0.339	-	0,3 - 2,3	99,01	-	U2e2a	1
Pal 7	Felsenbein	32809	100.07	0.430	-	1,2 - 2,8	98,1	-	J1c1	1
Pal6	Zahn	23853	135.81	0.418	0.460	0,0 - 0,5	99,9	93.86	K1a7	1
Klei 1	Knochen	27065	81.6	0,388	-	0,0 - 0,9	99,8	-	K1a4a1	1
Klei 1	Zahn	30759	93.66	0.338	-	0,3 - 1,5	99,8	-	K1a4a1	1
Klei 2	3M	24138	73.4	0.294	-	0,1 - 1,7	98,6	-	J1c1	1
Klei 3	Felsenbein	32976	100.72	0.334	-	0,0 - 0,7	99,9	-	K1a4	1
Klei 3	Knochen	10216	53.68	0.340	0.367	3,4 - 9,1	99,8	87,1	K1a4	1
Klei 5	Knochen	25675	161.85	0.327	0.292	4,4 - 8,0	94,1	103.85	HV	1
Klei 6	Knochen	20523	100.73	0.405	0.443	0,0 - 0,7	99,9	80.89	H29	0,8
Klei 7	Felsenbein	18491	55.32	0,373	-	1,1 - 3,7	97,9	-	H29	0,8
Klei 8	Felsenbein	6359	34.24	0.396	0.443	0,2 - 3,3	99,2	88.83	J1c2	1
Klei 10	Felsenbein	26146	128.22	0.340	0.369	1,7 - 3,4	97,6	80.97	K1a2	1
Fra 6	Knochen	1898	12	0.226	0.257	21,9 - 50,3	73,1		R1	0,6
Fra8	Knochen	22317	88.97	0.335	-	0,3 - 1,6	99,3	-	U3b3	1
Tha 2	Knochen	32935	293.88	0.209	0.246	0,0 - 0,5	99,8	147.71	H5	1

Die Proben St 4, Krk 6 und Fra 6 zeigten zu wenig angereicherte mitochondriale Fragmente um einen Dekontaminationsprozess durchzuführen. Bei der Probe Krk 5 konnte eine Extraktion bereinigt

werden (nach Dekontamination 2%, siehe Anhang Tabelle B.72), jedoch zeigt diese Probe eine recht niedrige Desaminierungsrate von 19% - eher untypisch für den Fundplatz Kremasti koilada, welcher im Schnitt eine Desaminierungsfrequenz von >30% zeigt. Diese Probe wird in weiteren Analysen ausgeschlossen. Die Probe St 6 zeigt eine Kontaminationslevel von 4,0 - 6,6%; bei Auswertung der Einzelanalysen wird eine leichte Kontamination in der zweiten Extraktion deutlich, so dass diese von der weiteren Analyse ausgeschlossen wird. Die Probe Tha 4 wurde zusätzlich zu C12 (Tabelle 4.35) durch eine dritte Extraktion angereichert, jedoch konnte kein authentisches Ergebnis erstellt werden (siehe Anhang, Tabelle B.71 & B.72)

Populationsgenetische Analyse der Sequenzdaten

6.1 Der prähistorische Datensatz aus Griechenland

Für eine populationsgenetische Analyse stehen 53 mitochondriale Genome zur Verfügung, welche mit 4000 Jahren den Zeitraum der Transition vom Mesolithikum zum Neolithikum in Griechenland abdecken (siehe Abbildung 6.32). Innerhalb dieses Projektes wurden drei Genome durch *Shotgun* Sequenzierung erstellt, deren Auswertung der Publikation Hofmanova, Kreuzer *et al.* (2016) entnommen werden kann (siehe Anhang B.7). Es wurden genetische Datensätze einer früh neolithische Probe (Revenia, 6350 v. Chr.) und zweier aus dem finalen Neolithikum (Paliambela, Kleitos, etwa 4000 v. Chr.) erstellt, welche mit einer Zeitspanne von 2500 Jahren in das griechische Neolithikum fallen. Diese Daten werden in der Diskussion dieser Arbeit in Abschnitt 7.2-7.7 aufgegriffen.

Die neolithischen Proben des mitochondrialen Datensatzes werden chronologisch in zwei Epochen unterteilt (früh & spät), um die genetische Zusammensetzung der einzelnen Epochen und mögliche Populationsumschwünge zu erkennen. Für diese Arbeit konnte ein umfangreicher Referenzdatensatz prähistorischer, mitochondrialer HVS-I Sequenzen aus Europa zusammengestellt werden, in dessen Rahmen die Position des griechischen Neolithikums während der Expansionswelle nach Europa untersucht werden soll.

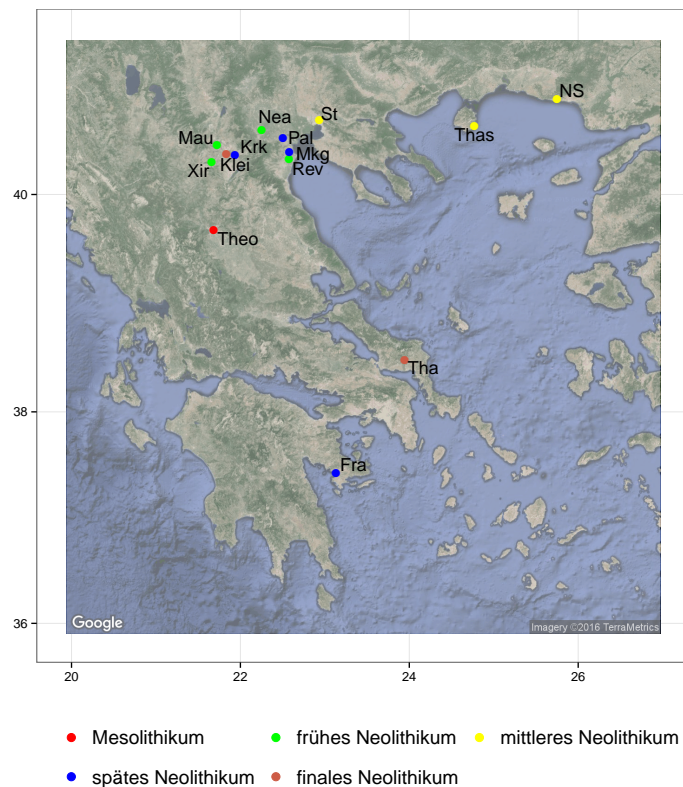


Abbildung 6.32: Der prähistorische Datensatz Griechenlandes abgebildet nach Fundplätzen. Insgesamt wurden 53 mitochondriale Genome aus der Zeitspanne des Mesolithikums bis zum Ende des Neolithikums erstellt. Die Farbgebung entspricht der chronologischen Einordnung der Fundplätze. Von den Fundplätzen Revenia (Rev), Paliambela (Pal) und Kleitos (Klei) wurden Datensätze mit einer Abdeckung von 1-2x der humanen Referenzsequenz erstellt. Abkürzungen siehe Tabelle 6.46.

Aufgrund des schlechten Erhaltungszustands des mesolithischen Probenmaterials (Theopetra, Franchti, Maroulas) konnten lediglich zwei authentische, mitochondriale Genome erstellt werden. Diese Probenanzahl ist für eine Gruppenbildung zur statistischen Analyse zu gering, da keine statistisch signifikanten Ergebnisse erhoben werden können. Es kann lediglich eine Aussage zur genetischen Linie getroffen werden. Die mitochondrialen Genome des Fundplatzes Theopetra (Linie K1c) sind bereits publiziert (siehe Anhang B.7). Bedingt durch die Probenverfügbarkeit und des vorhandenen, aber biomolekular mässig erhaltenen Probenmaterials, konnten für den Süden Griechenlands nur zwei mitochondriale Genome erstellt werden (Franchti (neolithisch), Tharrounia, Abbildung 6.32). Da sich die Gruppenbildung geographisch auf den Norden Griechenlands beschränkt, werden beide Proben, aufgrund der geographischen Distanz, aus dem Datensatz zur populationsgenetischen Analyse ausgeschlossen (siehe Karte in Abbildung 6.32, Fra, Tha).

In Folge dessen stehen 49 Proben aus dem Norden Griechenlands für eine populationsgenetische Analyse zur Verfügung, welche anhand von ^{14}C Datierungen oder der Stratigrafie des Fundplatzes gruppiert werden (siehe Tabelle 6.45 und Tabelle 6.46).

Tabelle 6.45: Gruppenbildung der prähistorischen Proben aus Griechenland
n= Anzahl Individuen je Gruppe

Gruppe	Abkürzung	n	Probenalter (kal. v.Chr.)	genomische Studien (siehe Anhang B.7)
Frühes Neolithikum Griechenland	NGR-EN	16	6500 - 6000	Rev 5
Spätes Neolithikum Griechenland	NGR-LN	33	6000 - 3300	Pal 7, Klei 10

Die Gruppe des frühen Neolithikums beschreibt die Ankunft der ersten Farmer im Norden Griechenlands und umfasst mit 16 Proben der Fundplätze Xirolimni, Revenia, Mavropigi und Nea Nikomedia einen Zeitraum von 6500 - 6000 v. Chr. (NGR-EN). Die folgenden Epochen des mittleren (n=7), späten (n=18) und finalen Neolithikums (n=8) werden als eine Gruppe ausgewertet, da eine Aufspaltung in einzelne Abschnitte mit geringer Probenzahl eine hohe Standardabweichung zeigte (Daten werden nicht gezeigt). Die Gruppe des fortschreitenden Neolithikums in Griechenland (NGR-LN) umfasst mit 33 Proben eine Zeitspanne von 6000 - 3300 v. Chr. Aufgrund des technologischen Fortschritts wird in dieser Epoche ein Populationswachstum angenommen, was aus archäologischer Sicht durch die Zunahme an Fundplätzen belegt wird (siehe Kapitel 2.3).

Tabelle 6.46: Gruppenbildung, detaillierte Aufschlüsselung.

HG = Haplogruppe erstellt über das komplette mitochondriale Genom, Arch. Zuordnung = Einordnung der Proben in den archäologischen Kontext

Fundplatz	Probe	HG	Alter (kal. v. Chr.)	Arch. Zuordnung	Region, Stadt	Breiten-grad	Längen-grad
Theopetra	Theo 1	K1c	7288-6771	Mesolithikum	Zentral-Griechenland, Thessalien	39,680	21,682
	Theo 5	K1c	7605-7529				
Revenia	Rev 4	H	6500 - 5800		Nord-Griechenland, Katerini, Makedonien	40,321	22,576
	Rev 5	X2b	6437 - 6264				
	Rev 6	H46	6500 - 5800				
Nea Nikomedia	Nea 1	K1a4b	6400 - 6000	frühes Neolithikum	Nord-Griechenland, Veria, Makedonien	40,584	22,253
	Nea 2	K1a	6223 - 6075				
	Nea 3	K1a2	6400 - 6000				
	Nea 4	H	6400 - 6000				
	Nea 5	H	6400 - 6000				
Mavropigi	Mau 1	H5	6590/6450-6200-6010		Nord-Griechenland, Kozani, Makedonien	40,448	21,722
	Mau 2	T2b23	6590/6450-6200-6010				
	Mau 4	H46	6590/6450-6200-6010				
	Mau 5	T2	6590/6450-6200-6010				
	Xir 1	T2c1d1	6000				
Xirolimni	Xir 2	T1a	6000		Nord-Griechenland, Kozani, Makedonien	40,294	21,661
	Xir 4	K1a1	6000				
	Xir 5	K1a24	6164				
Limnaria	Thas 1	K1b1a	5485 - 5372		Nord-Griechenland, Thassos	40,622	24,771
Nea Makri	NS 1	H1+152	5000 - 4000	mittleres Neolithikum	Ost-Griechenland, Alexandroupoli, Thrace	40,856	25,750
	NS 2	J1c	5000 - 4000				
Stavroupoli	St 2	N1a1a1a3	5800 -5400		Nord-Griechenland, Thessaloniki, Makedonien	40,672	22,934
	St 3	K1a	5800 -5400				
	St 5	H	5800 -5400				
	St 6	N1a1a1a3	5800 -5400				
Kremasti koilada	Krk 1	K1a1	5225 - 5014		Nord-Griechenland, Kozani, Makedonien	40,361	21,935
	Krk 2	H	5340 - 4930/4700				
	Krk 4	U8b1a1	5340 - 4930/4700				
	Krk 7	K1a3a	5340 - 4930/4700				
	Krk 8	N1a1a1a	5340 - 4930/4700				
	Krk 9	H	5340 - 4930/4700				
	Krk 10	K1a4a1	5340 - 4930/4700				
Makriyalos	Mkg 9	T2b	5500/5400 - 5000	spätes Neolithikum	Nord-Griechenland, Makryalos, Pierra	40,385	22,578
	Mkg 10	J1c3j	4787 - 4615				
	Mkg 11	K1a	5500/5400 - 5000				
	Mkg 12	U7b	5500/5400 - 5000				
	Mkg 13	K1a	5500/5400 - 5000				
Paliambela	Pal 1	U4b3	4900 - 4500		Nord-Griechenland, Katerini, Makedonien	40,511	22,506
	Pal 2	K1a7	4900 - 4500				
	Pal 4	U2e2a	4500 - 3200				
	Pal 5	U2e2a	4500 - 3200				
	Pal 6	K1a7	4900 - 4500				
	Pal 7	J1c1	4452 - 4350				
Kleitos	Klei 1	K1a4a1	5400 - 4000	finale Neolithikum	Nord-Griechenland, Kozani, Makedonien	40,368	21,832
	Klei 2	J1c	5400 - 4300				
	Klei 3	K1a4	5400 -4300				
	Klei 5	HV	5400 - 4300				
	Klei 6	H29	5400 - 4300				
	Klei 7	H29	5400 - 4300				
	Klei 8	J1c2	5400 - 4300				
	Klei 10	K1a2	4228 - 3995				
Tharounnia	Tha 2	H5	5300 - 3300	spätes Neolithikum	Zentralgriechenland, Evia, Tharounnia	38,485	23,946
Franchthi	Fra 8	U3b3	5000 - 4000	spätes Neolithikum	Süd-Griechenland, Peloponnes	37,423	23,132

6.2 Mitochondriale Genomanalyse

Das mitochondriale (mt) Genom umfasst mit 16569bp nur einen sehr geringfügigen Anteil der menschlichen Erbsubstanz, jedoch macht eine hohe Substitutionsrate dieses Molekül zu einem interessanten Marker in der Evolutionsforschung. Da es keine Möglichkeit der Rekombination bietet, häufen sich Mutationen innerhalb des mitochondrialen Genoms an und können in charakteristische Abstammungslinien eingegliedert werden. Das mt Genom umfasst verschiedene Kompartimente, von denen mehr als 90% Genregionen umfassen, welche unter anderem für Proteine der Atmungskette codieren (Position 500 - 16000). Lediglich die Kontrollregion (Position 16000 - 16569, 1 - 500), welche Ansatzstellen zur Reproduktion des mt Genoms beinhaltet, verhält sich neutral. Da sich hier einzelne Mutationen anhäufen, wird dieser Abschnitt als hypervariable Region (*hyper variable sequence*, HVS) bezeichnet. Die Kontrollregion hat im Vergleich zu den kodierenden Regionen des mt Genoms eine etwa 10-fach höhere Substitutionsrate (für eine detaillierte Darstellung siehe [44, 155]). Während kodierende Regionen (z.B. Cytochrom b) zur Identifikation einer Spezies herangezogen werden, liegt die HVS im Fokus populationsgenetischer Studien. Aufgrund der hohen Kopienzahl des Mitochondriums als Zellorganell zeigt das mitochondriale DNA Molekül im Vergleich zur DNA im Zellkern eine gute Verfügbarkeit, auch in stark degradierten Probenmaterial. In dieser Arbeit konnte zum Beispiel gezeigt werden, dass selbst bei einem endogenen DNA Gehalt von 0,05% ein authentisches, mitochondriales Genom isoliert werden kann (Probe: Theo 5). Einen autosomal-genomischen Datensatz aus Proben dieser Qualität zu erstellen ist nur bedingt möglich. Selbst wenn genug Probenmaterial zur Verfügung steht, dann stehen die Kosten der Sequenzierung in keiner Relation. Aus der Literatur ist bezüglich des mt Markers eine hohe Anzahl an Referenzdaten verfügbar, welche chronologisch und geographisch gruppiert mit den Daten dieser Arbeit ko-analysiert werden. Demographische Prozesse hinterlassen Signaturen innerhalb der genetischen Struktur dieser Gruppen, so dass Änderungen anhand der Haplotypendiversität und Nukleotiddiversität erfasst werden können. Durch die Berechnung genetischer Distanzen (F_{ST}) kann der genetische Zusammenhang der einzelnen, prähistorischen Populationen im Kontext des Referenzdatensatzes dargestellt werden.

Seit mehr als 10 Jahren werden mitochondriale Linien prähistorischer Populationen untersucht (Übersicht in Abschnitt 1.4), wobei die Analyse in der prä-NGS Ära auf die HVS-I Region durch PCR/Sanger Technologie eingeschränkt gewesen ist. Technischer Fortschritt erlaubt heutzutage die Erstellung mitochondrialer Genome als Nebenprodukt genomischer Projekte (Shotgun Sequenzierung) bzw. kann nach Anreicherung die vollständige Sequenz des mitochondrialen Genoms untersucht werden. Damit steht nun ein Referenzdatensatz zur Verfügung, welcher sich in Länge und Position der analysierten Region auf dem Mitochondrium unterscheidet, wobei der kleinste gemeinsame Nenner die HVS-I Region darstellt (Position 16051 - 16400). Populationsübergreifende Analysen können folglich nur im Rahmen der HVS-I Region durchgeführt werden, wofür die Consensus Sequenzen im FASTA Format verwendet werden (siehe Abschnitt 3.8). Da die HVS-I Region einen C-Stretch enthält (Position 16184-16193), an welchem während der Amplifikation durch die DNA Polymerase fehlerhafte Mutationen generiert werden können, wird die Position 16189 (C/T) in den folgenden Analysen ignoriert. Durch Literaturrecherchen steht ein Referenzdatensatz mit 1229 prähistorischen Individuen, eingeordnet in 28 Populationen, zur Verfügung (Übersicht in Tabelle 6.48). Zusätzlich werden die prähistorischen Daten mit 13 modernen Populationen aus der Ägäis und dem Balkan ko-analysiert (1640 Individuen, siehe Tabelle 6.47).

Verwendeten Methoden

Mitochondriale Haplogruppen wurden mit *haplogrep* (HVS Regionen) [86] oder *haplofind* (komplettes Genom) [196] bestimmt. Anhand des Programms Arlequin 3.5 [45] werden die populationsgenetischen Analysen durchgeführt. Innerhalb des Datensatzes aus Griechenland wird mit der kompletten Sequenz des Mitochondriums gearbeitet, um die genetischen Linien zu charakterisieren, gemeinsame Haplotypen in unterschiedlichen Individuen darzustellen (*shared haplotypes*) und die Frequenzen der Haplogruppen im griechischen Neolithikum zu berechnen. Anhand eines *Median Joining* Netzwerks werden die mutierten Positionen zwischen den einzelnen genetischen Linien des Datensatzes dargestellt (*Minimum spanning network* [7]). In dieser Analyse werden alle variablen Positionen über das gesamte mt Genom genutzt, wobei eine 10-fache Gewichtung auf Haplogruppen-definierende Positionen und auf der hypervariablen Region liegt. Die Astlänge entspricht der Anzahl variabler Positionen und stellt den Grad der Divergenz der einzelnen Linien/Individuen dar. Durch einen Abgleich mit der mitochondrialen SNP Datenbank Phylotree [192] kann anhand der definierenden Positionen die Authentizität der erstellten Linien überprüft werden.

Die Analysen der **molekularen Diversität** und genetischen Distanzen werden mit dem Datensatz auf Grundlage der HVS-I Region durchgeführt (siehe Tabelle 6.48). Es werden DNA Sequenzen (348bp) unter Verwendung des Kimura-2-Parameter Substitutionsmodells genutzt (Arlequin3.5 [45]). Analog zur Messung der Heterozygotität in diploiden Datensätzen wird für haploide Sequenzdaten die Gendiversität nach Nei bestimmt ([45], Seite 115). Die Gendiversität, oder auch Haplotypendiversität (H), beschreibt die Wahrscheinlichkeit, dass sich zwei zufällig gezogene Haplotypen aus einer Population unterscheiden. Um einen statistischen Fehler basierend auf der Probengröße (n) einer Population auszuschließen, wird durch die Einbeziehung der Probengröße der Parameter H normiert (\hat{H}). Eine weitere Beschreibung der molekularen Diversität ist die Messung der substituierten Nukleotide innerhalb einer Population, die Nukleotiddiversität π . Sie wird als Summe der Nukleotidifferenzen zwischen einem Sequenzpaar angegeben, gemittelt über die Gesamtzahl an Individuen der betrachteten Population.

Da mit zunehmender Evolutionszeit Mutationen auf dem Mitochondrium akkumuliert werden, gibt die Messungen der genetischen Distanz zweier Populationen Aufschluss über ihre gemeinsame Geschichte und Divergenzzeit. Unter Verwendung von Arlequin 3.5 wird eine Distanzmatrix zwischen sämtlichen Populationen des Referenzdatensatzes anhand der HVS-I Sequenzen gebildet, welche durch die Messung der Anzahl an Basenaustauschen zwischen zwei Populationen gebildet werden (*pairwise differences*). Eine Erweiterung bietet die **genetische Distanz nach Nei**, welche, unter Annahme, dass Mutation und Drift im Gleichgewicht stehen, aus der Zusammensetzung und Frequenz der Haplotypen zwischen zwei Populationen errechnet wird [45].

Geographische bzw. zeitliche Separation von Populationen verhindert die gleichbleibende Zusammensetzung des genetischen Materials, so dass anhand des **Fixierungsindex** F_{ST} das Maß der genetischen Unterschiede der Subpopulationen im Spektrum der Gesamtpopulation quantifiziert werden kann. In dieser Arbeit wird die genetische Distanz nach Reynolds (1983; Arlequin [45]) berechnet, welche für kurze evolutionäre Distanzen genutzt wird. Neumutationen können in einem kleinen Zeitfenster vernachlässigt werden, so dass genetische Drift als einzig wirkende evolutionäre Kraft angenommen wird. Die erhobenen Distanzen zwischen den Subpopulationen des Referenzdatensatzes werden durch eine euklidische Distanzmatrix abgebildet. Zur graphischen Darstellung werden die genetischen Distanzen durch eine multidimensionale Skalierung (MDS) reduziert (MASS Paket, R [194]) und die ermittelten Vektoren durch einen Plot abgebildet (R ggplot2 [202]).

Der mitochondriale Referenzdatensatz

Der mitochondriale Referenzdatensatz umfasst mit einem Zeitraum der letzten 40.000 Jahren die Kolonisation Europas durch den anatomisch modernen Menschen (AMM), die Ankunft der ersten Farmer in Europa über die danubische und mediterrane Route und die auf das Neolithikum folgende Bronzezeit. Die älteste mitochondriale Sequenz stammt aus dem Kontext des Protoaurignacien (Grotta di Fumane, Italien, 37855 kal. v. Chr. [11]) und beschreibt eingegliedert in eine Gruppe von 16 Individuen die voreiszeitliche mitochondriale Diversität (Pre-LGM). Während der Maximalvergletscherung besteht eine Datenlücke (24.000 - 18.000 v. Chr.). Die Jäger-Sammler Gesellschaft wird durch 6 weitere Gruppen abgebildet, welche unter anderem das späte, obere Paläolithikum nach der Einzeit (SOP-JS) sowie die mesolithische Population mit dem Beginn des Holozäns (Holo-JS) umfassen. Für die Mesolithiker Nord- und Nordosteuropas wurden die Gruppierungen aus der Dissertation von Anna Schulz [170] übernommen. Mit 17 Populationen aus dem europäischen Neolithikum deckt der Datensatz geographisch die Ägäis, den Balkan, Mittel-, West- und Nordeuropa ab. Dabei stellen die mitochondrialen Daten aus der Marmara Region (Nordwest Türkei) die älteste beprobte Region entlang der neolithischen Expansionsroute dar (ANA-EN) und werden mit den Daten dieser Arbeit im Kontext des "ägäischen Neolithikums" analysiert¹. Ausgehend von diesem Punkt können die zwei Ausbreitungsrichtungen des Neolithikums entlang der Donau (Ungarn/Kroatien, LBK) und der Mittelmeeroute (Spanien, Frankreich) analysiert werden². Auf beiden Expansionsrichtungen können die einzelnen neolithischen Gesellschaften in einzelne Epochen des Neolithikums gegliedert werden (früh, mittel, spät, final), so dass auch regionale demographische Prozesse beurteilt werden können. Vier Populationen aus der Bronzezeit stellen die jüngsten Proben dieses Datensatzes dar. In den folgenden Analysen werden zusätzlich moderne Populationen aus Süd-Ost-Europa berücksichtigt (siehe Tabelle 6.47).

Tabelle 6.47: Moderner, mitochondrialer Referenzdatensatz aus Südosteuropa.

Gruppe	Abk.	n	Referenz
Nordgriechenland	ngr	319	[75]
Albanien	alb	42	Belledi
Cypern	cyp	91	[75]
Bosien	bos	110	Harvey <i>et al.</i> (nicht publiziert)
Serbien	ser	56	Harvey <i>et al.</i> (nicht publiziert)
Kroatien	cro	59	Harvey <i>et al.</i> (nicht publiziert)
Türkei	tur	102	[29, 34, 169]
Bulgarien	bul	30	[29]
Sizilien	isi	118	[17]
Sardinien	isa	70	[17]
Adria	iad	182	[17]
Kreta	cre	180	[112]
Ungarn	hun	211	[74]

¹In der Publikation Hofmanova, Kreutzer *et al.* (2016) siehe Anhang B.7 konnte eine hohe genetische Ähnlichkeit zwischen den Individuen beider Seiten der Ägäis festgestellt werden.

²Eine graphische Darstellung zur Entstehung und Verbreitung des Neolithikums bietet Abbildung 1.3.

Tabelle 6.48: Übersicht des mitochondrialen HVS-I Datensatzes. Referenzen der verwendeten Daten sind in Tabelle 1.1 angegeben. Inklusiv der griechischen Gruppen NGR-EN (n=16) und NGR-LN (n=33) stehen 1278 Individuen eingegliedert in 30 Populationen zur Verfügung.

Gruppe	Abkürzung	n	Probenalter (kal. v.Chr.)	Kulturstufe
Oberes Paläolithikum	Pre-LGM	16	40.000 - 25.000	Gravettien/Aurignacien
Spätes oberes Paläolithikum (post LGM)	SOP-JS	16	17.000 - 11.000	Epigravettien/Magdalenien
Holozän Jäger-Sammler	Holo-JS	19	10.000 - 6000	Mesolithikum
Präneolithisch Nord-Ost Europa	NEE-JS	23	9000 - 5800	Mesolithikum
Präneolithisch Nordeuropa	NE-JS	25	7600 - 4800	Mesolithikum
Paraneolithisch südliches Baltikum	Neo-JS	18	4000 - 3000	Paraneolithische Mesolithiker
Paraneolithisch Skandinavien	PWC-JS	28	3000 - 2400	Pitted Ware Culture
Frühes Neolithikum Türkei	ANA-EN	27	6500 - 5600	Fikirtepe
Frühes Neolithikum Ungarn/Kroatien	STA-EN	44	6000 - 5500	Starčevo
Frühes Neolithikum Alföld (Ungarn)	KÖR-EN	18	6000 - 5600	Körös
Frühes Neolithikum Ungarn	LBK-T-EN	39	5300 - 4900	LBK Transdanubien
Mittleres Neolithikum Alföld I (Ungarn)	ALBK-I-MN	62	5500 - 5200	frühe Alföld LBK
Mittleres Neolithikum Alföld II (Ungarn)	ALBK-II-MN	119	5300 - 4900	späte Alföld LBK
Spätes Neolithikum Alföld (Ungarn)	TIS-LN	48	~ 4750	Tisza
Frühes Neolithikum Mitteleuropa	LBK-EN	120	5500 - 4775	LBK
Mittleres Neolithikum Mitteleuropa	CE-MN	49	4600 - 4000	Rössen, Schöningen
Spätes Neolithikum Mitteleuropa	CE-LN	64	4000 - 2800	Baalberge, Salzmünde
Finale Neolithikum Mitteleuropa	CE-FN	64	2800 - 2200	Corded Ware, Glockenbecher
Frühes Neolithikum Westeuropa	GUR	39	5000 - 4000	LBK- und Cardialelemente
Frühes Neolithikum Iberia	WE-EN	29	5475 - 4960	Cardialkultur
Spätes Neolithikum Iberia	WE-LN	61	4335 - 3635	
Finale Neolithikum Iberia	WE-FN	86	2900 - 1740	teilw. Glockenbecherkultur
Frühes Neolithikum Nordeuropa	NE-EN	22	3600 - 2500	Trichterbecher
Spätes Neolithikum Nordeuropa	NE-FN	11	2500 - 1800	Trichterbecher Kultur
Bronze Zeit Kreta	MIN-BA	37	2450 - 1750	Minoer
Bronze Zeit Steppe	YAM-BA	28	3300-2300	Yamnaya
Bronze Zeit Mitteleuropa	CE-BA	94	2200 - 1750	Unetice
Späte Bronzezeit Steppe	CAT-BA	23	2450-1950	Katakombengrabkultur

6.3 Ergebnisse der mitochondrialen Genomanalyse

Mitochondriale DNA Linien im prähistorischen Griechenland

Der prähistorische Datensatz aus Griechenland zeigt mitochondriale Linien der Haplogruppen K1, H, U*, T1, T2, J1c und X, welche in der Literatur bereits als charakteristische Linien des europäischen Neolithikums beschrieben worden sind [20, 181, 81]. Dies trifft ebenfalls für die zwei mesolithischen Individuen des Datensatzes zu. Bezeichnend für den Datensatz ist die hohe Frequenz der Haplogruppen K und H mit jeweils 36 und 25% (siehe Abbildung 6.33). Die Linien der K- und H Cluster zeigen zusätzlich eine geringe Anzahl variabler Positionen, was für eine geringe evolutionäre Divergenzzeit spricht. Definierende, mutierte Positionen für das K-Cluster sind 9055, 10398, 12308, 16224 und 16311, sowie für die H-Linien 73, 7028 und 14766. Ein *Minimum spanning network* zeigt die genetischen Linien der mitochondrialen Sequenzen aus dem griechischen Mesolithikum und Neolithikum (siehe Anhang B.59). Phylogenetisch aussagekräftige Mutationen sind zwischen den einzelnen Linien hervorgehoben. Die Darstellung eines Netzwerks kann auf phylogenetischer Ebene als ein weiteres Kriterium für die Authentizität des Datensatzes angesehen werden: Es wurden im gesamten Datensatz keine Haplogruppen-definierenden Mutationen gefunden, welche Abweichungen zu den bestimmten Linien zeigen. Ab dem späten Neolithikum treten Linien der Haplogruppe U* mit einer Frequenz von 11% im Datensatz auf (U2, U3, U4 und U7). Der gesamte Datensatz zeigt keine U5 definierenden Mutationen (U5: Position 3197, U5a: Position 14793, U5b: 14182), welche in westeuropäischen JS Gesellschaften in hoher Frequenz vorliegen und ebenfalls in den späten Epochen des zentral-europäischen Neolithikums auftreten.

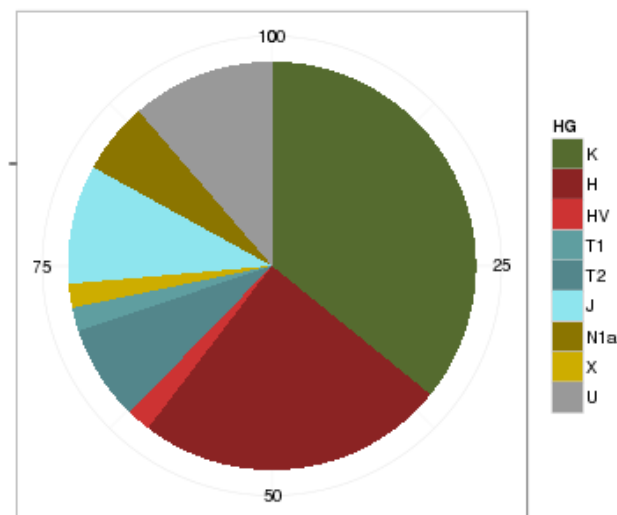


Abbildung 6.33: Frequenzen der mt Haplogruppen innerhalb des Datensatzes aus Griechenland.

Durch die *shared haplotype* Analyse, welche zu dem Standard Ergebnis von Arlequin gehört, konnten unter Einbeziehung des gesamten mitochondrialen Genoms 6 Haplotypen identifiziert werden, welche innerhalb des Datensatzes 2-3x vorkommen. Folgende Proben weisen über das gesamte mit-

ochondriale Genom die gleichen Mutationen auf^{1, 2}:

- Nea4 = Nea5 (HG: H)
- Klei1 = Klei3 = Krk10 (HG: K1a4a1)
- St5 = Krk2 = Krk9 (HG: H)
- Pal2 = Pal6 (HG: K1a7)
- Klei6 = Klei7 (HG: H29)
- St2 = St6 (HG: N1a1a1a3)

Das Neolithikum der Ägäis und des Balkans aus mitochondrialer Sicht

In der Region der Ägäis startete um 6500 v. Chr. die Expansion des Neolithikums auf den europäischen Kontinent. Im folgenden Abschnitt soll die genetische Zusammensetzung dieser Region besprochen sowie die Position des griechischen Neolithikums dargestellt werden. Mit dem Übertritt des Neolithikums auf den Balkan entwickelte sich die Vorstufe der LBK, welche in den folgenden Jahrhunderten nach Zentraleuropa expandierte. Zur Darstellung der Verteilung der mitochondrialen Linien wurden Haplogruppenfrequenzen innerhalb der neolithischen Gruppierungen ermittelt und in Abbildung 6.34 durch ein Balkendiagramm dargestellt. Es wird deutlich dass die hohe Frequenz an K1- und H-Linien besonders im griechischen Neolithikum betont ist (>50%) und die Frequenz mit dem Übertritt des Neolithikums auf den Balkan reduziert wird (<50%). Im Neolithikum der Ägäis und des Balkan können ähnliche genetische Linien identifiziert werden, jedoch in variierenden Frequenzen. Ab dem mittleren Neolithikum sind in der Alföld-LBK und der LBK in Transdanubien U5 Linien zu beobachten.

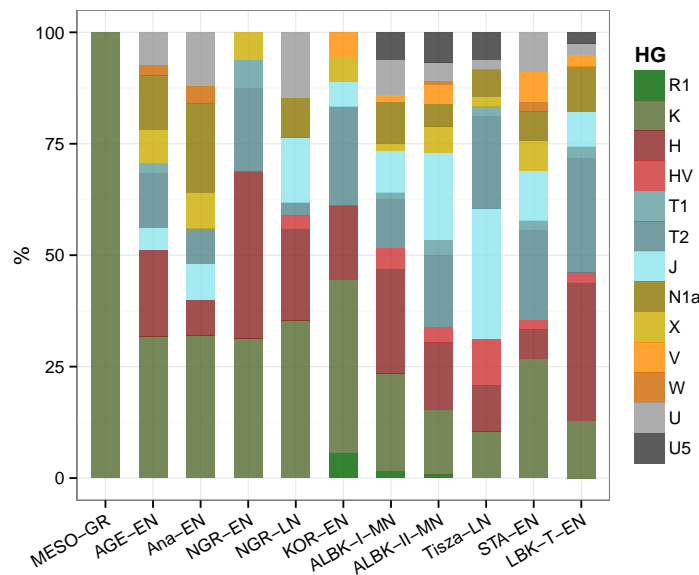


Abbildung 6.34: Zusammensetzung der mitochondrialen Linien einzelner Populationen der Ägäis und des Balkans.

Abkürzungen: MESO-GR - 2 Ind. aus der Theopetra Höhle, AGE-EN - ägäisches Frühneolithikum zusammengesetzt aus Ana-EN und NGR-EN), Ana-EN - Barcin, Mentese, NGR - Nord-Griechenland - frühes Neolithikum (EN), spätes/finales Neolithikum (LN), KOR - Körös Kultur, ALP - Alföld Linearbandkeramik, STA - Starčevo Kultur, LBK-T - Linearbandkeramik Transdanubien

¹Wird innerhalb eines Fundplatzes der gleiche Haplotyp gefunden, kann angenommen werden, dass es sich um das gleiche Individuum handelt. Die dargestellten Proben eines Fundplatzes sind aus unterschiedlichen Gräbern geborgen (vgl. Probenvorbereitung Abschnitt 2.3).

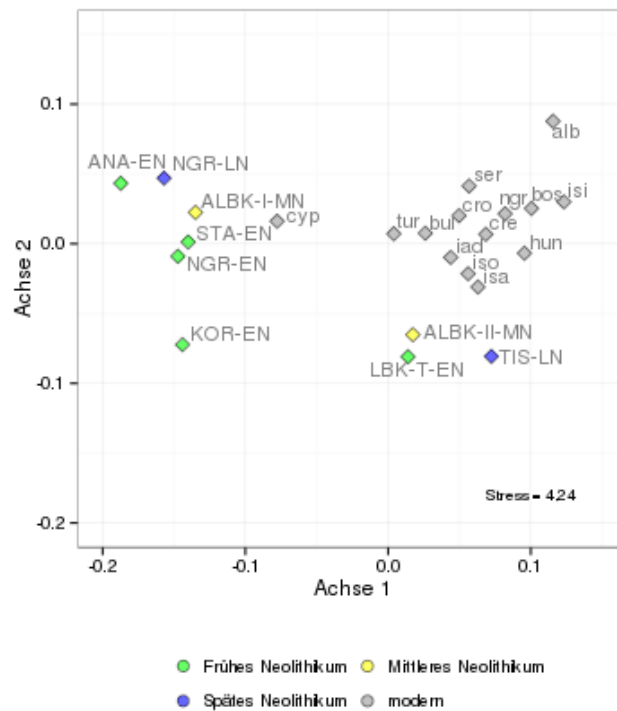
²Die Proben Pal 4 und Pal 5 sind aus archäologischem Kontext nicht eindeutig als zwei Individuen identifiziert. Die mt Genomsequenz der Probe Pal 4 unterscheidet sich an Position A16183C.

Die ermittelten F_{ST} Werte zwischen den Populationen der Ägäis (ANA-EN, NGR-EN, NGR-LN) und des frühen/mittleren Neolithikum auf dem Balkan (STA, KOR, ALBK-I) zeigen keine hohen Distanzen, jedoch ist diese Messung nicht signifikant (siehe Tabelle 6.35a). Erst mit dem fortschreitenden Neolithikum können zu der Alföld LBK (Phase II) und der LBK in Transdanubien signifikante F_{ST} -Werte mit einer Distanz >0.045 gemessen werden (p -Wert $<0,05$), was durch die Aufnahme der U5 Linien in den neolithischen Genpool erklärt werden kann. Das frühe Neolithikum aus Griechenland zeigt im Vergleich zu jenen aus der Marmara Region einen geringeren Grad an Differenzierung gegenüber den Populationen des fortschreitenden Neolithikums auf dem Balkan.

Die multidimensionale Skalierung der erhobenen Distanzen zeigt die Ausbildung zweier Cluster (siehe Abbildung 6.35b). Linksseitig im Plot angeordnet befinden sich die frühneolithischen Populationen der Marmara Region (ANA), Griechenlands (NGR; auch spätes Neolithikum), der Starčevo Kultur (STA) und der ersten Phase der Alföld LBK (ALBK-I-MN). Das zweite Cluster wird durch die modernen Populationen, der LBK aus Transdanubien (LBK-T), dem mittleren (ALBK-II-MN) und späten Alföld Neolithikum (TIS) gebildet. Dies spricht ab 5000 v.vChr. auf dem Balkan für in Kraft tretenden demographische Prozesse, wie z.B. die Aufnahme neuer Linien (U5) und einer Änderung der Zusammensetzung der Population im Gegensatz zu ihrem Ursprung (Cluster 1). In der Abbildung stellt die Körös Gruppe eine Aussengruppe dar.

	<i>ANA-EN</i>	<i>NGR-EN</i>	<i>NGR-LN</i>	<i>STA-EN</i>	<i>KOR-EN</i>	<i>ALBK-I</i>	<i>ALBK-II</i>	<i>TIS-LN</i>	<i>LBK-T-EN</i>
<i>ANA-EN</i>		0,262	0,376	0,293	0,145	0,210	0,002	0,003	0,025
<i>NGR-EN</i>	0,0117		0,240	0,693	0,945	0,291	0,075	0,044	0,123
<i>NGR-LN</i>	0,0004	0,0111		0,144	0,160	0,390	0,001	0,004	0,008
<i>STA-EN</i>	0,0051	0,0000	0,0133		0,387	0,410	0,206	0,117	0,278
<i>KOR-EN</i>	0,0268	0,0000	0,0208	0,0002		0,074	0,007	0,011	0,027
<i>ALBK-I</i>	0,0084	0,0062	0,0001	0,0000	0,0288		0,052	0,016	0,158
<i>ALBK-II</i>	0,0507	0,0254	0,0446	0,0047	0,0555	0,0102		0,286	0,228
<i>TIS-LN</i>	0,0696	0,0402	0,0530	0,0122	0,0670	0,0258	0,0022		0,148
<i>LBK-T-EN</i>	0,0469	0,0258	0,0531	0,0046	0,0582	0,0096	0,0046	0,0107	

(a) Genetische Distanzen zwischen neolithischen Populationen der Ägäis und des Balkans: F_{ST} Werte linearisiert nach Slatkin (unten) und p-value (oben) ermittelt mit Arlequin unter Anwendung der Kimura-2-P Substitutionsrate. Hervorgehobene F_{ST} Werte sind signifikant ($p < 0.05$). Übersicht der verwendeten Populationen siehe Tabelle 6.48.



(b) Multidimensionaler Skalierung der Reynolds F_{ST} -Werte zwischen den Populationen aus dem Neolithikum des Balkans und der Ägäis (mit modernen Populationen dieser Region).

Abbildung 6.35: F_{ST} Werte und multidimensionale Skalierung der genetischen Distanzen. Abkürzungen siehe Tabelle 6.48.

Die neolithische Expansion nach Europa

Im folgenden Abschnitt werden die genetischen Distanzen zwischen den Populationen des europäischen Neolithikums dargestellt. Die Reichweite der neolithischen Referenzdaten erlaubt die Darstellung der danubischen und mediterranen Ausbreitung des Neolithikums nach Europa, wobei die anatolischen und griechischen Neolithiker die ältesten Populationen des Datensatzes darstellen.

Im vorherigen Abschnitt konnte in den frühneolithischen Gesellschaften der Ägäis und des Balkan eine homogene Population beobachtet werden, in welcher keine signifikanten genetischen Distanzen zu sehen sind. Dieses Ergebnis setzt sich zwischen den Gruppen des ägäischen Frühneolithikums und den ersten Farmern in Europa fort. In Abbildung 6.36a und 6.36b sind die genetischen Distanzen der ägäischen Population (ANA, NGR) zu den Populationen der danubischen (STA, LBK, LBK-T) und mediterranen Ausbreitungsrouten (WE) dargestellt. In beiden Richtungen kann nur zwischen der Gruppierung des späten Neolithikums aus Griechenland und der LBK in Europa eine signifikante Distanz ermittelt werden ($F_{ST}=0.033$), welche als gering eingestuft werden kann. Im fortschreitenden Neolithikum nimmt die genetische Distanz zu den neolithischen Populationen beider Routen zu. Besonders deutlich sind die Distanzen zu der Gruppe des finalen Neolithikums in Zentraleuropa mit Werten von $F_{ST}(ANA - CE_FN)=0.115$, $F_{ST}(NGR_EN-CE_FN)=0.080$ und $F_{ST}(NGR_LN-CE_FN)=0.108$. Ähnliche Muster sind, wenn auch nicht so ausgeprägt, zu den Populationen des späten/finalen Neolithikums auf dem Balkan und Spanien zu sehen (siehe Tabelle 6.37).

Ein Großteil signifikanter Distanzen ist ausgehend von der Körös Gruppe (Ungarn) zu den spätneolithischen Populationen festzustellen (siehe Tabelle in Abbildung 6.37, Spalte KOR-EN). Der höchste Wert besteht zu der Population des finalen Neolithikums (CE-FN, $F_{ST}=0.144$), welcher die deutlichste Distanz im Kontext des gesamten Datensatzes darstellt. Verdeutlicht wird dieses Ergebnis durch die multidimensionale Skalierung der genetischen Distanzen, in welcher die Körös Gruppe separiert wird (siehe Abbildung 6.36a). Die Neolithiker der *Gurgy* Population (Paris, Frankreich) stellt die einzige Gruppierung des Datensatzes dar, zu welcher konstant signifikante genetische Distanzen berechnet werden können, wobei eine maximale Distanz zur Körös Gruppe ($F_{ST}=0,126$) besteht. Jene Population nimmt eine Aussenposition in der graphischen Darstellung ein (Abb. 6.36c). Dies zeigt bereits nach Ankunft in westlichen Teilen Europas eine deutliche Strukturierung der neolithischen Population.

In Abbildung 6.36c sind die genetischen Distanzen der Populationen des europäischen Neolithikums abgebildet. Wird das chronologische Auftreten der einzelnen neolithischen Kulturstufen während der Verbreitung in Europa berücksichtigt, kann eine Ausbreitungsrichtung beginnend mit den ersten Farmern aus Anatolien (ANA-EN) festgestellt werden. Dabei ordnen sich die Gruppen aus Griechenland entlang dieser Achse an. Es kann eine schrittweise Änderung der genetischen Diversität mit der Verbreitung nach Europa gezeigt werden. Im fortschreitenden Neolithikum sind Regionalisierungen zu erkennen, welche besonders ausgeprägt in den neolithischen Epochen Zentraleuropas ist ($F_{ST} \text{ LBK} - \text{CE-FN} = 0.056$). In Griechenland hingegen ist nur eine geringe, nicht signifikante genetische Distanz ($F_{ST} = 0.011$) zu sehen, was für eine genetisch homogene Population während des gesamten Neolithikums in Griechenland spricht.

Zur Beschreibung der mitochondrialen Diversität wurden deskriptiv statistische Parameter erhoben (Tabelle 6.49¹). Für die Verbreitung des Neolithikums ausgehend von Anatolien wäre eine stetige Abnahme der Diversität in Richtung des Balkan zu erwarten (*isolation by distance*). Interessanterweise

¹Die Parameter der molekularen Diversität aller prähistorischen Referenzen sind im Anhang B.73 zu finden.

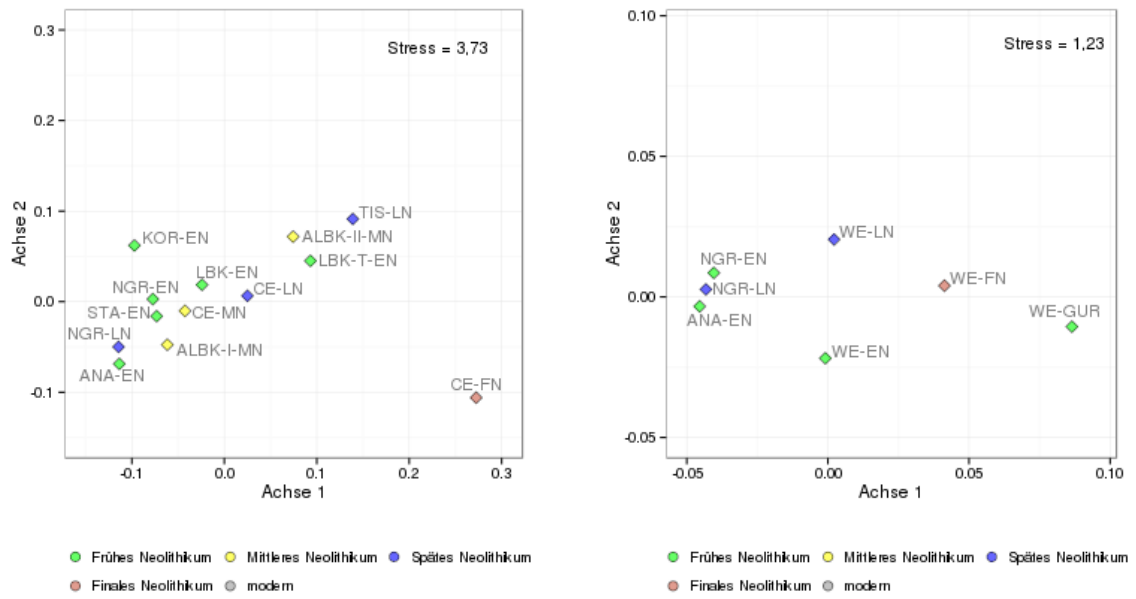
ist in den Populationen der Ägäis eine sehr geringe Haplotypdiversität feststellbar ($\hat{H} \sim 0,91$), welche mit dem Übergang des Neolithikums auf den Balkan sprunghaft ansteigt ($\hat{H} \sim 0,96$) und in der folgenden Verbreitung, als auch in späteren Epochen des Neolithikums, auf diesem Level verbleibt (siehe Tabelle B.73).

Zwischen den neolithischen Epochen in Griechenland (früh/spät) ist nur ein geringer Anstieg der Haplotypdiversität feststellbar. Das gleiche gilt für die Nukleotiddiversität, welche mit $\pi \sim 0,0120$ erniedrigt ist, im Gegensatz den anderen neolithischen Populationen ($\pi = 0.0132-0,0146$) - mit Ausnahme der Körös-Gruppe. Die Körös-Gruppe zeigt ebenfalls verringerte Diversitätswerte in der deskriptiven Statistik.

Tabelle 6.49: Deskriptiv statistische Parameter der genetischen Diversität. Auszug aus Tabelle B.73.

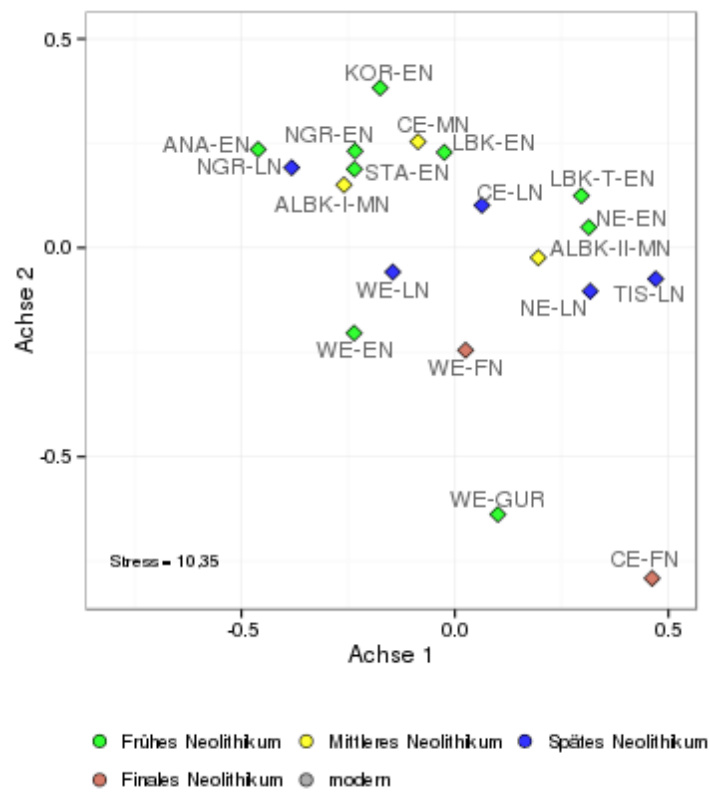
n: Anzahl Individuen, h: Anzahl Haplotypen, \hat{H} : Gendiversität (SD: Standardabweichung), n Pos: Anzahl analysierte Positionen, s: segregierende Positionen, Ts: Anzahl Transitionen, Tv: Anzahl Transversionen, π_n : Nukleotiddiversität, π : *pairwise differences*

Population	n	h	\hat{H}	SD \hat{H}	n Pos	s	Ts	Tv	π_n	SD π_n	π	SD π
ANA-EN	27	15	0,906	0,045	348	24	22	2	0,0144	0,008	5,01	2,51
NGR-EN	16	11	0,908	0,063	348	17	17	0	0,0115	0,0067	4,02	2,11
NGR-LN	33	16	0,918	0,030	348	25	22	3	0,0123	0,007	4,30	2,18
KOR-EN	18	12	0,895	0,065	348	14	14	0	0,0110	0,0065	3,84	2,02
STA-EN	44	30	0,965	0,017	343	32	31	1	0,0146	0,008	5,03	2,49
LBK-T-EN	39	28	0,978	0,011	343	32	32	1	0,0140	0,0078	4,82	2,4
ALBK-I-MN	62	38	0,976	0,008	348	40	37	3	0,0133	0,0073	4,64	2,31
ALBK-II-MN	119	67	0,977	0,006	348	48	44	4	0,0134	0,0073	4,68	2,31
TIS-LN	48	27	0,946	0,022	348	36	31	5	0,0132	0,0073	4,59	2,29



(a) Danubische Ausbreitungsrouten.

(b) Mediterrane Ausbreitungsrouten.



(c) Populationsstruktur des europäischen Neolithikums.

Abbildung 6.36: Multidimensionale Skalierung der genetischen Distanzen europäischer, neolithischer Populationen. Darstellung der Reynolds F_{ST} Werte.

	ANV-EN	NGR-EN	NGR-LN	STA-EN	KOR-EN	ALBK-I	ALBK-II	TIS-LN	LBK-T-EN	LBK-EN	CE-MN	CE-LN	CE-FN	WE-EN	WE-GTR	WE-LN	WE-FN	NE-EN	NE-LN
ANV-EN		0,252	0,384	0,281	0,152	0,271	0,002	0,002	0,022	0,100	0,169	0,033	0,000	0,101	0,005	0,040	0,005	0,032	0,095
NGR-EN	0,0117		0,242	0,687	0,942	0,283	0,071	0,049	0,122	0,230	0,401	0,160	0,003	0,164	0,004	0,269	0,032	0,105	0,155
NGR-LN	0,0004	0,0111		0,137	0,165	0,389	0,001	0,005	0,010	0,040	0,091	0,015	0,000	0,054	0,008	0,046	0,020	0,107	0,060
STA-EN	0,0055	0,0000	0,0137		0,381	0,389	0,195	0,116	0,281	0,577	0,875	0,204	0,000	0,087	0,001	0,089	0,010	0,331	0,272
KOR-EN	0,0268	0,0000	0,0209	0,0006		0,072	0,007	0,011	0,026	0,068	0,103	0,019	0,000	0,026	0,000	0,100	0,024	0,023	0,032
ALBK-I	0,0084	0,0062	0,0001	0,0002	0,0238		0,050	0,015	0,155	0,240	0,892	0,336	0,000	0,049	0,007	0,149	0,059	0,698	0,439
ALBK-II	0,0507	0,0234	0,0446	0,0031	0,0555	0,0102		0,287	0,220	0,024	0,570	0,399	0,003	0,005	0,002	0,004	0,001	0,962	0,422
TIS-LN	0,0696	0,0402	0,0530	0,0125	0,0670	0,0258	0,0022		0,155	0,056	0,130	0,017	0,000	0,001	0,000	0,000	0,001	0,552	0,103
LBK-T-EN	0,0470	0,0260	0,0533	0,0044	0,0584	0,0098	0,0047	0,0108		0,632	0,548	0,187	0,002	0,005	0,000	0,002	0,000	0,343	0,835
LBK-EN	0,0174	0,0105	0,0245	0,0000	0,0306	0,0033	0,0107	0,0146	0,0000		0,729	0,079	0,000	0,013	0,001	0,002	0,000	0,223	0,334
CE-MN	0,0122	0,0000	0,0168	0,0000	0,0244	0,0000	0,0000	0,0105	0,0000	0,0000		0,922	0,004	0,057	0,010	0,170	0,040	0,905	0,564
CE-LN	0,0277	0,0146	0,0304	0,0057	0,0475	0,0001	0,0001	0,0220	0,0069	0,0091	0,0000		0,018	0,020	0,025	0,104	0,050	0,990	0,716
CE-FN	0,1153	0,0797	0,1033	0,0603	0,1440	0,0408	0,0226	0,0541	0,0418	0,0556	0,0293	0,0151		0,000	0,001	0,000	0,001	0,993	1,000
WE-EN	0,0234	0,0216	0,0283	0,0239	0,0586	0,0225	0,0398	0,0679	0,0564	0,0384	0,0219	0,0265	0,0768		0,001	0,026	0,006	0,055	0,059
WE-GTR	0,0552	0,0731	0,0458	0,0490	0,1260	0,0299	0,0310	0,0603	0,0644	0,0492	0,0273	0,0186	0,0410	0,0499		0,003	0,009	0,229	0,042
WE-LN	0,0284	0,0074	0,0230	0,0135	0,0247	0,0072	0,0241	0,0535	0,0486	0,0317	0,0068	0,0090	0,0485	0,0294	0,0392		0,539	0,477	0,140
WE-FN	0,0502	0,0249	0,0275	0,0286	0,0448	0,0111	0,0246	0,0513	0,0535	0,0410	0,0151	0,0108	0,0341	0,0418	0,0267	0,0000		0,990	0,191
NE-EN	0,0486	0,0359	0,0245	0,0028	0,0796	0,0000	0,0000	0,0000	0,0034	0,0081	0,0000	0,0000	0,0000	0,0318	0,0079	0,0000	0,0000		0,051
NE-LN	0,0456	0,0324	0,0583	0,0108	0,0863	0,0000	0,0000	0,0314	0,0000	0,0041	0,0000	0,0000	0,0000	0,0458	0,0435	0,0247	0,0172	0,0533	

Abbildung 6.37: Genetische Distanzen zwischen den Populationen des europäischen Neolithikums (F_{ST} linearisiert nach Slatkin). Abkürzungen siehe Tabelle 6.48.

Der prähistorische Datensatz aus Europa

Im folgenden Abschnitt soll die Position des griechischen Neolithikums im Zusammenhang mit den verfügbaren mitochondrialen Daten aus dem europäischen Paläolithikum, Mesolithikum, Neolithikum, der Bronzezeit und modernen Populationen aus Süd-Ost-Europa untersucht werden. Abbildung 6.38a zeigt die genetischen Distanzen zwischen den Populationen des Datensatzes unter Anwendung dreier Methoden [45]: Die Unterschiede lassen sich (I) recht einfach durch die Anzahl der Basenaustausche innerhalb einer Population (orange) oder (II) zwischen zwei Populationen darstellen (grün, *pairwise differences*). Eine Erweiterung bietet die genetische Distanz nach Nei, welche, unter Annahme, dass Mutation und Drift im Gleichgewicht stehen, aus der Zusammensetzung und Frequenz der Haplotypen zwischen zwei Populationen errechnet wird (blau). Die erhobenen genetischen Distanzen auf Basis der F_{ST} Werte, linearisiert nach Slatkin, sind in Tabelle 6.38b in Form einer Distanzmatrix graphisch dargestellt. Diese zeigen keine signifikanten Abweichungen innerhalb der Ausprägungen zu den genetischen Distanzen nach Nei. Beide Analysen zeigen die höchsten Werte zwischen dem ägäischen Neolithikum und den JS Populationen des oberen Pleistozäns und Holozäns, welche in Abbildung 6.38a und 6.38b quadratisch hervorgehoben werden.

Eine hohe Anzahl an differenzierender Linien zwischen zwei Populationen führt zu einer genetischen Distanz beider Gruppierungen. Es benötigt eine gewisse Zeit, um diese Unterschiede durch Mutation und genetische Drift zu generieren, sprich der gemeinsame Vorfahre ist zeitlich weit entfernt. In dem vorliegenden Datensatz ist die größte Distanz (nach Nei und Slatkin F_{ST}) zwischen den ersten Farmergesellschaften (ANA-EN, NGR, STA-EN) und den Mesolithiker Zentraleuropas (Holo-JS, NE-JS) zu finden (siehe Abbildung 6.38a, blau und 6.38b, Tabelle 6.50). Die prägnanteste Distanz besteht zwischen der Körös Gruppe und den JS aus dem Holozän ($F_{ST}=0.57$). Die frühen Neolithiker aus Griechenland zeigen zu den JS aus dem Holozän eine Distanz von $F_{ST}=0.49$. Zu den ersten Farmern Europas konnte eine Distanz von F_{ST} (Holo_JS - LBK)=0.30 und F_{ST} (Holo_JS - WE_EN)=0.37 ermittelt werden.

Tabelle 6.50: F_{ST} Werte linearisiert nach Slatkin zwischen den JS Populationen und den ersten Farmern Europas. Auszug aus Abbildung 6.38b. Die dargestellten Distanzen sind mit einem p-Wert <0.05 signifikant. Die zwei höchsten Distanzen sind hervorgehoben.

	Pre-LGM	SOP-JS	Holo-JS	NEE-JS	NE-JS	Neo-JS	PWC-JS
ANA-EN	0.116	0.162	0.400	0.145	0.359	0.157	0.130
NGR-EN	0.163	0.236	0.496	0.145	0.431	0.099	0.106
NGR-LN	0.123	0.169	0.450	0.152	0.404	0.158	0.116
STA-EN	0.095	0.121	0.358	0.135	0.323	0.089	0.110
KOR-EN	0.243	0.317	0.571	0.198	0.519	0.160	0.148
LBK-T-EN	0.082	0.108	0.333	0.153	0.297	0.069	0.138
LBK-EN	0.075	0.097	0.307	0.139	0.274	0.086	0.111
WE-EN	0.074	0.137	0.374	0.133	0.325	0.130	0.123
NE-EN	0.050	0.071	0.444	0.161	0.392	0.098	0.142

Im vorherigen Kapitel wurde bereits eine Zunahme der F_{ST} Werte zwischen dem ägäischen Neolithikum und den neolithischen Epochen des fortschreitenden Neolithikums in Europa beschrieben. Die Analyse der F_{ST} Werte zwischen den den Epochen des mittleren, späten und finalen Neolithikums in Zentraleuropa zeigen eine kontinuierliche Abnahme der genetischen Distanz im Vergleich mit der JS Population aus dem Holozän (siehe Tabelle 6.51). Ein ähnlicher Trend ist im fortschreitenden Neolithikum auf der iberischen Halbinsel ersichtlich, jedoch nicht im Neolithikum der Alföld LBK. Im Vergleich des zentraleuropäischen Neolithikums zu der Yamnaya Population (Bronzezeit der Steppe)

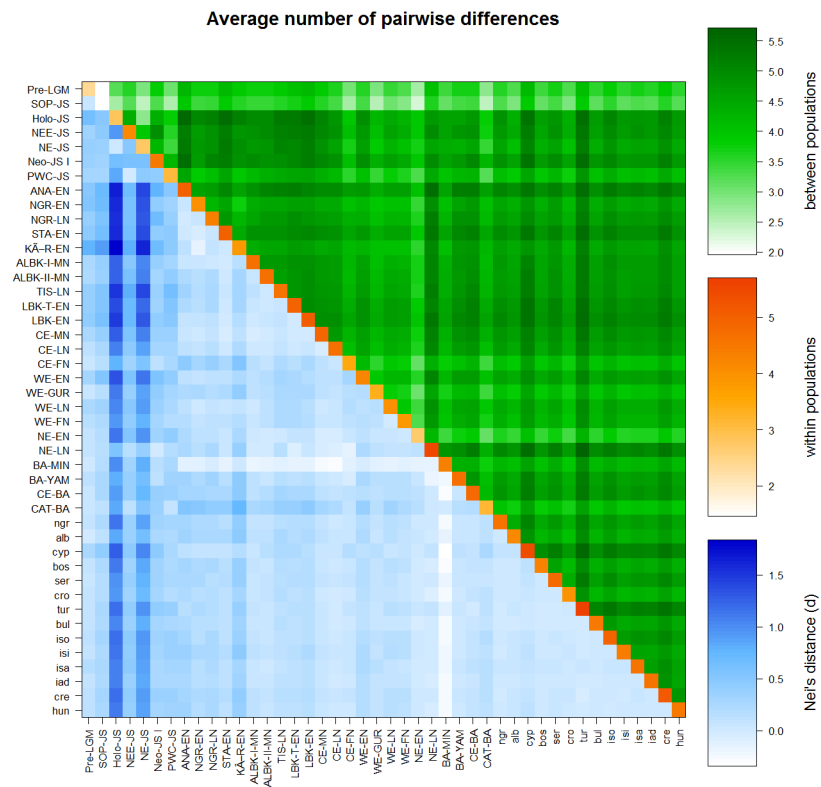
ist ebenfalls eine Abnahme der Distanzen zu sehen ($F_{ST}(\text{BA-YAM} - \text{CE-MN})=0.012$, $F_{ST}(\text{BA-YAM} - \text{CE-LN})=0.006$, $F_{ST}(\text{BA-YAM} - \text{CE-FN})=0.000$). Dies bestätigt vorherige Resultate aus der paläogenetischen Literatur [20, 61].

Tabelle 6.51: F_{ST} Werte linearisiert nach Slatkin zwischen den JS Populationen und den Populationen des fortschreitenden Neolithikums in Europa. Hervorgehobene Werte zeigen signifikante Distanzen.

	Pre-LGM	SOP-JS	Holo-JS	NEE-JS	NE-JS	Neo-JS	PWC-JS
ALBK-I-MN	0.051	0.068	0.290	0.107	0.247	0.084	0.077
ALBK-II-MN	0.038	0.048	0.266	0.125	0.238	0.062	0.103
TIS-LN	0.091	0.101	0.363	0.181	0.351	0.077	0.161
CE-MN	0.047	0.071	0.297	0.117	0.261	0.072	0.093
CE-LN	0.017	0.036	0.245	0.098	0.207	0.062	0.077
CE-FN	0.015	0.025	0.231	0.090	0.170	0.037	0.081
WE-GUR	0.015	0.027	0.349	0.114	0.279	0.120	0.097
WE-LN	0.063	0.065	0.281	0.125	0.245	0.092	0.078
WE-FN	0.038	0.029	0.244	0.107	0.209	0.081	0.059
NE-LN	0.047	0.078	0.199	0.048	0.153	0.005	0.064

In Abbildung 6.39a ist die Populationsstruktur anhand der mitochondrialen Diversität in einem Zeitfenster vom späten Paläolithikum bis in die heutige Zeit dargestellt. Grundlage sind die zuvor besprochenen genetischen Distanzen (Reynolds, Arlequin). In der graphischen Darstellung bilden die JS aus dem Holozän neben den JS Nordeuropas die deutlichste Aussengruppe. Die Populationen aus dem oberen Pleistozän hingegen zeigen geringere Distanzen und fallen in der Darstellung nahe die Gruppierung der modernen Populationen und jenen aus der Bronzezeit. Die modernen Populationen Südosteuropas können kaum voneinander differenziert werden und bilden ein homogenes Cluster. Die genetische Struktur des Datensatzes separiert paläolithische/mesolithische Populationen von den ersten Farmern Europas durch eine entgegengesetzte Platzierungen innerhalb des Plots. Die Strukturierung der Daten spiegelt die zuvor beschriebene genetische Distanz der früh-neolithischen Populationen zu allen Subpopulationen der JS Kultur wieder. Regionale Differenzierungen innerhalb der Populationen des fortschreitenden Neolithikums zeigen eine Annäherung an die heutige, moderne Population. Durch die Platzierung der spät-neolithischen Population aus Griechenland (NGR-LN) wird deutlich, dass diese Prozess in dem ägäischen Neolithikum nicht stattgefunden haben.

Mit Fokus auf die Kulturen des Neolithikums wurde eine vergrößerte Darstellung auf Grundlage der Berechnung der genetischen Distanzen aus Abbildung 6.39a erstellt (siehe Abbildung 6.39b). Die früh-neolithischen Populationen ordnen sich auf einer querverlaufenden Achse (rechts oben - mitte) entlang der Ausbreitungsrichtung des Neolithikums an (NGR-EN, ANA-EN, STA-EN, LBK-EN, LBK-T-EN). Die Populationen des mittleren, späten und finalen Neolithikums des Balkans und Zentraleuropas folgen dieser Achse und enden in einem Cluster bronzezeitlicher und moderner Populationen. Diese Achse beschreibt die Migration der frühen Farmer Kulturen. Durch aufeinanderfolgende Gründereffekte sowie Verpaarung mit lokalen JS änderte sich die Zusammensetzung der einzelnen neolithischen Gesellschaften, so dass eine genetische Distanz zum Ursprung des Neolithikums entstanden ist.



(a) Paarweise Unterschiede und genetische Distanz nach Nei.

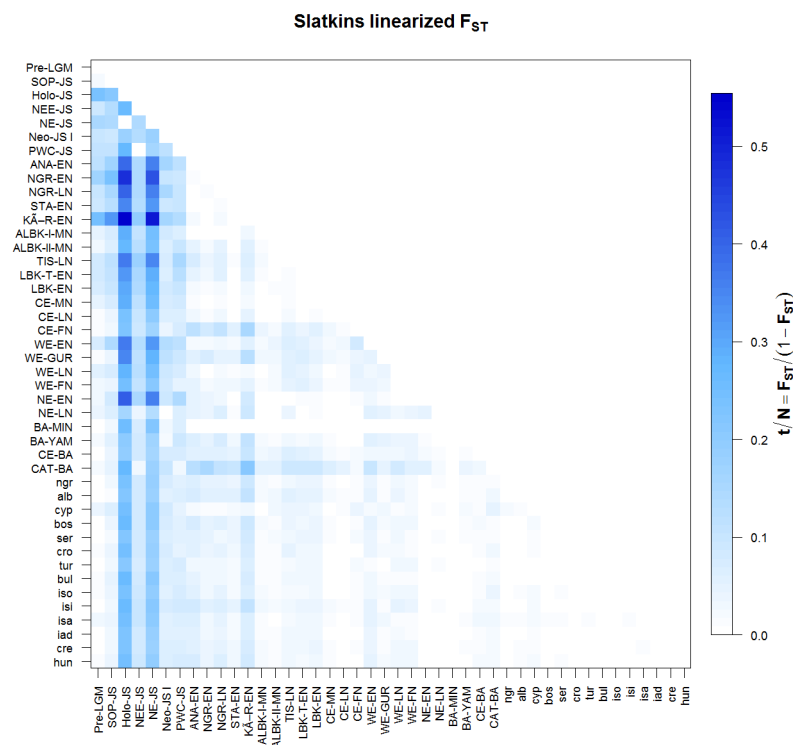
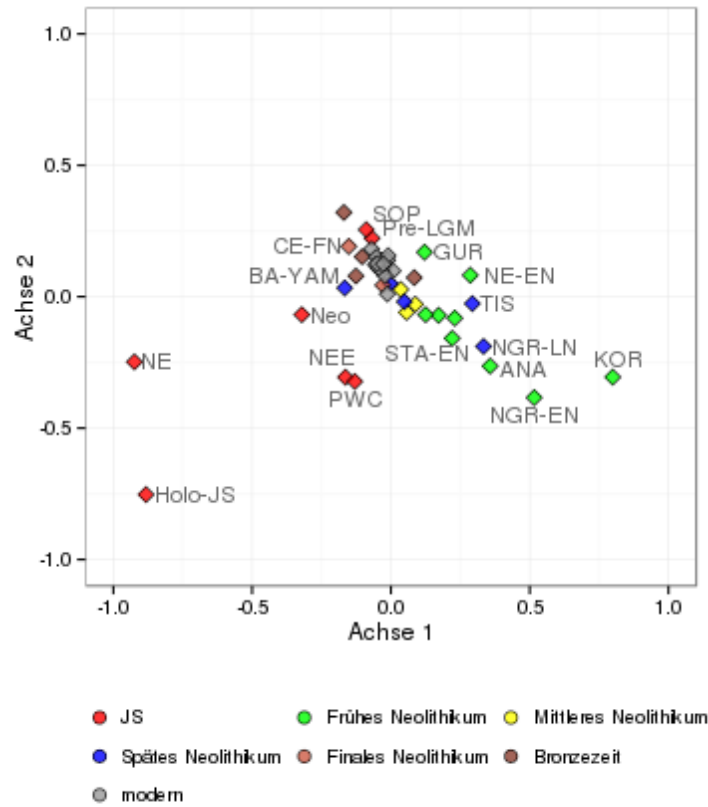
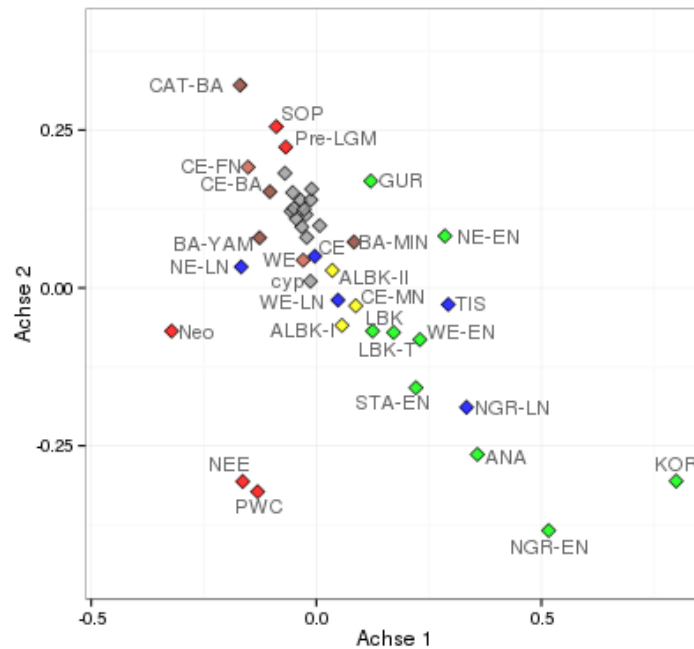
(b) Abbildung der paarweisen F_{ST} -Werte linearisiert nach Slatkin in Form einer Distanzmatrix.

Abbildung 6.38: Abbildung der genetischen Distanzen zwischen einzelnen Populationen des prähistorischen Datensatzes aus Europa inklusive ausgewählter moderner Populationen. Die Abbildungen wurden durch Arlequin erstellt [45].



(a) Genetische Distanzen der prähistorischen Populationen aus Europa



(b) Vergrößerte Darstellung auf Populationen der neolithischen Expansion nach Europa (Grundlage ist MDS aus Abbildung 6.39a).

Abbildung 6.39: Strukturierung der Populationen des prähistorischen Datensatzes anhand des mitochondrialen Markers (HVS-I). Gezeigt werden die genetischen Distanzen (nach Reynolds) nach multidimensionaler Skalierung. Prähistorische Populationen sind anhand der Kodierung in Tabelle 6.48 beschriftet und anhand von Geographie (Symbol) und Chronologie (Farbgebung) gekennzeichnet (siehe Legende).

6.4 Räumlich-explizite Simulationen

Anhand der Erfassung der genetischen Diversität in prähistorischen und modernen Populationen können demographische Parameter durch Simulationen von Populationsmodellen zwischen beiden Zeitstufen getestet werden. Innerhalb dieser Arbeit wurde dies in einer Kooperation durch Kollegen der Universität Genf, Nuno Da Silva und Mathias Currat, durchgeführt. Das folgende Kapitel referiert ihre Ergebnisse unter Anwendung des Programms SPLATCHE2 (*SPatiaL And Temporal Coalescences in Heterogeneous Environments*, Version 2.0) [147], welches für die räumlich-explizite Simulation des mitochondrialen Datensatzes angewendet wird. Gegenüber den geläufigen Methoden hat dieser Algorithmus den Vorteil, dass er die räumliche und zeitliche Komponente prähistorischer Populationen als auch Migration berücksichtigt. Durch die zeitliche Trennung eines prähistorischen Datensatzes kann die erste Kolonisation einer Region mit folgenden Migrationen in mehreren Schichten dargestellt werden, wobei eine Rate des genetischen Austauschs durch Paarung der Individuen beider Schichten (*Admixture Rate*, γ) ermittelt werden kann. Die räumliche Trennung der Regionen des Datensatzes in einzelne Deme ermöglicht die Darstellung einer Migrationsrate zwischen einzelnen Regionen. Für die Deme wird eine Populationsdichte (K , *carrying capacity*) und Wachstumsrate (r), sowie eine Migrationsrate (m) zwischen einzelnen Demen, angenommen. Es können Populationsmodelle unter Anwendung variabler demographischer Parameter angenommen und durch Simulation getestet werden. Auf Basis der alten DNA Datensätze werden neue Generationen an Pseudodaten unter Berücksichtigung des zu testenden Populationsmodells erstellt (Anzahl Generationen wird durch Probenalter festgelegt). Anschliessend wird jenes Modell bestimmt, welches nach der Simulation die maximale Übereinstimmung innerhalb der molekularen Diversität zu den beobachteten, modernen Daten zeigt.

Bezüglich des mitochondrialen Datensatzes aus Griechenland stellt sich zum einen die Frage nach Populationskontinuität zwischen den einzelnen Epochen des Neolithikums und zwischen dem Neolithikum und der heutigen modernen Population in Griechenland. Zum anderen soll die *Admixture Rate* mit lokalen JS Populationen während der Verbreitung des Neolithikums analysiert werden. Die folgenden Daten sind unveröffentlichte Ergebnisse, deren Publikation im nächsten Jahr erfolgt.

Test auf Populationskontinuität in Griechenland

Die Frage nach Kontinuität in der Bevölkerungsgeschichte einer Populationen unter Annahme von genetischer Drift als alleinig wirkende Kraft impliziert, dass keine massiven Populationsbewegungen stattgefunden haben. Im Rahmen dieser Arbeit kann diese Frage durch die prähistorischen Daten Griechenlands und aus der Literatur verfügbare moderne Daten der Region getestet werden (ngr, Tabelle 6.47 [75]). Der prähistorische Datensatz aus Griechenland wird in zwei Populationen geteilt:

1. Das frühe Neolithikum zur Abbildung der Ankunft der ersten Farmer (NGR-EN) und
2. die folgenden Epochen des Neolithikums (NGR-MN + NGR-LN).

Es werden vier mögliche Szenarien beschrieben und durch räumlich-expliziten Simulation analysiert. Ausgehend von Anatolien wird die neolithische Expansion während der letzten 10.000 Jahre nachgestellt (400 Generationen), wobei grundlegende, demographische Parameter anhand archäologischen Indizien festgelegt worden sind. Um die räumliche Komponente in die Simulation einfließen zu lassen, wird die geographische Region in einzelne Kompartimente (Deme) gegliedert - hier erfolgt die Einteilung Europas in einzelne Deme von 100 x 100 km. Jedem Dem wird eine Populationsdichte

(K) von 1000 Individuen zugeschrieben¹, welche zur Zeit des Römischen Reichs auf 24.000 Individuen erhöht wurde. Die Migrationsrate ($m = 0,4$) und Wachstumsrate ($r = 0,45$) wurden anhand erhobener Daten innerhalb der neolithischen Transition in Europa bestimmt. Der prähistorische Datensatz fließt anhand GPS Daten räumlich und durch die Datierung der einzelnen Proben zeitlich in die Simulation ein (siehe Tabelle 6.46). Innerhalb der räumlich-expliziten Simulation wurden Populationsmodelle für vier Szenarien formuliert. Das erste Szenario beschreibt ein Basismodell unter Anwendung der oben beschriebenen Parameter K, m & r (Model 1 "SP basic"). In einem zweiten Szenario wird eine geringere Populationsdichte ($K = 500$) angenommen, welche durch eine höhere Wachstumsrate ($r = 0,53$) kompensiert wird (Model 2 "SP low K"). Das dritte Szenario nimmt eine Reduktion der Migrationsrate um 1/10 zwischen einzelner Deme an, um ein Populationsgleichgewicht nach Ankunft des Neolithikums darzustellen (Model 3 "SP low m"). In einem vierten Szenario wird die räumliche Komponente außer Acht gelassen. Das bedeutet, dass die Region Griechenlands als ein Deme angesehen wird (Panmixie) und die Proben chronologisch (nicht räumlich) in die Simulation einfließen. Die Wachstumsrate wurde angepasst, um die Populationsdichte vor dem Neolithikum ($r = 0,003$), nach dem Neolithikum ($r = 0,07$) und zur Zeit des Römischen Reichs ($r = 0,4$) zu erreichen (Model 4 "NSP"). Auf Kontinuität wurde zwischen beiden Zeitstufen des Neolithikums, sowie beider hinsichtlich der modernen Populationen geprüft.

Zwischen dem frühen und späten Neolithikum kann die Kontinuität in keinem der vier formulierten Szenarien widerlegt werden. Die genetische Diversität beider Zeitstufen gibt keinen Anlass eine Änderung der genetischen Zusammensetzung beider Populationen anzunehmen. Jedoch sind Populationsverschiebungen innerhalb des neolithischen Genpools nicht ausgeschlossen (z.B. trans-ägäische Migration), da sich in diesem Fall die molekularen Signaturen gleichartig verhalten. Unter Anwendung von Szenario 4 kann eine Kontinuität zwischen dem Neolithikum und der modernen Population Griechenlands widerlegt werden, jedoch nicht unter Anwendung der räumlich-expliziten Szenarien (Ausnahme spätes Neolithikum - modern; Szenario 1). Da räumlich-explizite Simulationen Populationsstruktur und Migration zweier benachbarter Kompartimente berücksichtigen, sind diese Modellierungen näher an der Realität. Folglich ist die Kernaussage, dass Populationskontinuität zwischen der Zeitstufe des Neolithikums und der heutigen Bevölkerung Griechenlands nicht ausgeschlossen werden kann.

Wird diese Analyse mit den Daten des zentraleuropäischen Neolithikums durchgeführt, dann kann für Modell 3 die Kontinuität zwischen mehreren Zeitstufen widerlegt werden²: Ausgehend vom Frühen und Mittleren Neolithikum kann keine Kontinuität in die heutige Zeit bestätigt werden. Schon während des Neolithikums können zwei Populationsumbrüche identifiziert werden. Der erste ist zwischen dem Mittleren und Finalen Neolithikum zu sehen. An diesem Punkt beschreiben vorhergehende Analysen die Aufnahme von JS Linien in den neolithischen Genpool [20, 61]. Ein weiterer Populationsshift erfolgte zwischen dem Späten und Finalen Neolithikum, was auf die massive Einwanderung der Yamnaya Kultur aus der Steppe zurückzuführen ist (siehe [20, 61]). Im Gegensatz zur Analyse nach [20] erfolgte in dieser Studie eine feinere Untergliederung der neolithischen Zeitstufen, so dass beide Events voneinander getrennt dargestellt werden können. Neben Analysen der mitochondrialen Diversität [20] und genomisch-statistischen Analysen [61], kann nun anhand einer räumlich-expliziten Simulation die Populationsbewegungen innerhalb des zentraleuropäischen Neolithikums bestätigt werden.

¹kalkuliert anhand der Dichte von LBK Fundplätzen ($\sim 0,6$ Individuen/km²)

²Die Gruppenbildung folgt jener in Tabelle 6.48, so dass auf Kontinuität zwischen dem frühen Neolithikum (LBK), dem mittleren Neolithikum (CE-MN), dem späten Neolithikum (CE-LN) und dem finalen Neolithikum (CE-FN) getestet werden kann. Als moderner Referenzdatensatz dienen die modernen Daten aus Bramanti *et al.* (2009). [19]

Räumlich-explizite Simulation der neolithischen Expansion

Zur Aufklärung der *Admixture* Rate während der Verbreitung Neolithikums wurden weitere Modelle formuliert, welche ausgehend von Anatolien über Griechenland und den Balkan nach Zentraleuropa die neolithische Expansionswelle nachbilden und zusätzliche Interaktion mit europäischen JS Populationen erlauben. Die Simulation wurde auf einen Zeitraum von 40.000 Jahren ausgedehnt, so dass die JS Population Europas berücksichtigt werden kann (Tabelle 6.48, Holo-JS). Die Grundlage dieses Modells beschreibt zwei Ebenen, welche jeweils eine Welle der Besiedlung Europas darstellen. Die erste Ebene bildet die paläolithische Population ab und stellt die Kolonisation des europäischen Kontinents durch den AMM dar. Auf einer zweiten Ebene findet die neolithische Expansion beginnend in Anatolien vor etwa 8.000 Jahren statt. Für die JS Population können eigene Parameter bestimmt werden ($m=0,15$, $r=0,2$, $K=100$), während für die Ebene des Neolithikums die Parameter aus dem vorherigen Abschnitt übernommen werden.

Neben den bereits besprochenen prähistorischen Datensätzen aus Griechenland und Zentraleuropa (siehe oben) wurden die zwei früh-neolithischen Populationen der Starčevo Kultur und der LBK Transdanubien [181] dem Datensatz hinzugefügt (siehe Tabelle 6.48, STA-EN, LBK-T-EN). Mit der vorliegenden Simulation soll die Interaktion in Form von Paarungsereignissen zwischen JS in Europa und ankommenden Farmern beurteilt werden. Durch die Aufnahme der räumlichen Komponente in die Simulation, können diverse Ausprägungen der *Admixture*-Prozesse innerhalb einzelner Regionen berechnet werden. Als Basismodell werden die Parameter m , r und K konstant verwendet. Interaktionen zwischen beiden Ebenen können anhand des *Admixture* Parameters (γ) untersucht werden, weshalb 4 verschiedene Szenarien zur Analyse formuliert wurden:

1. Eine konstante *Admixture* Rate, beginnend ab der Ankunft des Neolithikums in einer Region, von Anatolien bis nach Zentraleuropa
2. Eine zunehmende *Admixture* Rate nach Ankunft des Neolithikums in einer Region, wobei die Rate in allen Regionen den gleichen Wert hat
3. Die *Admixture* Rate ist nur in Zentraleuropa ab der Ankunft der Neolithiker vorhanden (konstant)
4. Nach Ankunft des Neolithikums in Zentraleuropa nimmt die *Admixture* Rate stetig zu. Auf dem Weg von Anatolien nach Griechenland ist kein *Admixture* vorhanden

Szenario 1 & 3 können durch die Simulation mit hoher Signifikanz widerlegt werden. Damit ist eine konstante *Admixture* Rate, welche direkt nach der Ankunft des Neolithikums beginnt, nicht wahrscheinlich, sondern es werden die Modelle favorisiert, welche zunächst eine parallele Entwicklung beider Populationen suggerieren. Eine Vermischung zwischen beiden Ebenen findet erst im Verlauf des Neolithikums statt. Ob die Paarungsereignisse auf die Region Zentraleuropas begrenzt gewesen sind (Szenario 4) oder auf der gesamten Route stattgefunden hat (Szenario 2), ist aus der vorliegenden Simulation nicht signifikant zu unterscheiden, jedoch scheint das Signal vermehrt Szenario 2 zu widerlegen (p-Wert: Szenario 2 $>39\%$, Szenario 4 $>57\%$). Basierend auf Szenario 4 kann ein *Admixture* Parameter von $16,3\%$ (Abweichung: $8,5 - 27,8\%$) für die Rate an Austausch von mitochondrialen Linien zwischen beiden Ebenen errechnet werden. Diese Analyse bestätigen den demischen Diffusionsprozess der Neolithiker nach Zentraleuropa. Die genetische mitochondriale Diversität suggeriert eine Aufnahme an Individuen aus der JS Population, welche jedoch erst nach einiger Zeit der parallelen Existenz beider Populationen in derselben Region stattgefunden hat.

Diskussion

7.1 Diskussion der Laborergebnisse

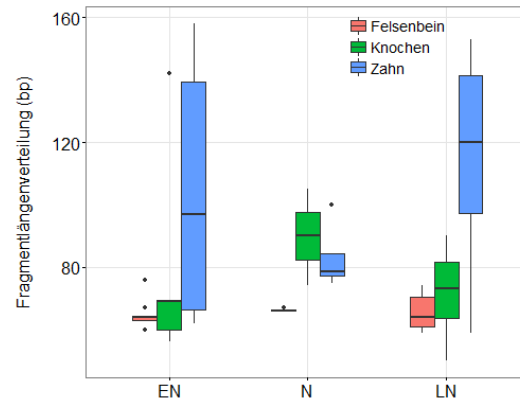
Charakterisierung des Probenmaterials hinsichtlich Qualität

Neben der parallelen Sequenzierung des gesamten Genoms einer Probe bieten NGS Methoden weitere Vorteile, welche besonders in dem Gebiet der Paläogenetik genutzt werden, um durch eingehende Charakterisierung der alten DNA Moleküle einen Datensatz zu authentifizieren. Die meisten alten DNA Extrakte sind durch fluoreszenzbasierte Methoden nicht quantifizierbar, da die Konzentrationen den Messbereich unterschreiten. Folglich werden die Schritte der Probenvorbereitung für eine Sequenzierung "blind" durchgeführt. Zu Zeiten der PCR/ Sanger Technologie bestand ein Probenscreening auf dem Erfolg einer PCR (Gelbild), welche im Anschluss meist problemlos durch die Sanger Technologie sequenziert werden konnte. Dies verhält sich unter Anwendung von NGS Methoden anders, da jedes Molekül in eine Library eingebaut wird und ein positives Ergebnis (z.B. durch Bioanalyser Messung) nicht mit einer erfolgreichen Sequenzierung endogener DNA Moleküle gleichgesetzt werden kann. Damit war zu Beginn der NGS Ära erst nach einer Sequenzierung ersichtlich, ob sich eine Probe für eine molekulargenetische Analyse geeignet hat. Wie in Abschnitt 3.1 ersichtlich, waren prähistorische Proben aus wärmeren Regionen wie Griechenland und Iran eher ungeeignet bzw. waren die angewendeten Verfahren nicht optimal. Es mussten folglich zwei Probleme gelöst werden: Zum einen sollte eine Probe charakterisiert werden können, bevor zeit- und kostenintensive Labortechniken wie Capture Verfahren und Sequenzierung angewendet werden und zum anderen musste die Sensitivität der Laborabläufe gesteigert werden (siehe nächster Diskussionspunkt).

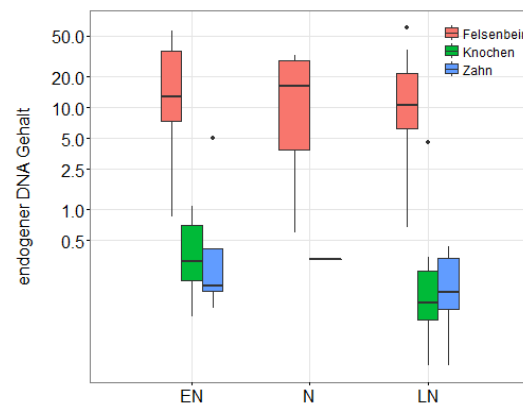
Die erste Stufe der Charakterisierung des Probenmaterials ist die Quantifizierung der umgesetzten Moleküle in eine Library nach dem Protokoll von Meyer und Kircher (2010) [120]. An dieser Stelle werden Proben ausgeschlossen, welche einen geringen Abstand zu den Molekülzahlen der Leerkontrolle zeigen. Anhand Versuche dieser Arbeit konnte gezeigt werden, dass eine Mindestmenge an einzigartigen Molekülen einer Library notwendig ist um ein authentisches altes DNA Molekül nach Sequenzierung darzustellen (Probe Krk1 & Krk6, Abschnitt 4.5). Auf dieser Basis wurde eine Vorauswahl an Proben für eine Sequenzierung zur Quantifizierung des endogenen DNA Gehalts getroffen (MiSeq 50bpSE). Die Kombination beider Parameter resultiert in der Quantifizierung der einzigartigen endogenen DNA Moleküle (uniq) einer Probe, welche aussagekräftig für die Wahl der Analysestrategie ist. Zum Beispiel ist eine einzigartige endogene Molekülzahl von $>2 \cdot 10^5$ notwendig, um im Anreicherungsverfahren ein Mitochondrium mit ausreichender Abdeckung zu sequenzieren (siehe Abschnitt 4.5). Weitere Parameter, welche Auskunft über die Qualität einer Probe geben sind die Fragmentlängenverteilung innerhalb der Library (Bioanalyser Messung) und die Desaminierungsfrequenzen am 5'-Ende eines Moleküls nach MiSeq Sequenzierung.

Die Charakterisierung der Proben dieser Arbeit ist in Abschnitt 4.3 dargestellt. In Abbildung 7.40 werden die durchschnittlich ermittelten Parameter des endogenen DNA Gehaltes, der "uniq" Molekülzahlen und der Fragmentlängenverteilung in den drei neolithischen Zeitstufen (früh/mittel/final) gezeigt. Es wird deutlich, dass nicht die Zeit ausschlaggebend für die Qualität einer Probe ist, sondern das beprobte Skelettelement. DNA extrahiert aus Felsenbeinen zeigt durchweg eine konstante Fragmentlängenverteilung, welche etwa 60bp beträgt. In Knochen- und Zahnmaterialien findet man hingegen eine breite Verteilung der Fragmentlängen, was für eine poröse Matrix spricht (Abbildung 7.40). Dies begünstigt das Eindringen von Bodenlebewesen aus dem umgebenden Milieu, welche ko-

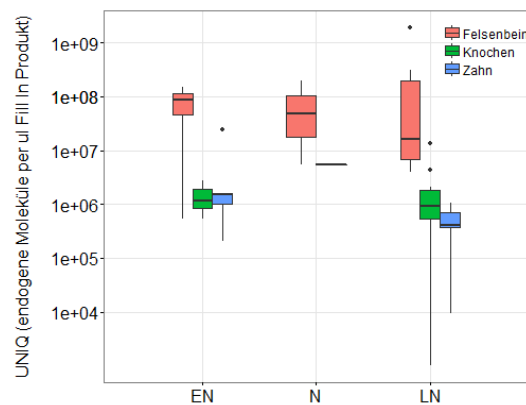
extrahiert einen grossen Anteil an exogener DNA einer Probe ausmachen. Dies spiegelt sich direkt in den Werten des endogenen DNA Gehaltes (Abbildung 7.40b) und der Komplexität einer Probe (Abbildung 7.40c) wieder.



(a) Fragmentlängenverteilung.



(b) Endogener DNA Gehalt.



(c) Komplexität der Proben.

Abbildung 7.40: Charakterisierung der Qualität innerhalb des nordgriechischen Probenspektrums. Gruppiert anhand der verfügbaren Skelettelemente Knochen, Zahn und Felsenbein. EN: frühes Neolithikum, N: mittleres Neolithikum, LN: spätes Neolithikum

Protokolloptimierung für biomolekular mäßig erhaltenes Probenmaterial

Da das griechische Probenmaterial mit Ausnahme der Felsenbeine eine geringe Anzahl einzigartiger endogener Moleküle besitzt, ist das Ziel einer Protokolloptimierung, den Einbau hochfragmentierter und stark geschädigter DNA Moleküle in eine Sequenzierlibrary zu steigern, sowie die Hybridisierungsreaktion während des Anreicherungsverfahrens an kurze Moleküle anzupassen. Um eine höhere molekulare Komplexität der einzelnen Sequenzierlibraries zu erhalten, wurde, beginnend mit der Extraktion, der gesamte Protokollablauf modifiziert.

Für die Extraktion konnte gezeigt werden, dass eine komplette Lyse des Skelettmaterials zu einer Steigerung der Molekülzahl führt. Nach initialer Testung konnte dies unter Verwendung von 0.2g Skelettmaterial und 6.7ml Lysepuffer am effizientesten umgesetzt werden. Eine Steigerung des eingesetzten Materials mit gleichzeitiger Erhöhung des Volumens an Lysepuffer zeigte einen negativen Effekt, da EDTA als Komponente des Lysepuffers nicht vollständig entfernt werden konnte und nachfolgende enzymatische Reaktionen inhibiert (siehe Abschnitt 4.2, Abbildung 4.22). Zusätzlich wird das limitierte Probenmaterial eingespart. Weitere Vorteile birgt die Reduktion der Zentrifugalkraft auf 2500g während der Aufkonzentration der Lysate, da prozentual ein grösserer Anteil an kürzeren Fragmentlängen zurückgehalten werden kann. Es wurde eine Variante der Silika Extraktion getestet (nach Brotherton (2013) [23]), jedoch zeigte diese Methode im Vergleich eine Tendenz zur Gewinnung längerer Fragmente (siehe Anhang Abbildung A.49). Vielversprechend ist eine neue Variante der Extraktion, welche Isopropanol statt Ethanol im Bindepuffer nutzt, dessen biochemische Eigenschaften die Bindung kleinerer Moleküle an eine Silika Matrix favorisieren (siehe Dabney *et al.*, [37]).

Zur Steigerung der umgesetzten Molekülzahlen in eine Sequenzierlibrary wurden diverse Protokollvarianten getestet. Die höchste Ausbeute konnte durch eine Variante mit einem Hitzeinaktivierungsschritt nach der *Fill In* Reaktion erzielt werden (siehe Anhang A.2). Der grösste Fortschritt liegt in der Modifizierung des Protokolls zur Anreicherung des mitochondrialen Genoms. Da eine gewisse Sättigung der RNA-*baits* gegeben sein muss und hierfür eine gewisse Anzahl endogener Moleküle vorausgesetzt wird, konnte nach der Vereinigung mehrerer Libraries einer Probe für die Hybridisierungsreaktion eine enorme Steigerung der *coverage* der Zielregion erreicht werden (siehe Dissertation Melanie Groß [58]). Ist diese Anzahl endogener Moleküle nicht vorhanden, gehen die RNA-*baits* unspezifische Bindungen ein, sodass die Effizienz des Anreicherungsverfahrens gesenkt wird. Umgekehrt existiert ein oberer Schwellenwert, da zu viele Moleküle die Reaktionskinetik negativ beeinflussen. Des Weiteren arbeitet das verwendete Programm zur Entfernung der PCR Duplikate (Mark Duplicates) mit Sequenzkoordinaten des 5'-Endes eines alignierten DNA Moleküls. Folglich können nach Duplikatentfernung während der mitochondrialen Datenanalyse maximal 33138 Sequenzen zurückbleiben. Je höher die Komplexität einer Probe, desto weniger Librarymoleküle werden in der Hybridisierungsreaktion benötigt. Werden komplexe Libraries verwendet, können ebenfalls geringere Rohdatenmengen während der Sequenzierung antizipiert werden (siehe Abschnitt 4.5). Nach Anpassung an die Grösse der sequenzierten Zielregion kann die in Abschnitt 4.5 aufgeführte Kalkulation der antizipierten Rohdatenmenge analog für Anreicherungsverfahren genomischer Loci, sowie Shotgun Sequenzierungen angewendet werden. Der wichtigste Aspekt der Anpassung der antizipierten Rohdatenmenge an die Komplexität einer Probe ist die Reduktion des Duplikationslevels nach Sequenzierung. Ab einem Duplikationslevel von 90% ist der Zuwachs an neuer Sequenzinformation gering (siehe Abbildung 4.29a), was besonders im Rahmen von Shotgun Sequenzierungen vermieden werden sollte.

Stark fragmentierte und hoch desaminierte DNA Moleküle zeigen eine verringerte Bindungsstärke zu den RNA-*baits* während der Hybridisierungs- und Waschreaktion. Durch das Herabsetzen der

Waschtemperatur auf 57°C konnte die Ausbeute an Sequenzdaten für stark geschädigte Proben bis zu 28% gesteigert werden (siehe Abschnitt 4.5). Da besonders DNA Moleküle aus Felsenbeinen starke DNA Schäden aufweisen, obgleich ein hoher endogener DNA Gehalt besteht (siehe folgender Abschnitt), sollten besonders für die Bearbeitung dieser Proben sensitivierte Protokollvarianten, beginnend mit der Extraktion, angewendet werden. Besondere Vorsicht sollte bei der Bearbeitung dieser Proben bezüglich der Wahl der Polymerase zur ersten Amplifikation der *Library* Moleküle getroffen werden. In der paläogenetischen Literatur werden oft Polymerasen des Phusion Stammes verwendet (z.B. AccuPrime), da diese Desaminierungsraten an den Enden alter DNA Moleküle herab setzen. Diese Reduktion basiert darauf, dass jene Polymerasen nicht über Uracil hinweg amplifizieren und ein Strangabbruch erfolgt. Hinsichtlich der Amplifikation eines Librarymoleküls führt dies zum Verlust des Moleküls, da einseitig der Adapter entfernt und dieses Molekül seine Sequenzierfähigkeit verliert. Damit erfolgt eine Reduktion der Komplexität einer Probe, was wiederum das Duplikationslevel während der Sequenzierung erhöht (siehe Abschnitt 4.5). Diese Polymerase dürfen nur in Kombination mit einer vorherigen Behandlung durch den USER¹™ Mix eingesetzt werden. Dieser Enzym-Mix² schneidet das DNA Molekül an einer Uracil-tragenden Position. Da in dieser Arbeit gezeigt werden kann, dass die Desaminierungsrate mit der Fragmentierung einer Probe korreliert (siehe folgender Abschnitt), ergibt sich für stark geschädigte Proben der Nachteil, dass die Fragmentlängen extrem gekürzt werden und in der Datenauswertung ein hoher Anteil über den angewendeten 30bp Filter verloren gehen. Dies hat bei der Probe Rev 5 (md5 = 56%) in der Shotgun Sequenzierung einen Verlust an Sequenzdaten von bis zu 30% zur Folge und die Fragmentlängenverteilung innerhalb dieser Probe wurde von 52bp auf 43bp reduziert (siehe Tabelle 4.29). Während der Datenanalyse besteht der Vorteil, dass Transitionen in der Auswertung berücksichtigt werden können, jedoch können Dekontaminationsverfahren, welche mit den Desaminierungsraten arbeiten, nicht angewendet werden.

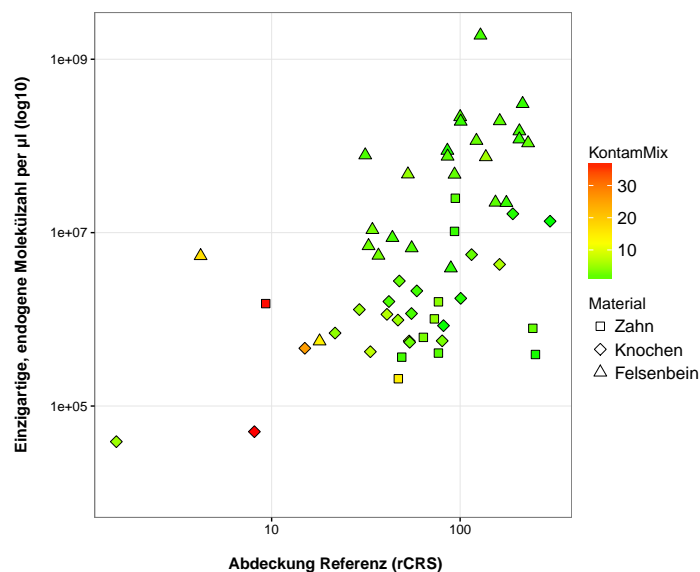


Abbildung 7.41: Darstellung des *Capture* Erfolgs durch die Relation der Abdeckung der rCRS mit der einzigartigen, endogenen Molekülzahl einer Probe (uniq). Die Skelettelement sind durch die Form der Symbole codiert und die Farbgebung entspricht der Kontaminationsrate ermittelt mit KontamMix. Beide Achsen sind logarithmisch skaliert.

Durch Quantifizierung des einzigartigen, endogenen DNA Gehalts konnte eine Vorauswahl der Proben erfolgen, welche durch optimierte Protokollabläufe hinsichtlich der mitochondrialen Genomsequenz angereichert wurden. Abbildung 7.41 verdeutlicht dieses Ergebnis: Proben mit geringen "uniq" Werten zeigen eine schwache Abdeckung der Referenzsequenz und sind anfällig für Kontaminationen.

¹ *Uracil-Specific Excision Reagent*

² Uracil DNA Glykosylase (UDG), DNA Glykosylase-Lyase Endonuklease VIII

Zwei der Proben (rote Farbgebung) zeigen ein Kontaminationslevel $>20\%$ - hier kann davon ausgegangen werden, dass lediglich eine exogene Kontamination sequenziert wurde (Proben Theo 3 und Fra 1 in Tabelle B.71), welche in der Vorabquantifizierung nicht erkannt werden kann. Im oberen, rechten Abschnitt des Plots ordnen sich Proben guter Qualität an. An dieser Stelle wird die Auswirkung von *Mark Duplicates* zur Entfernung der PCR Duplikate deutlich, da keine höhere Abdeckung als $\sim 300x$ erreicht werden kann. Es sind etwa 200.000 einzigartige, endogene Moleküle per μl *Fill In* Produkt notwendig, um ein authentisches, mitochondriales Genom zu sequenzieren (siehe Abschnitt 4.5).

Analyse der Desaminierungs und Fragmentierungsmuster alter DNA Moleküle

Während der Beurteilung der Probenqualität durch ein MiSeq *screening* konnte keine Korrelation zwischen dem Probenalter und der Degradierungsmuster innerhalb des Probenspektrums dieser Arbeit festgestellt werden (siehe Abbildung 4.28). Für eine erweiterte Untersuchung bezüglich dieser Fragestellung kann der Datensatz nach Anreicherung der mitochondrialen Genome herangezogen werden. Es werden Datensätze nach einer 100bp PE bzw. 150bp SE Sequenzierung genutzt, da jene das alte DNA Molekül in voller Länge repräsentieren. Aus den Datensätzen werden die durchschnittlich alignierte Fragmentlänge und Desaminierungsrate einer Probe extrahiert. Als Desaminierungsrate dient das 3'-Ende eines Moleküls, da an diesem Ende keine Nukleotide durch die T4 DNA Polymerase während des *Blunt-End repairs* abgeschnitten werden - es wird gewährleistet, dass das Molekül in Originalform verwendet wird¹. Beide Parameter der Degradierung werden in einem Biplot gegeneinander aufgetragen (Abbildung 7.42).

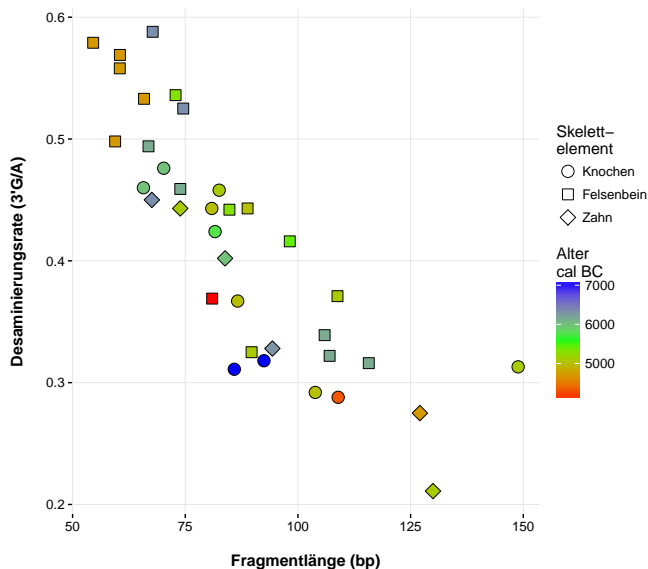


Abbildung 7.42: Analyse der Degradierungsmuster innerhalb des griechischen Probenspektrums. Gezeigt sind die Desaminierungsrate in Abhängigkeit ihrer durchschnittlichen Fragmentlänge. Die Form eines Datenpunktes gibt das Skelettelement an, die Farbgebung entspricht dem Alter einer Probe (kal. v. Chr.). Korrelationskoeffizient (nach *Pearson*): $r=-0.84$.

Diese Analyse verdeutlicht, dass beide Parameter der Degradierung negativ miteinander korrelieren (Korrelationskoeffizient nach *Pearson* $r=-0.84$). Je stärker ausgeprägt die Desaminierung, desto kürzer ist die Fragmentlänge innerhalb einer Probe. Zusätzlich verdeutlicht diese Abbildung die extrem geringe Fragmentlängenverteilung der DNA Moleküle aus manchen Felsenbeinen, welche mit Desaminierungsraten $>50\%$ zu den höchsten in der paläogenetischen Literatur zählen. Ähnliche Werte sind bisher nur in einem spät-paläolithischen Datensatz aus Europa [48], sowie in den Skelettelementen des *Homo heidelbergensis* [119] gefunden worden.

¹die 3'-Enden zeigten in allen Analysen stets eine höhere Rate an Desaminierungen.

Wird in diesem Zusammenhang der Faktor des Probenalters betrachtet, kann keine temporale Abhängigkeit der Degradierungsmuster festgestellt werden. In dieser Darstellung befinden sich die mesolithischen Proben mit vergleichsweise geringen Degradierungsmustern im unteren Drittel des Probenpektrums. Weitaus jüngere Proben aus dem späten Neolithikum (Felsenbeine, Mkg) zeigen annähernd eine Verdopplung der Desaminierungsfrequenz. Der biomolekulare Erhaltungszustand einer Probe ist in erster Linie abhängig von den Umgebungskonditionen der Grabstätte. Theopetra ist z.B. einer der wenigen Höhlen-Fundstellen, welche recht konstante Temperaturen während der Jahreszeiten bieten und die Proben vor Hitze schützen.

Ausblick

Die eingehende Charakterisierung des vorliegenden Probenmaterials zeigte, dass innerhalb des Projektes noch einige Proben guter Qualität vorhanden sind, welche für eine tiefere Sequenzierung des Genoms herangezogen werden können. Für zukünftige Projekte sollten die Proben aus Tabelle 7.52 berücksichtigt werden. Die Proben sind absteigend anhand ihrer einzigartigen, endogenen Molekülzahl (uniq) dargestellt.

Tabelle 7.52: Proben aus dem griechischen Neolithikum zur Sequenzierung des gesamten Genoms. Bereits sequenzierte Proben sind grau hinterlegt und dienen als Maßstab für zukünftige Experimente.

FragL: Fragmentlänge ermittelt durch Bioanalyzer Messung, endo= endogener DNA Gehalt, uniq=einzigartige, endogene Molekülzahl, mD5=Desaminierungsrate 5'-Ende

Probe	Fundstelle	Material	FragL	endo	uniq	mD5	KontamMix (%)
Klei 10	Kleitos	Felsenbein	74	59,9	1886850000	0,300	0,4 - 1,8
Krk 8	Toumba Kremasti koilada	Felsenbein	65	14,4	306294000	0,383	0,2 - 1,0
Pal 7	Paliambela	Felsenbein	70	36,9	216020225	0,430	0,1 - 2,5
St 5	Stavroupouli	Felsenbein	66	32,5	194071275	0,449	0,9 - 2,7
Klei 3	Kleitos	Felsenbein	73	28,5	191679675	0,350	0,1 - 1,2
Nea 2	Nea Nikomedia	Felsenbein	64	51,8	148233800	0,337	0,5 - 1,5
Nea 3	Nea Nikomedia	Felsenbein	64	56,7	119594800	0,345	0,2 - 1,1
Rev 4	Revenia	Felsenbein	76	13,5	115090000	0,480	0,9 - 3,6
Nea 5	Nea Nikomedia	Felsenbein	64	35,4	108000500	0,317	1,2 - 2,8
Nea 1	Nea Nikomedia	Felsenbein	60	12,3	88466700	0,518	0,0 - 0,5
Mkg 10	Makrygialos	Felsenbein	59	19,3	78320341	0,586	0,0 - 2,3
Rev 5	Revenia	Felsenbein	64	14,1	75743850	0,540	0,4 - 2,9
St 6	Stavroupouli	Felsenbein	67	4,9	74613750	0,504	4,0 - 6,6
Nea 4	Nea Nikomedia	Felsenbein	63	6,8	47192000	0,496	2,1 - 7,1
Rev 6	Revenia	Felsenbein	67	7,4	46926500	0,560	0,1 - 3,9
Xir 1	Xirolimni	Zahn	64	5,0	24924750	0,456	0,8 - 2,5
Krk 10	Toumba Kremasti koilada	Felsenbein	61	10,3	22307600	0,320	1,1 - 2,4
Thas 1	Thassos	Felsenbein	66	27,8	22120840	0,399	0,2 - 1,2
Tha 2	Tharrounia	Knochen	62	4,1	16482000	0,238	0,00 - 0,6
Krk 1	Toumba Kremasti koilada	Knochen	80	4,6	13545025	0,356	0,0 - 0,04

7.2 Die genetische Struktur im griechischen Mesolithikum

Die archäologische Beschreibung des griechischen Mesolithikums zeichnet ein mannigfaltiges Bild, welches eine Lebensweise beschreibt, die stark differenziert zu jener der JS Populationen Westeuropas ist. Dies ist zum einen durch die Art und Weise der Herstellung der Werkzeuge belegt [150], aber auch klimatische und geographische Begebenheiten bieten einen vielfältigen Lebensraum mit bewaldeten Gebieten im Norden und einer steppenartigen Ausprägung in den südlichen Teilen Griechenlands. Der direkte Anschluss an das Mittelmeer bietet des weiteren eine maritime Lebensgrundlage und die Beherrschung der Meere ist nachweislich aus dem griechischen Mesolithikum bekannt. Benachbarte Populationen aus dem Südosten, wie z.B. die Natufian Kultur der Levante, zeigen ab 10.000 v. Chr. technische und kulturelle Fortschritte, welche die Region des östlichen Mittelmeers und der Ägäis beeinflussten. Eine Darstellung der mesolithischen Lebensweise in der heutigen Region Griechenlands ist in Abschnitt 2.2 dieser Arbeit gegeben. Es stellt sich die Frage ob die zu den westlichen JS beschriebene Differenzen ebenfalls auf genetischer Ebene ersichtlich sind.

Aus einer prä-neolithischen Zeitstufe in Griechenland konnten zwei mitochondriale Genome isoliert werden, welche die mitochondriale Abstammungslinie K1 zeigen, eine Linie welche im Spektrum des zentral-europäischen Mesolithikums (WHG) bisher nicht identifiziert werden konnte. Der Großteil der erhobenen Daten aus der Kulturstufe des Paläolithikums/Mesolithikums in Zentraleuropa zeigen die Haplogruppe U5 [19, 143], so dass die gemeinsamen Vorfahren der WHG und der zwei Individuen der Theopetra Höhle vor der Abspaltung der WHG zu den Vorfahren der ersten Farmer zu suchen sind. Die ^{14}C Daten der Individuen umfassen eine Zeitspanne von 900 Jahren, ein Anzeichen, dass die Höhle von Theopetra kontinuierlich durch Individuen einer Gruppierung besucht wurde. Mit einer ^{14}C Datierung von 7600 v. Chr. gehört Theo 1 zu einer der ältesten Proben der Linie K1. Eine Probe aus dem präkeramischen Neolithikum Zentralanatoliens (Bon002, 8279-7977 v. Chr.) konnte ebenfalls der Linie K1 zugeschrieben werden [82]. Da eine parallele Entwicklung der Linie K1 auf beiden Seiten der Ägäis recht unwahrscheinlich ist, spricht dies für ein trans-ägäisches Netzwerk zur Zeit des Mesolithikums. Dies bestätigt die Sichtweise von Runnels und Van Andel (1995), welche einen prä-neolithischen Kontakt und Interaktion der Gesellschaften beider Seiten der Ägäis beschreiben [191].

Auffällig ist, dass im gesamten neolithischen Datensatz, welcher sich geographisch auf Nord-Griechenland (NGR-EN, NGR-LN) und die Ägäis (ANA-EN) bezieht, kein Individuum der Linie U5 zu finden ist. Auf der weiteren Verbreitungsrouten über den Balkan nach Zentraleuropa erscheinen im neolithischen Kontext vereinzelt die Linien U5 (siehe Abbildung 6.34), was auf eine Aufnahme von Individuen aus der JS Population spricht (1x in LBK-T [181], verstärkt in der LBK des Alfölds [81, 53]). Besonders deutlich ist dieses Muster in spät-neolithischen Gesellschaften Zentraleuropas, in welchen sich eine starke Zunahme an JS Linien beobachten lässt (Event B dargestellt in Brandt *et al.* (2013) [20]). Lässt man den Probennahmefehler außer Acht, stellt dies einen weiteren Hinweis dar, dass im griechischen Mesolithikum keine oder nur wenige U5 Linien vorhanden waren. Auf Ebene der genomischen Analysen der neolithischen Individuen Rev 5, Pal 7 und Klei 10 bestätigt sich dieses Szenario. In der f_4 Statistik der Form $f_4(\text{Ägäis}, \text{Ägäis}, \text{WHG}; \text{Khomani})$ ergeben sich keine signifikanten Abweichungen von 0 (siehe *Dataset S2*, #11, *Z-score* <3). Das bedeutet, dass in den Proben aus der Ägäis keine bedeutende JS Komponente (WHG) festgestellt werden kann. Dieses Ergebnis wird ebenfalls durch *ADMIXTURE* Analysen bestätigt. Interessanterweise kann aber eine gemeinsame genetische Drift mit den Individuen der JS aus dem Kaukasus (CHG) identifiziert werden ($f_4(\text{Ägäis}, \text{Ägäis}, \text{CHG}; \text{Khomani})$, siehe SI Tabelle 27 in [71]), ein Aspekt, welcher in folgenden Abschnitten vertieft werden soll.

Aktuelle Studien der JS Population aus der Region der "Iron Gates" zeigen in 22% der untersuchten Individuen die Haplogruppe K1 (neben 75% Haplogruppe U5) [113]. Diese Indizien lassen die Annahme einer strukturierten JS Population im Holozän zu. Mit der Rekolonisation Europas nach der Eiszeit dominierte Haplogruppe U5 [143], wobei die südlichsten Individuen dieser JS Population die Individuen aus Villabruna (12230-11830 v. Chr. [48]) und der adriatischen Insel Vela Spila (7308-7027 v. Chr. [181]) sowie jene der *Iron Gates* sind. Geographisch bilden das Balkan Gebirge und die Rhodopen eine Abgrenzung der westlichen JS Population, wobei die *Iron Gates* eine Kontaktzone zur JS Gesellschaft der Ägäis und des Nahen Ostens darstellen. Um das ägäische Mesolithikum im Detail definieren zu können, bedarf es weiterer genetischer Analysen mesolithischer Individuen aus der Ägäis.

7.3 Die genetische Struktur des griechischen Neolithikums

Mitochondriale Sequenzdatendaten des griechischen Neolithikums zeigen die Linien K1, H, U*, T1, T2, J1c und X, welche zuvor im Kontext des frühen Neolithikums NW Anatoliens (Barçın, Menteşe, [114, 71]), des Balkans (Starcévo Kultur [181]) und in Zentraleuropa (LBK [60]) beschrieben worden sind (Abbildung 6.34). Die früh-neolithische Populationen der Ägäis, des Balkans und Zentraleuropas zeigen eine hohe genetische Ähnlichkeit, was durch die erhobenen genetische Distanzen anhand des mitochondrialen Markers belegt wird (siehe F_{ST} -Werte in Tabelle 6.35a). Werden die Daten in Relation zur Ankunft des Neolithikums in den Regionen der Ägäis, des Balkan und Zentraleuropas gesetzt, dann kann aus der graphischen Darstellung der genetischen Distanzen eine Migrationsrichtung angegeben werden. Dargestellt in Abbildung 7.43 kann die Migrationsrichtung parallel zu Achse 1 der MDS abgebildet werden, ausgehend von der Population der Marmara Region (ANA-EN) in Richtung der ersten Farmer in Nordeuropa (NE-EN). In dieser Darstellung ordnen sich die Neolithiker Griechenlands (NGR) innerhalb der Gruppierung früh-neolithischer Populationen an, so dass das griechische Neolithikum aus Sicht des mitochondrialen Markers als Teil der demischen Expansionswelle des Neolithikums gesehen werden kann. Dies weist eindeutig eine autochthone Entstehung eines parallelen Neolithikums auf dem griechischen Festland zurück, obgleich die identifizierte mitochondriale Linie des Mesolithikums konsistent zu jenen der ersten Farmer ist [71].

Das Modell des griechischen Neolithikums beschreibt eine enge Verbindung nord-ägäischer Küstenregionen anhand neolithischer Artefakte [152]. Während des Neolithikums wird dieses Netzwerk über die gesamte Ägäis erweitert [72] und die Populationsdichte auf dem griechischen Festland nimmt zu, was anhand der Fundplatzdichte und -größe zu erkennen ist. Mitochondriale Daten belegen eine genetische Kontinuität der Population im griechischen Neolithikum über einen Zeitraum von 3000 Jahren. Die Analyse der genetischen Distanz zeigt einen geringen F_{ST} -Wert zwischen dem frühen und späten Neolithikum in Nord-Griechenland (0,011, nicht signifikant). Unter Anwendung eines Populationsmodells zur Simulation der genetischen Drift über den Zeitraum des griechischen Neolithikums kann die Kontinuität nicht widerlegt werden (siehe Abschnitt 6.4), da keine Aufnahme von signifikant differenzierten genetischen Material festgestellt werden kann. Wenn Migration über die Ägäis stattgefunden hat, dann aus einer benachbarten Population hoher genetischer Identität. Ein stetiger Genfluss aus Nachbarregionen impliziert eine Zunahme der genetischen Diversität, messbar durch die Haplotypendiversität innerhalb der neolithischen Population. Es ist lediglich eine schwache Zunahme zwischen dem frühen und späten Neolithikum zu verzeichnen (NGR-EN $\hat{H}=0.908 \pm 0,063$, NGR-LN $\hat{H}=0.911 \pm 0,030$, siehe Tabelle 6.49). Im anatolischen Frühneolithikum der Marmara Region konnte eine genetische Diversität von $\hat{H}=0,906 \pm 0,063$ 0,045 errechnet werden und die genetische Distanz zwischen beiden Seiten der Ägäis beträgt $F_{ST}=0,0117$ (ANA_EN-NGR_EN) und $F_{ST}=0,0004$ (ANA_EN-NGR_LN, beide nicht signifikant, siehe Tabelle 6.35a). Zusammengefasst beschreiben die

Analysen des mitochondrialen Markers eine genetisch homogene Population im ägäischen Neolithikum mit geringer genetischer Diversität. Eine geringe genetische Diversität kann als Hinweis für ein kürzlich erfolgtes *Bottleneck* mit anschließendem Populationswachstum herangezogen werden oder durch Gründereffekte während der Migration einer Population. In den berücksichtigten Populationen des griechischen und anatolischen Neolithikums sind 50% der genetischen Linien auf die Haplogruppe K und H zurückzuführen, welche eine geringe Anzahl variabler Positionen auf der HVS-I zeigen (Abbildung B.59). Als Erklärung für die geringe genetische Diversität des ägäischen Neolithikums soll das Kolonisationsmodell der südwest-anatolischen Küstenregion, erstellt durch Barbara Horeis (2015), herangezogen werden [72]. Dieses Modell beschreibt eine maritime Kolonisation der anatolischen Küstenregion ausgehend von östlichen Küstenregionen des Mittelmeers (Levante, Zentralanatolien), im Gegensatz zu einer Route aus dem Inland Anatoliens. Das Ergebnis ist eine zügige Besiedlung der Ägäis ab 6700 v. Chr. in SW Anatolien, jedoch impliziert der Seeweg eine begrenzte Transportkapazität [142]. Wenn nicht eine direkte Migration aus der Levante/Zentralanatolien nach Griechenland stattgefunden hat, sondern mehrere Stationen in der Ägäis eingelegt worden sind, ist es durch aufeinander folgende Gründereffekte zu einer Reduktion der genetischen Diversität innerhalb der Population der Seefahrerrekolonisten gekommen.

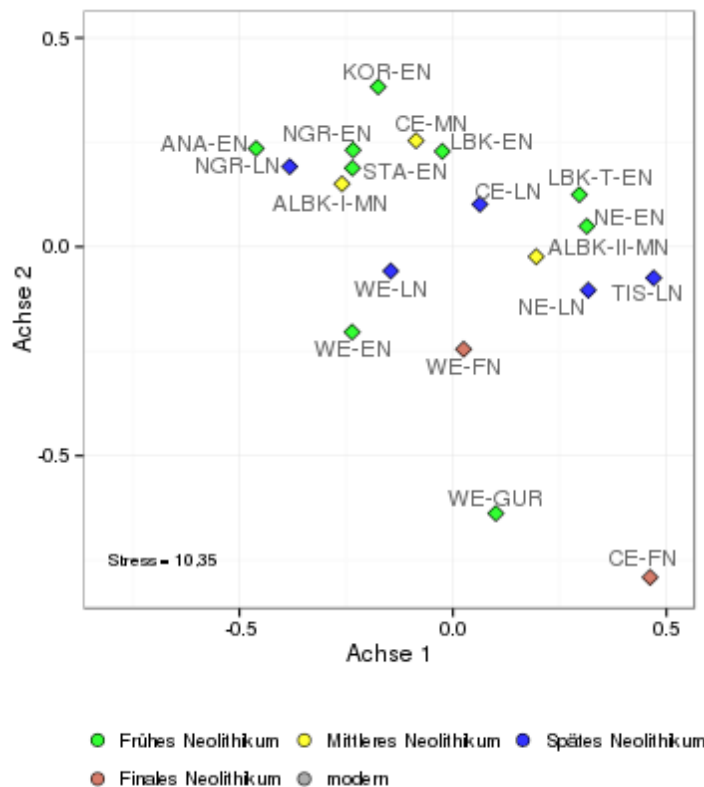


Abbildung 7.43: Die genetischen Distanzen zwischen den Populationen des Neolithikums nach multidimensionaler Skalierung.

Abkürzungen: ANA-EN= Frühes Neolithikum Türkei, NGR-EN=Frühes Neolithikum Griechenland, NGR-MN=Mittleres Neolithikum Griechenland, NGR-LN=Spätes Neolithikum Griechenland, STA-EN=Frühes Neolithikum Ungarn/Kroatien, KÖR-EN=Frühes Neolithikum Alföld (Ungarn), LBK-T-EN=Frühes Neolithikum Transdanubien (Ungarn), ALBK-I-MN=Mittleres Neolithikum Alföld I (Ungarn), TIS-LN=Spätes Neolithikum Alföld/ Tisza Kultur (Ungarn), LBK-EN=Frühes Neolithikum Mitteleuropa, CE-MN=Mittleres Neolithikum Mitteleuropa (Schöningen), CE-LN=Spätes Neolithikum Mitteleuropa (Baalberge, Salzminde), CE-FN=Finales Neolithikum Mitteleuropa (CWC, Glockenbecher), WE-EN=Frühes Neolithikum Iberia, WE-LN= Spätes Neolithikum Iberia, WE-FN=Finales Neolithikum Iberia, WE-GUR=Gurgy Population

Die Analysen auf genomischer Ebene bestätigen die genetische Ähnlichkeit der ersten Farmer aus Griechenland und NW Anatolien. Neben der PCA kann durch f -Statistiken verdeutlicht werden, dass diese Individuen einer Herkunft angehören und genetische Differenzen durch Drift erklärt werden können. Dies zeigen hohe Werte in der f_3 Statistik zwischen den Individuen aus Nord Griechenland und NW Anatolien (siehe Abbildung 7.44), sowie gleiche ancestrale Komponenten in den *ADMIXTURE* Analysen (siehe *Supplementary information*, S18 in [71]). Auf dieser Basis werden diese Proben seit der Veröffentlichung der Publikation Hofmanova, Kreutzer *et al.* (2016) in eine Population des

ägäischen Neolithikums eingegliedert. Damit belegen zum ersten Mal genetische Daten die Theorien aus der Archäologie dieser Region, welche eine enge Verknüpfung beider Seiten der östlichen Ägäis zeigen, da übereinstimmende Komponenten des neolithischen Paketes auf beiden Seiten geborgen werden konnten (siehe Abbildung 2.14). Die Migrationsrichtung kann auf genomischer Ebene durch die *ChromoPaint* Analyse belegt werden, welche in Abschnitt 7.6 der Diskussion besprochen werden soll. Damit sind Ansätze eigenständiger Domestikationsprozesse für den Norden Griechenlands ebenfalls durch genomische Daten widerlegt¹.

Für den Süden Griechenlands konnten in dieser Arbeit zwei mitochondriale Genome aus dem späten Neolithikum erstellt werden, welche die Haplogruppe H5 (Tharrounia) und U3b3 (Franchthi) zeigem. Zwei weiteren Studien aus diesem Jahr publizierten Daten für das späte Neolithikum aus dem Süden Griechenlands. Von der Fundstelle Diros (Alepotrypa Höhle) konnten vier Individuen analysiert werden, welche die Haplogruppen K1a26 [100], T1a, K1b1a und H [113] tragen, sowie ein weiteres Individuum aus der Bestattung von Franchthi mit der Haplogruppe H2. Die mitochondrialen Haplogruppen zeigen keine Abweichungen zu jenen aus dem Norden Griechenlands. Auf genomischer Ebene beschreiben Mathieson und Kollegen [113]² eine differenzierte Ursprungsregion für das südliche Neolithikum in Griechenland im Vergleich zum anatolischen Neolithikum. In einer Modellierung von Admixture Komponenten setzt sich das Neolithikum der Peloponnes aus 85,1% Herkunft aus dem anatolischen Neolithikum, 13,1 % CHG- und 1,8 % WHG-Herkunft zusammen (*Supplementary Table 3.2.3*). Die genomischen Daten aus dem griechischen Neolithikum aus dem Norden werden in der Darstellung der Ergebnisse dieser Studie nicht berücksichtigt, so dass keine Aussagen bezüglich der Migrationsrouten nach Griechenland und der Interaktion beider Regionen im späten Neolithikum getroffen werden können. Für eine detaillierte Darstellung der Prozesse während des ägäischen Neolithikums bietet sich die Probe Tha 2 für genomische Analysen an, da diese Probe einen endogenen DNA Gehalt von 4% hat (siehe Tabelle 7.52).

7.4 Das trans-ägäische Netzwerk und Konnektivität

Hinsichtlich der Fragestellung einer anhaltenden trans-ägäischen Konnektivität bietet die aktuelle genomische Analysen den Vorteil, dass beide Seiten der Ägäis durch Individuen des frühen (Barcin - Bar 31 - 6419-6238 kal. v. Chr., Revenia - Rev 5 - 6438-6264 kal. v. Chr.) und späten/finalen Neolithikums (Paliambela, Kleitos - Klei 10 - 4230-3995 kal. v. Chr., Kumtepe6) bzw. durch eine Probe aus dem Chalkolithikum (Kumtepe4 - 3500-2800 v. Chr.) abgedeckt sind [71]. Folglich kann ein Zeitraum von 3500 Jahren analysiert werden. Die Analyse der genomischen Daten zeigt nur minimale Veränderungen des genetischen Materials zwischen den neolithischen Proben der Ägäis über einen Zeitraum von 2500 Jahren (Bar 31 - Klei 10/Kumtepe6). In der PCA kann im griechischen Neolithikum keine Abgrenzung zwischen dem frühen und späten/finalen Neolithikum festgestellt werden, wie dies z.B. aus dem zentraleuropäischen Neolithikum bekannt ist [61]. Die *Outgroup-f₃* Statistik bestätigt die genetische Ähnlichkeit innerhalb der Ägäis über einen Zeitraum von 2500 Jahren, hervorgehoben durch einen hohen Wert an gemeinsamer genetischer Drift zwischen dem ältesten Individuum Rev5 und dem spät-neolithischen Individuum Kumtepe6 auf der gegenüberliegenden Küstenseite (siehe Abbildung 7.44).

F_4 -Statistiken der Form (NGR,NGR,früher Neolithiker Europa, Khomani) zeigen keine signifikante

¹Diese Ansätze gelten innerhalb der Archäologie für Thessalien und den Süden Griechenlands, so dass darauf hingewiesen werden soll, dass die Daten dieser Arbeit diese Region nicht abdecken.

²BioRxiv, Stand Oktober 2017

Abweichungen von 0, was darauf hinweist, dass nach der Abwanderung aus Griechenland keine Rückwanderung aus dem Westen stattgefunden hat. Wird diese Analyse innerhalb der Region der Ägäis durchgeführt (f_4 (Ägäis, Kumtepe, Ägäis, Khomani) siehe Tabelle S23 in [71]), werden für das Individuum Kumtepe6 signifikant negative Werte und für das Individuum Kumtepe4 signifikant positive Werte identifiziert. Dies beweist zum einen, dass Kumtepe4 genetisches Material aufgenommen hat, welches nicht aus dem ägäischen Genpool stammt, und zum anderen eine hohe genetische Übereinstimmung zwischen Kumtepe6 und den Individuen aus der Ägäis. Deutlich negative Werte sind mit der spät-neolithischen Probe Pal 7 (f_4 (Bar8,Kumtepe6,Pal7,Khomani), $Z\text{-score}=-6.821$) zu beobachten. Dies zeigt im späten Neolithikum eine starke Verknüpfung des Norden Griechenlands mit der anatolischen Küstenregion, welche zuvor schon durch Keramiken im Kumtepe-Stil auf einem makedonischen Fundplatz (Limenaria auf Thassos) angenommen wurde (siehe [131]). Der Fundplatz Paliambela (Pal 7) liegt nahe der Küste und könnte damit Teil eines Handelsnetzwerks innerhalb der östlichen Ägäis gewesen sein. Auf diesem Fundplatz sind erstmals mitochondriale Sequenzen der Linien U4 und U2 zu sehen, welche erhöhte Frequenzen in den modernen Populationen des Kaukasus zeigen (Referenz: Eupedia). Weitere U-Linien sind in Griechenland im späten Neolithikum auf Fundstellen aufgetreten (Mkg12 U7, Fra 8 U3), welche nahe an der Küste liegen.

Die zweite Probe, Kumtepe 4, wird einem post-neolithischen Kontext zugeordnet und zeigt die Aufnahme genetischen Materials, welches keine Ähnlichkeit zu den Farmern aus der Ägäis besitzt, so dass auf Genfluss aus einer anderen Richtung zu schliessen ist. Diese Annahme wird durch die Analyse der Form f_4 (Ägäis, Kumtepe, CHG, Khomani) mit signifikant negativen Werten für die Probe Kumtepe6 bestätigt, was auf Genfluss aus dem Kaukasus hinweist (siehe Tabelle S25 in [71]). Dieser Effekt wird ebenfalls in der f_3 -Statistik deutlich (Abbildung 7.44), in welcher die spät-neolithischen Individuen aus Griechenland eine größere Ähnlichkeit zu den CHG Individuen (SATP-Satusurblia, KK1-Kotias) im Vergleich zu Rev 5 zeigen.

In den vorherigen Abschnitten der Diskussion konnte gezeigt werden, dass (1.) ein trans-ägäisches Netzwerk im Mesolithikum etabliert und für vereinzelte Migrationen genutzt wurde. Das bestehende Netzwerk wurde (2.) für die Verbreitung des Neolithikums in der Ägäis genutzt. Ab 6700 v. Chr. entstehen erste Pioniersiedlungen an der westanatolische Küstenregion (Fundplatz Cukurici Höyü) [72] und etwa zeitgleich kommt es zu ersten neolithischen Siedlungen in Makedonien [87]. Ab dem fortschreitenden Neolithikum kann auf beiden Seiten der Küste melischer Obsidian als Material für die Werkzeugproduktion nachgewiesen werden. Die vorliegenden Analysen [71] bestätigen die Konnektivität der ägäischen Küstenregionen auf genetischer Basis über einen Zeitraum von 2500 Jahren. Es kann eine genetisch homogene Population beschrieben werden, welche sich isoliert zu den westlichen Teilen des Neolithikums aus Europa verhält, da kein Genfluss aus dem Nordwesten durch Rückmigration identifiziert werden kann. Zum Ende des Neolithikums hingegen kann Genfluss aus dem Osten wahrgenommen werden. Dies zeigt, dass nach Ankunft der ersten Farmer in der Ägäis differenzierte demographische Prozesse im Vergleich zum fortschreitenden Neolithikum in Zentraleuropa stattgefunden haben, welche im Detail im folgenden Abschnitt besprochen werden sollen.

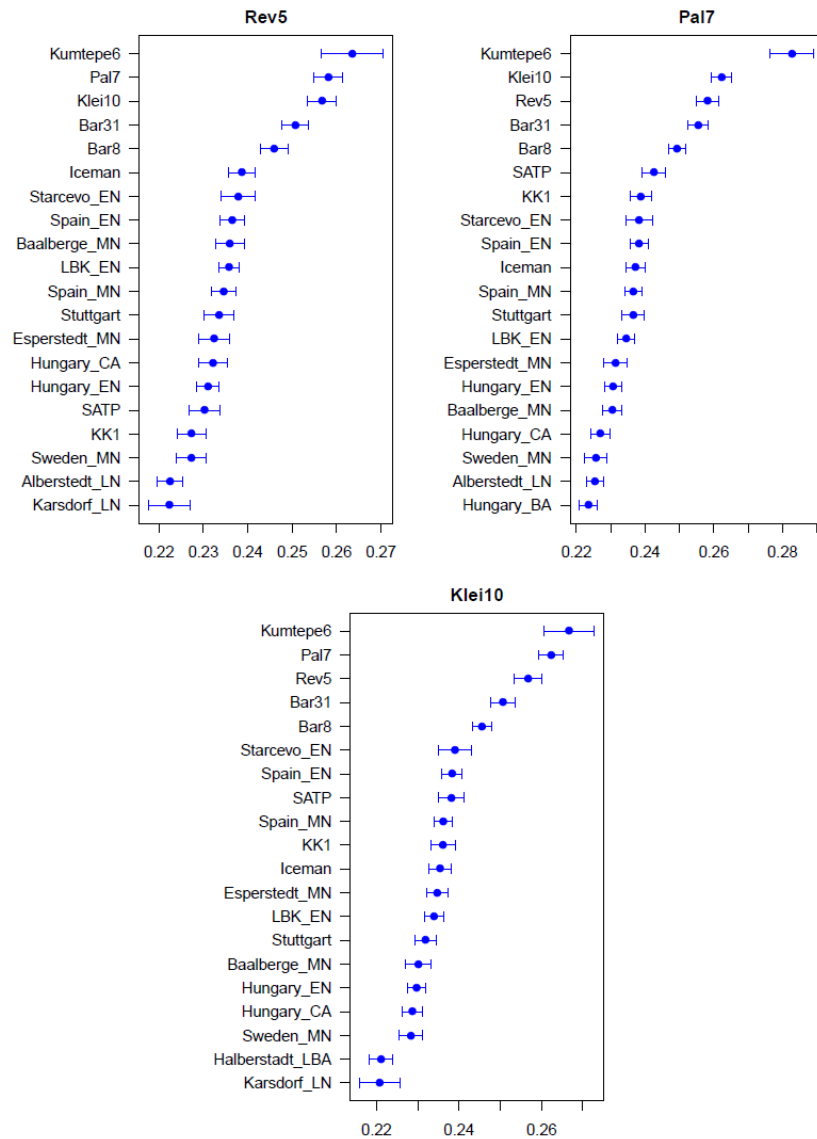


Abbildung 7.44: Darstellung der genetisch ähnlichsten prähistorischen Populationen anhand f_3 (Khomani; prähistorische Population, Neolithikum Griechenland). Entnommen aus [71] *Supplements*, SI 7.

7.5 Das ägäische Neolithikum im überregionalen Vergleich

Die Populationen des ägäischen Neolithikums wurden in dieser Arbeit im Rahmen eines prähistorischen Datensatzes mit 1278 Individuen eingegliedert in 30 Populationen anhand der mitochondrialen HVS-I Region analysiert (siehe Tabelle 1.1). Es konnte gezeigt werden, dass das ägäische Neolithikum keine signifikanten Distanzen zu den ersten Farmern des Balkan und der LBK aufweist, jedoch extreme Werte zu den JS des Holozäns ($F_{ST} (NGR-EN - Holo-JS)=0.49$, siehe Abschnitt 6.3). Dies wird in der multidimensionalen Skalierung der Distanzen des gesamten Datensatz wiedergegeben (Abbildung 7.45). Die frühen Neolithiker sind weitläufig im linken Abschnitt des Plots platziert, während die JS aus dem Holozän auf der gegenüberliegenden Seite des Plots abgebildet werden. Diese starke Differenzierung basiert auf dem eiszeitlichen *bottleneck* mit anschließenden Populationswachstum, so dass auf einem Diversitätsverlust agierende genetische Drift zu einer im Gesamtbild des prähistorischen Datensatzes stark separierten Teilpopulationen der JS aus dem Holozän führte [143]. Im Vergleich zur Analyse auf Grundlage der neolithischen Populationen (Abbildung 7.43) wird das griechische Neolithikum ebenfalls im Bereich der ersten Farmer aus Anatolien (ANA-EN) und jener des Balkans (STA-EN) platziert, jedoch erscheint das griechische Früneolithikum etwas separiert. Das griechische Spätneolithikum hingegen wird nahe zur Starčevo Population abgebildet. Aus archäologischer Sicht wurde bereits eine enge Verknüpfung des späten Neolithikums (Makryalos [88]) in Griechenland mit der Vinca Kultur festgestellt, welche auf Basis der Gestaltung der Keramiken basiert.

Der in dieser Arbeit dargestellte Datensatz beinhaltet erstmalig frühe Neolithiker, welche zum Beginn der Migrationsroute stehen (NGR, ANA) und zusätzlich konnten weitere neolithische Populationen der danubischen (Alföld LBK) und mediterranen (Iberia) Route hinzugefügt werden. Dies bietet die Möglichkeit die Geschichte der Kolonisation Europas durch den mitochondrialen Marker im Detail darzustellen. Neben der klaren Separation der JS zu den ersten Farmern Europas kann ebenfalls die in Abschnitt 7.3 beschriebene Migrationsrichtung ausgehend vom ägäischen Neolithikum dargestellt werden (Abbildung 7.45). Des weiteren bestätigt diese Analyse die Darstellung der Besiedlungsgeschichte durch Brandt und Kollegen (2013) [20], jedoch bietet eine feinere Aufgliederung des fortschreitenden Neolithikums in die Gruppen mittleres (4600 - 4000 v. Chr.), spätes (4000 - 2800 v. Chr.) und finales (2800 - 2200 v. Chr.) Neolithikum die Möglichkeit, einzelne demographische Ereignisse separat zu erörtern. Während das mittlere Neolithikum eine geringe Distanz zur LBK aufweist (nicht signifikant), können ab dem späten Neolithikum signifikante Distanzen bestimmt werden (siehe Tabelle 6.37). In der graphischen Darstellung der Distanzen kann gezeigt werden, dass sich die Populationen des fortschreitenden Neolithikums in Europa (geographisch und chronologisch) in ihrer genetischen Komposition der heutigen europäischen Bevölkerung annähern. Dies impliziert Populationsumschwünge nach der Ankunft des Neolithikums [61] durch Paarungsereignisse im späten Neolithikum (CE-LN) mit den Jäger-Sammler Westeuropas, sowie im finalen Neolithikum (CE-FN) nach Einwanderung der Yamnaya Population aus der Steppe (F_{ST} -Wert siehe Abschnitt 6.3). Eine im Rahmen dieser Arbeit durchgeführte räumlich-explizite Simulation der neolithischen Expansion anhand ausgewählter, mitochondrialer Referenzdaten zeigt¹, dass eine Aufnahme des genetischen Materials des WHG nach der Ankunft des Neolithikums in Zentraleuropa und auf dem Balkan stattgefunden hat. In Griechenland sind jedoch keine Anzeichen dafür zu finden. Der Kontinuitätstest, basierend auf dem mitochondrialen Datensatz, zeigt keine signifikanten Hinweise für einen Populationsumschwung zwischen dem frühen und späten Neolithikum in Griechenland, welche eindeutig für die verschiedenen Epochen des Neoli-

¹Diese Analyse wurde durch Kollegen der Universität Genf, Nuno Da Silva und Mathias Currat, anhand des Programmes SPLATCHE2 durchgeführt.

thikums in Zentraleuropa nachgewiesen werden können (siehe Abschnitt 6.4).

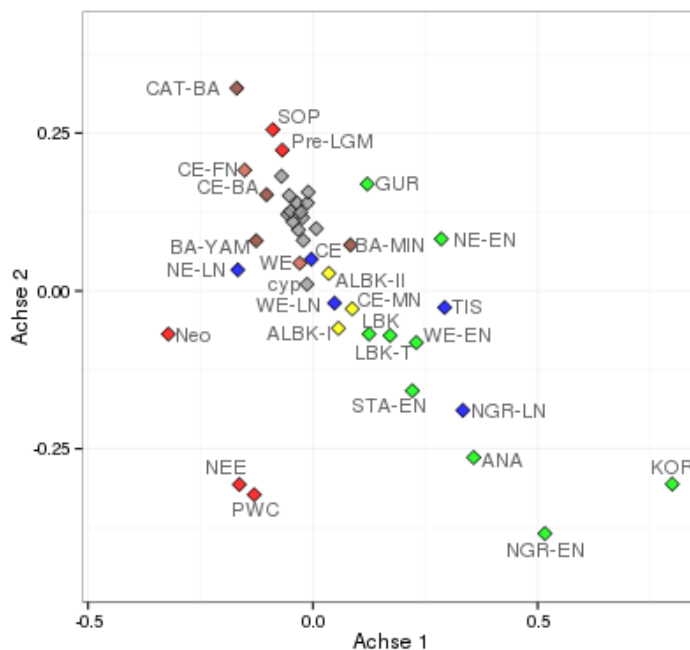


Abbildung 7.45: Die genetischen Distanzen zwischen den Populationen des prähistorischen Datensatzes Zentraleuropas, des Balkans, Iberias und der Ägäis nach multidimensionaler Skalierung.

Abkürzungen: Pre-LGM=Oberes Paläolithikum, SOP=Spätes oberes Paläolithikum (post-LGM), NEE=Präneolithisch Nord-Ost Europa, Neo=Paraneolithisch südliches Baltikum, PWC=Paraneolithisch Skandinavien, ANA= Frühes Neolithikum Türkei, NGR-EN=Frühes Neolithikum Griechenland, NGR-LN=Spätes Neolithikum Griechenland, STA-EN=Frühes Neolithikum Ungarn/Kroatien, KOR=Frühes Neolithikum Alföld (Ungarn), LBK-T=Frühes Neolithikum Transdanubien (Ungarn), ALBK-I=Mittleres Neolithikum Alföld I (Ungarn), TIS=Spätes Neolithikum Alföld/ Tisza Kultur (Ungarn), LBK=Frühes Neolithikum Mitteleuropa, CE-MN=Mittleres Neolithikum Mitteleuropa (Rössen, Schöningen), CE=Spätes Neolithikum Mitteleuropa (Baalberge, Salzmünde), CE-FN=Finiales Neolithikum Mitteleuropa (CWC, Glockenbecher), NE-EN/FN=Frühes/Spätes Neolithikum Nordeuropa, MIN-BA=Bronze Zeit Kreta (Minoer)

Die Aufnahme von JS Genen in den neolithischen Genpool konnte auch im südosteuropäischen Neolithikum entlang der danubischen Ausbreitungsrouten und der LBK des Alfölds gezeigt werden (siehe Abbildung 6.34, Zunahme der Linie U5). Ein direkter Beweis stellt ein Individuum dar (KO 1, 5780-5650 kal. v. Chr. [53]), dessen Genmaterial zu 100% mesolithischen Ursprungs ist (bestätigt durch PCA, f_4 Statistik und *Admixture* Analysen in [53, 71]), jedoch im früh-neolithischen Körös Kontext geborgen wurde. In dem Gebiet des ungarischen Alfölds ist dieses Ergebnis auf mesolithische Populationen aus den umgebenden Karpathen zurückzuführen [81]. Nach Aufnahme einzelner Individuen in die Farmerpopulation können durch Rekombinationsereignisse in den folgenden Stufen des Neolithikums des Alfölds JS Komponenten in neolithischen Genomen nachgewiesen werden (NE1, NE3, NE4 [53]; siehe auch *Admixture* Analyse in [71], SI 8).

Die Quantifizierung der geteilten genetischen Drift zeigt einen hohen Anteil genetischer Übereinstimmung zwischen den Individuen der Ägäis, wobei das Starčevo Individuum einen der genetisch ähnlichsten Individuen außerhalb der Ägäis darstellt (*Outgroup- f_3* Statistik, Abbildung 7.44). Analysen der Proben Rev 5, Pal 7 und Klei 10 zeigen neben der Starčevo Probe eine hohe Ähnlichkeit zu einer früh-neolithischen Probe aus Spanien (siehe Abbildung 7.44, Spain_EN, Els Trocs, 5300-5200 kal. v. Chr.). Dies bestätigt die Theorie der mediterranen Verbreitungswelle des Neolithikums im Zusammenhang mit der Impresso Kultur über die Adria nach Italien und weiterführend nach Frankreich und Spanien. In diesem Zuge sind die italienischen Alpen besiedelt worden - hier stellt das Genom des "Ötzi" ein Relikt dieser Expansionswelle dar. Datiert auf etwa 3300 vor Chr. zeigt sein Genom keine Signaturen von post-neolithischen *Admixture* Ereignissen, worauf die hohe genetische Ähnlichkeit zum griechischen Neolithikum zurückzuführen ist (hervorgehoben im früh-neolithischen Individuum Rev 5, Abbildung 7.44).

In der Publikation Hofmanova, Kreutzer *et al.* (2016) konnte eindeutig gezeigt werden, dass die Migration der ersten Farmer nach Europa auf Nordgriechenland und Anatolien zurückgeführt werden kann. Jedoch beschreibt dies nur den Anfang der Migration nach Europa und es fehlt der genetische

Link zur Kernzone des Neolithikums. Genetische Daten aus dem iranischen Neolithikum (Broushaki *et al.* (2016)) [24] und aus dem Epipaläolithikum und Neolithikum der Levante (Lazaridis *et al.* (2016) [101]) hingegen decken weiträumig den Entstehungsort der neolithischen Kultur ab. Diese Studien zeigen eine ausgeprägte Populationsstruktur in der Kernzone des Neolithikums. Das iranische Neolithikum zeigt eine hohe genetische Ähnlichkeit zu den JS des Kaukasus (CHG) und bildet mit jenen eine Gruppe in der PCA, separiert zu den ägäischen Farmern [24]. Die Neolithiker der Levante und Jordanien bilden eine weitere, genetische differenzierte Gruppe. In dieser Region kann eine genetische Kontinuität zwischen den JS des Epipaläolithikums zu den ersten Farmern gezeigt werden [101]. In einzelne Individuen der Bronzezeit aus Anatolien und der Levante konnte eine Mixtur beider früh-neolithischen Komponenten festgestellt werden, was auf Populationsbewegungen im fortschreitenden Neolithikum hinweist [101]. In diesem Kontext ist es auch zu Genfluss in die Ägäis gekommen, welcher in den Individuen des ägäischen Spätneolithikums in Form von CHG Herkunft messbar ist [71, 24].

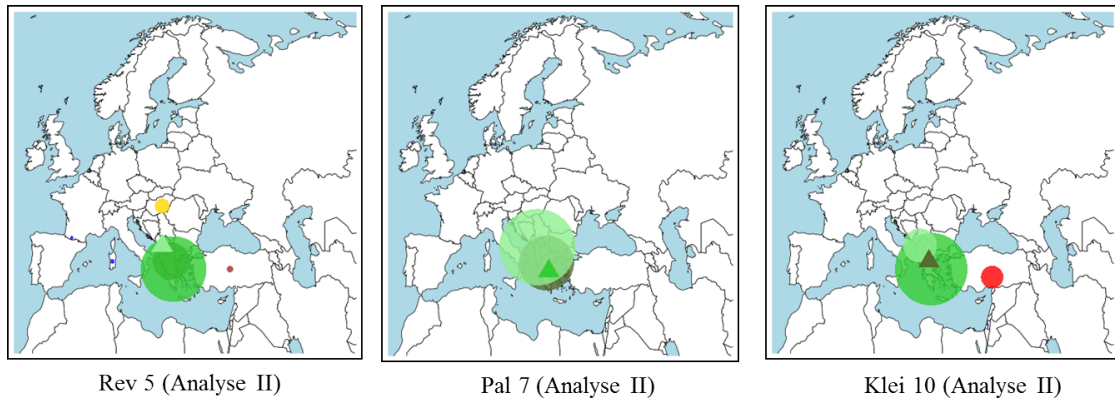
7.6 Die Rolle Griechenlands in der Verbreitung des Neolithikums

Durch PCA, f_4 Statistik und *ADMIXTURE* Analysen konnte gezeigt werden, dass die Individuen des ägäischen Neolithikums genetisch sehr ähnlich sind, jedoch sind diese Methoden limitiert, so dass der Einfluss der ägäischen Individuen auf die Populationen der danubischen und mediterranen Ausbreitungsrouten nicht prozentual quantifizierbar ist. Anhand des mitochondrialen Markers können vom frühen Neolithikum in Griechenland im Vergleich zum türkischen Neolithikum geringere genetische Distanzen zu den Gruppen des iberischen Neolithikums berechnet werden (siehe Tabelle 6.37), jedoch sind diese Werte nicht aussagekräftig (p -Wert >0.05). Auf genomischer Ebene kann anhand der *Chromosome painting* Analyse gezeigt werden, dass das genetische Material der ägäischen Individuen in unterschiedlichen Ausmaß nach Europa transferiert wurde. Diese Analyse benötigt Datensätze mit einer Abdeckung des humanen Genoms von mind. 1x, so dass nur Individuen aus *Shotgun* Projekten berücksichtigt werden können: Bar 31 (6419 – 6238), Bar 8 (6212 – 6030), Stuttgart (5100 - 4800 v. Chr., LBK), Loschbour (6220 - 5990 v. Chr., Holo-JS) [102], NE1 (5310 - 5070 v. Chr., ALBK-MN-I), BR2 (1270 - 1110 v. Chr., Bronzezeit) [53] und moderne Daten aus Busby *et al.* (2015) [28]. Diese Analyse ist stark abhängig von den Daten, welche zur Modellierung der Donorkomponenten genutzt werden¹.

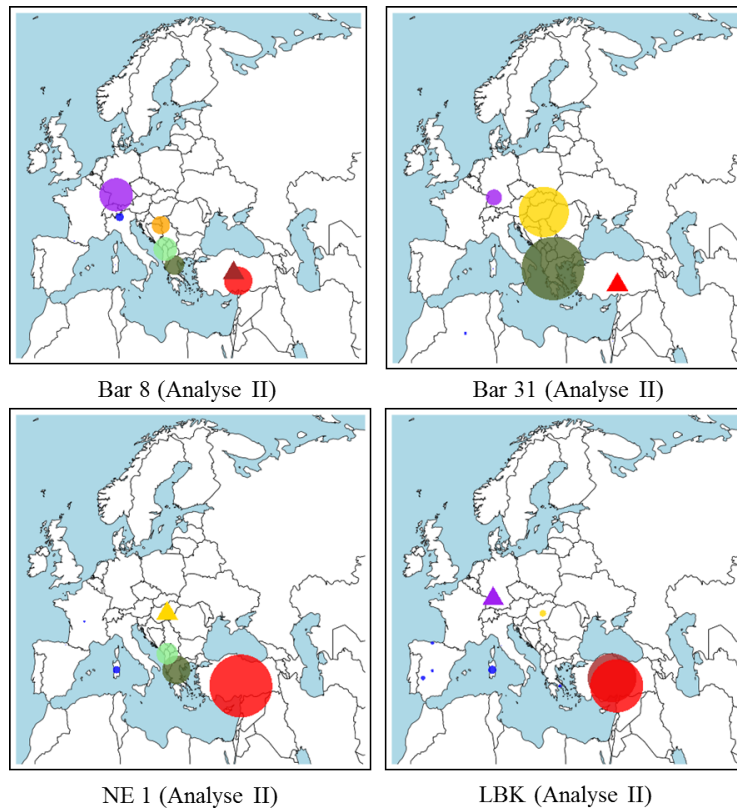
Abbildung 7.46a zeigt die Analyse der griechischen Individuen Rev 5 (6438 - 6264 kal. v. Chr.), Pal 7 (4452 - 4350 kal. v. Chr.) und Klei 10 (4230 - 3995 kal. v. Chr.) als Empfänger im Kontext der modernen und prähistorischen Genome als Donorkomponenten (Analyse II aus Hofmanova, Kreutzer (2016), SI 10). Die griechischen Individuen kopieren jeweils sehr stark voneinander, so dass die Signale der anderen Donoren kaum auszumachen sind. Lediglich die Probe Rev 5 zeigt Übereinstimmungen mit der neolithischen Probe des Balkan (NE1, gelb), was auf der Weitergabe des genetischen Materials innerhalb der neolithischen Expansion zurückzuführen ist. Dieses Signal kann in den spät-neolithischen Genomen Griechenlands nicht identifiziert werden. Hier kann in der Probe Klei 10 ein Donorsignal ausgehend aus dem früh-neolithischen Genom Bar 31 (rot) gesehen werden. Die selbige Analyse der anatolischen Individuen Bar 8 & Bar 31 zeigt hingegen deutliche Donorsignale auf der danubischen Ausbreitungsrouten, ausgehend von den Genomen der Individuen Rev 5, Klei 10, NE1 und LBK (siehe

1

Analyse I - moderne Daten als Donoren
 Analyse II - moderne Daten + aDNA Genome als Donoren
 Analyse III - aDNA Genome + Yoruba
 Analyse IV - aDNA Genome (ohne BR2) + Yoruba



(a) Donorkomponenten in den Empfängerindividuen Rev 5, Pal 7 und Klei 10 aus dem frühen und späten/finalen Neolithikum im Norden Griechenlands.



(b) Donorkomponenten in den Empfängerindividuen Bar 8 und Bar 31 aus dem frühen Neolithikum in NW Anatolien sowie den ersten Farmern der danubischen Ausbreitungsroute NE 1 und LBK.

Abbildung 7.46: Donorkomponenten innerhalb der neolithischen Individuen Anatoliens, des Balkans und Zentraleuropas (Analyse II). Abbildung entnommen aus [71], SI 10, Abbildung S21 - S24. Farbkodierung: Rev 5 = hellgrün, Pal 7 = grün, Klei 10 = olivgrün, NE1 = gelb, Bar 31 = rot, Bar 8 = dunkelrot, LBK = lila.

Abbildung 7.46b). Zusammengenommen bedeutet dies, dass die Ausbreitung des Neolithikums verstärkt aus Anatolien stattgefunden hat und Griechenland zwar als Transit genutzt wurde, jedoch das genetische Material recht verdünnt im zentraleuropäischen Neolithikum angekommen ist. An dieser Dilution können eine verminderte Transportkapazität über die Ägäis, sowie die Ausbreitung in zwei Richtungen (mediterranean und danubisch) beteiligt gewesen sein.

Anhand des *Chromosome paintings* kann die Richtung des Genflusses identifiziert werden, da chronologisch später auftretende Genome beim Vergleich invertierter Analysen immer zu einem geringeren Prozentsatz von ihren jüngeren Donoren kopieren. Besonders deutlich wird dies bei dem Vergleich der Genome von Bar 31 und LBK (Analyse IV, siehe Tabelle 7.53). Während in dem LBK Genom 50% der Gene auf Bar 31 zurückzuführen sind, kann in dem Genom von Bar 31 nur 22% LBK Herkunft nachgewiesen werden. Dies stellte in der Publikation Hofmanova, Kreutzer *et al.* (2016) einen der stärksten Beweise für eine Expansionswelle dar, welche in Anatolien startete und auf diesem Wege das genetische Material des Neolithikums nach Zentraleuropa gebracht hat [71]. Anhand der Kalkulation der Donorkomponenten kann die genetische Ähnlichkeit zwischen den Neolithikern aus Griechenland quantifiziert werden. Wie in Tabelle 7.53 dargestellt, kann in der Probe Rev 5 75% genetisch identisches Material zu seinen Nachfahren Pal 7 (54,72%) und Klei 10 (20,26%) identifiziert werden. Die 2000 Jahre jüngeren Genome von Pal 7 und Klei 10 zeigen respektive 98% und 80% der Komponenten aus den jeweilig anderen Genome aus Griechenland. Dies bestätigt die genetische Kontinuität im Neolithikum Griechenlands über einen Zeitraum von 2500 Jahren ohne massive Populationsbewegungen sowie signifikante WHG *Admixture*¹ auf Basis einer weiteren Statistik. Eintreffender Genfluss im späten Neolithikum kann nicht wiedergegeben werden, da die Daten der CHG JS, des iranischen Neolithikums und den Kumtepe Individuen in dieser Analyse nicht beteiligt sind.

Tabelle 7.53: Kalkulation der Donorkomponenten in den Empfängergenomen aus Analyse IV (Auszug aus Dataset S4 aus [71]).

Empf.	Donoren							
	Loschbour	LBK	Bar31	Bar8	Klei10	Pal7	Rev5	NE1
Loschbour	0	0	0	19.81	0	78.28	0	0
LBK	7.99	0	50.43	14.43	4.01	21.36	0	0
Bar31	0	22.01	0	0	38.37	0	0	38.84
Bar8	5.63	8.31	7.36	0	34.23	24.62	18.65	0
Klei10	0	0	20.42	0	0	52.61	26.98	0
Pal7	2.64	0	0	0	38.82	0	58.54	0
Rev5	0	0	0	13.04	20.26	54.72	0	11.98
BR2	45.57	0	0	17.55	36.68	0	0	0
NE1	8.54	0	52.75	0	20.87	0	17.82	0

Hinsichtlich der Fragestellung der Ausbreitungsrouten und des genetischen Einflusses sollen die Donorkomponenten des Individuums NE1 (Alföld LBK) anhand Analyse IV betrachtet werden (Tabelle 7.53, Abbildung 7.46b): Es setzt sich aus den Donoren Bar 31 (53%), Rev 5 (18%) und Klei 10 (21%) zusammen, jedoch sind anteilig nur 38% NE1 Herkunft in dem Genom von Bar 31 nachweisbar (NE1 Komponente in Rev 5 ist 12% und 0% in Klei 10). Dies belegt, dass die danubische Ausbreitungsrichtung stärker durch das türkische Neolithikum geprägt ist. In diesem Zusammenhang wäre die Analyse eines frühen Farmers aus dem mediterranen Raum interessant. Ein Indiz für eine Verbreitung des griechischen Neolithikums über den Mittelmeerraum zeigt Analyse AI, bei welcher nur moderne

¹ Lediglich in der Probe Pal 7 können 2,64% Übereinstimmung mit Loschbour verzeichnet werden.

Populationen als Donoren genutzt wurden (siehe Abbildung 7.47 und Tabelle 7.54). Die griechischen Neolithiker empfangen >81% aus den Donoren der mediterranen Populationen (Italien, Sardinien, Spanien, Südfrankreich), während in den Genomen der türkischen Neolithiker <71% gemessen werden kann.

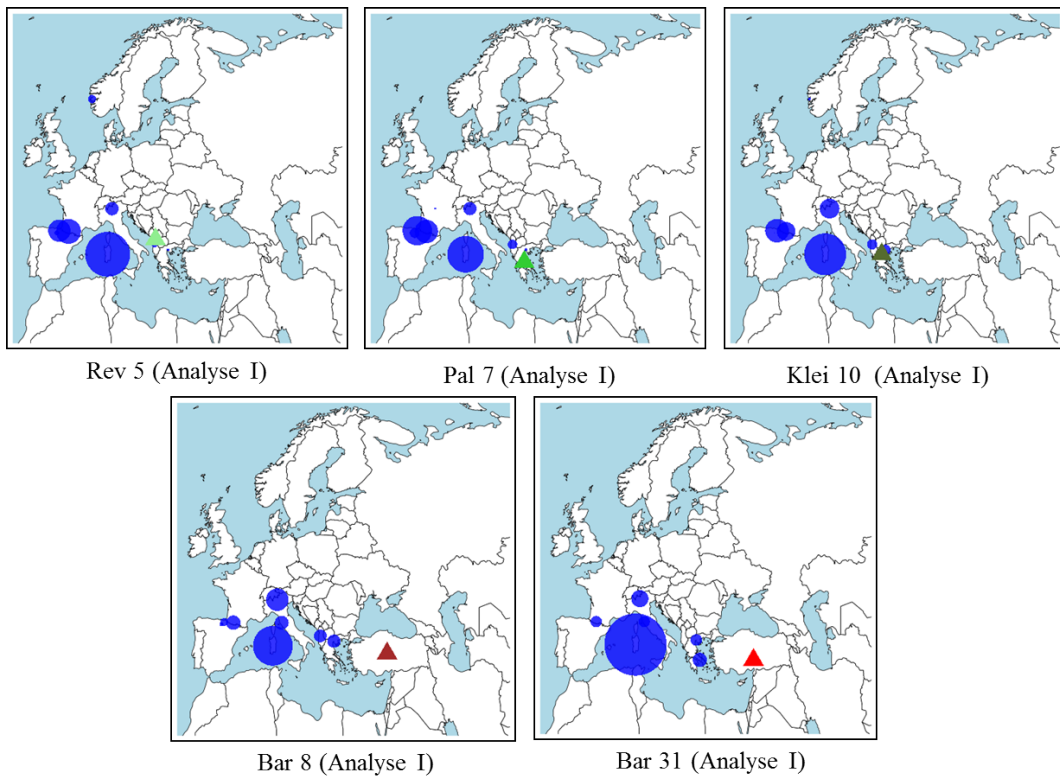


Abbildung 7.47: Donorkomponenten moderner europäischer Populationen innerhalb der neolithischen Individuen Griechenlands und Anatoliens (Analyse I). Abbildung entnommen aus [71], SI 10, Abbildung S21 & S22.

Tabelle 7.54: Herkunftsspezifische Donorkomponenten (modern, Analyse I) in den prähistorischen Empfänger Individuen. Entspricht Tabelle S30 aus SI 10, Hofmanova, Kreutzer *et al.* (2016).

Donoren	Empfänger								
	Loschbour	LBK	Bar31	Bar8	Klei10	Pal7	Rev5	BR2	NE1
<i>Südeuropa</i>	0	71.14	71.05	68.98	81.92	88.67	86.71	0	86.89
<i>Sardinien</i>	0	39.54	48.7	31	32.97	28.84	35.36	0	40.5
<i>Nord-/Zentraleuropa</i>	63.04	0	0	0	2.2	0	6.41	69.48	0
<i>Levante</i>	0	0	0	0	0	0	0	0	0
<i>Kaukasus</i>	0	0	0	0	0	0	0	6.66	0
<i>andere</i>	36.96	28.86	28.95	31.02	15.88	11.33	6.88	23.86	13.11
<i>Griechenland</i>	0	17.19	20.15	10.23	7.78	2.37	2.09	0	6.23

7.7 Das griechische Neolithikum im Kontext der heutigen europäischen Bevölkerung

Auf mitochondrialer Ebene konnte eine genetische Distanz von $F_{ST} = 0.04$ (signifikant) zwischen der Gruppierung des frühen und späten Neolithikums in Griechenland zur heutigen Population in Nordgriechenland berechnet werden. Innerhalb der multidimensionalen Skalierung der genetischen Distanzen kann eine klare Separation beider Zeitstufen gezeigt werden (siehe Abbildung 7.45). Dies verdeutlicht auf genomischer Ebene die PCA, in welcher das griechische Neolithikum deutlich separiert zur modernen griechischen Bevölkerung dargestellt wird (siehe Anhang B.7, Abbildung 2), aber auch die f_3 -Statistik, in welcher ein stärkerer Wert an gemeinsamer genetischer Drift zu mediterranen Populationen (Sardinien, Italien, Albanien, Spanien) im Vergleich zur modernen griechischen Population gemessen werden kann (siehe Abbildung 7.48). In den *Chromosome painting* Analysen kann im Vergleich zu den mediterranen Populationen ebenfalls ein reduziertes Donorsignal in den neolithischen Empfängerindividuen Rev 5, Pal 7 und Klei 10 dargestellt werden (siehe Tabelle 7.54). Mit der modernen Population aus Griechenland stimmen nur 2-7% der Allele mit den griechischen Neolithikern überein, während >80% übereinstimmende Allele aus den modernen Populationen Südeuropas (Italien, Sardinien, Spanien, Südfrankreich) stammen. Ziel der Analyse prähistorischer Daten im Vergleich zu einer regionalen, modernen Population ist der Nachweis der Kontinuität des genetischen Materials zwischen den Populationen beider Zeitstufen. Ist diese vorhanden kann daraus geschlossen werden, dass keine signifikanten demographischen Prozesse innerhalb dieser Region stattgefunden haben. Diskontinuität kann z.B. durch Einwanderung einer Population differenzierten Ursprungs hervorgerufen werden, welche die regionale Population ersetzt bzw. mit jener genetisches Material austauscht. Diese Analyse wurde separat auf mitochondrialer und genomischer Ebene durchgeführt.

Auf mitochondrialer Ebene konnten 49 Proben, eingeordnet in zwei neolithischen Zeitstufen, untersucht werden. Mit der Hypothese einer anhaltenden Kontinuität des genetischen Materials zwischen den Populationen beider Zeitstufen wurde anhand des mitochondrialen Markers eine Simulation durchgeführt, bei welcher die genetische Drift als einzig wirkende Kraft über die Zeit simuliert wurde (Durchführung und Ergebnis siehe Abschnitt 6.4). Die Populationskontinuität konnte nur innerhalb eines panmiktischen Modells widerlegt werden (Szenario 4, Abschnitt 6.4). Wird eine Strukturierung der Population durch die räumliche Komponente der Individuen einbezogen, kann die Kontinuität nicht widerlegt werden (Szenario 1-3). Dies bedeutet, dass die zuvor berechneten genetischen Distanzen auf genetische Drift zurückzuführen sind und in den letzten 8000 Jahren keine signifikanten Prozesse wie Migration und *Admixture* innerhalb der Region des heutigen Griechenlands stattgefunden haben. Da dieses Ergebnis auf der HVS-I Region (348bp) basiert, könnte eine geringe genetische Variation zwischen der lokalen griechischen Population und einer einwandernden Population mit geringer Divergenz dazu führen, dass die genetische Kontinuität nicht mit statistischer Signifikanz widerlegt werden kann.

Im Rahmen der Publikation Hofmanova, Kreutzer (2016) wurde die Populationskontinuität auf Basis der vorhandenen Genome aus dem griechischen Neolithikum (Rev 5, Pal 7 und Klei 10) in Bezug auf die heutige Bevölkerung Griechenlands untersucht (siehe SI 9 in [71]). Diese Analyse wurde mit 200.000 - 400.000 Positionen des Genoms durchgeführt. Die Kontinuität konnte für die drei neolithischen Proben widerlegt werden, lediglich unter Annahme einer recht geringen effektiven Populationsgröße der neolithischen Population kann die Kontinuität nicht widerlegt werden (Rev 5: 728 - 1887 Individuen, Pal 7 & Klei 10 \sim 170 - 450 Individuen). Die Annahme einer geringen effektiven Populationsgröße stellt eine Übereinstimmung zur mitochondrialen Analyse auf Kontinuität dar.

Jedoch ist eine Abnahme der Populationsgrösse im Verlauf des Neolithikums unwahrscheinlich, da archäologische Untersuchungen eine starke Zunahme an Fundplätzen und damit auch Bewohnern dieser Region beschreiben. Zusammengenommen beschreiben mitochondriale und genomische Ergebnisse sowie Kontinuitätsanalysen, dass in der heutigen Region Griechenlands ein Populationsumschwung stattgefunden hat.

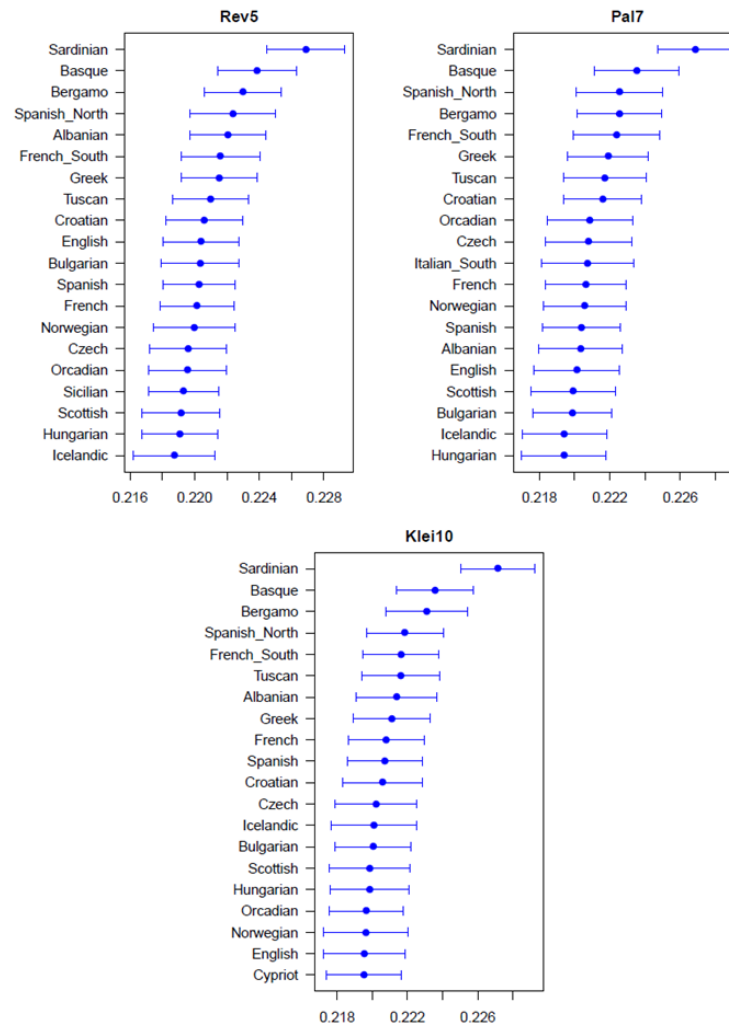


Abbildung 7.48: Darstellung der genetisch ähnlichsten modernen Populationen anhand f_3 (Khomani; moderne Population, Neolithikum Griechenland). Entnommen aus [71] *Supplements*, SI 7.

In einer aktuellen, genetischen Studie bronzzeitlicher Individuen der Ägäis [100] werden Komponenten des anatolischen Neolithikums und nahöstliche Einflüsse aus den JS des Kaukasus, sowie dem iranischen Neolithikum festgestellt. In der PCA (*Figure 1* in [100]) werden die Individuen der ägäischen Bronzezeit separiert zu jenen des anatolischen und europäischen Neolithikums dargestellt - es erfolgte eine Verschiebung parallel zur PC2 in Richtung des iranischen Neolithikums. Als Herkunftskomponenten für die Minoer konnten anhand einer Admixture Modellierung 46,7% für CHG, 50,3 % für das anatolische Neolithikum und 2,9 % für WHG errechnet werden [113]. Folglich hat während der Transformation zur Bronzezeit und der Ausbildung einer ägäischen Zivilisation ein enormer Populationsumschwung stattgefunden, welcher durch Genfluss aus nahöstlichen Populationen hervorgerufen wurde.

7.8 Schlussfolgerung

Während die Kolonisation des europäischen Kontinents und die Ausbildung der heutigen modernen Population in ihren Grundpfeilern aufgeklärt ist, haben aktuelle Studien das Ziel regionale, demographische Prozesse im Detail aufzuklären. Die Siedlungsgeschichte des anatomisch modernen Menschen ist stark abhängig von geographischen Barrieren und klimatischen Begebenheiten, welche regional differenzierte Lebensräume schaffen. In diesem Rahmen kann die eindeutig als demisch definierte Expansionswelle des Neolithikums über den europäischen Kontinent nicht als graduell verlaufender Prozess angesehen werden, der sich gleichmässig über Europa ausbreitete. Vielmehr wird dieser Prozess durch klimatische Bedingungen herausgefordert und soziale Interaktionen mit lokalen Jäger Sammler Gesellschaften haben in unterschiedlichen Wirkungsmass stattgefunden [18]. Das Auftreten der ersten Farmer auf dem europäischen Kontinent ist durch die Archäologie seit etwa 100 Jahren dokumentiert und das Neolithikum in der Ägäis kann durch zahlreiche Fundstellen abgebildet werden. In der vorliegenden Arbeit wird erstmalig der genetische Hintergrund prähistorischer Individuen aus der heutigen Region Griechenlands zur Zeit des Mesolithikums und des Neolithikums dargestellt. Skelettmaterialien aus dem mesolithischen Kontext (Theopetra, Franchthi) sowie dem frühen (Makedonien) und späten (Nord-/Südgriechenland) Neolithikum bieten die Möglichkeit archäologische Thesen durch die genetische Struktur zu überprüfen. Für die Analyse stehen 53 mitochondriale Genome zur Verfügung, welche im Rahmen eines Referenzdatensatzes von ~ 1200 prähistorischen Individuen eingegliedert in 28 Populationen (oberes Paläolithikum - Beginn Bronzezeit) analysiert worden sind. Für drei Individuen aus dem frühen (Revenia, 6438-6264 kal. v. Chr.) und späten/finalen (Kleitos 4230-3995 kal. v. Chr., Paliambela 4452 - 4350 kal. v. Chr.) Neolithikum wurden genomischen Sequenzen mit einer Abdeckung von 1-2x erstellt. Die Gruppenbildung der mitochondrialen Analyse und die tiefer sequenzierten Genome umfassen die Regionen Makedonien und Thrakien. Folglich gelten die Erkenntnisse dieser Arbeit für den "nordägäischen Kulturkomplex" [150].

Geringe genetische Distanzen des mitochondrialen Markers zwischen den Populationen des frühen Neolithikums aus Nordgriechenland, der Marmara Region und Europa zeigen, dass das griechische Neolithikum Teil der Migration nach Europa gewesen ist. Hohe Werte geteilter genetischer Drift zwischen den genomischen Datensätzen der Individuen aus der Ägäis sprechen für eine hohe genetische Übereinstimmung der Populationen dieser Region. Es können eindeutige Beweise für die Ankunft des Neolithikums im Norden Griechenlands über eine Route ausgehend von der west-anatolischen Küste geliefert werden. Dies belegt, dass neolithische Artefakte und Domestikate mit einer neolithischen Population über die Ägäis nach Griechenland gekommen sind, jedoch liefern die Daten dieser Arbeit keine genetische Analyse hinsichtlich der Verbindung zu den Populationen der Kernzone des Neolithikums (Zentralanatolien, Levante und Iran). Eine hohe genetische Identität mit der Starčevo Population bestätigt einen engen Kontakt zu dem frühen Neolithikum auf dem Balkan, welcher auf archäologischer Basis bereits durch die Gestaltung von Keramiken der Vinca Tradition thematisiert worden ist [88]. Diese Identität ist stärker ausgeprägt in der Population bzw. in den Individuen des späten Neolithikums.

Durch die genomische Analyse konnte der genetische Einfluss des griechischen Neolithikums hinsichtlich der Ausbreitung nach Europa quantifiziert werden. Es kann gezeigt werden, dass der Einfluss auf die danubischen Ausbreitungsrouten zu einem größeren Anteil auf das nordwest-anatolische Neolithikum zurückzuführen ist. Ausgehend vom griechischen Neolithikum sind hingegen verstärkt Signale in modernen, mediterranen Populationen zu sehen. Der Nachweis von Obsidianhandel über große Distanzen zeigt, dass das ägäische Neolithikum eine maritim-orientierte Gesellschaft gewesen ist, welche

ihre Kenntnisse bereits zum Beginn des Holozäns erworben hatte [22, 72]. Dies prägte die Lebensweise der Individuen in dieser Region bereits im Mesolithikum. Schon zu dieser Zeit ist die Beherrschung des Meers durch Handelsnetzwerke dokumentiert [22]. In dieser Arbeit kann durch die Identifikation der mitochondrialen Linie K1 in den Individuen von Theopetra gezeigt werden, dass im Mesolithikum bereits eine genetische Identität zu den Populationen Zentralanatoliens und des östlichen Mittelmeerraumes bestand. Diese Linie wird zudem in einer hohen Frequenz in der neolithischen Population aus Griechenland gefunden ($\sim 30\%$), was die Theorie eines prä-neolithischen Kontakts in Form von vereinzelten Migrationen und einer zweiten, ausgeprägten Migrationswelle zu Beginn des Neolithikums unterstützt [162]). Die mitochondrialen K1 Linien innerhalb eines mesolithischen Kontexts bieten einen Kontrast zu den genetischen Linien der Jäger und Sammler aus dem Westen und Osten Europas, jedoch bieten diese auch keinen Anlass zur Annahme eines massiven Populationsumschwungs zwischen dem Meso- und Neolithikum in Griechenland. Aufgrund der geringen Probenzahl konnte keine statistische Analyse durchgeführt werden und genomische Datensätze aus dem ägäischen Mesolithikum sind notwendig, um bestehende Thesen, wie z.B. der Nachweis einer über größere Distanzen interagierende Population und Genfluss aus kontemporären östlichen Populationen, nachzuweisen. Dafür sollten Genome von Individuen und/oder Populationen der ägäischen Inseln (z.B. Maroulas), des griechischen Festlands (Theopetra, Franchthi, Sidari) in Kombination mit präkeramischen Populationen aus Zentralanatolien und der PPN Kultur zur Verfügung stehen. Unter der Annahme eines prä-neolithischen Kontakts des ägäischen Mesolithikums und kontemporären Populationen aus dem Osten bieten mitochondriale Datensätze aufgrund der angenommen, kurzen Divergenzzeit keine ausreichende statistische Signifikanz, um Unterschiede zu identifizieren. Tief sequenzierte genomische Datensätze (Abdeckung $\sim 10x$) bieten durch die Analyse von Millionen von Marker die Möglichkeit diese Unterschiede aufzudecken. Von großem Interesse ist auch die Bestimmung der effektiven Populationsgrösse der mesolithischen aber auch neolithischen Population, eine Analyse die im Rahmen dieser Arbeit nicht durchgeführt werden konnte, da die genomischen Datensätze keine ausreichende Abdeckung zeigen.

Wird das griechische Neolithikum hingehend demographischer Prozesse analysiert, wie sie aus dem zentraleuropäischen Neolithikum bekannt sind [61], dann kann im griechischen Neolithikum keine Aufnahme von Jäger Sammler Genen aus Westeuropa nachgewiesen werden. Im Gegenteil, die genetische Kontinuität kann auf Basis des mitochondrialen Datensatzes durch Simulation der genetischen Drift über einen Zeitraum von 2500 Jahren nicht widerlegt werden. Es sind folglich keine signifikanten Populationsumschwünge durch Migration einer genetisch differenzierten Population nach Griechenland zu verzeichnen. Dieses Ergebnis kann durch die neolithischen Genome der Ägäis verifiziert werden (f_3/f_4 -Statistik, PCA, *Chromosome Painting*) [71]. Dem hingegen kann keine genetische Kontinuität zur heutigen Bevölkerung in Griechenland nachgewiesen werden. Folglich haben vom Ende des Neolithikums bis heute verstärkt Populationsbewegungen stattgefunden. Innerhalb des trans-ägäischen Netzwerkes kann im fortschreitenden Neolithikum zunehmender Genfluss aus den Jäger Sammler Genomen des Kaukasus identifiziert werden (f_3/f_4 -Statistik). Dies zeigt, dass maritime Kenntnisse während des Neolithikums erweitert, Handelsnetzwerke ausgebaut wurden [72] und soziale Interaktionen mit Populationen des Mittelmeerraumes stattgefunden haben. Diese Prozesse zeigen innerhalb des maritimen Netzwerkes im Neolithikum zunächst schwache Signale auf genetischer Ebene, jedoch bieten sie eine Basis für die Transformation der ägäischen Populationen in der Bronzezeit [100].

Literaturverzeichnis

- [1] David H Alexander, John Novembre, and Kenneth Lange. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9):1655–1664, 2009.
- [2] Morten E Allentoft, Martin Sikora, Karl-Göran Sjögren, Simon Rasmussen, Morten Rasmussen, Jesper Stenderup, Peter B Damgaard, Hannes Schroeder, Torbjörn Ahlström, Lasse Vinner, et al. Population genomics of Bronze Age Eurasia. *Nature*, 522(7555):167–172, 2015.
- [3] Stelios Andreou, Michael Fotiadis, and Kostas Kotsakis. Review of aegean prehistory v: the neolithic and bronze age of northern greece. *American Journal of Archaeology*, 100(3):537–597, 1996.
- [4] Richard M Andrews, Iwona Kubacka, Patrick F Chinnery, Robert N Lightowlers, Douglass M Turnbull, and Neil Howell. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics*, 23(2):147–147, 1999.
- [5] J Lawrence Angel. Early Neolithic people of Nea Nikomedeia. *Die Anfänge des Neolithikums vom Orient bis Nordeuropa*, 3:103–112, 1973.
- [6] Erik Aronesty. Comparison of sequencing utility programs. *The Open Bioinformatics Journal*, 7(1), 2013.
- [7] Hans-Jurgen Bandelt, Peter Forster, and Arne Röhl. Median-joining networks for inferring intraspecific phylogenies. *Molecular biology and evolution*, 16(1):37–48, 1999.
- [8] Ofer Bar-Yosef. The natufian culture in the levant, threshold to the origins of agriculture. *Evolutionary Anthropology: Issues, News, and Reviews*, 6(5):159–177, 1998.
- [9] Frédéric Baudat, Yukiko Imai, and Bernard de Massy. Meiotic recombination in mammals: localization and regulation. *Nature Reviews Genetics*, 14(11):794–806, 2013.
- [10] Stefano Benazzi, Katerina Douka, Cinzia Fornai, Catherine C Bauer, Ottmar Kullmer, Jiří Svoboda, Ildikó Pap, Francesco Mallegni, Priscilla Bayle, Michael Coquerelle, et al. Early dispersal of modern humans in europe and implications for neanderthal behaviour. *Nature*, 479(7374):525–528, 2011.
- [11] Stefano Benazzi, Viviane Slon, Saha Talamo, F Negrino, M Peresani, Shara E Bailey, S Sawyer, D Panetta, G Vicino, E Starnini, et al. The makers of the protoaurignacian and implications for neandertal extinction. *Science*, 348(6236):793–796, 2015.
- [12] Manthos Besios, Fotini Adaktylou, and Dimitris Kontogiorgos. *From Tells to Extended Settlements: On-site Geoarchaeology and Cultural Formation Process at the Extended Neolithic Settlement at Korinos (Northern Greece)*, chapter 5. Nova Science Publishers, Inc., New York, 2009.
- [13] M. Bessios and F. Adaktylou. A neolithic settlement at revenia near korinos. *AEMTH*, 18(2004):357–366, 2006.
- [14] Paolo Biagi and Michela Spataro. New observations on the radiocarbon chronology of the starčevo-criş and körös cultures. *Prehistoric Archaeology & Anthropological Theory and Education. Reports of Prehistoric Research Projects*, pages 6–7, 2005.
- [15] JOHN BiNTLiFF et al. The plain of western macedonia and the neolithic site of nea nikomedeia. In *Proceedings of the Prehistoric Society*, volume 42, pages 241–262. Cambridge Univ Press, 1976.
- [16] Brendan Blumenstiel, Kristian Cibulskis, Sheila Fisher, Matthew DeFelice, Andrew Barry, Tim Fennell, Justin Abreu, Brian Minie, Maura Costello, Geneva Young, et al. Targeted exon sequencing by in-solution hybrid selection. *Curr Protoc Hum Genet*, 166, 2010.
- [17] Alessio Boattini, Begoña Martínez-Cruz, Stefania Sarno, Christine Harmant, Antonella Useli, Paula Sanz, Daniele Yang-Yao, Jeremy Manry, Graziella Ciani, Donata Luiselli, et al. Uniparental markers in italy reveal a sex-biased genetic structure and different historical strata. *PLoS one*, 8(5):e65441, 2013.
- [18] Ruth Bollongino, Olaf Nehlich, Michael P Richards, Jörg Orschiedt, Mark G Thomas, Christian Sell, Zuzana Fajkošová, Adam Powell, and Joachim Burger. 2000 years of parallel societies in stone age central europe. *Science*, 342(6157):479–481, 2013.
- [19] B Bramanti, MG Thomas, W Haak, M Unterlaender, P Jores, K Tambets, I Antanaitis-Jacobs, MN Haidle, R Jankauskas, C-J Kind, et al. Genetic discontinuity between local hunter-gatherers and central europe’s first farmers. *science*, 326(5949):137–140, 2009.
- [20] Guido Brandt, Wolfgang Haak, Christina J Adler, Christina Roth, Anna Szécsényi-Nagy, Sarah Karimnia, Sabine Möller-Rieker, Harald Meller, Robert Ganslmeier, Susanne Friederich, et al. Ancient dna reveals key stages in the formation of central european mitochondrial genetic diversity. *Science*, 342(6155):257–261, 2013.
- [21] Marcus R Breese and Yunlong Liu. Ngsutils: a software suite for analyzing and manipulating next-generation sequencing datasets. *Bioinformatics*, 29(4):494–496, 2013.

- [22] Cyprian Broodbank. The origins and early development of mediterranean maritime activity. *Journal of Mediterranean Archaeology*, 19(2):199, 2006.
- [23] Paul Brotherton, Wolfgang Haak, Jennifer Templeton, Guido Brandt, Julien Soubrier, Christina Jane Adler, Stephen M Richards, Clio Der Sarkissian, Robert Ganslmeier, Susanne Friederich, et al. Neolithic mitochondrial haplogroup h genomes and the genetic origins of europeans. *Nature communications*, 4:1764, 2013.
- [24] Farnaz Broushaki, Mark G Thomas, Vivian Link, Saioa López, Lucy van Dorp, Karola Kirsanow, Zuzana Hofmanová, Yoan Diekmann, Lara M Cassidy, David Diez-del Molino, et al. Early neolithic genomes from the eastern fertile crescent. *Science*, 353(6298):499–503, 2016.
- [25] Michael Bunce, Charlotte L Oskam, and Morten E Allentoft. Quantitative real-time pcr in adna research. *Ancient DNA: methods and protocols*, pages 121–132, 2012.
- [26] J Burger. Zwischen kontamination und authentizität. der nachweis menschlicher dna aus archäologischen skeletten. *Brennpunkte und Perspektiven der aktuellen Anthropologie. Mitteilungen der Berliner Gesellschaft für Anthropologie, Ethnologie und Urgeschichte, Beiheft*, 1:53–59, 2006.
- [27] J Burger and R Bollongino. Richtlinien zur bergung, entnahme und archivierung von skelettproben für palaeogenetische analysen [guidelines for the recovery, acquisition and storage of skeletal samples for palaeogenetical analyses]. *Bulletin der Schweizerischen Gesellschaft für Anthropologie*, 16(1-2):71–78, 2010.
- [28] George BJ Busby, Garrett Hellenthal, Francesco Montinaro, Sergio Tofanelli, Kazima Bulayeva, Igor Rudan, Tatijana Zemunik, Caroline Hayward, Draga Toncheva, Sena Karachanak-Yankova, et al. The role of recent admixture in forming the contemporary west eurasian genomic landscape. *Current Biology*, 25(19):2518–2526, 2015.
- [29] F Calafell, P Underhill, A Tolun, D Angelicheva, and L Kalaydjieva. From asia to europe: mitochondrial dna sequence variability in bulgarians and turks. *Annals of human genetics*, 60(1):35–49, 1996.
- [30] Meredith L Carpenter, Jason D Buenrosto, Cristina Valdiosera, Hannes Schroeder, Morten E Allentoft, Martin Sikora, Morten Rasmussen, Simon Gravel, Sonia Guillén, Georgi Nekhrizov, et al. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient dna sequencing libraries. *The American Journal of Human Genetics*, 93(5):852–864, 2013.
- [31] V Gordon Childe. The dawn of european civilization (kegan paul, london). *AMERICAN ANTIQUITY*, 674, 1925.
- [32] Çiler Çilingiroglu. The concept of “neolithic package” considering its meaning and applicability. *Documenta Praehistorica*, 32:1–13, 2005.
- [33] John E Coleman. *Kephala: a late Neolithic settlement and cemetery*, volume 1. ASCSA, 1977.
- [34] David Comas, Francesc Calafell, Eva Mateu, Anna Perez-Lezaun, and Jaume Bertranpetit. Geographic variation in human mitochondrial dna control region sequence: the population history of turkey and its relationship to the european populations. *Molecular biology and evolution*, 13(8):1067–1077, 1996.
- [35] 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature*, 526(7571):68–74, 2015.
- [36] Tracey Cullen. Mesolithic mortuary ritual at franchthi cave, greece. *Antiquity*, 69(263):270, 1995.
- [37] Jesse Dabney, Michael Knapp, Isabelle Glocke, Marie-Theres Gansauge, Antje Weihmann, Birgit Nickel, Cristina Valdiosera, Nuria García, Svante Pääbo, Juan-Luis Arsuaga, et al. Complete mitochondrial genome sequence of a middle pleistocene cave bear reconstructed from ultrashort dna fragments. *Proceedings of the National Academy of Sciences*, 110(39):15758–15763, 2013.
- [38] Jesse Dabney and Matthias Meyer. Length and gc-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern dna sequencing libraries. *Biotechniques*, 52(2):87–94, 2012.
- [39] Jesse Dabney, Matthias Meyer, and Svante Pääbo. Ancient dna damage. *Cold Spring Harbor perspectives in biology*, 5(7):a012567, 2013.
- [40] Mark A DePristo, Eric Banks, Ryan Poplin, Kiran V Garimella, Jared R Maguire, Christopher Hartl, Anthony A Philippakis, Guillermo Del Angel, Manuel A Rivas, Matt Hanna, et al. A framework for variation discovery and genotyping using next-generation dna sequencing data. *Nature genetics*, 43(5):491–498, 2011.
- [41] Clio Der Sarkissian, Oleg Balanovsky, Guido Brandt, Valery Khartanovich, Alexandra Buzhilova, Sergey Koshel, Valery Zaporozhchenko, Detlef Gronenborn, Vyacheslav Moiseyev, Eugen Kolpakov, et al. Ancient dna reveals prehistoric gene-flow from siberia in the complex human population history of north east europe. *PLoS Genet*, 9(2):e1003296, 2013.
- [42] Marine Dumousseau, Nicolas Rodriguez, Nick Juty, and Nicolas Le Novere. Melting, a flexible platform to predict the melting temperatures of nucleic acids. *BMC bioinformatics*, 13(1):1, 2012.
- [43] N Efstratiou and María Pilar Fumanal García. Excavations at the neolithic settlement of makri, thrace, greece (1988-1996): A preliminary report. *SAGVNTVM. Papeles del Laboratorio de Arqueología de Valencia*, 31:11–62, 1998.

- [44] Phillip Endicott and Simon YW Ho. A bayesian evaluation of human mitochondrial substitution rates. *The American Journal of Human Genetics*, 82(4):895–902, 2008.
- [45] Laurent Excoffier, Guillaume Laval, and Stefan Schneider. Arlequin ver 3.5 user manual. *Computational and Molecular Population Genetics Lab (CMPG), Institute of Zoology, University of Berne*, 2009.
- [46] Qiaomei Fu, Heng Li, Priya Moorjani, Flora Jay, Sergey M Slepchenko, Aleksei A Bondarev, Philip LF Johnson, Ayinuer Aximu-Petri, Kay Prüfer, Cesare de Filippo, et al. Genome sequence of a 45,000-year-old modern human from western siberia. *Nature*, 514(7523):445–449, 2014.
- [47] Qiaomei Fu, Alissa Mittnik, Philip LF Johnson, Kirsten Bos, Martina Lari, Ruth Bollongino, Chengkai Sun, Liane Giemisch, Ralf Schmitz, Joachim Burger, et al. A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology*, 23(7):553–559, 2013.
- [48] Qiaomei Fu, Cosimo Posth, Mateja Hajdinjak, Martin Petr, Swapan Mallick, Daniel Fernandes, Anja Furtwängler, Wolfgang Haak, Matthias Meyer, Alissa Mittnik, et al. The genetic history of ice age europe. *Nature*, 2016.
- [49] Karamitrou-Mentessidi G. *Xirolimni Kozani Prefecture 1998*. AEMTH, 1998.
- [50] Karamitrou-Mentessidi G. *About prehistoric sites in western Macedonia: prefectures of Kozani and Grevena*, volume A century of Research in Prehistoric Macedonia. Archaeological Museum of Thessaloniki, Thessaloniki, 2014.
- [51] Nena Galanidou. Advances in the palaeolithic and mesolithic archaeology of greece for the new millennium. *Pharos*, 20(1):1–40, 2014.
- [52] Cristina Gamba, Eva Fernández, M Tirado, MF Deguilloux, MH Pemonge, P Utrilla, M Edo, M Molist, R Rasteiro, L Chikhi, et al. Ancient dna from an early neolithic iberian population supports a pioneer colonization by first farmers. *Molecular Ecology*, 21(1):45–56, 2012.
- [53] Cristina Gamba, Eppie R Jones, Matthew D Teasdale, Russell L McLaughlin, Gloria Gonzalez-Fortes, Valeria Mattiangeli, László Domboróczi, Ivett Kővári, Ildikó Pap, Alexandra Anders, et al. Genome flux and stasis in a five millennium transect of european prehistory. *Nature communications*, 5, 2014.
- [54] Marie-Theres Gansauge and Matthias Meyer. Single-stranded dna library preparation for the sequencing of ancient or damaged dna. *Nature protocols*, 8(4):737–748, 2013.
- [55] Fernando García-Alcalde, Konstantin Okonechnikov, José Carbonell, Luis M Cruz, Stefan Götz, Sonia Tarazona, Joaquín Dopazo, Thomas F Meyer, and Ana Conesa. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics*, 28(20):2678–2679, 2012.
- [56] DB Grammenos and S Kotsos. Sostikes anaskafes sto neolithiko oikismo stavroupolis thessalonikis, meros ii (1998–2003). *Rettende Ausgrabungen in der neolithischen Siedlung von Stavroupoli bei Thessaloniki Teil II (1998-2003)*, 2004.
- [57] Richard E Green, Johannes Krause, Adrian W Briggs, Tomislav Maricic, Udo Stenzel, Martin Kircher, Nick Patterson, Heng Li, Weiwei Zhai, Markus Hsi-Yang Fritz, et al. A draft sequence of the neandertal genome. *science*, 328(5979):710–722, 2010.
- [58] Melanie Groß. *Veröffentlichung ausstehend*. Doktorarbeit, Johannes Gutenberg-Universität Mainz, 2017.
- [59] Wolfgang Haak, Oleg Balanovsky, Juan J Sanchez, Sergey Koshel, Valery Zaporozhchenko, Christina J Adler, Clio SI Der Sarkissian, Guido Brandt, Carolin Schwarz, Nicole Nicklisch, et al. Ancient dna from european early neolithic farmers reveals their near eastern affinities. *PLoS Biol*, 8(11):e1000536, 2010.
- [60] Wolfgang Haak, Peter Forster, Barbara Bramanti, Shuichi Matsumura, Guido Brandt, Marc Tänzler, Richard Villems, Colin Renfrew, Detlef Gronenborn, Kurt Werner Alt, et al. Ancient dna from the first european farmers in 7500-year-old neolithic sites. *Science*, 310(5750):1016–1018, 2005.
- [61] Wolfgang Haak, Iosif Lazaridis, Nick Patterson, Nadin Rohland, Swapan Mallick, Bastien Llamas, Guido Brandt, Susanne Nordenfelt, Eadaoin Harney, Kristin Stewardson, et al. Massive migration from the steppe was a source for indo-european languages in europe. *Nature*, 2015.
- [62] P. Halstead and K. Kotsakis. Paliambela. In *Archaeological Reports for 2001-2002*, page 80. Society for the Promotion of Hellenic Studies and the British School at Athens, 2002.
- [63] P. Halstead and K. Kotsakis. Paliambela. In *Archaeological Reports for 2002-2003*, pages 49–66. Society for the Promotion of Hellenic Studies and the British School at Athens, 2003.
- [64] Jennifer Hammann. Optimierung einer extraktionsmethode für die behandlung von prähistorischen, bodengelagerten skelettproben zur weiteren nutzung für das *Next Generation Sequencing*. Magisterarbeit, Johannes Gutenberg-Universität Mainz, 2012.
- [65] Oliva Handt, Matthias Krings, Ryk H Ward, and S Pääbo. The retrieval of ancient human dna sequences. *American journal of human genetics*, 59(2):368, 1996.

- [66] Katerina Harvati, Andreas Darlas, Shara E Bailey, Thomas R Rein, Sireen El Zaatari, Luca Fiorenza, Ottmar Kullmer, and Eleni Psathi. New neanderthal remains from mani peninsula, southern greece: the kalamakia middle paleolithic cave site. *Journal of human evolution*, 64(6):486–499, 2013.
- [67] Garrett Hellenthal, Adam Auton, and Daniel Falush. Inferring human colonization history using a copying model. *PLoS Genet*, 4(5):e1000078, 2008.
- [68] Garrett Hellenthal, George BJ Busby, Gavin Band, James F Wilson, Cristian Capelli, Daniel Falush, and Simon Myers. A genetic atlas of human admixture history. *Science*, 343(6172):747–751, 2014.
- [69] GJ Hennig, W Herr, E Weber, and N Iu Xirotiris. Esr-dating of the fossil hominid cranium from petralona cave, greece. 1981.
- [70] Tom Higham, Tim Compton, Chris Stringer, Roger Jacobi, Beth Shapiro, Erik Trinkaus, Barry Chandler, Flora Gröning, Chris Collins, Simon Hillson, et al. The earliest evidence for anatomically modern humans in northwestern europe. *Nature*, 479(7374):521–524, 2011.
- [71] Zuzana Hofmanová, Susanne Kreutzer, Garrett Hellenthal, Christian Sell, Yoan Diekmann, David Díez-del Molino, Lucy van Dorp, Saïoa López, Athanasios Kousathanas, Vivian Link, et al. Early farmers from across europe directly descended from neolithic aegeans. *Proceedings of the National Academy of Sciences*, page 201523951, 2016.
- [72] B Horejs, B Milić, F Ostmann, U Thanheiser, B Weninger, and A Galik. The aegean in the early 7th millennium bc: maritime networks and colonization. *Journal of world prehistory*, 28(4):289–330, 2015.
- [73] Jeffery R Hughey, Peristera Paschou, Petros Drineas, Donald Mastropaolo, Dimitra M Lotakis, Patrick A Navas, Manolis Michalodimitrakis, John A Stamatoyannopoulos, and George Stamatoyannopoulos. A european population in minoan bronze age crete. *Nature communications*, 4:1861, 2013.
- [74] Jodi Irwin, Balazs Egyed, Jessica Saunier, Gabriella Szamosi, Jennifer O’Callaghan, Zsolt Padar, and Thomas J Parsons. Hungarian mtdna population databases from budapest and the baranya county roma. *International journal of legal medicine*, 121(5):377–383, 2007.
- [75] Jodi Irwin, Jessica Saunier, Katharine Strouss, Carla Paintner, Toni Diegoli, Kimberly Sturk, Leda Kovatsi, Anita Brandstätter, Marios A Cariolou, Walther Parson, et al. Mitochondrial control region sequences from northern greece and greek cypriots. *International journal of legal medicine*, 122(1):87–89, 2008.
- [76] TW Jacobsen and Tracey Cullen. A consideration of mortuary practices in neolithic greece: Burials from franchthi cave. *Mortality and immortality: The anthropology and archaeology of death*, pages 79–101, 1981.
- [77] Eppie R Jones, Gloria Gonzalez-Fortes, Sarah Connell, Veronika Siska, Anders Eriksson, Rui Martiniano, Russell L McLaughlin, Marcos Gallego Llorente, Lara M Cassidy, Cristina Gamba, et al. Upper palaeolithic genomes reveal deep roots of modern eurasiens. *Nature communications*, 6, 2015.
- [78] Hákon Jónsson, Aurélien Ginolhac, Mikkel Schubert, Philip LF Johnson, and Ludovic Orlando. mapdamage2. 0: fast approximate bayesian estimates of ancient dna damage parameters. *Bioinformatics*, page btt193, 2013.
- [79] Karamitrou-Mentessidi. *Kozani Prefecture 2000: excavations of the main national roads*, volume 14. AEMTH, 2000.
- [80] G Karamitrou-Mentessidi, N Efstratiou, JK Kozłowski, M Kaczanowska, Y Maniatis, A Curci, S Michalopoulou, A Papatheanasiou, and SM Valamoti. New evidence on the beginning of farming in greece: the early neolithic settlement of mavropigi in western macedonia (greece). *Antiquity*, 87:336, 2013.
- [81] Victoria Keerl. *A River Runs Through It: Ancient DNA Data on the Neolithic Populations of the Great Hungarian Plain*. Doktorarbeit, Johannes Gutenberg-Universität Mainz, 2015.
- [82] Gülşah Merve Kılınç, Ayça Omrak, Füsün Özer, Torsten Günther, Ali Metin Büyükkarakaya, Erhan Bıçakçı, Douglas Baird, Handan Melike Dönertaş, Ayshin Ghalichi, Reyhan Yaka, et al. The demographic development of the first farmers in anatolia. *Current Biology*, 26(19):2659–2666, 2016.
- [83] Martin Kircher. Analysis of high-throughput ancient dna sequencing data. *Ancient DNA: methods and protocols*, pages 197–228, 2012.
- [84] Martin Kircher, Susanna Sawyer, and Matthias Meyer. Double indexing overcomes inaccuracies in multiplex sequencing on the illumina platform. *Nucleic acids research*, page gkr771, 2012.
- [85] Mirna Kirin, Ruth McQuillan, Christopher S Franklin, Harry Campbell, Paul M McKeigue, and James F Wilson. Genomic runs of homozygosity record population history and consanguinity. *PLoS one*, 5(11):e13996, 2010.
- [86] Anita Kloss-Brandstätter, Dominic Pacher, Sebastian Schönherr, Hansi Weissensteiner, Robert Binna, Günther Specht, and Florian Kronenberg. Haplogrep: a fast and reliable algorithm for automatic classification of mitochondrial dna haplogroups. *Human mutation*, 32(1):25–32, 2011.
- [87] KOSTAS KOTSAKIS. Domesticating the periphery. *Pharos*, 20(1):41–73, 2014.

- [88] Stavros Kotsos and Dushka Urem-Kotsou. *Filling the Neolithic Landscape of Central Macedonia, Greece*. Serbian Academy of Science and Arts, 2006.
- [89] Margarita Koumouzelis, Boleslaw Ginter, Janusz K Kozłowski, Maciej Pawlikowski, Ofer Bar-Yosef, Rosa Maria Albert, Maria Litynska-Zajac, Ewa Stworzewicz, Piotr Wojtal, Grzegorz Lipecki, et al. The early upper palaeolithic in greece: the excavations in klisoura cave. *Journal of Archaeological Science*, 28(5):515–539, 2001.
- [90] Athanasios Kousathanas, Christoph Leuenberger, Vivian Link, Christian Sell, Joachim Burger, and Daniel Wegmann. Inferring heterozygosity from ancient and low coverage genomes. *bioRxiv*, page 046748, 2016.
- [91] Johannes Krause, Adrian W Briggs, Martin Kircher, Tomislav Maricic, Nicolas Zwyns, Anatoli Derevianko, and Svante Pääbo. A complete mtDNA genome of an early modern human from Kostenki, Russia. *Current Biology*, 20(3):231–236, 2010.
- [92] Raiko Krauß, Nedko Elenski, Bernhard Weninger, Lee Clare, Canan Çakırlar, and Petăr Zidarov. Beginnings of the neolithic in southeast Europe: the early neolithic sequence and absolute dates from Džuljuniča-smārdeš (Bulgaria). *Documenta Prähistorica*, 41:51–77, 2014.
- [93] Raiko Krauß, Elena Marinova, Hanne De Brue, and Bernhard Weninger. The rapid spread of early farming from the Aegean into the Balkans via the sub-Mediterranean-Aegean vegetation zone. *Quaternary International*, 2017.
- [94] Susanne Kreutzer. Etablierung eines Protokolls zur selektiven Anreicherung degradierter, humaner mitochondrialer DNA aus Skelettfunden durch Sure Select Target Enrichment und anschließendem Next Generation Sequencing auf der Illumina Plattform. Diplomarbeit, Johannes Gutenberg-Universität Mainz, 2011.
- [95] Martin Kuhlwilm, Ilan Gronau, Melissa J Hubisz, Cesare de Filippo, Javier Prado-Martinez, Martin Kircher, Qiaomei Fu, Hernán A Burbano, Carles Lalueza-Fox, Marco de La Rasilla, et al. Ancient gene flow from early modern humans into eastern Neanderthals. *Nature*, 530(7591):429–433, 2016.
- [96] Marie Lacan, Christine Keyser, François-Xavier Ricaut, Nicolas Brucato, Josep Tarrús, Angel Bosch, Jean Guilaine, Eric Crubézy, and Bertrand Ludes. Ancient DNA suggests the leading role played by men in the Neolithic dissemination. *Proceedings of the National Academy of Sciences*, 108(45):18255–18259, 2011.
- [97] Kurt Lambeck. Sea-level change and shore-line evolution in Aegean Greece since upper palaeolithic time. *Antiquity*, 70(269):588–611, 1996.
- [98] Eric S Lander, Lauren M Linton, Bruce Birren, Chad Nusbaum, Michael C Zody, Jennifer Baldwin, Keri Devon, Ken DeWaver, Michael Doyle, William FitzHugh, et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.
- [99] N Laskaris, A Sampson, F Mavridis, and Ioannis Liritzis. Late Pleistocene/early Holocene seafaring in the Aegean: new obsidian hydration dates with the SIMS-SS method. *Journal of Archaeological Science*, 38(9):2475–2479, 2011.
- [100] Iosif Lazaridis, Alissa Mittnik, Nick Patterson, Swapan Mallick, Nadin Rohland, Saskia Pfrenkle, Anja Furtwängler, Alexander Peltzer, Cosimo Posth, Andonis Vasilakis, et al. Genetic origins of the Minoans and Mycenaeans. *Nature*, 548(7666):214, 2017.
- [101] Iosif Lazaridis, Dani Nadel, Gary Rollefson, Deborah C Merrett, Nadin Rohland, Swapan Mallick, Daniel Fernandes, Mario Novak, Beatriz Gamarra, Kendra Sirak, et al. Genomic insights into the origin of farming in the ancient Near East. *Nature*, 536(7617):419–424, 2016.
- [102] Iosif Lazaridis, Nick Patterson, Alissa Mittnik, Gabriel Renaud, Swapan Mallick, Karola Kirsanow, Peter H Sudmant, Joshua G Schraiber, Sergi Castellano, Mark Lipson, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, 513(7518):409–413, 2014.
- [103] Chenhong Li, Michael Hofreiter, Nicolas Straube, Shannon Corrigan, and Gavin JP Naylor. Capturing protein-coding genes across highly divergent species. *Biotechniques*, 54(6):321–326, 2013.
- [104] Heng Li and Richard Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754–1760, 2009.
- [105] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009.
- [106] Tomas Lindahl. Instability and decay of the primary structure of DNA. *Nature*, 362(6422):709–715, 1993.
- [107] Yontao Lu, Nick Patterson, Yiping Zhan, Swapan Mallick, and David Reich. Technical design document for a SNP array that is optimized for population genetics. Available at ftp://ftp.cephb.fr/hgdp_supp10, 2011.
- [108] V Lykousis. Sea-level changes and shelf break prograding sequences during the last 400ka in the Aegean margins: subsidence rates and palaeogeographic implications. *Continental Shelf Research*, 29(16):2037–2044, 2009.
- [109] Helena Malmström, M Thomas P Gilbert, Mark G Thomas, Mikael Brandström, Jan Storå, Petra Molnar, Pernille K Andersen, Christian Bendixen, Gunilla Holmlund, Anders Götherström, et al. Ancient DNA reveals lack of continuity between Neolithic hunter-gatherers and contemporary Scandinavians. *Current Biology*, 19(20):1758–1762, 2009.

- [110] Y. Maniatis. Radiocarbon dating of the major cultural changes in Prehistoric Macedonia: recent developments. In Stefani, Merousis, and Dimoula, editors, *A century of Research in Prehistoric Macedonia. International Conference Proceedings 22-24 November 2012*, volume 10, pages 205–222. Archaeological Museum of Thessaloniki, Thessaloniki, 2014.
- [111] Sotiris K Manolis and Helen J Stravopodi. An assessment of the human skeletal remains in the mesolithic deposits of theopetra cave: a case study. *British School at Athens Studies*, pages 207–216, 2003.
- [112] Laisel Martinez, Sheyla Mirabal, Javier R Luis, and Rene J Herrera. Middle eastern and european mtDNA lineages characterize populations from eastern crete. *American journal of physical anthropology*, 137(2):213–223, 2008.
- [113] Iain Mathieson, Songül Alpaslan Roodenberg, Cosimo Posth, et al. The genomic history of southeastern europe. 2017.
- [114] Iain Mathieson, Iosif Lazaridis, Nadin Rohland, Swapan Mallick, Nick Patterson, Songül Alpaslan Roodenberg, Eadaoin Harney, Kristin Stewardson, Daniel Fernandes, Mario Novak, et al. Genome-wide patterns of selection in 230 ancient eurasians. *Nature*, 528(7583):499–503, 2015.
- [115] Linea Melchior, Niels Lynnerup, Hans R Siegismund, Toomas Kivisild, and Jørgen Dissing. Genetic diversity among ancient nordic populations. *PLoS one*, 5(7):e11898, 2010.
- [116] Paolo Menozzi, Alberto Piazza, and L Cavalli-Sforza. Synthetic maps of human gene frequencies in europeans. *Science*, 201(4358):786–792, 1978.
- [117] Matthias Meyer, Juan-Luis Arsuaga, Cesare de Filippo, Sarah Nagel, Ayinuer Aximu-Petri, Birgit Nickel, Ignacio Martínez, Ana Gracia, José María Bermúdez de Castro, Eudald Carbonell, et al. Nuclear dna sequences from the middle pleistocene sima de los huesos hominins. *Nature*, 531(7595):504–507, 2016.
- [118] Matthias Meyer, Adrian W Briggs, Tomislav Maricic, Barbara Höber, Barbara Höffner, Johannes Krause, Antje Wehmann, Svante Pääbo, and Michael Hofreiter. From micrograms to picograms: quantitative pcr reduces the material demands of high-throughput sequencing. *Nucleic Acids Research*, 36(1):e5–e5, 2008.
- [119] Matthias Meyer, Qiaomei Fu, Ayinuer Aximu-Petri, Isabelle Glocke, Birgit Nickel, Juan-Luis Arsuaga, Ignacio Martínez, Ana Gracia, José María Bermúdez de Castro, Eudald Carbonell, et al. A mitochondrial genome sequence of a hominin from sima de los huesos. *Nature*, 505(7483):403–406, 2014.
- [120] Matthias Meyer and Martin Kircher. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc*, 2010(6):t5448, 2010.
- [121] Matthias Meyer, Martin Kircher, Marie-Theres Gansauge, Heng Li, Fernando Racimo, Swapan Mallick, Joshua G Schraiber, Flora Jay, Kay Prüfer, Cesare de Filippo, et al. A high-coverage genome sequence from an archaic denisovan individual. *Science*, 338(6104):222–226, 2012.
- [122] Kornelija Minichreiter and Ines Krajcar Bronić. New radiocarbon dates for the early starčevo culture in croatia. *Prilozi Instituta za arheologiju u Zagrebu*, 23(1):5–16, 2007.
- [123] A Nafplioti. The mesolithic occupants of maroulas on kythnos: skeletal isotope ratio signatures of their geographic origin. *The Prehistory of the Island of Kythnos (Cyclades, Greece) and the Mesolithic Settlement at Maroulas. The Polish Academy of Arts and Sciences*, pages 207–215, 2010.
- [124] Christina Nielsen-Marsh, A Gernaey, Gordon Turner-Walker, Robert Hedges, Alistair Pike, and Matthew Collins. The chemical degradation of bone. *Human osteology in archaeology and forensic science. London: Greenwich Medical Media*, pages 439–454, 2000.
- [125] John Novembre and Matthew Stephens. Interpreting principal component analyses of spatial population genetic variation. *Nature genetics*, 40(5):646–649, 2008.
- [126] Iñigo Olalde, Hannes Schroeder, Marcela Sandoval-Velasco, Lasse Vinner, Irene Lobón, Oscar Ramirez, Sergi Civit, Pablo García Borja, Domingo C Salazar-García, Sahra Talamo, et al. A common genetic origin for early farmers from mediterranean cardial and central european lbc cultures. *Molecular Biology and Evolution*, 32(12):3132–3142, 2015.
- [127] Ludovic Orlando, Aurélien Ginolhac, Guojie Zhang, Duane Froese, Anders Albrechtsen, Mathias Stiller, Mikkel Schubert, Enrico Cappellini, Bent Petersen, Ida Moltke, et al. Recalibrating equus evolution using the genome sequence of an early middle pleistocene horse. *Nature*, 499(7456):74–78, 2013.
- [128] Krisztián Oross and Eszter Bánffy. Three successive waves of neolithisation: Lbc development in transdanubia. *Documenta Praehistorica*, 36:175–189, 2009.
- [129] M Özdoğan, P Matthiae, F Pinnock, L Nigro, and N Marchetti. Westward expansion of the neolithic way of life: Sorting the neolithic package into distinct packages. In *Proceedings of the 6th International Congress on the Archaeology of the Ancient Near East*, volume 1, pages 883–897. Harrassowitz Verlag Wiesbaden, 2010.
- [130] Mehmet Özdoğan. A new look at the introduction of the neolithic way of life in southeastern europe. changing paradigms of the expansion of the neolithic way of life. *Documenta Praehistorica*, 41:33–49, 2014.

- [131] S Papadopoulos and D Malamidou. Limenaria. a neolithic and early bronze age settlement at thassos, north-east aegean. In *The Aegean in the Neolithic, Chalcolithic and the Early Bronze Age. Proceedings of the International Symposium at Urla-Izmir*, pages 13–19, 1997.
- [132] A Papatthanasidou. A bioarchaeological analysis of neolithic alepotrypa cave, greece, british archaeological reports international series, 2001.
- [133] Anastasia Papatthanasidou. Stable isotope analysis in neolithic greece and possible implications on human health. *International Journal of Osteoarchaeology*, 13(5):314–324, 2003.
- [134] Richards M. Papatthanasidou A. *Anthropological remains from the Early neolithic sites of Mauropigi, Xirolimni and Pontokomi at Western Macedonia*, volume 1. The archaeological Work in Upper Macedonia, 2009.
- [135] GA Papatthanasopoulos. Spelea dirou, 1971 (ai anaskafai tou 1970-1971). *Athens Annals of Archaeology*, 4:12–26, 1971.
- [136] M Pappa. Neolithic societies: recent evidence from northern greece. In *The Struma/Strymon River Valley in Prehistory, Proceedings of the International Symposium Strymon Praehistoricus, Bulgaria-Greece*, pages 27–09, 2004.
- [137] M Pappa, P Halstead, K Kotsakis, A Bogaard, R Fraser, V Isaakidou, I Mainland, D Mylona, K Skourtopoulou, S Triantaphyllou, et al. The neolithic site of makriyalos, northern greece: reconstruction of social and economic structure of the settlement through comparative study of the finds. *Subsistence, Economy and Society in the Greek World: improving the integration of archaeology and science*, pages 77–88, 2013.
- [138] Maria Pappa and Manthos Besios. The makriyalos project: rescue excavations at the neolithic site of makriyalos, pieria, northern greece. *Neolithic society in Greece*, 2:108–120, 1999.
- [139] Nick Patterson, Alkes L Price, and David Reich. Population structure and eigenanalysis. *PLoS genet*, 2(12):e190, 2006.
- [140] Daniel A Peiffer and Kevin L Gunderson. Design of tag snp whole genome genotyping arrays. *DNA Microarrays for Biomedical Research: Methods and Protocols*, pages 51–61, 2009.
- [141] C Perlès. From the near east to greece: let’s reverse the focus. cultural elements that didn’t transfer. *How did farming reach Europe?*, pages 275–290, 2005.
- [142] Catherine Perlès. *The early Neolithic in Greece: the first farming communities in Europe*. Cambridge University Press, 2001.
- [143] Cosimo Posth, Gabriel Renaud, Alissa Mittnik, Dorothée G Drucker, Hélène Rougier, Christophe Cupillard, Frédérique Valentin, Corinne Thevenet, Anja Furtwängler, Christoph Wißing, et al. Pleistocene mitochondrial genomes suggest a single major dispersal of non-africans and a late glacial population turnover in europe. *Current Biology*, 26(6):827–833, 2016.
- [144] Kay Prüfer, Fernando Racimo, Nick Patterson, Flora Jay, Sriram Sankararaman, Susanna Sawyer, Anja Heinze, Gabriel Renaud, Peter H Sudmant, Cesare de Filippo, et al. The complete genome sequence of a neanderthal from the altai mountains. *Nature*, 505(7481):43–49, 2014.
- [145] Maanasa Raghavan, Pontus Skoglund, Kelly E Graf, Mait Metspalu, Anders Albrechtsen, Ida Moltke, Simon Rasmussen, Thomas W Stafford Jr, Ludovic Orlando, Ene Metspalu, et al. Upper palaeolithic siberian genome reveals dual ancestry of native americans. *Nature*, 505(7481):87–91, 2014.
- [146] Morten Rasmussen, Xiaosen Guo, Yong Wang, Kirk E Lohmueller, Simon Rasmussen, Anders Albrechtsen, Line Skotte, Stinus Lindgreen, Mait Metspalu, Thibaut Jombart, et al. An aboriginal australian genome reveals separate human dispersals into asia. *Science*, 334(6052):94–98, 2011.
- [147] Nicolas Ray, Mathias Currat, Matthieu Foll, and Laurent Excoffier. Splat2: a spatially explicit simulation framework for complex demography, genetic admixture and recombination. *Bioinformatics*, 26(23):2993–2994, 2010.
- [148] David Reich, Richard E Green, Martin Kircher, Johannes Krause, Nick Patterson, Eric Y Durand, Bence Viola, Adrian W Briggs, Udo Stenzel, Philip LF Johnson, et al. Genetic history of an archaic hominin group from denisova cave in siberia. *Nature*, 468(7327):1053–1060, 2010.
- [149] David Reich, Kumarasamy Thangaraj, Nick Patterson, Alkes L Price, and Lalji Singh. Reconstructing indian population history. *Nature*, 461(7263):489–494, 2009.
- [150] Agathe Reingruber. The argissa magoula and the beginning of the neolithic in thessaly. *How did farming reach Europe*, pages 155–71, 2005.
- [151] Agathe Reingruber. *Die Argissa-Magoula - Das frühe und das beginnende mittlere Neolithikum im Lichte transägäischer Beziehungen*. Rudolf Habelt, 2008.
- [152] Agathe Reingruber. Early neolithic settlement patterns and exchange networks in the aegean. *Documenta Praehistorica*, 38:291–305, 2011.
- [153] Agathe Reingruber. Pre-ceramic, aceramic or early ceramic? the radiocarbon dated beginning of the neolithic in the aegean. *Documenta Praehistorica*, 42:147–158, 2015.

- [154] Gabriel Renaud, Viviane Slon, Ana T Duggan, and Janet Kelso. Schmutzi: estimation of contamination and endogenous mitochondrial consensus calling for ancient dna. *Genome biology*, 16(1):1, 2015.
- [155] Adrien Rieux, Anders Eriksson, Mingkun Li, Benjamin Sobkowiak, Lucy A Weinert, Vera Warmuth, Andres Ruiz-Linares, Andrea Manica, and François Balloux. Improved calibration of the human mitochondrial clock using ancient genomes. *Molecular biology and evolution*, 31(10):2780–2792, 2014.
- [156] Maïté Rivollat, Fanny Mendisco, Marie-Hélène Pemonge, Audrey Safi, Didier Saint-Marc, Antoine Brémont, Christine Couture-Veschambre, Stéphane Rottier, and Marie-France Deguilloux. When the waves of european neolithization met: first paleogenetic evidence from early farmers in the southern paris basin. *PloS one*, 10(4):e0125521, 2015.
- [157] ROBERT J Rodden. recent discoveries from prehistoric macedonia: An interim report"(with eight plates). *Balkan Studies*, 5:109–124, 1964.
- [158] Robert J Rodden. An early neolithic village in greece. *Scientific American*, 212(4):82–88, 1965.
- [159] Robert J Rodden, GW Dimbleby, AC Western, EH Willis, ES Higgs, and WJ Clench. Excavations at the early neolithic site at nea nikomedeia, greek macedonia (1961 season). *Proceedings of the Prehistoric Society (New Series)*, 28:267–288, 1962.
- [160] Christina Roth. *Once upon a time in the West : paleogenetic analyses on Mesolithic to Early Bronze Age individuals from the Iberian Peninsula*. Doktorarbeit, Johannes Gutenberg-Universität Mainz, 2016.
- [161] C Runnels. The stone age of greece from the paleolithic to the advent of the neolithic: addendum 1995–1999. *Aegean prehistory: a review*, pages 225–254, 2001.
- [162] Curtis Runnels. Review of aegean prehistory iv: the stone age of greece from the palaeolithic to the advent of the neolithic. *American Journal of Archaeology*, 99(4):699–728, 1995.
- [163] Adamantios Sampson. Late neolithic remains at tharrounia, euboea: a model for the seasonal use of settlements and caves. *The Annual of the British School at Athens*, 87:61–101, 1992.
- [164] ADAMANTIOS SAMPSON. The mesolithic settlement at maroulas, kythn os. 2002.
- [165] Adamantios A Sampson, Małgorzata Kaczanowska, Janusz Krzysztof Kozłowski, and Stefan Witold Alexandrowicz. *The prehistory of the island of Kythnos (Cyclades, Greece) and the Mesolithic settlement at Maroulas*. Polish Academy of Arts and Sciences, 2010.
- [166] Juan J Sanchez, Claus Børsting, Charlotte Hallenberg, Anders Buchard, Alexis Hernandez, and Niels Morling. Multiplex per and minisequencing of snps—a model with 35 y chromosome snps. *Forensic science international*, 137(1):74–84, 2003.
- [167] Susanna Sawyer, Johannes Krause, Katerina Guschanski, Vincent Savolainen, and Svante Pääbo. Temporal patterns of nucleotide misincorporations and dna fragmentation in ancient dna. *PloS one*, 7(3):e34131, 2012.
- [168] Carl Schildkraut and Shneior Lifson. Dependence of the melting temperature of dna on salt concentration. *Biopolymers*, 3(2):195–208, 1965.
- [169] Anna Schönberg, Christoph Theunert, Mingkun Li, Mark Stoneking, and Ivan Nasidze. High-throughput sequencing of complete human mtDNA genomes from the caucasus and west asia: high diversity and demographic inferences. *European Journal of Human Genetics*, 19(9):988–994, 2011.
- [170] Anna Schulz. *Paläo- und Populationsgenetik von Jäger-Sammler-Fischern und Ackerbauern in Nordzentraleuropa und angrenzenden Regionen*. Doktorarbeit, Johannes Gutenberg-Universität Mainz, 2015.
- [171] Andaine Seguin-Orlando, Thorfinn S Korneliusson, Martin Sikora, Anna-Sapfo Malaspinas, Andrea Manica, Ida Moltke, Anders Albrechtsen, Amy Ko, Ashot Margaryan, Vyacheslav Moiseyev, et al. Genomic structure in europeans dating back at least 36,200 years. *Science*, 346(6213):1113–1118, 2014.
- [172] Christian Sell. *Addressing Challenges of Ancient DNA Sequence Data Obtained with Next Generation Methods*. Doktorarbeit, Johannes Gutenberg-Universität Mainz, 2017.
- [173] Christian Sell, Melanie Strobel, and Susanne Kreutzer. Maximizing the information of dna extracts obtained from skeletal remains. Dublin, 2012. Society of Molecular Biology and Evolution - *Conference poster*.
- [174] Pontus Skoglund, Helena Malmström, Ayça Omrak, Maanasa Raghavan, Cristina Valdiosera, Torsten Günther, Per Hall, Kristiina Tambets, Jüri Parik, Karl-Göran Sjögren, et al. Genomic diversity and admixture differs for stone-age scandinavian foragers and farmers. *Science*, 344(6185):747–750, 2014.
- [175] Pontus Skoglund, Helena Malmström, Maanasa Raghavan, Jan Storå, Per Hall, Eske Willerslev, M Thomas P Gilbert, Anders Götherström, and Mattias Jakobsson. Origins and genetic legacy of neolithic farmers and hunter-gatherers in europe. *Science*, 336(6080):466–469, 2012.
- [176] Pontus Skoglund, Bernd H Northoff, Michael V Shunkov, Anatoli P Derevianko, Svante Pääbo, Johannes Krause, and Mattias Jakobsson. Separating endogenous ancient dna from modern day contamination in a siberian neandertal. *Proceedings of the National Academy of Sciences*, 111(6):2229–2234, 2014.

- [177] Pontus Skoglund, Jan Storå, Anders Götherström, and Mattias Jakobsson. Accurate sex identification of ancient human remains using dna shotgun sequencing. *Journal of Archaeological Science*, 40(12):4477–4482, 2013.
- [178] Montgomery Slatkin. Statistical methods for analyzing ancient dna from hominins. *Current Opinion in Genetics & Development*, 41:72–76, 2016.
- [179] Chris Stringer. The status of homo heidelbergensis (schoetensack 1908). *Evolutionary Anthropology: Issues, News, and Reviews*, 21(3):101–107, 2012.
- [180] P Sümeği, R Kertész, I Juhász, G Timár, and S Gulyás. Mesolithic/neolithic transition in the carpathian basin—was there an ecological trap during the neolithic. *Comparative Archaeology and Palaeoclimatology. Socio-cultural response to a changing world*, pages 109–177, 2013.
- [181] Anna Szécsényi-Nagy, Guido Brandt, Wolfgang Haak, Victoria Keerl, János Jakucs, Sabine Möller-Rieker, Kitti Köhler, Balázs Gusztáv Mende, Krisztián Oross, Tibor Marton, et al. Tracing the genetic origin of europe’s first farmers reveals insights into their social organization. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1805):20150339, 2015.
- [182] Francesca Tassi, Silvia Ghirotto, Massimo Mezzavilla, Sibelle Torres Vilaça, Lisa De Santi, and Guido Barbujani. Early modern human dispersal from africa: genomic evidence for multiple waves of migration. *Investigative genetics*, 6(1):1, 2015.
- [183] R Core Team. R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria, 2012, 2014.
- [184] Vangelis Turloukis and Panagiotis Karkanis. The middle pleistocene archaeological record of greece and the role of the aegean in hominin dispersals: new data and interpretations. *Quaternary Science Reviews*, 43:1–15, 2012.
- [185] S. Triantaphyllou and F. Adaktylou. The treatment of the dead during the early neolithic period in macedonia a first insight in revenia, korinou of northern pieria. *AEMTH*, 28, 2014. In press (in Greek).
- [186] Sevasti Triantaphyllou. A bioarchaeological approach to prehistoric cemetery populations from western and central greek macedonia. 1999.
- [187] Sevi Triantaphyllou. Living with the dead: a re-consideration of mortuary practices in the greek neolithic. In V. Isaakidou and P. Tomkins, editors, *Escaping the Labyrinth: The Cretan Neolithic in Context*, volume 8, pages 139–157. Oxbow Monographs, Oxford, 2008.
- [188] MARTina Unterländer. *Populationsgenetik eisenzeitlicher Reiternomaden der Eurasischen Steppe*. Doktorarbeit, Johannes Gutenberg-Universität Mainz, 2014.
- [189] D. Urem-Kotsou, A. Papaioannou, T. Papadakou, N. Saridaki, and Z. Intze. *Pottery and stylistic boundaries. Early and middle neolithic pottery in Macedonia*. ARCHAEOLOGICAL MUSEUM OF THESSALONIKI, 2014.
- [190] D. Urem-Kotsou, A. Papaioanou, T. Silva-Gracia, F. Adaktylou, and M. Besios. Settlement of early and middle neolithic in revenia korinou. preliminary results of the ceramic analysis. *AEMTH*, 25, 2011. In press (in Greek).
- [191] Tjeerd H Van Andel and Curtis N Runnels. The earliest farmers in europe. *Antiquity*, 69(264):481–500, 1995.
- [192] Mannis van Oven. Phylotree build 17: Growing the human mitochondrial dna tree. *Forensic Science International: Genetics Supplement Series*, 5:e392–e394, 2015.
- [193] Mannis Van Oven and Manfred Kayser. Updated comprehensive phylogenetic tree of global human mitochondrial dna variation. *Human mutation*, 30(2):E386–E394, 2009.
- [194] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth edition, 2002. ISBN 0-387-95457-0.
- [195] Alexander Verpoorte. Limiting factors on early modern human dispersals: The human biogeography of late pleniglacial europe. *Quaternary International*, 201(1):77–85, 2009.
- [196] Dario Vianello, Federica Sevini, Gastone Castellani, Laura Lomartire, Miriam Capri, and Claudio Franceschi. Haplofind: A new method for high-throughput mtdna haplogroup assignment. *Human mutation*, 34(9):1189–1194, 2013.
- [197] Nicolas von Ahsen, Carl T Wittwer, and Ekkehard Schütz. Oligonucleotide melting temperatures under pcr conditions: nearest-neighbor corrections for mg²⁺, deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. *Clinical Chemistry*, 47(11):1956–1961, 2001.
- [198] Chaolong Wang, Xiaowei Zhan, Jennifer Bragg-Gresham, Hyun Min Kang, Dwight Stambolian, Emily Y Chew, Kari E Branham, John Heckenlively, The FUSION Study, Robert Fulton, et al. Ancestry estimation and control of population stratification for sequence-based association studies. *Nature genetics*, 46(4):409–415, 2014.
- [199] Spencer Wells. *Deep ancestry: Inside the genographic project*. National Geographic Books, 2006.

- [200] Bernhard Weninger, Lee Clare, Fokke Gerritsen, Barbara Horejs, Raiko Krauß, Jörg Linstädter, Rana Özbal, and Eelco J Rohling. Neolithisation of the aegean and southeast europe during the 6600–6000 calbc period of rapid climate change. 41:1–31, 12 2014.
- [201] Bernhard Weninger, Clare Lee, Gerritsen Fokke, Horejs Barbara, Krauß Raiko, Linstädter Jörg, Özbal Rana, et al. Neolithisation of the aegean and southeast europe during the 6600-6000 calbc period of rapid climate change. *Documenta Praehistorica*, 41, 2014.
- [202] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.
- [203] Sandra Wilde. *Populationsgenetik kupfer- und bronzzeitlicher Bevölkerungen der osteuropäischen Steppe*. Doktorarbeit, Johannes Gutenberg-Universität Mainz, 2014.
- [204] Sandra Wilde, Adrian Timpson, Karola Kiršanow, Elke Kaiser, Manfred Kayser, Martina Unterländer, Nina Hollfelder, Inna D Potekhina, Wolfram Schier, Mark G Thomas, et al. Direct evidence for positive selection of skin, hair, and eye pigmentation in europeans during the last 5,000 y. *Proceedings of the National Academy of Sciences*, 111(13):4832–4837, 2014.
- [205] Sewall Wright. Size of population and breeding structure in relation to evolution. *Science*, 87(2263):430–431, 1938.
- [206] Maniatis Y. Radiocarbon dating of the major cultural changes in prehistoric macedonia: recent developments in a century of research in prehistoric macedonia. 10:205 – 222, 2012.
- [207] Melinda A Zeder. Domestication and early agriculture in the mediterranean basin: Origins, diffusion, and impact. *Proceedings of the national Academy of Sciences*, 105(33):11597–11604, 2008.
- [208] C. Ziota. The neolithic settlement of “kleitos i“ and the new updates for prehistoric research. In G. Karamitrou-Mentessidi, editor, *The Archaeological Work in Upper Macedonia in 2009*, pages 211–230. Archaeological Museum of Aiani, 2011. (in Greek).
- [209] Magkouretsiou E Ziota C, Chondrogianni-Metochi A. The archaeological research at kleitos kozanis in 2009. *AEMTH*, 23:36–52, 2013.

Anhang I - Methoden

A.1 Experiment zur Silika Extraktion

Da sich die erwarteten Fragmentlängen der Proben in einem Bereich von 40 - 80 bp befinden, soll eine weitere Methode zur Aufreinigung des lysierten Knochenpulvers getestet werden. Diese bedient sich der Bindung der DNA Moleküle unter dem Vorkommen von chaotropen Salzen an Silika Partikel in Lösung. Es wurden 3 Proben mit einem Einsatz von 200mg Knochenpulver verwendet (Label:XII, Proben: Krk 7, Kea 1, Tha 1), welche zu Vergleichszwecken in einer parallelen Extraktion durch Phenol-Chloroform aufgereinigt wurden (Label XI). Je Probe wurde eine Leerkontrolle mitgeführt, um neben der allgemeinen Kontamination von Chemikalien auch Effekte von Kreuzkontaminationen messen zu können. Leerkontrollen wurden durch PCR Amplifikationen (Primer G3 - Tabelle 3.11, 42 Zyklen) auf Kontamination überprüft. Zur Auswertung der Silika Extraktion wurden die Extrakte dieser Versuchsreihe unter identischen Bedingungen in *Libraries* überführt und anschließend hinsichtlich des mitochondrialen Genoms angereichert. Die Auswertung erfolgt über die Ermittlung der Komplexität der einzelnen *Libraries* und der *Capture* Effizienz bzw. weiterer Parameter wie der Fragmentlänge.

Ergebnis

Für die Probe Krk 7 wurden aus jedem Extrakt jeweils 3 *Library* Parallelen erstellt und über die unverzerrte Molekülzahl nach qPCR ausgewertet. Schon bei dieser Messung wird die Steigerung der Komplexität der *Library* Moleküle deutlich: Im Vergleich zur Standardextraktion (Krk 7 XI) erhält man durch die Silikaextraktion das 229-fache an Molekülen (Tabelle A.55, Spalte "qPCR FI"). Anhand einer Bioanalyser Messung konnte eine Zunahme der durchschnittlichen Fragmentlänge in den *Libraries* nach Silika Extraktion festgestellt werden (Probe Krk 7, Spalte "Frag [bp]"). Die Proben Tha 1 und Kea 1 wurden über jeweils eine *Library* getestet. Der endogene DNA Gehalt wurde durch einen MiSeq Lauf bestimmt. Für die stark fragmentierten Proben Kea 1 und Tha 1 kann eine leichte Erhöhung der Molekülzahl festgestellt werden, jedoch beinhalten diese Libraries keine endogenen DNA Moleküle.

Tabelle A.55: *Library* Parameter zur Auswertung der Silika Extraktion

Extr.: Extraktion, Frag [bp]: Fragmentlänge in Basenpaaren ohne Adapter (Bioanalyser), qPCR FI: Anzahl an Molekülen in der *Library*, FI Kont.: theoretisches Kontaminationslevel errechnet anhand der Moleküle der *Library* Leerkontrollen (qPCR FI), endo [%]: endogener DNA Gehalt; Extr. V = 500mg Pulver, Standardextraktion Phenol-Chloroform; Extr. XI: 200mg Pulver, Standardextraktion Phenol-Chloroform; Extraktion XII: 200mg Pulver, Silika Extraktion nach Brotherton (2013, [23])

Library #	Probe	Extr.	Frag [bp]	qPCR FI	FI Kont. Level	endo [%]
LII.1	Krk 7.1	V	58	5,08E+08	0,0101%	
LII.2	Krk 7.2	V	51	6,17E+08	0,0083%	0,07
LII.3	Krk 7.3	V	48	5,75E+08	0,0089%	
LII.4	Krk 7.4	XI	56	7,91E+07	0,0647%	
LII.5	Krk 7.5	XI	40	2,21E+07	0,2313%	0,15
LII.6	Krk 7.6	XI	44	1,15E+08	0,0446%	
LII.7	Krk 7.7	XII	68	1,48E+10	0,0003%	
LII.8	Krk 7.8	XII	66	2,30E+10	0,0002%	
LII.9	Krk 7.9	XII	64	1,17E+10	0,0004%	
LIV.1	Kea 1	XI	64	2,37E+08	0,05%	0,02
LIV.2	Kea 1	XII	59	7,35E+08	0,02%	0,00
LIV.3	Tha 1	XI	65	5,67E+07	0,23%	0,13
LIV.4	Tha 1	XII	62	9,80E+08	0,01%	0,01

Weitere Untersuchungen können erst nach der Sequenzierung durchgeführt werden. Da der endogene DNA Gehalt dieser Proben recht gering ist (0 - 0,15%), wurde ein mitochondrialer Sequenzdatensatz für die Extrakte der Probe Krk 7 erstellt. Dass bisherige Sequenzierungen des mitochondrialen Ge-

noms (nach Anreicherung) auf 1/8 HiSeq Lane überdimensioniert waren (weiteres siehe Abschnitt 4.5), lässt sich daraus schließen, dass schon mit einer Ausbeute von 300.000 Sequenzen (PE) eine volle Abdeckung des Genoms erzielt werden kann (Tabelle A.56). Mit der Silika Extraktion konnte nach Entfernung der PCR Duplikate eine deutlich höhere Abdeckung des Genoms erreicht werden (Tabelle A.56, ALN Mkdup, cov), was direkt mit der gesteigerten Komplexität der *Libraries* zusammenhängt. Der Vorteil der 229-fachen Ausbeute an einzigartigen Molekülen wird in diesem Experiment besonders in der Duplikationsrate ("DUP [%]") und Kontaminationsbestimmung ("Kont. [%]") deutlich. Während die Sequenzierreaktion von Extraktion V mit 7776 alignierten Sequenzen bereits gesättigt ist (DUP = 99,8%), ist bei Extraktion XII mit 11444 alignierten Sequenzen erst ein Duplikationslevel von 80,31 % erreicht. Die gesteigerte Komplexität der *Libraries* in Extraktion XII zeigen ein maximales Kontaminationslevel von 1,3%, wohingegen in Extraktion V ein maximales Level von 6.7% besteht. Dies bestätigt das theoretische Kontaminationslevel errechnet auf Basis der qPCR auf *Fill In* Ebene (siehe Tabelle A.55).

Das eigentliche Ziel war mit Hilfe der Extraktion kleiner Fragmentlängen in stark degradierten Proben wie Kea 1 und Tha 1 zu favorisieren, was nicht erreicht werden konnte. Die Fragmentlängen der Silika Extraktionen liegen etwa 20 bp über der Standardextraktion (Tabelle A.56, Frag BAM [bp] & Abbildung A.49).

Tabelle A.56: Auswertung der Sequenzdaten nach *Capture*. Krk 7 V: Phenol-Chloroform Extraktion, Krk 7 XII: Silika Extraktion

(a) Parameter Qualität und Alignment

Seq: Anzahl Sequenzen (Rohdaten), Qual: Anzahl Sequenzen nach Qualitätsfilter (q15), join: Anzahl überlappende Sequenzen (PE), ALN mt: alignierte Sequenzen (Seq) und Verhältnis zu *join* [%], ALN Mkdup: nach Duplikatsentfernung

Probe	Seq	Qual	join	ALN mt	ALN mt [%]	ALN Mkdup
Krk 7 V	18254675	16782698	7692072	3935594	51%	7776
Krk 7 XII	300372	284345	133265	58126	44%	11444

(b) Fortsetzung

DUP %: Duplikationslevel, cov: mittlere Abdeckung der Zielregion (*coverage*), mD5/mD3: Deaminierungsrate am 5' - bzw. 3' Ende der Moleküle, Kont.: Kontaminationsangabe in %, Frag BAM: Fragmentlänge der alignierten Moleküle

Probe	DUP [%]	cov	mD5	mD3	Kont. [%]	Frag BAM [bp]
Krk 7 V	99,80%	31,6	0.419	0.464	0,7 - 6,7	67.21 (± 26.42)
Krk 7 XII	80,31%	60,1	0.407	0.455	0,1 - 1,3	86.96 (± 27.35)

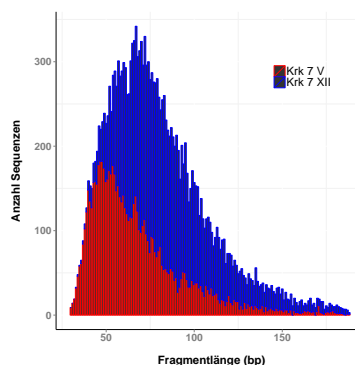


Abbildung A.49: Alignierte Fragmentlängen nach *Capture*

A.2 Library Erstellung

Untersuchung verschiedener *Library* Protokolle

Standardmäßig wurde in Phase A das *Library* Protokoll nach Meyer & Kircher 2010 ([120]) mit einer Modifizierung verwendet, bei welcher der Aufreinigungsschritt nach dem *blunt-end Repair* durch eine Inaktivierung des Enzyms ersetzt wird (siehe [18]). Der Hintergrund liegt im Cut-Off Wert des Aufreinigungskits (MinElute PCR Purification Kit - Qiagen, Hilden), welcher bei 70bp liegt und folglich kleinere Moleküle verloren gehen. Als Folge findet die Ligation in einem größeren Volumen statt, um die Konzentration der Chemikalien und Enzyme aufrecht zu erhalten. Meyer und Kircher modifizierten ihr bestehendes Protokoll ebenfalls durch eine Inaktivierung des Enzyms nach dem *Fill In* Schritt [84]. Der folgende Versuch dient der Untersuchung verschiedener Protokollabläufe während der *Library* Erstellung und umfasst:

- das Original Protokoll nach Meyer und Kircher ([120], Protokoll A),
- die modifizierte Version, bei welcher der erste Aufreinigungsschritt durch eine Inaktivierung ersetzt wird ([18], Protokoll B),
- die Weiterentwicklung durch Meyer und Kircher 2012, bei welcher der dritte Aufreinigungsschritt durch eine Inaktivierung ersetzt wird ([84], Protokoll C)
- und einer Kombination der Protokolle B/C (Protokoll D).

Ein schematischer Aufbau des Versuches ist in Abbildung A.50 skizziert.

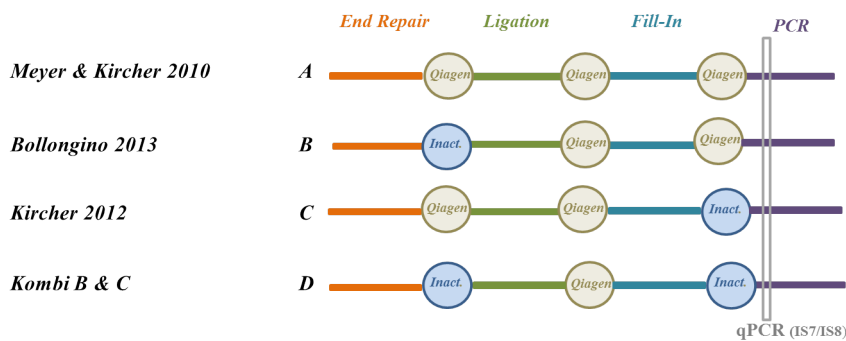


Abbildung A.50: Versuchsaufbau zur Testung verschiedener Protokolle zur *Library* Erstellung

Diese Versuchsreihe wurde mit drei Proben durchgeführt, die in Vorversuchen unterschiedliche Moleküleigenschaften bezüglich Fragmentlänge und Molekülzahl zeigten. Probe Vra 2 zeigte eine geringe Molekülzahl (10^6) mit recht großer Fragmentlänge ($180\text{bp} \pm 25\text{bp}$), bei Probe Tem 6 ist die Molekülzahl erhöht (10^7) mit einer Fragmentlänge von $150\text{bp} \pm 22\text{bp}$ und die Probe Kru 3 zeigt die meisten Moleküle (10^{8-9}) und eine Fragmentlänge von $60\text{bp} \pm 19\text{bp}$. Die Auswertung erfolgt mittels quantitativer Real-Time PCR nach dem Adapter *Fill In* (siehe Anhang, Abschnitt A.3), um eine Verzerrung des Ergebnis durch Amplifikation auszuschließen. Mittels einer Bioanalyzer Messung nach PCR werden Fragmentlängen und das Auftreten von Primer-/Adapterdimeren bestimmt. Jedes Protokoll wurde als Duplikat pipettiert und dies von zwei Personen. Je Probe wurde durch Mehrfachextraktion ein sogenanntes "Megaextrakt" erstellt um eine gleiche Ausgangssituation für alle Protokolle zu gewährleisten (Extraktionsprotokoll: Standard Phase A - leichte Erhöhung EDTA Menge auf 4ml, Inkubation

über Nacht, siehe Tabelle A.57). Testproben: Vra 2 = Vra 2, Frankreich (Bronzezeit), TEM 6a (siehe [203]), Kru 3 (siehe [170]).

Tabelle A.57: Zusammensetzung der Megaextrakte der drei Testproben.

(# Extr.: Anzahl Extraktionen, Extr. Vol.: Extraktionsvolumen Megaextrakt, Norm.: Normalisierungsfaktor zum Vergleich unterschiedlicher Extraktionen/Proben (siehe qPCR Abschnitt))

Probe	# Extr.	Pulver [g]	Extr. Vol. μl	Norm [mg/ μl]
Vra 2	11	5,5	3110	1,7
TEM 6a	11	5,34	3025	1,76
Kru 3	11	5,53	2770	1,99

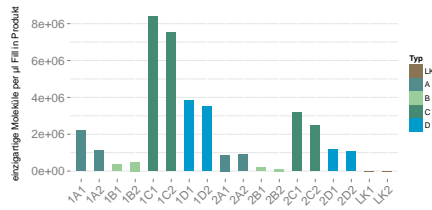
Die drei Testproben wurden jeweils an einem eigenen Versuchstag bearbeitet und für Protokoll A/C/D wurde mit einer dritten *Library* Parallele ein Polymerasetest durchgeführt. Der Polymerasetest umfasste folgende Varianten:

- Vra 2 – PCR Bedingungen nach Kircher *et al.* 2012 (100 μl Ansatz, AmpliTaq, Thermo Pol Puffer (NEB), Primerkonzentration 400nm)
- TEM 6a – KAPA Hifi Uracil⁺ Polymerase (PeqLab, nach Herstellerangaben)
- Kru3 – Pfu Turbo Cx Polymerase (Agilent, nach Herstellerangaben)

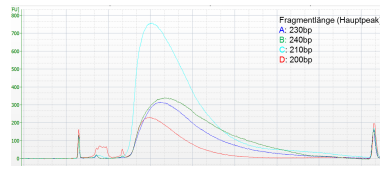
Zusätzlich werden die Effekte der *bead* Aufreinigung (SPRI-*beads*, Beckmann Coulter) im Vergleich zur Säulenaufreinigung (MinElute Purification Kit, Qiagen) anhand des Protokoll A getestet.

Ergebnisse

Alle drei Proben zeigen in der Versuchsdurchführung für Protokoll C eine erhöhte Zahl an ligierten Molekülen (siehe Abbildung A.51a,A.52a,A.53a). Hinzu kommen Auswirkungen auf die Fragmentlängenverteilungen in amplifizierten Libraries. Liegt eine breite Verteilung vor, Vra 2 & Tem 6, zeigt sich bei Protokoll C/D eine leichte Verschiebung in Richtung kleinerer Moleküle (siehe Abbildung A.51b,A.52b,A.53b). Bei der Prüfung verschiedener PCR Protokolle konnte durch das verwendete Protokoll nach Kircher *et al.* 2012 kein PCR Produkt erzielt werden (Abbildung A.54a). Die Anwendung zweier modifizierter Polymerasen zeigten ein positives PCR Produkt (Abbildung A.54b, A.54c). Die Effizienz unterschiedlicher Methoden der Aufreinigung zwischen einzelnen enzymatischen Reaktionen der *Library* Erstellung ist stark abhängig von der Fragmentlänge einer Probe, da beide Methoden unterschiedliche Schwellenwerte für die Retention von Moleküllängen besitzen.

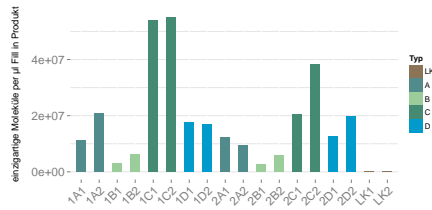


(a) Einzigartige Molekülzahl in Einzellibraries der Fill In Produkte

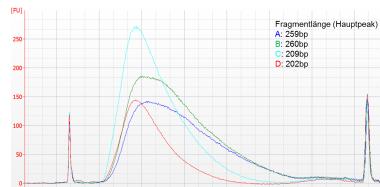


(b) Fragmentlängenbestimmung mittels Bioanalyzer Messung nach PCR

Abbildung A.51: Auswertung Vra 2

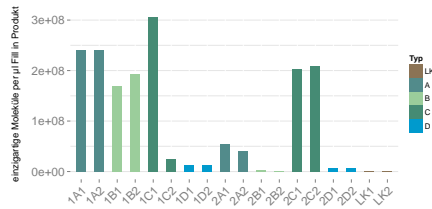


(a) Einzigartige Molekülzahl in Einzellibraries der Fill In Produkte

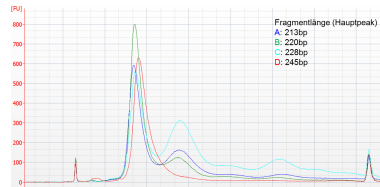


(b) Fragmentlängenbestimmung mittels Bioanalyzer Messung nach PCR

Abbildung A.52: Auswertung Tem 6



(a) Einzigartige Molekülzahl in Einzellibraries der Fill In Produkte



(b) Fragmentlängenbestimmung mittels Bioanalyzer Messung nach PCR

Abbildung A.53: Auswertung Kru3

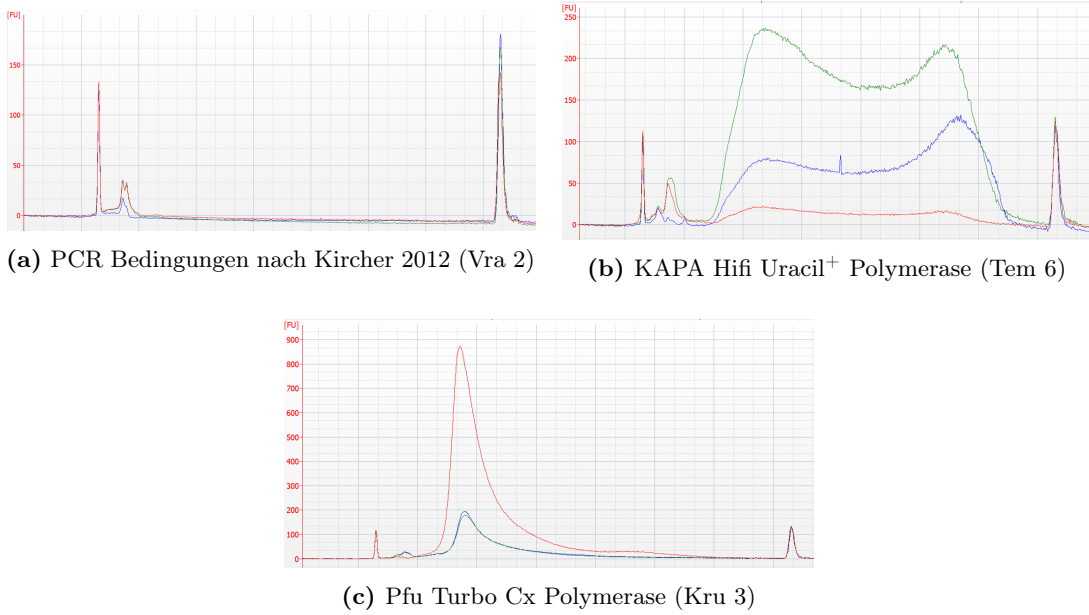


Abbildung A.54: Bioanalyser Messung nach Anwendung diverser PCR Protokolle und Polymerase (als Vergleich für Fragmentlängenverteilung siehe parallele Amplifikationen mit Taq Polymerase in Abbildungen A.51, A.52 & A.53)

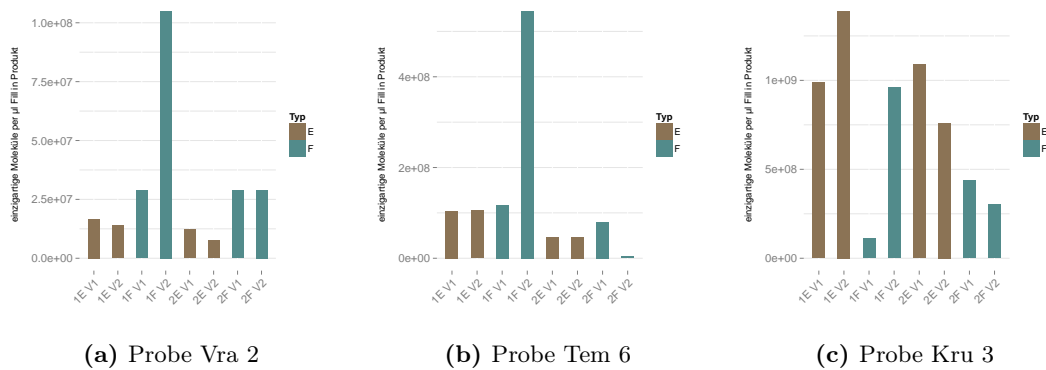


Abbildung A.55: Test Aufreinigungsmethoden für die *Library* Erstellung. E: Silika, F: SPRIbeads

Nonsense Hybrid - Sequenz und Ansatz

Tabelle A.58: Sequenzen der Positivkontrolle (*Nonsense Hybrid*).

Oben: alte Sequenz, wurde bis Juli 2014 genutzt; Unten: neue Sequenz mit 5' - und 3' Überhangen

<i>Nonsense Sequenz (5'-3')</i>	
Nonsense 1	CCATCTCATCCCTGCGTGTCAACTGACTAAACTAGGTGCCACGTCGTGAAAGTCTGACAACCTATCCCTGTGTGCCTTG
Nonsense 2	CAAGGCACACAGGGGATAGGTTGTCTAGACTTTCACGACGTGGCACCTAGTTTGTAGTCAGTTGACACGCAGGGATGAGATGG
<i>NEU_NH(14) - mit 3'- und 5' - Überhang</i>	
NH(14)_minus10bp	ACTAACCATCTCATCCCTGCGTGTCACTAGGTGCCACGTCGTGACCTATCCCTGTGTGCCTTGAAGTCT
NH(14)_revcomp	TTGTCAGACTTCAAGGCACACAGGGATAGGTCACGACGTGGCACCTAGTGACACGCAGGGATGAGATGGTTAGTCAGTT

Tabelle A.59: Ansatz des *Nonsense Hybrid (NH)*

1. Äquimolare Kombination von beiden Strängen		
	Nonsense 1	Nonsense 2
Stocklösung	100 μ M	100 μ M
Volumen aus Stocklsg.	2 μ l	2 μ l
H ₂ O	198 μ l	198 μ l
Finale Konzentration	1 μ M	1 μ M
2. Hybridisierung		
Stufe	Temperatur	Zeit
Initiale Denaturierung	95°C	10 Sek
graduelles Absenken	0,1°C/sec to 12°C	14 Min
Hold	12°C	-
3. für 100ng in 50 μ l (=1 Library Ansatz): 4 μ l NH + 46 μ l H ₂ O)		

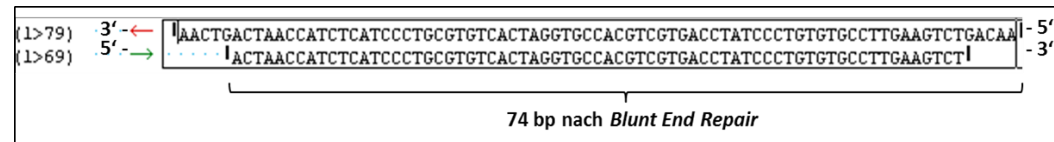


Abbildung A.56: Alignment neue Positivkontrolle zur Darstellung der überhängenden Enden und Produktlänge nach *Blunt End Repair*

A.3 Quantifizierung

qPCR

Bei der quantitativen PCR wird die exponentielle Phase einer ablaufenden PCR Reaktion genutzt, um die Molekülzahl einer Probe zu ermitteln. Je mehr Ausgangsmoleküle eine Probe besitzt, desto schneller hebt sich die Fluoreszenz aus dem Hintergrund ab und erreicht den exponentiellen Bereich einer PCR Reaktion. Unter gleichen Reaktionsbedingungen (Reaktionsansatz, Temperatur, PCR Zyklusbedingungen) wird eine Standardreihe mitgeführt, welche aus einer Serie von 10-fach Verdünnungen einer bekannten Molekülzahl besteht. Abhängig ihrer Ausgangskonzentration erreichen die Standards der Reihe mit einem Abstand von 3,34 Zyklen den exponentiellen Bereich. Eine gute Übersicht und Erklärung der Versuchsdurchführung bietet [25].

Reaktionsbedingungen

Ein künstlich synthetisiertes Oligonukleotid, welches die Adapterstrukturen eines *Library* Moleküls trägt, kommt in einer qPCR Reaktion mit einer Konzentration des ersten Standards von 10^8 Molekülen zum Einsatz (siehe Abbildung A.57). Die Standardmoleküle liegen als vorwärts und revers komplementäre Stränge in der Reaktion vor. Da der erste Schritt einer PCR Reaktion die Denaturierung ist, entfällt eine Hybridisierungsreaktion zum Ansatz eines doppelsträngigen Standardmoleküls. Anhand dem Schema dargestellt in Tabelle ?? werden beide Stränge äquimolar in einen Reaktionsansatz gebracht und anschließend auf die Konzentrationen der Standardreihe verdünnt.

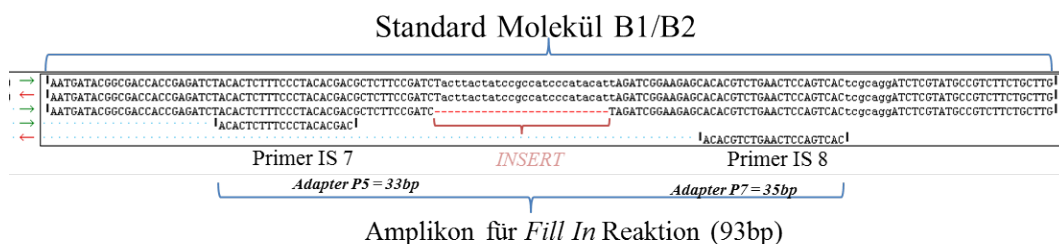


Abbildung A.57: Sequenz des verwendeten Standardmoleküls mit Abbildung der Primerbindungsstellen für IS7 und IS8.

Zur Erstellung der Standardreihe werden die einzelnen Stufen (S1-S5, 10^8 - 10^4) als Duplikate gemessen. Die Bestimmung der Molekülzahl in Proben erfolgt in einer Tripletmessung. Als Leerkontrolle der qPCR Reaktionen dienen zwei Reaktionsansätze ohne Zugabe von DNA (*no template control*, NTC). Für die Amplifikation der Moleküle während der qPCR wird der *KAPA SYBR[®] FAST* qPCR Master Mix (2X) Universal verwendet. Es enthält eine modifizierte Taq-Polymerase, welche aufgrund ihrer effektiven Arbeitsweise die Fähigkeit besitzt 1000 Basen pro Minute zu ligieren. Der Hersteller garantiert die Amplifikation von Uracil-tragenden Molekülen. Die Konzentrationen der Primer IS7/IS8 beträgt 200nM (für die *Library* Primer wurden diverse Testreihen bezüglich Annealing Temperatur und Primerkonzentration durchgeführt - Daten werden nicht gezeigt). Die qPCR Reaktion wird nach Herstellerangaben auf einer *Step One Plus* der Firma Applied Biosystems/Life Technologies durchgeführt. Die Probennahme für die qPCR Reaktion erfolgt während der *Library* Erstellung nach der Enzyminaktivierung der *Fill In* Reaktion. Es wird $1\mu\text{l}$ *Fill In* Produkt in $39\mu\text{l}$ Wasser überführt - dies stellt die Verdünnung dar, welche in der qPCR Reaktion zum Einsatz kommt.

Auswertung

Nach einer qPCR Reaktion wird anhand des Amplifikationsplots der Schwellenwert auf einen Fluoreszenzwert von 0,2 gesetzt. Die ermittelten Ct-Werte der Standardreihe werden gegen den Logarithmus (\log_{10}) der bekannten Kopienzahl in einem Diagramm aufgetragen und zwischen den einzelnen Standards wird eine Regressionsgerade abgebildet. Über die Geradengleichung ($y=mx+b$) der Regressionsgerade kann nun anhand der Ct-Werte (y) der einzelnen Proben die unbekannte, logarithmische Molekülzahl (x) ermittelt werden. Unter Anwendung der Exponentialfunktion zur Basis 10 wird die Molekülzahl umgewandelt. Um den tatsächlichen Wert der Zahl an sequenzierfähigen *Library* Molekülen pro μl *Fill In* Produkt zu bestimmen muss das Ergebnis mit dem Verdünnungsfaktor (40) multipliziert werden. Da sich der Fluoreszenzfarbstoff SYBR Green I gleichmäßig in die DNA Doppelhelix einlagert ist die gemessene Molekülzahl abhängig von der Fragmentlängenverteilung der DNA Moleküle innerhalb einer Probe. Bei gleicher Fluoreszenzstärke besitzen kleiner fragmentierte Proben eine höhere Zahl an DNA Strängen als weniger stark fragmentierte Proben. Fragmentlängenverteilungen bzw. eine Mittelwert kann durch eine Bioanalyser Messung bestimmt werden. Für jede Probe kann auf diese Weise ein Korrekturfaktor eingerechnet werden, welcher in Abhängigkeit zu der Länge der Standardmoleküle (89bp) steht. Dabei muss für Standard und *Library* Molekül der Probe die Adapterstruktur, gegeben durch IS7/IS8, mit einbezogen werden.

A.4 Capture des mitochondrialen Genoms

Optimiertes Design des mitochondrialen *Capture* (Mai 2012)

Bei der Erstellung der *bait* Sequenzen des ersten mitochondrialen *Capture* konnte die von Agilent bereitgestellte Software lediglich das Mitochondrium von hg19 (GRCh37/hg19 = African Yorouban) als Grundlage zur Verfügung stellen. Um das Spektrum der Haplogruppen Diversität abzudecken, wurden 29 Haplotypen per Hand erstellt (siehe Diplomarbeit Kreutzer (2011) [94]). Erweiterung auf dem *eArray* Server boten für das neue Design die Möglichkeit einzelne FASTA Dateien einzuladen („*Sure-Select RNA enrichment*“), über welche nun das *tiling* (überlappende Abdeckung des mitochondrialen Genoms) erstellt werden konnte. Es wurde eine *tiling* Frequenz von 8x gewählt, was bedeutet, dass bei einer *bait* Länge von 120bp alle 15bp eine neue *bait* Sequenz erstellt wird und jedes Base 8-fach abgedeckt ist. Als Grundstock für die Abdeckung des kompletten mitochondrialen Genoms dienten übergeordnete Haplogruppen der Linien M*, D, N*, R*, R0 und U (siehe Tabelle A.60), welche aus dem phylogenetischen Stammbaum entnommen wurden (siehe Tabelle A.60, [193], *mtDNA tree Build 14*, 5. April 2012). Zusätzliche *bait* Sequenzen wurden eingefügt, um den Ringschluss zu gewährleisten. Da die Diversität in der Kontrollregion des Mitochondriums am höchsten ist, wurde ein zusätzliches *tiling* dieser Region mit 19 verschiedenen Haplotypen durchgeführt (Position 15900-16569, 1-600, *tiling* Frequenz 10x, siehe Table A.61).

Tabelle A.60: Genutzte FASTA Sequenzen mit Referenz für die Erstellung der *bait* Sequenzen über das mitochondriale Genom

Haplotyp	NCBI Acc.
HG M*	JQ702247
HG D	JN253391
HG N1a	GU290213
HG R1	HM852841
HG R0ab	JF717361
HG U1a'c	HM852844

Tabelle A.61: Referenz der verwendeten FASTA Sequenzen zur Erstellung des zusätzlichen *tilings* über die Kontrollregion (erweiterte CR: 15900 - 600)

Haplotyp	Vorgänger HGs	NCBI Nummer
<i>M</i>-Untergruppen		
C1a	aus M*/M8	EU007858
Z	aus C	AY519493
E1a	aus M*/M9	EF093536
Q1	aus M*/M29	AY289090
<i>U</i>-Untergruppen		
U5a1		GU296570
K1a	aus U8/U8b1	HM043706
<i>N</i>-Untergruppen		
I	aus N1a/N1e'I	JF298212
W	aus N2/N2a	AY339475
A	aus N22	AY255144
X1a	aus A/O/S	EU600318
<i>R</i>*-Untergruppen		
B4	aus R11/B6	EU597559
F1a1	aus R9	AP008906
R2	Vorläufer JT	AY714007
J1b	aus R2	JQ064573
T2	aus J	JF707633
<i>R0</i>-Untergruppen		
HV	aus R0	HM776579
V	aus HV/HV0	HQ711364.1
H	aus HV	JQ705953
H2a2a1	rCRS	NC_012920

Anhang II - Abbildungen, Tabellen und Publikation

B.1 Probenvorbereitung

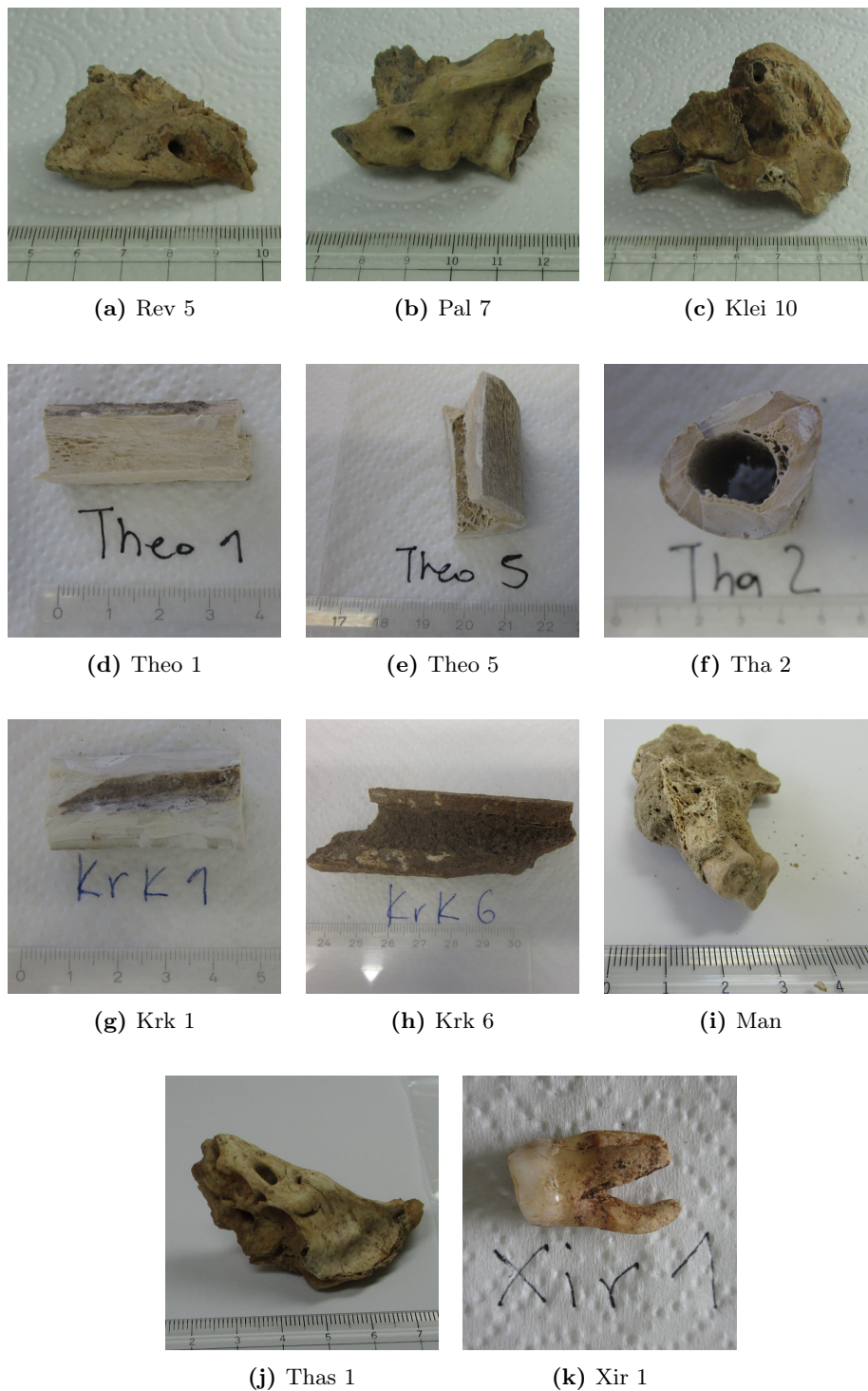


Abbildung B.58: Darstellung ausgewählter Proben Proben

B.2 Ergebnisse PCR und Sanger Sequenzierung Phase A/B

Tabelle B.62: Ergebnisse der PCR Daten erstellt in Phase A und B.

(Extr.: verwendeten Extraktionen, HG: Haplogruppe, Abdeckung HVS-I: 16013-16409, X: 6336-6403, U/K: 12303-12352, K/J1c: 14783-14839)

Probe	Extr.	HG	Abdeckung	SNPs
Fra 1	I - III	X	HVS-I; X	6371T 16189C 16223T 16274A 16278T
Theo 1	I + II	K	HVS-I; U/K; K/J1c	12308G 14798C 16224C 16311C
Theo 4	I + II	T2k	HVS-I	16126C 16291T 16294T 16296T
Theo 5	I - III	T2k	HVS-I	16126C 16291T 16294T 16296T
Mau 1	I + II	H5	HVS-I; U/K; K/J1c	12308A 14798T 16304C
Mau 2	I	T2b23	HVS-I	16126C 16147T 16294T 16296T 16297C 16304C
Mau 4	I	rCRS	16013-16250;16285-16409	
Mau 5	I	T2	HVS-I	16126C 16294T 16296T
Xir 1	I - III	H14b1	HVS-I	16126C
Xir 2	I + II	T1a	HVS-I	16126C 16163G 16186T 16189C 16294T
Xir 3	I	T2c1d1	HVS-I	16126 16292 16294
Xir 4	I		HVS-I	16224C 16311C
Xir 5	I	<i>zu geringe Erfolgsrate in PCR und Sequenzierung, Datenmenge zu gering für Analyse</i>		
Tha 4	I - III	T2k	HVS-I	16126C 16291T 16294T 16296T
Fra 6	I - III	<i>zu geringe Erfolgsrate in PCR und Sequenzierung, Datenmenge zu gering für Analyse</i>		
Fra 8	I - IV	H1aj1	HVS-I; U/K; K/J1c	12308G 16192T 16295T 16343G
Klei 1	I + II	K	HVS-I; U/K; K/J1c	12308G 14798C 16224C 16311C
Klei 2	I + II	J1c	HVS-I; U/K; K/J1c	12308A 14798C 16069T 16126C
Klei 3	I - III	K1a1	HVS-I	16048A 16093C 16224C 16311C
Klei 4	I	K1a4a1	HVS-I	16224 16311
Klei 5	I	H+16311	HVS-I	16311C
Klei 6	I	H29	HVS-I	16355T
Krk 1	I + II	K1a1	HVS-I; U/K; K/J1c	12308G 14798C 16093C 16224C 16311C
Krk 2	I	rCRS		
Krk 4	I - IV	U8b1a1	HVS-I; K/J1c	12308G 16066G 16129A 16183C 16189C 16234T
Krk 7	I - III	K1a1	HVS-I; U/K; K/J1c	12308G 14798C 16093C 16224C 16311C
Krk 8	I	N1a1a1a	HVS-I	16147A 16172C 16223T 16248T 16320T 16355T
Krk 9	I	rCRS		
Pal 1	I	U4b3		16294T 16356C 16362C
Pal 4	I + II	U2e	HVS-I	16051G 16129C 16183C 16189C 16362C
Pal 5	I + II	U2e	HVS-I	16051G 16129C 16183C 16189C 16362C
St 2	I	N1a1a1a3	HVS-I	16147A 16154C 16172C 16223T 16248T 16320T 16355T
St 3	I	<i>zu geringe Erfolgsrate in PCR und Sequenzierung, Datenmenge zu gering für Analyse</i>		

B.3 DNA Screening (Leerkontrollen)

Tabelle B.63: Molekülzahl, Fragmentlänge (Bioanalyzer; BA) und MiSeq *Screening* Daten einzelner Leerkontrollen aus dem Versuchsablauf (in Abbildung 4.23 sind alle Leerkontrollen graphisch dargestellt)

Probe	Phase	Material	qPCR FI	qPCR FI (N)	endo	uniq	uniq (N)
LK 1	C	LK	3,37E+05	3,37E+05	0,78	2,63E+03	2,63E+03
LK 2	C	LK	1,28E+05	1,28E+05	0,05	6,40E+01	6,40E+01
LK 3	C	LK	6,97E+03	6,97E+03	0,3	2,09E+01	2,09E+01
LK 4	C	LK	4,43E+04	4,43E+04	0,4	1,77E+02	1,77E+02
LK 5	B	LK	9,26E+04	9,26E+04	0,8	7,41E+02	7,41E+02
LK 6	B	LK	3,64E+05	3,64E+05	0,23	8,37E+02	8,37E+02
ExKO 1	B	ExKO	1,22E+05	1,22E+05	1,1	1,34E+03	1,34E+03
ExKO 2	C	ExKO	6,33E+04	6,33E+04	2,31	1,46E+03	1,46E+03
ExKO 3	C	ExKO	1,09E+05	1,09E+05	2,95	3,21E+03	3,21E+03
ExKO 4	C	ExKO	3,91E+05	3,91E+05	1,89	7,39E+03	7,39E+03
MKO 1	C	MKO	1,07E+05	1,07E+05	1,12	1,20E+03	1,20E+03
MKO 2	C	MKO	1,14E+05	1,14E+05	5,81	6,62E+03	6,62E+03
MKO 3	C	MKO	7,21E+04	7,21E+04	2,24	1,62E+03	1,62E+03
MKO 4	C	MKO	3,04E+05	3,04E+05	1,5	4,56E+03	4,56E+03
MKO 5	C	MKO	4,02E+05	4,02E+05	2,868	1,15E+04	1,15E+04

B.4 Durchführung und Parameter einzelner *Capture* Experimente

Polymerasetest

Tabelle B.64: Datenauswertung Polymerasetest nach 50bp SE Sequenzierung (MiSeq).

Abkürzungen siehe Tabelle 4.31. A: AccuPrime Polymerase, K: KAPA HiFi Polymerase, T: AmpliTaq Gold[®] Polymerase, T/H: Kombination T mit Herculase Polymerase

Probe-Pol	Seq	ALN mt	ALN mt [%]	ALN Mkdup	DUP [%]	mD5	mD3	auth. [%]	Daten
NS1A-A	37598	19586	52,1%	2636	86,5%	0,205	0,080	98,5	
NS1A-K	37598	18793	50,0%	2529	86,5%	0,406	0,141	93,9	
NS1A-T	37598	18339	48,8%	3428	81,3%	0,350	0,080	99,1	
NS1A-TH	37598	19048	50,7%	3831	79,9%	0,357	0,102	97,7	
NS2A-A	22899	10287	44,9%	613	94,0%	0,187	0,112	99,6	
NS2A-K	22899	9345	40,8%	687	92,6%	0,392	0,085	96,4	
NS2A-T	22899	10359	45,2%	998	90,4%	0,331	0,088	89,7	
NS2A-TH	22899	10750	46,9%	3711	65,5%	0,384	0,046	99,2	
Pal2-A	56849	41729	73,4%	2630	93,7%	0,252	0,080	97,5	
Pal2-K	56849	34026	59,9%	2484	92,7%	0,361	0,107	98,0	
Pal2-T	56849	33541	59,0%	3946	88,2%	0,360	0,100	99,8	
Pal2-TH	56849	33112	58,2%	4148	87,5%	0,390	0,083	99,5	
Xir-A	17507	13253	75,7%	1568	88,2%	0,289	0,206	75,5	
Xir-K	17507	9798	56,0%	1341	86,3%	0,430	0,333	99,5	
Xir-T	17507	8622	49,2%	1535	82,2%	0,453	0,271	99,3	
Xir-TH	17507	12287	70,2%	1266	89,7%	0,468	0,281	99,5	

B.5 Durchführung und Parameter einzelner Capture Experimente

Ermittlung Sequenzdatenmenge

Tabelle B.65: Downsample Experiment

Probe	Seq	Seqs kept	[%] kept	ALN mt	ALN mt [%]	ALN Mkdup	DUP [%]	cov
1-Pal4b	10000	9703	97,0%	5768	59,4%	2805	51,37%	8,54
2-Pal4b	20000	19468	97,3%	11434	58,7%	3990	65,10%	12,14
3-Pal4b	50000	48792	97,6%	28550	58,5%	5691	80,07%	17,28
4-Pal4b	100000	97564	97,6%	57286	58,7%	6969	87,83%	21,15
5-Pal4b	200000	195142	97,6%	114732	58,8%	8057	92,98%	24,41
6-Pal4b	500000	487403	97,5%	287703	59,0%	9412	96,73%	28,42
7-Pal4b	1000000	974718	97,5%	579833	59,5%	10262	98,23%	30,92
1-Pal5	10000	9646	96,5%	3898	40,4%	2870	26,37%	8,68
2-Pal5	20000	19349	96,7%	7749	40,0%	4577	40,93%	13,85
3-Pal5	50000	48394	96,8%	19543	40,4%	7241	62,95%	21,86
4-Pal5	100000	96726	96,7%	39144	40,5%	9311	76,21%	28,04
5-Pal5	200000	193502	96,8%	78445	40,5%	11085	85,87%	33,27
6-Pal5	500000	483283	96,7%	197764	40,9%	12870	93,49%	38,42
7-Pal5	834403	807178	96,7%	332640	41,2%	13620	95,91%	40,53
1-Xir1a	10000	9836	98,4%	5149	52,3%	4368	15,17%	12,85
2-Xir1a	20000	19714	98,6%	10307	52,3%	7550	26,75%	22,29
3-Xir1a	50000	49378	98,8%	25597	51,8%	13578	46,95%	40,26
4-Xir1a	100000	98459	98,5%	51347	52,2%	18752	63,48%	55,82
5-Xir1a	200000	196140	98,1%	102405	52,2%	23153	77,39%	69,2
6-Xir1a	500000	491541	98,3%	257837	52,5%	27199	89,45%	81,56
7-Xir1a	1000000	983218	98,3%	518100	52,7%	29043	94,39%	87,15
1-Klei3	10000	9500	95,0%	4117	43,3%	2619	36,4%	14,26
2-Klei3	20000	19144	95,7%	8207	42,9%	3861	53,0%	21,01
3-Klei3	50000	47930	95,9%	20636	43,1%	5649	72,6%	30,32
4-Klei3	100000	95822	95,8%	41109	42,9%	6935	83,1%	37
5-Klei3	200000	191584	95,8%	82335	43,0%	8027	90,3%	42,34
6-Klei3	500000	480000	96,0%	207235	43,2%	9130	95,6%	47,89
7-Klei3	1000000	960485	96,0%	417905	43,5%	9848	97,6%	51,51
1-Krk9	10000	9304	93,0%	5139	55,2%	3359	34,6%	16,31
2-Krk9	20000	18793	94,0%	10375	55,2%	5020	51,6%	24,21
3-Krk9	50000	47225	94,5%	26110	55,3%	7212	72,4%	34,17
4-Krk9	100000	94664	94,7%	52292	55,2%	8531	83,7%	39,72
5-Krk9	200000	189297	94,6%	104724	55,3%	9519	90,9%	43,65
6-Krk9	500000	474314	94,9%	262207	55,3%	10395	96,0%	46,96
7-Krk9	1000000	948818	94,9%	526246	55,5%	10860	97,9%	48,75
1-Mau2	10000	9268	92,7%	4635	50,0%	3437	25,8%	18,62
2-Mau2	20000	18791	94,0%	9285	49,4%	5438	41,4%	29,32
3-Mau2	50000	47170	94,3%	23324	49,4%	8482	63,6%	45,3
4-Mau2	100000	94385	94,4%	46400	49,2%	10491	77,4%	55,4
5-Mau2	200000	188925	94,5%	93060	49,3%	11976	87,1%	62,37
6-Mau2	500000	472905	94,6%	232660	49,2%	13366	94,3%	68,61
7-Mau2	1000000	947683	94,8%	467902	49,4%	14117	97,0%	71,94
1-St2B	10000	9429	94,3%	3986	42,3%	3316	16,8%	16
2-St2B	20000	19044	95,2%	8163	42,9%	5924	27,4%	28,87
3-St2B	50000	47839	95,7%	20802	43,5%	10568	49,2%	52,15
4-St2B	100000	95682	95,7%	41583	43,5%	14373	65,4%	71,45
5-St2B	200000	191036	95,5%	82712	43,3%	17676	78,6%	88,35
6-St2B	500000	478699	95,7%	207455	43,3%	20814	90,0%	103,56
7-St2B	1000000	956767	95,7%	417807	43,7%	22358	94,6%	110,83
1_Mau1	10000	9204	92,0%	5161	56,1%	3055	40,8%	17,8
2_Mau1	20000	18677	93,4%	10436	55,9%	4375	58,1%	25,1
3_Mau1	50000	47301	94,6%	26374	55,8%	5797	78,0%	32,72
4_Mau1	100000	94857	94,9%	52771	55,6%	6515	87,7%	36,3
5_Mau1	200000	188918	94,5%	104921	55,5%	7036	93,3%	38,9
6_Mau1	500000	473199	94,6%	263626	55,7%	7584	97,1%	41,8
7_Mau1	1000000	946400	94,6%	528952	55,9%	7981	98,5%	44,2
1_Xir4	10000	9038	90,38%	2195	24,29%	912	58,45%	5,2
2_Xir4	20000	18409	92,05%	4583	24,90%	1176	74,34%	6,63
3_Xir4	50000	46761	93,52%	11650	24,91%	1446	87,59%	7,9
4_Xir4	100000	93819	93,82%	23446	24,99%	1613	93,12%	8,7
5_Xir4	200000	186624	93,31%	46920	25,14%	1758	96,25%	9,3
6_Xir4	500000	467562	93,51%	117973	25,23%	1951	98,35%	10,2
7_Xir4	1000000	936808	93,68%	236619	25,26%	2133	99,10%	11,1

Schmelztemperatur von *Capture* Komponenten**Tabelle B.66:** Parameter zur Berechnung der Schmelztemperatur der einzelnen Duplexkomponenten während einer *Capture* Reaktion. SeqL = Sequenzlänge

Hybridisierung (Hyb): 650mM Na ⁺				
Komponente	Sequenz	GC Gehalt [%]	SeqL	MELTING
BO4.p7	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	53	34	82.69 deg C
BO6.p7	CAAGCAGAAGACGGCATAACGAGAT	50	24	75.36 deg C
BO.p5	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	52	33	81.66 deg C
DNA 30bp (GCn)	TATAGTTTAAACAAAACGAATGATTTTCGAC	27	30	65.86 deg C
DNA 60bp (GCn)	TATAGTTTAAACAAAACGAATGATTTTCGACTCATTAAATTATGATAATCATATTTACCAA	22	60	70.19 deg C
DNA 30bp (GCm)	GCTATAGCAATAAAAAC TAGGAATAGCCCC	43	30	79.19 deg C
DNA 60bp (GCm)	GCTATAGCAATAAAAAC TAGGAATAGCCCCCTTTCAC T TCTGAGTCCCAGAGGTTACCCAA	45	60	88.86 deg C
DNA 30bp (GCm)	AAGGCCCCACC CCAGTCTCAGCCCTACTCC	67	30	97.86 deg C
DNA 60bp (GCm)	TCTCCACCCCTGACTCCCCTCAGCCATAGAAGGCCCCACCC CAGTCTCAGCCCTACTCCA	63	60	103.52 deg C
Waschschrift 1 (W1): 150mM Na ⁺				
Komponente	Sequenz	GC Gehalt [%]	SeqL	MELTING
BO4.p7	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	53	34	74.10 deg C
BO6.p7	CAAGCAGAAGACGGCATAACGAGAT	50	24	66.77 deg C
BO.p5	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	52	33	73.07 deg C
DNA 30bp (GCn)	TATAGTTTAAACAAAACGAATGATTTTCGAC	27	30	57.27 deg C
DNA 60bp (GCn)	TATAGTTTAAACAAAACGAATGATTTTCGACTCATTAAATTATGATAATCATATTTACCAA	22	60	61.60 deg C
DNA 30bp (GCm)	GCTATAGCAATAAAAAC TAGGAATAGCCCC	43	30	70.60 deg C
DNA 60bp (GCm)	GCTATAGCAATAAAAAC TAGGAATAGCCCCCTTTCAC T TCTGAGTCCCAGAGGTTACCCAA	45	60	80.27 deg C
DNA 30bp (GCm)	AAGGCCCCACC CCAGTCTCAGCCCTACTCC	67	30	89.27 deg C
DNA 60bp (GCm)	TCTCCACCCCTGACTCCCCTCAGCCATAGAAGGCCCCACCC CAGTCTCAGCCCTACTCCA	63	60	94.94 deg C
Waschschrift 2 (W2): 15mM Na ⁺				
Komponente	Sequenz	GC Gehalt [%]	SeqL	MELTING
BO4.p7	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	53	34	58.15 deg C
BO6.p7	CAAGCAGAAGACGGCATAACGAGAT	50	24	50.81 deg C
BO.p5	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT	52	33	57.12 deg C
DNA 30bp (GCn)	TATAGTTTAAACAAAACGAATGATTTTCGAC	27	30	41.31 deg C
DNA 60bp (GCn)	TATAGTTTAAACAAAACGAATGATTTTCGACTCATTAAATTATGATAATCATATTTACCAA	22	60	45.65 deg C
DNA 30bp (GCm)	GCTATAGCAATAAAAAC TAGGAATAGCCCC	43	30	54.65 deg C
DNA 60bp (GCm)	GCTATAGCAATAAAAAC TAGGAATAGCCCCCTTTCAC T TCTGAGTCCCAGAGGTTACCCAA	45	60	64.31 deg C
DNA 30bp (GCm)	AAGGCCCCACC CCAGTCTCAGCCCTACTCC	67	30	73.31 deg C
DNA 60bp (GCm)	TCTCCACCCCTGACTCCCCTCAGCCATAGAAGGCCCCACCC CAGTCTCAGCCCTACTCCA	63	60	78.98 deg C

Capture Extrakte Phase A

Tabelle B.67: Datenauswertung der Anreicherung von Extrakten aus Phase A

Seq: Anzahl Sequenzen (Rohdaten), Qual: Anzahl Sequenzen nach Qualitätsfilter (q15), ALN mt: alignierte Sequenzen, ALN Mkdup: nach Duplikatsentfernung, DUP %: Duplikationslevel, cov: mittlere Abdeckung der Zielregion (*coverage*), mD5: Deaminierungsrate am 5' Ende der Moleküle, Kont.: Kontaminationsangabe in Prozent, HG (CAP): Haplogruppe nach Capture, HG (PCR): Haplogruppe nach PCR

Probe	Extr	Seq	Qual	ALN mt	ALN mt [%]	ALN Mkdup	DUP [%]	cov	mD5	Kont [%]	multi	HG (CAP)	HG (PCR)
Klei 2	CPXI	661856	650221	404975	62,3%	24138	94,04%	73,4	0.294	0,1 - 1,7	1	J1c1	
Klei 2	CPI	401303	394327	242676	61,5%	7969	96,72%	24,0	0.297	4,0 - 10,9	2	J1c1	J1c
Klei 2	merge					32107		97,44	0,295	1,19-3,24	1	J1c1	
Klei 1	CPI	403450	394846	215978	54,7%	3898	98,20%	11,8	0.281	1,1 - 7,2	7	K1a4a1	
Klei 1	CPX	737520	724223	422342	58,3%	30669	92,74%	93,4	0.351	0,1 - 1,2	1	K1a4a1	K
Klei 1	merge					34567		105,16	0,342	0,39 - 1,55	1	K1a4a1	
Xir 1	CPVII	1564576	1539134	813043	52,8%	29758	96,34%	89,4	0.472	0,6 - 2,4	1	T2c1d1	
Xir 1	CPII	881118	864290	474198	54,9%	24458	94,84%	72,6	0.447	0,9 - 3,2	1	T2c1d1	H14b1
Xir 1	merge					54216		161,97	0,461	1,1-2,5	2	T2c1d1	
Tha 4	CPIV	267140	260903	112375	43,1%	10927	90,28%	33,0	0.272	46,8 - 68,4	10	R	
Tha 4	CPVII	122184	119450	47173	39,5%	1924	95,92%	5,7	0.285	4,2 - 24,3	16	HV0f	T2k
Tha 4	merge					12851		38,74	0,274	46,4-65,2	7	R1	
Fra 1	CPVI	310096	302764	130962	43,3%	10352	92,10%	31,5	0.217	38,3 - 56,2	11	U	
Fra 1	CPVIII	851515	832561	398428	47,9%	8402	97,89%	25,5	0.279	33,9 - 54,7	10	R0	X
Fra 1	merge					18754		57,02	0,244	51,6 - 66,8	8	R1	
Xir 2	CPII	609076	596415	294812	49,4%	20249	93,13%	61,2	0.349	0,7 - 3,1	2	T1a5	
Xir 2	CPVII	779096	764771	353425	46,2%	17262	95,12%	51,9	0.350	3,2 - 7,0	2	T1a5	T1a
Xir 2	merge					37511		113,08	0,350	2,2-4,5	1	T1a5	
Theo 5	CPX	674596	659474	329103	49,9%	8662	97,37%	25,6	0.350	7,2 - 13,4	3	K1c	
Theo 5	CPVIII	939832	919790	437289	47,5%	2896	99,34%	8,6	0.285	11,5 - 26,4	22	K1	T2k
Theo 5	merge					11558		34,17	0,335	9,7-15,4	4	K1c	
Theo 4	CPVIII	504657	492293	217229	44,1%	3774	98,26%	11,4	0.245	44,3 - 77,8	24	R	
Theo 4	CPVI	405334	394895	165771	42,0%	2268	98,63%	6,9	0.220	15,1 - 37,9	21	U	T2k
Theo 4	merge					6042		18,23	0,236	37,1-61,1	19	R1	
Fra 6	XXIV	326671	314932	85977	27,3%	2016	97,66%	6,1	0.228	14,9 - 42,4	9	R1	-
Theo 1	XXIV	3047966	3001085	1726866	57,5%	7359	99,57%	21,7	0.322	1,7 - 6,8	4	K1c	K
Fra 8	XXIV	1836878	1802758	1066755	59,2%	18424	98,27%	55,2	0.327	0,2 - 1,9	0	U3b3	H1aj1

Tabelle B.68: Nachsequenzierung kontaminierter Produkte aus C12 (HiSeq 100bp PE)

Seq: Anzahl Sequenzen (Rohdaten), Qual: Anzahl Sequenzen nach Qualitätsfilter (q15), ALN mt: alignierte Sequenzen, ALN Mkdup: nach Duplikatsentfernung, DUP %: Duplikationslevel, cov: mittlere Abdeckung der Zielregion (*coverage*), mD5: Deaminierungsrate am 5' Ende der Moleküle, Kont.: Kontaminationsangabe in Prozent, HG (CAP): Haplogruppe nach Capture, HG (PCR): Haplogruppe nach PCR

Probe	Extr	Seq	Qual	join	ALN mt	ALN mt [%]	ALN Mkdup	DUP [%]	cov	mD5	mD3	Frag (BAM)	STD	Kont [%]	HG (CAP)	score
Fra 1	CPVI	841476	802506	352726	144896	41,1%	10372	92,842%	47.7	0,229	0,2608	107.90	33,8	43,5 - 57,4	R	1
Fra 1	CPVIII	2674288	2558752	1139266	502611	44,1%	8404	98,328%	48	0,268	0,3209	94.45	33,19	29.3 - 46.8	R	1
Fra 1	merge						14525		93,57	0.242	0.275	106.35	34.68	53,9 - 68,7	R	1
Tha 4	CPIV	1241104	1150809	518974	176272	34,0%	11501	93,475%	69,6	0,269	0,3115	99.91	37.12	65.5 - 86.1	R	1
Tha 4	CPVII	447724	410748	182691	78204	42,8%	2008	97,432%	11	0,288	0,3202	90.80	38.11	5.4 - 23.9	HV0f	0,7
Tha 4	merge						12342		74.96	0.270	0.311	100.17	37.48	59,4 - 78,0	R	1
Theo 4	CPVIII	1514124	1453793	634330	274080	43,2%	2286	99,166%	13,9	0,224	0,2591	97.64	37,36	34.4 - 62.8	K1c	0,4
Theo 4	CPVI	1822060	1766908	805946	364674	45,2%	3881	98,936%	22,9	0,242	0,2881	101.09	36,74	44.9 - 73.0	R12'21	0,4
Theo 4	merge						5353		32.51	0.228	0.268	100.24	37,33	43,9 - 65,3	R1	0,6
Theo 5	CPX	3989364	3816029	1746769	801869	45,9%	2900	99,638%	15	0,279	0,3114	75.08	32.03	5.9 - 18.3	K1c	0,4
Theo 5	CPVIII	2635458	2527454	1176964	572594	48,7%	8973	98,433%	40,7	0,346	0,3774	85.88	37.04	5.1 - 10.8	K2	0,5
Theo 5	merge						10339		48,74	0,332	0,365	77.79	33.70	0,7 - 8,7	K1c	0,4
Fra 6	XXIV	1114524	1015803	438959	120992	27,6%	1898	98,431%	12	0,226	0,2576	104.68	36,98	21.9 - 50.3	R1	0,6

Dekontamination des HiSeq Laufs (100bp PE)**Tabelle B.69:** Dekontamination durch "schmutzi" (HiSeq 100bp SE). Kontaminationsbestimmung erfolgt durch das Programm .

Abkürzungen Tabelle siehe 4.37

Probe	Extr	Start KontEst (min/max)	Seq endo	endo hg	Seq cont	cont hg	Final KontEst (min/max)
Fra 1	CPVI	3 (0,5 - 5,5)	9303	U5a ^a b	515	K1a2a1	2 (1 - 3)
Fra 1	CPVIII	23 (21 - 25)	7260	R0	568	K1a4	12 (11-13)
Fra 1	merge	13,5 (11,5-15,5)	-	-	829	-	*
Tha 4	CPIV	19,5 (17,5-21,5)	9365	R	652	K1a4a1e	11 (10-12)
Tha 4	CPVII	15,5 (11-20)	2246	HV0f	312	H76a	7 (5-9)
Tha 4	merge	18 (16 - 20)	10323	(*)	778	T2c1d1	11 (10-12)
Theo 4	CPVIII	26 (21,5 - 30,5)	2024	U	237	H76a	12 (9 - 15)
Theo 4	CPVI	1,5 (0 - 5,5)	3510	R	338	K1a4a1e	10 (9 - 11)
Theo 4	merge	2,5(0-5,5)	4779	U2'3'4'7'8'9	451	H1ap1	13(12-14)
Theo 5	CPX	2,5 (0 - 6)	2581	K1c	277	H56a	17 (15 - 19)
Theo 5	CPVIII	21 (19,5 - 22,5)	8165	K1c	549	T2i2	9 (8 - 10)
Theo 5	merge	19 (17,5-20,5)	9116	K1c	670	T2k	11(10-12)
Fra 6	XXIV	45 (42 - 49)	1684	R1	196	H5e1a1	6 (4 - 8)

Tabelle B.70: Anwendung von *PMDtools* und resultierende Kontaminationsbestimmung (HiSeq 100bp PE).

Abkürzungen Tabelle siehe 4.38

(a) *PMDtools* (t1)

Probe	Extr	cov	Kont [%]	mD5	Frag Bam [bp]	HG
Fra 1	CPVI	22.05	41,8 - 68,3	0,582	109.78	R1
Fra 1	CPVIII	17.33	16,9 - 42,2	0,621	93.18	R1
Fra 1	merge	31.70	50,2 - 73,5	0.593	106.34	R1
Tha 4	CPIV	24.82	55,8 - 88,1	0,607	98.02	R12'21
Tha 4	CPVII	3.94	6,0 - 24,2	0,639	85.40	HV0f
Tha 4	merge	26.64	54,8 - 82,1	0.610	97.74	R12'21
Theo 4	CPVIII	4.63	10,9 - 40,1	0,594	102.61	U2c'd
Theo 4	CPVI	7.92	5,3 - 27,4	0,594	97.83	U8b
Theo 4	merge	10.98	22,4 - 59,1	0.582	100.62	U8b'c
Theo 5	CPX	4.93	9,5 - 34,3	0,619	85.41	K2
Theo 5	CPVIII	15.03	0,2 - 5,6	0,691	74.66	K1c
Theo 5	merge	17.67	1,8 - 9,4	0.672	77.23	K1c
Fra 6	XXIV	3.84	13,4 - 39,9	0,604	103.33	H27

(b) *PMDtools* (t3)

Probe	Extr	cov	Kont [%]	mD5	Frag Bam [bp]	HG	haploscore
Fra 1	CPVI	9.94	19,6 - 53,8	0,746	104.84	U	0,4
Fra 1	CPVIII	8.31	1,46 - 25,4	0,767	89.38	R1	0,6
Fra 1	merge	14.46	25,7 - 58,9	0,754	102.15	R1	0,6
Tha 4	CPIV	11.56	33,7 - 74,8	0,764	90.95	R12'21	0,4
Tha 4	CPVII	2.19	5,13 - 21,6	0,740	80.68	HV9c	0,5
Tha 4	merge	12.59	33,7 - 68,1	0,763	90.78	R12'21	0,4
Theo 4	CPVIII	2.50	1,5 - 16,4	0,770	98.55	U8b1	0,7
Theo 4	CPVI	3.66	5,5 - 29,3	0,783	90.26	H2a1n	0,5
Theo 4	merge	5.33	11,0 - 36,3	0,772	94.57	H2a3	0,6
Theo 5	CPX	2.77	8,3 - 34,7	0,767	81.25	K1c	0,7
Theo 5	CPVIII	7.83	2,86 - 18,2	0,812	72.33	K	0,8
Theo 5	merge	9.34	0,5 - 11,2	0,799	74.81	K	0,8
Fra 6	XXIV	1.94	1,7 - 16,8	0,801	95.54	H2a1m	0,5

B.6 Capture Ergebnisse kontaminierter Proben

Tabelle B.71: Detailliertes Sequenzierergebnis einzelner Extraktion von kontaminierten Proben. In der folgenden Tabelle ist der Dekontaminationsprozess anhand "schmutzi" dargestellt

beschreibung

Probe	Extr	Seq	Qual	join	ALN mt	ALN mt [%]	ALN Mkdup	DUP [%]	cov	mD5	mD3	Frag (BAM)	STD	Kont [%]	HG (CAP)	score
Theo5	XXIX	467175	461050	-	362928	78,7%	5738	98,4%	29,99	0.352	0.365	85.87	34.61	0,4 - 9,0	K1c	0,4
Theo 3	CPVII	1565510	1522035	709834	252423	35,56%	6978	97,2%	39,07	0.299	0.318	92.46	34.58	8,6 - 30,4	R11'B6	0,5
Theo 3	XXIII	357908	343212	159830	38917	24,35%	1876	95,2%	10,06	0.286	0.360	88.95	31.86	9,1 - 38,3	R1	0,6
Theo 3	XXVI	2435292	2348685	1119713	25902	2,31%	753	97,1%	3,22	0.456	0.517	71.16	27.77	0,9 - 38,09	R	0,6
Fra 1 (C14)	V	5396294	5086686	2260329	267712	12%	1226	99,5%	6,87	0.123	0.071	92.34	39.99	6,8 - 28,8	X2	1
Fra 1 (C15)	V	504188	496581	-	359436	72%	3122	99,1%	17,62	0.253	0.254	92.95	35.00	16,61 - 35,32	R	0,6
Fra 3	CPVIII	290354	275455	127649	37171	29,12%	7870	78,8%	40,02	0.267	0.305	84.05	31.62	5,1 - 29,1	H5	1
Fra 3	CPV	172990	161826	73902	23901	32,34%	1694	92,9%	7,24	0.339	0.423	71.39	29.79	4,6 - 39,8	T	0,4
Fra 3	CPX	785632	720693	323881	35459	10,95%	3325	90,6%	19,22	0.273	0.311	95.46	33.07	1,7 - 16,9	R	1
Fra 3	V	80970	76081	34124	8052	23,60%	712	91,2%	3,8	0.381	0.416	88.79	34.27	0,1 - 6,6	HV	1
Rev 7 (C14)	XXVII	4112046	3685467	1610099	856725	53,21%	1029	99,9%	3,31	0.393	0.357	52.99	18.12	0,8 - 18,1	K1a	0,8
Rev 7 (C14)	XXX	555988	514625	234711	124825	53,18%	3148	97,5%	15,07	0.274	0.313	79.08	31.72	8,1 - 20,9	H5	1
Rev 7 (C15)	XXVII	908830	882097	-	721019	81,74%	1387	99,8%	-	0.369	0.326	52.7	18,12	-	-	-
Rev 7 (C15)	XXX	1497596	1451482	-	1047946	72,20%	1114	99,9%	-	-	-	-	-	-	-	-
St6 30k-10	XXXIII	414162	406533	-	254556	62,62%	23614	90,7%	86,07	0.506	0.541	60.42	16.93	0,2 - 1,8	N1a1a1a3	1.00
St6 30k 40	XXXIII	549299	538298	-	345070	64,10%	28716	91,7%	112,19	0.513	0.551	64.66	16.50	0,1 - 1,0	N1a1a1a3	1.00
St6 merge 30k	XXXIII	963461	944831	-	599626	63,46%	30326	94,9%	124,62	0.509	0.543	67.94	17.44	0,2 - 1,3	N1a1a1a3	1.00
St6-50k	XXXIII	547300	536918	-	325105	60,55%	21291	93,5%	85,77	0.481	0.508	66.78	23.33	7,0 - 11,0	N1a1a1a3	1.00
Krk 5	XVIII	1097936	1075590	-	699790	65,06%	535	99,9%	1,55	0.201	-	-	-	0,2 - 9,8	J	0,5
Krk 5	XX	213280	208230	-	103461	49,69%	6655	93,6%	20,23	0.190	-	-	-	22,9 - 46,3	R	0,6
Tha 4	XX	776072	758702	-	346165	45,63%	2669	99,2%	8,06	0.286	-	-	-	23,6 - 50,4	U2'3'4'7'8'9	1

Tabelle B.72: Dekontamination belasteter Proben durch "schmutzi"

Probe	Extr	Start KontEst (min/max)	Seq endo	endo hg	score	Seq cont	cont hg	score	Final KontEst (min/max)
Theo 3	CPVII	0 (0-2,5)	6380	JT	0,4	519	T2b23	1	2 (1-3)
Theo 3	XXIII	0 (0-1,5)	1608	K	0,4	254	H10a1	1	9 (7-11)
Theo 3	XXVI	22,5 (18,5-26,5)	169	N1b	0,3	516	X2m	1	99
Fra 1 (C14)	V	43,5 (27,5-59,5)	865	X2b	0,5	154	T1a1r	1	14 (11-17)
Fra 1 (C15)	V	35 (32-38)	2739	U8	1	261	X2b	1	15 (13-17)
Fra 3	CPVII	12(9,5-14,5)	7183	H5	1	597	U3a	0,8	6(5-7)
Fra 3	CPV	0 (0-2,5)	1355	T2c1	1	282	H7a1a	0,5	11 (9-13)
Fra 3	CPX	0 (0-2,5)	2974	U	0,4	336	T2f1a	0,8	11 (10-12)
Fra 3	V	41 (36,5-45,5)	161	H3h2	1	527	U2'3'4'7'8'9	1	99
Krk 2	XIII	0 (0-1,5)	9858	H	1	1259	H1c2a	1	2 (1 - 3)
Krk 5	XVIII	56,5 (48,5-64,5)	100	K1	1	228	J2	0,8	99
Krk 5	XX	41,5(38,5-44,5)	5842	HV0	0,5	483	K1a4a1e	0,6	2(1-3)
Tha 4	XX	23,5(18,5-28,5)	2246	HV0	0,6	312	K1a4a1b2	1	9(7-11)
Rev 7 (C14)	XXVII	12,5 (6,5-18,5)	177	K1a	0,8	617	K1a	0,8	99
Rev 7 (C14)	XXX	13 (9,5-16,5)	2808	H5	1	318	K1a30	0,5	7(5-9)
Rev 7 (C15)	XXVII	14 (,5-19,5)		U	0,7		K1a	1	99
Rev 7 (C15)	XXX								

Tabelle B.73: Deskriptiv statistische Parameter der genetischen Diversität.

n: Anzahl Individuen, h: Anzahl Haplotypen, \hat{H} : Gendiversität (SD: Standardabweichung), n Pos: Anzahl analysierte Positionen, s: segregierende Positionen, Ts: Anzahl Transitionen, Tv: Anzahl Transversionen, π_n : Nukleotiddiversität, π : pairwise differences

Population	n	h	\hat{H}	SD \hat{H}	n Pos	s	Ts	Tv	π_n	SD π_n	π	SD π
Pre-LGM	16	11	0,950	0,036	347	12	10	2	0,0071	0,0045	2,49	1,41
SOP-JS	15	7	0,800	0,083	347	7	5	2	0,0041	0,003	1,45	0,93
Holo-JS	19	12	0,942	0,033	347	16	14	2	0,0082	0,005	2,87	1,57
NEE-HG	23	12	0,925	0,030	344	20	18	2	0,0122	0,007	4,2	2,16
NE-HG	25	11	0,863	0,050	316	11	10	1	0,0084	0,0051	2,65	1,46
PWC	28	14	0,918	0,031	339	15	15	0	0,0088	0,0053	2,98	1,6
Neo-JS	18	16	0,987	0,023	334	19	17	2	0,0128	0,0074	4,23	2,2
ANA-EN	27	15	0,906	0,045	348	24	22	2	0,0144	0,008	5,01	2,51
NGR-EN	16	11	0,908	0,063	348	17	17	0	0,0115	0,0067	4,02	2,11
NGR-MN	7	6	0,952	0,096	348	14	13	1	0,0167	0,0104	5,83	3,17
NGR-LN	26	14	0,911	0,040	347	23	20	3	0,0111	0,0064	3,85	2
KOR-EN	18	12	0,895	0,065	348	14	14	0	0,0110	0,0065	3,84	2,02
STA-EN	44	30	0,965	0,017	343	32	31	1	0,0146	0,008	5,03	2,49
LBK-T-EN	39	28	0,978	0,011	343	32	32	1	0,0140	0,0078	4,82	2,4
ALBK-I-MN	62	38	0,976	0,008	348	40	37	3	0,0133	0,0073	4,64	2,31
ALBK-II-MN	119	67	0,977	0,006	348	48	44	4	0,0134	0,0073	4,68	2,31
TIS-LN	48	27	0,946	0,022	348	36	31	5	0,0132	0,0073	4,59	2,29
LBK-EN	120	53	0,969	0,006	348	37	53	2	0,0145	0,0078	5,04	2,46
CE-MN	49	36	0,980	0,010	335	37	34	3	0,0142	0,0078	4,78	2,37
CE-LN	64	44	0,981	0,007	348	47	43	4	0,0135	0,0074	4,72	2,34
CE-FN	64	42	0,983	0,006	344	35	32	3	0,0106	0,006	3,65	1,87
NE-EN	22	17	0,970	0,024	335	15	15	0	0,0078	0,0048	2,63	1,46
NE-LN	11	11	1,000	0,039	326	22	20	2	0,0167	0,0098	5,44	2,83
MIN-BA	37	28	0,981	0,012	309	33	29	4	0,0127	0,0072	3,94	2,02
YAM-BA	28	23	0,984	0,014	348	33	32	2	0,0142	0,0079	4,95	2,48
CE-BA	94	62	0,986	0,004	343	48	44	7	0,0145	0,0079	5	2,45
CAT	23	11	0,885	0,048	348	18	18	0	0,0099	0,0058	3,45	1,83

B.7 Publikation Hofmanova, Kreutzer *et al.*

Aus dem vorliegenden Projekt konnten aus drei neolithischen Individuen aus Nord-Griechenland genomische Sequenzen mit einer Abdeckung des gesamten Genoms von 1-2x erstellt werden (siehe Tabelle 4.28). Zusammen mit zwei neolithischen Individuen aus der Marmara Region (Türkei) sind der westliche und östliche Teil der Ägäis zur Analyse der genetischen Übereinstimmungen abgedeckt. Bereits in ersten Analysen wurde der hohe Verwandtschaftsgrad der Individuen beider Seiten gezeigt, so dass diese Individuen einer Population zuzuordnen bzw. ihre gemeinsamen Vorfahren zeitlich nahe zu suchen sind. Die Analyse der ägäischen Individuen deckt ein Zeitfenster von 6500 - 4000 vor Chr. ab, wobei beide anatolischen Individuen und ein griechisches Individuum innerhalb des frühen Neolithikums einzuordnen sind. Zwei weitere Individuen aus Griechenland aus dem späten Neolithikum bilden die Grundlage, um neolithische Prozesse innerhalb der Ägäis und folgende Beziehungen zum Neolithikum auf dem europäischen Kontinent zu analysieren.

Early farmers from across Europe directly descended from Neolithic Aegeans

Zuzana Hofmanová^{a,1}, Susanne Kreutzer^{a,1}, Garrett Hellenthal^b, Christian Sell^a, Yoan Diekmann^b, David Díez-del-Molino^b, Lucy van Dorp^b, Saïoa López^b, Athanasios Kousathanas^{c,d}, Vivian Link^{c,d}, Karola Kiršanow^a, Lara M. Cassidy^e, Rui Martiniano^e, Melanie Strobel^a, Amelie Scheu^{a,e}, Kostas Kotsakis^f, Paul Halstead^g, Sevi Triantaphyllou^f, Nina Kyparissi-Apostolika^h, Dushka Urem-Kotsouⁱ, Christina Ziota^j, Fotini Adaktylou^k, Shyamalika Gopalan^l, Dean M. Bobo^l, Laura Winkelbach^a, Jens Blöcher^a, Martina Unterländer^a, Christoph Leuenberger^m, Çiler Çilingiroğluⁿ, Barbara Horejs^o, Fokke Gerritsen^p, Stephen J. Shennan^q, Daniel G. Bradley^e, Mathias Currat^r, Krishna R. Veeramah^l, Daniel Wegmann^{c,d}, Mark G. Thomas^b, Christina Papageorgopoulou^{s,2}, and Joachim Burger^{a,2}

^aPalaeogenetics Group, Johannes Gutenberg University Mainz, 55099 Mainz, Germany; ^bDepartment of Genetics, Evolution, and Environment, University College London, London WC1E 6BT, United Kingdom; ^cDepartment of Biology, University of Fribourg, 1700 Fribourg, Switzerland; ^dSwiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; ^eMolecular Population Genetics, Smurfit Institute of Genetics, Trinity College Dublin, Dublin 2, Ireland; ^fFaculty of Philosophy, School of History and Archaeology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; ^gDepartment of Archaeology, University of Sheffield, Sheffield S1 4ET, United Kingdom; ^hHonorary Ephor of Antiquities, Hellenic Ministry of Culture & Sports, 106 82 Athens, Greece; ⁱDepartment of History and Ethnology, Democritus University of Thrace, 69100 Komotini, Greece; ^jEphorate of Antiquities of Florina, Hellenic Ministry of Culture & Sports, 106 82 Athens, Greece; ^kEphorate of Antiquities of Chalkidiki and Mount Athos, Hellenic Ministry of Culture & Sports, 106 82 Athens, Greece; ^lDepartment of Ecology and Evolution, Stony Brook University, Stony Brook, New York 11794-5245; ^mDepartment of Mathematics, University of Fribourg, 1700 Fribourg, Switzerland; ⁿDepartment of Protohistory and Near Eastern Archeology, Faculty of Letters, Ege University, 35100 Bornova, Izmir, Turkey; ^oInstitute of Oriental and European Archaeology, Austrian Academy of Sciences, 1010 Vienna, Austria; ^pNetherlands Institute in Turkey, Beyoglu 34433, Istanbul, Turkey; ^qInstitute of Archaeology, University College London, London WC1H 0PY, United Kingdom; ^rAnthropology Unit, Department of Genetics & Evolution, University of Geneva, 1211 Geneva, Switzerland; and ^sLaboratory of Anthropology, Department of History and Ethnology, Democritus University of Thrace, 69100 Komotini, Greece

Edited by Eske Willerslev, University of Copenhagen, Copenhagen, Denmark, and approved April 29, 2016 (received for review May 18, 2015)

Farming and sedentism first appeared in southwestern Asia during the early Holocene and later spread to neighboring regions, including Europe, along multiple dispersal routes. Conspicuous uncertainties remain about the relative roles of migration, cultural diffusion, and admixture with local foragers in the early Neolithization of Europe. Here we present paleogenomic data for five Neolithic individuals from northern Greece and northwestern Turkey spanning the time and region of the earliest spread of farming into Europe. We use a novel approach to recalibrate raw reads and call genotypes from ancient DNA and observe striking genetic similarity both among Aegean early farmers and with those from across Europe. Our study demonstrates a direct genetic link between Mediterranean and Central European early farmers and those of Greece and Anatolia, extending the European Neolithic migratory chain all the way back to southwestern Asia.

paleogenomics | Neolithic | Mesolithic | Greece | Anatolia

It is well established that farming was introduced to Europe from Anatolia, but the extent to which its spread was mediated by demic expansion of Anatolian farmers, or by the transmission of farming technologies and lifeways to indigenous hunter-gatherers without a major concomitant migration of people, has been the subject of considerable debate. Paleogenetic studies (1–4) of late hunter-gatherers (HG) and early farmers indicate a dominant role for migration in the transition to farming in central and northern Europe, with evidence of only limited hunter-gatherer admixture into early Neolithic populations, but increasing toward the late Neolithic. However, the exact origin of central and western Europe's early farmers in the Balkans, Greece, or Anatolia remains an open question.

Recent radiocarbon dating indicates that by 6,600–6,500 calibrated (cal) BCE sedentary farming communities were established in northwestern Anatolia at sites such as Barcın, Menteşe, and Aktopraklık C and in coastal western Anatolia at sites such as Çukuriçi and Ulucak, but did not expand north or west of the Aegean for another several hundred years (5). All these sites show material culture affinities with the central and southwestern Anatolian Neolithic (6).

Early Greek Neolithic sites, such as the Franchthi Cave in the Peloponnese, Knossos in Crete, and Mauropigi, Paliambela, and

Revenia in northern Greece date to a similar period (7–9). The distribution of obsidian from the Cycladic islands, as well as similarities in material culture, suggest extensive interactions since the Mesolithic and a coeval Neolithic on both sides of the Aegean (8). Although it has been argued that in situ Aegean Mesolithic hunter-gatherers played a major role in the “Neolithization” of Greece (7), the

Significance

One of the most enduring and widely debated questions in prehistoric archaeology concerns the origins of Europe's earliest farmers: Were they the descendants of local hunter-gatherers, or did they migrate from southwestern Asia, where farming began? We recover genome-wide DNA sequences from early farmers on both the European and Asian sides of the Aegean to reveal an unbroken chain of ancestry leading from central and southwestern Europe back to Greece and northwestern Anatolia. Our study provides the coup de grâce to the notion that farming spread into and across Europe via the dissemination of ideas but without, or with only a limited, migration of people.

Author contributions: Z.H., S.K., G.H., K. Kiršanow, K.R.V., D.W., M.G.T., C.P., and J. Burger designed research; Z.H., S.K., L.v.D., S.L., L.W., M.U., D.W., and C.P. performed research; K. Kotsakis, P.H., S.T., N.K.-A., D.U.-K., C.Z., F.A., Ç.Ç., B.H., F.G., and C.P. provided archaeological background information; K. Kotsakis, S.T., N.K.-A., D.U.-K., C.Z., F.A., and F.G. provided samples; K. Kotsakis, S.T., D.U.-K., Ç.Ç., B.H., F.G., S.J.S., and C.P. provided text; M.S., A.S., D.W., and M.G.T. contributed new reagents/analytic tools; Z.H., S.K., G.H., C.S., Y.D., D.D.-d.-M., L.v.D., S.L., A.K., V.L., K. Kiršanow, L.M.C., R.M., J. Blöcher, C.L., K.R.V., D.W., and J. Burger analyzed data; and Z.H., S.K., G.H., D.D.-d.-M., L.v.D., S.L., K. Kiršanow, S.G., D.M.B., Ç.Ç., B.H., S.J.S., D.G.B., M.C., K.R.V., M.G.T., C.P., and J. Burger wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Mitochondrial genome sequences have been deposited in the GenBank database (KU171094–KU171100). Genomic data are available at the European Nucleotide Archive under the accession no. PRJEB11848 in BAM format.

¹Z.H. and S.K. contributed equally to this work.

²To whom correspondence may be addressed. Email: cpapage@he.duth.gr or jburger@uni-mainz.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1523951113/-DCSupplemental.

Principal Component Analysis, *f*-Statistics, and Mixture Modeling

The first two dimensions of variation from principal component analysis (PCA) reveal a tight clustering of all five Aegean Neolithic genomes with Early Neolithic (EN) genomes from central and southern Europe (2, 3, 13) (Fig. 2). This cluster remains well-defined when the third dimension of variation is also considered (https://figshare.com/articles/Hofmanova_et_al_3D_figure_S4/3188767). Two recently published pre-Neolithic genomes from the Caucasus (20) appear to be highly differentiated from the genomes presented here and most likely represent a forager population distinct from the Epipaleolithic/Mesolithic precursors of the early Aegean farmers.

To examine this clustering of Early Neolithic farmers in more detail, we calculated outgroup *f*₃ statistics (26) of the form *f*₃(\ddagger Khomani; TEST, Greek/Anatolian), where TEST is one of the available ancient European genomes (*SI Appendix, SI7. Using *f*-statistics to Infer Genetic Relatedness and Admixture Amongst Ancient and Contemporary Populations* and Figs. S8–S10; Dataset S2); \ddagger Khomani San were selected as an outgroup as they are considered to be the most genetically diverged extant human population. Consistent with their PCA clustering, the northern Aegean genomes share high levels of genetic drift with each other and with all other previously characterized European Neolithic genomes, including early Neolithic from northern Spain, Hungary, and central Europe. Given the archaeological context of the different samples, the most parsimonious explanation for this shared

drift is migration of early European farmers from the northern Aegean into and across Europe (12).

To better characterize this inferred migration, we modeled ancient and modern genomes as mixtures of DNA from other ancient and/or modern genomes, a flexible approach that characterizes the amount of ancestry sharing among multiple groups simultaneously (18, 27) (Fig. 3; *SI Appendix, SI10. Comparing Allele Frequency Patterns Among Samples Using a Mixture Model*). Briefly, we first represented each ancient or modern “target” group by the (weighted) number of alleles that they share in common with individuals from a fixed set of sampled populations (i.e., the “unlinked” approach described in ref. 27), which we refer to as the “allele-matching profile” for that target group. To cope with issues such as unequal sample sizes, we then used a linear model (28) to fit the allele-matching profile of the target group as a mixture of that of other sampled groups. Sampled groups that contribute most to this mixture indicate a high degree of shared ancestry with the target group relative to other groups. Under this framework the oldest Anatolian genome (Bar31) was inferred to contribute the highest amount of genetic ancestry (39–53%) to the Early Neolithic genomes from Hungary (13) and Germany (2) compared with any other ancient or modern samples, with the next highest contributors being other ancient Aegean genomes (Klei10, Pal7, Bar8) (*SI Appendix, Figs. S23, S24, and S29*). This pattern is not symmetric in that we infer smaller contributions from the German (<26%) and Hungarian (<43%) Neolithic genomes to any of the Anatolian or Greek ancient genomes. Furthermore, in this analysis

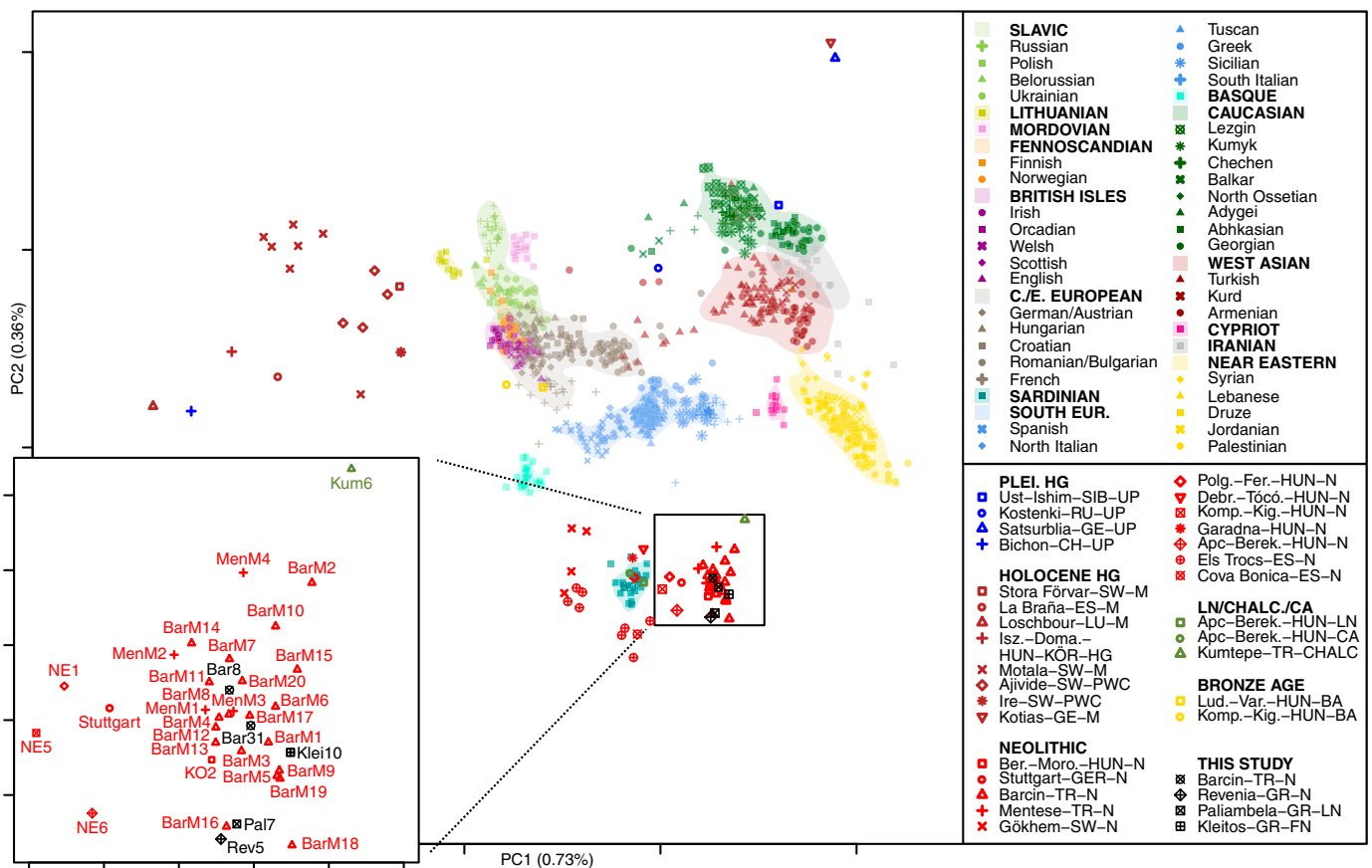


Fig. 2. PCA of modern reference populations (18, 19) and projected ancient individuals. The Greek and Anatolian samples reported here cluster tightly with other European farmers close to modern-day Sardinians; however, they are clearly distinct from previously published Caucasian hunter-gatherers (20). This excludes the latter as a potential ancestral source population for early European farmers and suggests a strong genetic structure in hunter-gatherers of Southwest Asia. Central and East European (C./E. European), South European (South Eur.). Ancient DNA data: Pleistocene hunter-gatherer (Plei. HG) (20, 21, 22), Holocene hunter-gatherer (Holocene HG) (2, 4, 13, 20, 23), Neolithic (2, 4, 12, 13, 24), Late Neolithic/Chalcolithic/Copper Age (LN/Chalc./CA) (13, 25), and Bronze Age (13). Ancient samples are abbreviated consistently using the nomenclature “site-country code-culture”; see *SI Appendix, Table S14* and *Dataset S1* for more information. A 3D PCA plot can be viewed as a 3D figure (https://figshare.com/articles/Hofmanova_et_al_3D_figure_S4/3188767).

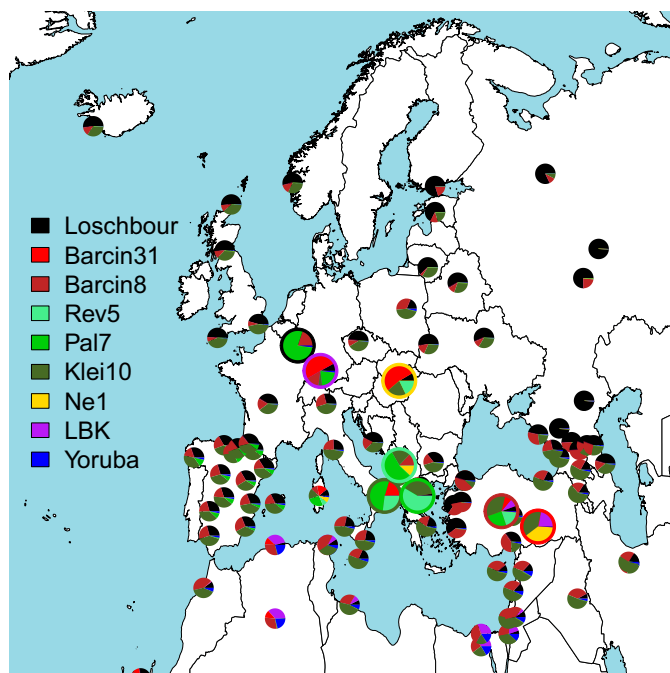


Fig. 3. Inferred mixture coefficients when forming each modern (small pies) and ancient (large pies, enclosed by borders matching key at left) group as a mixture of the modern-day Yoruba from Africa and the ancient samples shown in the key at left.

modern samples from Europe and surrounding regions are inferred to be relatively more genetically related to the Aegean Neolithic genomes than to the Neolithic genomes from Germany and Hungary (Fig. 3; *SI Appendix, S110. Comparing Allele Frequency Patterns Among Samples Using a Mixture Model*). These patterns are indicative of founder effects (29) in the German and possibly Hungarian Neolithic samples from a source that appears to be most genetically similar to the Aegean Neolithic samples (specifically, Bar31) and that distinguishes them from the ancestors of modern groups. Consistent with this, we found fewer short runs of homozygosity (ROH) (between 1 and 2 Mb) in our high-coverage Anatolian sample (Bar8) than in Early Neolithic genomes from Germany and Hungary (*SI Appendix, S111. Runs of Homozygosity* and Fig. S31). However, it is not possible to infer a direction for dispersal within the Aegean with statistical confidence because both the Greek and Anatolian genomes copy from each other to a similar extent. We therefore see the origins of European farmers equally well represented by Early Neolithic Greek and northwestern Anatolian genomes.

Ongoing gene flow into and across the Aegean is also indicated in the genome of a Chalcolithic individual from Kumtepe [Kum6 (25)], a site geographically close to Barcin but dating to ~1,600 y later. Although archaeological evidence indicates a cultural break in many Aegean and West Anatolian settlements around 5,700/5,600 cal BCE [i.e., spanning this 1,600-y period (30)], Kum6 shows affinities to the Barcin genomes in “outgroup” f_3 -statistics in the form f_3 (‡Khomani; TEST, Greek/Anatolian). The shared drift between Kum6 and both the early and late Neolithic Aegeans is similar in extent to the drift that Aegeans share with one another. However, f_4 statistics of the form f_4 (Aegean, Kum6, Early farmer, ‡Khomani) were often significantly positive (*SI Appendix, Table S22; Dataset S2*), suggesting that European Neolithic farmers [namely, Linearbandkeramik (LBK), Starcevo, and Early Hungarian Neolithic farmers] share some ancestry with early Neolithic Aegeans that is absent in Kum6. This is consistent with population structure in the Early Neolithic Aegean or with Kum6 being sampled from a population that differentiated from early Neolithic Aegeans after they expanded into the rest of Europe. Accordingly, compared with Barcin, Kum6 shares unique drift with the Late Neolithic genomes from Greece (Klei10 and Pal7), consistent with ongoing gene flow

across the Aegean during the fifth millennium and with archaeological evidence demonstrating similarities in Kumtepe ceramic types with the Greek Late Neolithic (31). Finally, the Kum6, Klei10, and Pal7 genomes show signals of Caucasus hunter-gatherer (20) admixture that is absent in the Barcin genomes, suggesting post early Neolithic gene flow into the Aegean from the east.

It is widely believed that farming spread into Europe along both Mediterranean and central European routes, but the extent to which this process involved multiple dispersals from the Aegean has long been a matter of debate (32). We calculated f_4 statistics to examine whether the Aegean Neolithic farmers shared drift with genomes from the Spanish Epicardial site Els Trocs in the Pyrenees (3, 12) that is distinct from that shared with Early Neolithic genomes from Germany and Hungary. In a test of the form f_4 (Germany/Hungary EN, Spain EN, Aegean, ‡Khomani), we infer significant unique drift among Neolithic Aegeans (not significantly in Bar8) and Early Neolithic Spain to the exclusion of Hungarian and German Neolithic genomes (*SI Appendix, Table S21*). The best explanation for this observation is that migration to southwestern Europe started in the Aegean but was independent from the movement to Germany via Hungary. This is also supported by other genetic inferences (24) and archaeological evidence (33). An alternative scenario is a very rapid colonization along a single route with subsequent gene flow back to Greece from Spain. Potentially, preexisting hunter-gatherer networks along the western Mediterranean could have produced a similar pattern, but this is not supported by archaeological data. Interestingly, Ötzi the Tyrolean Iceman (11) shows unique shared drift with Aegeans to the exclusion of Hungarian Early Neolithic farmers and Late and Post Neolithic European genomes and feasibly represents a relict of Early Neolithic Aegeans (*SI Appendix, S17. Using f -statistics to Infer Genetic Relatedness and Admixture Amongst Ancient and Contemporary Populations* and Table S18).

Hunter-Gatherer Admixture

Given that the Aegean is the likely origin of European Neolithic farmers, we used Bar8 and Bar31 as putative sources to assess the extent of hunter-gatherer admixture in European farmers through the Neolithic. f_4 statistics of the form f_4 (Neolithic farmer, Anatolian, HG, ‡Khomani) indicated small but significant amounts of hunter-gatherer admixture into both Spanish and Hungarian early farmer genomes, and interestingly, the Early Neolithic Greek genome. Our mixture modeling analysis also inferred a small genetic contribution from the Loschbour hunter-gatherer genome (3–9%) to each of the Early Neolithic Hungarian and German genomes, but evidence of a smaller contribution to any Aegean genomes (0–6%). These results suggest that mixing between migrating farmers and local hunter-gatherers occurred sporadically at low levels throughout the continent even in the earliest stages of the Neolithic. However, consistent with previous findings (3), both f_4 statistics and ADMIXTURE analysis indicate a substantial increase in hunter-gatherer ancestry transitioning into the Middle Neolithic across Europe, whereas Late Neolithic farmers also demonstrate a considerable input of ancestry from steppe populations (*SI Appendix, S18. Proportions of Ancestral Clusters in Neolithic Populations of Europe* and Fig. S32).

Relation to Modern Populations

Most of the modern Anatolian and Aegean populations do not appear to be the direct descendants of Neolithic peoples from the same region. Indeed, our mixture model comparison of the Aegean genomes to >200 modern groups (2) indicates low affinity between the two Anatolian Neolithic genomes and six of eight modern Turkish samples; the other two were sampled near the Aegean Sea at a location close to the site of the Neolithic genomes. Furthermore, when we form each Anatolian Neolithic genome as a mixture of all modern groups, we infer no contributions from groups in southeastern Anatolia and the Levant, where the earliest Neolithic sites are found (*SI Appendix, Figs. S22 and S30 and Table S30; Dataset S3*). Similarly, comparison of allele sharing between ancient and modern genomes to those expected under population continuity indicates Neolithic-to-modern discontinuity in Greece and western Anatolia, unless ancestral populations were unrealistically small

(*SI Appendix, S19. Population Continuity*). Instead, our mixing analysis shows that each Aegean Neolithic genome closely corresponds to modern Mediterraneans (>68% contributions from southern Europe) and in particular to Sardinians (>25%), as also seen in the PCA and outgroup f_3 statistics with few substantial contributions from elsewhere. Modern groups matching the Neolithics—mostly from the Mediterranean and North Africa—strikingly match more to Bar8 from northwestern Anatolia than to the LBK genome from Stuttgart in Germany, indicating that the LBK genome experienced processes such as drift and admixture that were independent from the Mediterranean expansion route, consistent with the dual expansion model.

Concluding Remarks

Over the past 7 years, ancient DNA studies have transformed our understanding of the European Neolithic transition (1–4, 12, 13), demonstrating a crucial role for migration in central and southwestern Europe. Our results further advance this transformative understanding by extending the unbroken trail of ancestry and migration all of the way back to southwestern Asia.

The high levels of shared drift between Aegean and all available Early Neolithic genomes in Europe, together with the inferred unique drift between Neolithic Aegeans and Early Neolithic genomes from Northern Spain to the exclusion of Early Neolithic genomes from central Europe, indicate that Aegean Neolithic populations can be considered the root for all early European farmers and that at least two independent colonization routes were followed.

A key remaining question is whether this unbroken trail of ancestry and migration extends all the way back to southeastern Anatolia and the Fertile Crescent, where the earliest Neolithic sites in the world are found. Regardless of whether the Aegean early farmers ultimately descended from western or central Anatolian, or even Levantine hunter-gatherers, the differences between the ancient genomes presented here and those from the Caucasus (20) indicate that there was considerable structuring of forager populations in southwestern Asia before the transition to farming. The dissimilarity and lack of continuity of the Early Neolithic Aegean genomes to most modern Turkish and Levantine populations, in contrast to those of early central and southwestern European farmers and modern Mediterraneans, is best explained by subsequent gene flow into Anatolia from still unknown sources.

Methods

Ancient DNA Extraction and Sequencing. Five Neolithic and two Mesolithic samples from both sides of the Aegean were selected for ancient DNA extraction and sequencing (Table 1). DNA was extracted, and Next Generation Sequencing libraries were constructed in dedicated ancient DNA facilities as previously described (1, 34) with slight modifications. DNA quality and quantity of all samples were derived from the combination of estimates of endogenous DNA content based on the percentage of reads mapping to the reference genome (GRCh37/hg19) after shallow Illumina Miseq sequencing and estimating the DNA copy number of extracts by quantitative PCR. The five Neolithic samples (Bar8, Bar31, Rev5, Klei10, and Pal7) showed endogenous DNA contents between 8.80 and 60.83% and underwent deep Illumina whole-genome resequencing. The two Mesolithic samples (Theo 1 and Theo 5) showed endogenous DNA content of only 0.05% and 0.62%, respectively, and were used to capture the full mitochondrial genome. *SI Appendix, Fig. S1*, displays the relationship between endogenous DNA content and copy number for each sample and DNA extraction. The enrichment of the mitochondrial genome was carried out with Agilent's SureSelect^{XT} in-solution target enrichment kit. The protocol for the preparation of further libraries for shotgun sequencing and capture was modified according to previously estimated sample quality, whereby some libraries from samples Bar8, Bar31, and Rev5 were prepared with USER treated DNA extract. Detailed information about the experimental setup is described in *SI Appendix, S12. Sample Preparation*.

Bioinformatics. All sequence reads underwent 3' adapter trimming and were filtered for low-quality bases. For paired-end sequences only pairs with overlapping sequence were retained and merged into a single sequence. All sequences were aligned against the human reference build GRCh37/hg19 using BWA (35) and realigned using the Genome Analysis Toolkit (36) (*SI Appendix, S13. Read Processing*). For genotyping, we developed a novel method to

recalibrate quality scores and call genotypes that probabilistically accounts for postmortem damage patterns as estimated in mapDamage2.0 (37). For low-coverage genomes, we further developed a Bayesian haploid caller to reliably identify the most likely allele call for each site (code available on request from D.W.). For further details see *SI Appendix, S15. Genotype Calling for Ancient DNA*.

Ancient DNA Authenticity. The assessment of ancient DNA authenticity was performed using the sequence reads mapping to the mitochondrial genome following the likelihood approach described in Fu et al. (38) (*SI Appendix, S14. Analysis of Uniparental Markers and X Chromosome Contamination Estimates*). Post-mortem damage deamination rates were estimated using mapDamage 2.0 (37) and are displayed together with distribution of DNA fragment lengths of each sample (*SI Appendix, Fig. S3*). We used ANGSD (39) to determine X-chromosome contamination in male samples (*SI Appendix, S14. Analysis of Uniparental Markers and X Chromosome Contamination Estimates*).

Analysis of Uniparental Markers. Mitochondrial haplogroups were determined using HaploFind (40). Consensus sequences in FASTA format were created from alignments with SAMtools (41) (*SI Appendix, S14. Analysis of Uniparental Markers and X Chromosome Contamination Estimates*).

To determine patrilineal lineages in ancient samples, we used clean_tree (42). This software requires BAM format files as input, and alleles are called with SAMtools mpileup at given SNP positions. These SNP positions were provided with the clean_tree software and contain 539 SNPs used for haplogroup determination (*SI Appendix, S14. Analysis of Uniparental Markers and X Chromosome Contamination Estimates*).

PCA. Principal component analysis was performed with LASER v2.02 (43). First, a reference space was generated on genotype data of modern individuals. For Fig. 2, we used European and Middle Eastern populations from a merged dataset published as part of Hellenthal et al. and Busby et al. (18, 19). In a second step, ancient samples provided as BAM files were projected into the reference space via a Procrustes analysis. See *SI Appendix, S16. PCA*, for details.

D-Statistics and ADMIXTURE Analysis. f_3 and f_4 statistics and the associated Z-scores (via block jackknife with default options) were calculated using the ADMIXTOOLS package (44) on haploid calls (*SI Appendix, S17. Using f-statistics to Infer Genetic Relatedness and Admixture Amongst Ancient and Contemporary Populations*). Samples from this study were compared with the Haak et al. (3) dataset containing 2,076 contemporary and ancient individuals. Additionally, ADMIXTURE analysis (45) was performed on a subset of these data containing all Eurasian ancient samples that predate the Bronze Age ($n = 77$) and additionally with Caucasus hunter-gatherers ($n = 79$) and prehistoric individuals from the eastern European steppe ($n = 89$) (*SI Appendix, S18. Proportions of Ancestral Clusters in Neolithic Populations of Europe*).

Mixture Modeling. To compose a target group as mixtures of other sampled groups, we used the following two-step procedure. First, we used a previously described technique (27) to infer an allele-matching profile for each target group by comparing its allele frequencies independently at each SNP to that of a set of "donor" groups. In particular, at a given SNP for each chromosome in our target group, we identified all X nonmissing donor chromosomes that shared the same allele type as the target and assigned each of these donors a score of $1/X$ and all other donors a score of 0. We did this for each SNP and each target individual and then summed up these scores across SNPs and target individuals to give an allele-matching profile for the target group conditional on that set of donors. For each target group, the contributions from each donor group were rescaled to sum to 1. For analyses presented here, our donor groups consisted of modern individuals (2) (plus Neanderthal and Denisova). Our target groups included all modern and ancient groups. Next, analogous to ref. 18, we performed a multiple linear regression using the target group's allele-matching profile as a response and a set of allele-matching profiles of different "surrogate" groups as predictors. In all analyses, we used three different sets of surrogate groups: (i) all (or a subset of) modern groups, (ii) all ancient groups and all (or a subset of) modern groups, and (iii) the modern Yoruba plus all (or a subset of) ancient groups. Mixture coefficients were inferred using nonnegative least squares in R with a slight modification to ensure that the coefficients sum to 1 (*SI Appendix, S110. Comparing Allele Frequency Patterns Among Samples Using a Mixture Model*).

Population Continuity Test. We used a forward-simulation approach to test for population continuity. For our purposes a continuous population is defined as "a single panmictic population without admixture from other populations." Our approach is designed to test continuity using a single ancient genome

and a set of modern genomes. We designate alleles as ancestral or derived by comparing them to the chimpanzee genome (panTro2) and consider only haploid calls for the ancient genome to avoid genotype calling biases. We examine the proportion of allele sharing between the ancient haploid and modern diploid genome calls that fall into each of the following six classes: A/AA; D/DD; A/DD; D/AA; A/AD; D/AD (where A = ancestral and D = derived alleles in the ancient haploid/modern diploid genomes, respectively). To generate expected proportions of these allele-sharing classes, we forward-simulate genetic drift by binomial sampling from a set of allele frequency vectors based on the modern site frequency spectrum. Finally, we use Fisher's method to combine two-tailed P values for the observed sharing class fractions falling into the simulated ranges and compare the resultant χ^2 values to those obtained by comparing each simulation against the set of all other simulations (46) to obtain a P value. We performed this test for a range of assumed ancient and modern effective population sizes (*SI Appendix, S19. Population Continuity*).

Runs of Homozygosity. The distribution of ROH for 5 ancient (2, 13) and 2,527 modern individuals (47) was determined with PLINK v1.90 (48) following the specifications used in ref. 13, with a set of 1,447,024 transversion SNPs called securely across all ancient samples (*SI Appendix, S11. Runs of Homozygosity*).

Functional Markers. Genotypes were determined using the diploid genotyping method described in *SI Appendix, S15. Genotype Calling for Ancient DNA*, and further verified through direct observation of BAM files using samtools tview

([htslib.org](#)). We included sites having $\geq 2\times$ coverage in the analysis (*SI Appendix, S12. Functional Markers*).

ACKNOWLEDGMENTS. We thank Songül Alpaslan for help with sampling in Barcin and Eleni Stravopodi for help with sampling in Theopetra. Z.H. and R.M. are supported by a Marie Curie Initial Training Network (BEAN/Bridging the European and Anatolian Neolithic, GA 289966) awarded to M.C., S.J.S., D.G.B., M.G.T., and J. Burger. C.P., J. Burger and S.K. received funding from DFG (BU 1403/6-1). C.P. and J. Burger received funding from the Alexander von Humboldt Foundation. C.S. and M.S. were supported by the European Union (EU) SYNTHESYS/Synthesis of Systematic Resources GA 226506-CP-CSA-INFRA, DFG: (BO 4119/1) and Volkswagenstiftung (FKZ: 87161). L.M.C. is funded by the Irish Research Council (GOIPG/2013/1219). A.S. was supported by the EU CodeX Project 295729. K. Kotsakis, S.T., D.U.-K., P.H., and C.P. were cofinanced by the EU Social Fund and Greek national funds research funding program THALES. C.P., M.U., K. Kotsakis, S.T., and D.U.-K. were cofinanced by the EU Social Fund and the Greek national funds research funding program ARISTEIA II. M.C. was supported by Swiss NSF Grant 31003A_156853. A.K. and D.W. were supported by Swiss NSF Grant 31003A_149920. S.L. is supported by the BBSRC (Grant BB/L009382/1). L.v.D. is supported by CoMPLEX via EPSRC (Grant EP/F500351/1). G.H. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant 098386/Z/12/Z) and by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. M.G.T. and Y.D. are supported by a Wellcome Trust Senior Research Fellowship Grant 100719/Z/12/Z (to M.G.T.). J. Burger is grateful for support by the University of Mainz and the HPC cluster MOGON (funded by DFG; INST 247/602-1 FUGG). F.G. was supported by Grant 380-62-005 of the Netherlands Organization for Scientific Research.

- Bramanti B, et al. (2009) Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* 326(5949):137–140.
- Lazaridis I, et al. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513(7518):409–413.
- Haak W, et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522(7555):207–211.
- Skoglund P, et al. (2014) Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* 344(6185):747–750.
- Weninger B, et al. (2014) Neolithisation of the Aegean and Southeast Europe during the 6600–6000 calBC period of Rapid Climate Change. *Documenta Praehistorica* 41:1–31.
- Özdoğan M (2014) A new look at the introduction of the Neolithic way of life in Southeastern Europe. Changing paradigms of the expansion of the Neolithic way of life. *Documenta Praehistorica* 41:33–49.
- Perles C, Quiles A, Valladas H (2013) Early seventh-millennium AMS dates from domestic seeds in the Initial Neolithic at Franchthi Cave (Argolid, Greece). *Antiquity* 87(338):1001–1015.
- Kotsakis K (2014) Domesticating the periphery: New research into the Neolithic of Greece. *Pharos* 20(1):41–73.
- Maniatis Y (2014) Radiocarbon dating of the major cultural changes in Prehistoric Macedonia: Recent developments. 1912–2012. A Century of Research in Prehistoric Macedonia. *Proceedings of the International Conference, Archaeological Museum of Thessaloniki*, 22–24 November 2012, eds Stefani E, Merousis N, Dimoula A (Archaeological Museum of Thessaloniki, Thessaloniki, Greece), pp 205–222.
- Lacan M, et al. (2011) Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. *Proc Natl Acad Sci USA* 108(24):9788–9791.
- Keller A, et al. (2012) New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nat Commun* 3:698.
- Mathieson I, et al. (2015) Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* 528(7583):499–503.
- Gamba C, et al. (2014) Genome flux and stasis in a five millennium transect of European prehistory. *Nat Commun* 5:5257.
- Allentoft ME, et al. (2015) Population genomics of Bronze Age Eurasia. *Nature* 522(7555):167–172.
- Itan Y, Jones BL, Ingram CJ, Swallow DM, Thomas MG (2010) A worldwide correlation of lactase persistence phenotype and genotypes. *BMC Evol Biol* 10(1):36.
- Helgason A, et al. (2007) Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution. *Nat Genet* 39(2):218–225.
- Raj T, et al. (2013) Common risk alleles for inflammatory diseases are targets of recent positive selection. *Am J Hum Genet* 92(4):517–529.
- Hellenthal G, et al. (2014) A genetic atlas of human admixture history. *Science* 343(6172):747–751.
- Busby GB, et al. (2015) The role of recent admixture in forming the contemporary West Eurasian genomic landscape. *Curr Biol* 25(19):2518–2526.
- Jones ER, et al. (2015) Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun* 6:8912.
- Fu Q, et al. (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* 514(7523):445–449.
- Seguin-Orlando A, et al. (2014) Paleogenomics. Genomic structure in Europeans dating back at least 36,200 years. *Science* 346(6213):1113–1118.
- Olalde I, et al. (2014) Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* 507(7491):225–228.
- Olalde I, et al. (2015) A common genetic origin for early farmers from Mediterranean Cardial and Central European LBK cultures. *Mol Biol Evol* 32(12):3132–3142.
- Omrak A, et al. (2016) Genomic evidence establishes Anatolia as the source of the European Neolithic gene pool. *Curr Biol* 26(2):270–275.
- Raghavan M, et al. (2014) Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* 505(7481):87–91.
- Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of population structure using dense haplotype data. *PLoS Genet* 8(1):e1002453.
- Leslie S, et al.; Wellcome Trust Case Control Consortium 2; International Multiple Sclerosis Genetics Consortium (2015) The fine-scale genetic structure of the British population. *Nature* 519(7543):309–314.
- van Dorp L, et al. (2015) Evidence for a common origin of blacksmiths and cultivators in the Ethiopian Ari within the last 4500 years: Lessons for clustering-based inference. *PLoS Genet* 11(8):e1005397.
- Schoop UD (2005) Das anatolische Chalkolithikum. Eine chronologische Untersuchung zur vorbronzezeitlichen Kultursequenz im nördlichen Zentralanatolien und den angrenzenden Gebieten [The Anatolian Chalcolithic Period. A chronological analysis of the Early Bronze Age cultural-sequence in northern Central Anatolia and neighbouring regions]. (Urgeschichtliche Studien I) (Bernhard Albert Greiner, Remshalden, Germany). German.
- Korfmann M, et al. (1995) Kumtepe 1993. Bericht über die Rettungsgrabung [Kumtepe 1993. Report on the rescue excavation]. *Stud Troica* 5:237–289. German.
- Whittle AWR (1996) *Europe in the Neolithic: The Creation of New Worlds* (Cambridge Univ. Press, Cambridge, UK), pp 1–460.
- Özdoğan M (2010) Westward expansion of the Neolithic way of life: Sorting the Neolithic package into distinct packages. *Proceedings of the 6th International Congress of the Archaeology of the Ancient Near East*, eds Matthiae P, Pinnock F, Nigro L, Marchetti N (Harrassowitz, Wiesbaden, Germany), Vol 1, pp 883–897.
- Kircher M, Sawyer S, Meyer M (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 40(1):e3.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760.
- DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43(5):491–498.
- Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L (2013) mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.
- Fu Q, et al. (2013) A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr Biol* 23(7):553–559.
- Korneliusen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics* 15:356.
- Vianello D, et al. (2013) HAPLOFIND: A new method for high-throughput mtDNA haplogroup assignment. *Hum Mutat* 34(9):1189–1194.
- Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079.
- Ralf A, van Oven M, Zhong K, Kayser M (2015) Simultaneous analysis of hundreds of Y-chromosomal SNPs for high-resolution paternal lineage classification using targeted semiconductor sequencing. *Hum Mutat* 36(1):151–159.
- Wang C, Zhan X, Liang L, Abecasis GR, Lin X (2015) Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *Am J Hum Genet* 96(6):926–937.
- Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192(3):1065–1093.
- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.
- Voight BF, et al. (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci USA* 102(51):18508–18513.
- The 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74.
- Chang CC, et al. (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4(1):7.

Anhang IV - Geräte und Chemikalien

C.1 Geräte

Tabelle C.74: Geräte

Gerät		Hersteller
2100 Expert Bioanalyzer	Quantifizierung und Fragmentlängenverteilung	Agilent Technologies
Electer Emax IH-300	Säge	MAFRA
Eppendorf Concentrator plus	Speed-vac	Eppendorf AG
Hettich Zentrifugen Mikro 200	Zentrifuge	Hettich Lab Technology
Hettich Zentrifugen Universal 320	Zentrifuge	Hettich Lab Technology
HORICO Diaflex®	Diamantschleifscheibe	HORICO DENTAL Hopf, Ringleb Ringleb & Co. GmbH & Cie
KERN EMB 600-2	Waage	KERN & SOHN GmbH
MasterCycler®	Thermocycler	Eppendorf AG
Punktstrahler P-G400	Sandstrahl Maschine	Harnisch+Rieth GmbH+Co.KG
Qubit® Fluorometer	Quantifizierung	Invitrogen™
Schwingmühle MM 200	Mahlmaschine	RETSCH GmbH
StepOnePlus™ Real-Time PCR System	qPCR Gerät	Applied Biosystems
Thermomixer comfort	Heizblock	Eppendorf AG

C.2 Verbrauchsmaterialien und Reinigungsequipment

Tabelle C.75: Verbrauchsmaterialien und Reinigungsequipment

Material	Hersteller
Amicon® Ultra-15 (30, 50 kDa)	Merck Millipore Corporation
Astronaut caps	HANSA - TRADING HTH GmbH & Co. KG
DNA LoBind Tubes 0,5 ml, 1,5 ml	Eppendorf AG
DuPont™ Tyvek® Protective Apparel	DuPont (via VWR International)
Eppendorf Safe-Lock Tubes 0,5 ml, 1,5 ml	Eppendorf AG
Falcon 15 ml, 50 ml	SARSTEDT AG & Co
Einweg Handschuhe	HANSA - TRADING HTH GmbH & Co. KG
Gesichtsmaske	HANSA - TRADING HTH GmbH & Co. KG
Reinigungsmittel	Hersteller
Alconox® detergent IF BC	Sigma-Aldrich Co.
DanKlorix Hygiene Reiniger	DanKlorix
DNA-ExitusPlus™ IF BC	AppliChem GmbH

C.3 Chemikalien und Kits

Tabelle C.76: Chemikalien

Chemikalie	Hersteller
Agencourt AMPure XP beads	Beckman Coulter
Agilent High Sensitivity DNA Reagent	Agilent Technologies
Aluminiumoxyd-Edelkorund	Harnisch+Rieth GmbH+Co.KG
AmpliTaq Gold [®] Puffer II (10x)	Life Technologies [™]
AmpliTaq Gold [®] DNA Polymerase	Life Technologies [™]
ATP (100 mM)	Life Technologies [™]
Bovine Serum Albumin (BSA) (20 mg/ml)	Roche Diagnostics
Bst Polymerase, Large Fragment (8 U/μl)	New England Biolabs GmbH
dNTPs (each 10 mM)	Qiagen
dNTPs (each 25 mM)	Agilent Technologies
Dynabeads [®] MyOne [™] Streptavidin T1 magnetische <i>beads</i>	Thermo Fisher Scientific Inc.
EDTA (0.5 M), pH 8.0	Ambion [®]
Ethanol 96% (vergällt)	Carl Roth GmbH + Co. KG
Herculase II fusion DNA Polymerase	Agilent Technologies
Herculase II Reaktionspuffer	Agilent Technologies
HPLC-H ₂ O	Fisher Scientific GmbH
Hydroxyapatite	Sigma-Aldrich Co.
KAPA SYBR [®] FAST QPCR MasterMix Universal	Kapa Biosystems/ Peqlab
MgCl ₂ (25 mM)	Life Technologies [™]
N-Laurylsarcosine	Merck Millipore Corporation
Nuclease-freies H ₂ O	Life Technologies [™]
PEG-4000	Thermo Scientific [™]
Proteinase-K	Roche Diagnostics
Roti [®] -Phenol/Chloroform/Isoamylalkohol	Roche Diagnostics
Seesand	Carl Roth GmbH + Co. KG
T4 DNA Ligase (5 U/μl)	Thermo Scientific [™]
T4 DNA Ligase Buffer (10X)	Thermo Scientific [™]
T4 DNA Polymerase (5 U/μl)	Thermo Scientific [™]
T4 Polynukleotide Kinase	Invitrogen [™]
Tango Buffer (10x)	Life Technologies [™]
ThermoPol Buffer (10X)	New England Biolabs GmbH
Trichlormethan/Chloroform	Carl Roth GmbH + Co. KG
USER [™] Enzym	New England Biolabs GmbH

Tabelle C.77: Kits

Kit	Hersteller
High Sensitivity DNA Kit	Agilent Technologies
MinElute [®] PCR Purification Kit	Qiagen
MSB [®] Spin PCRapace Kit	Stratec
Agilent [®] <i>Sure Select Target enrichment Kit</i>	Agilent
QIAquick PCR Purification Kit	Qiagen
Qubit [®] HS Assay Kit	Thermo Fisher Scientific Inc.

Tabelle C.78: Blocking Oligos.

Oligio ID	Sequenz (5' - 3')
Meyer.Bo4.P7.part1.R [120]	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-Pho
Meyer.Bo6.P7.part2.R [120]	CAAGCAGAAGACGGCATACGAGAT-Pho
BOIs1.revcomp	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-Pho
BO2.P5.part2_R	GTGTAGATCTCGGTGGTCGCCGTATCATT-Pho

Pho bezeichnet ein 3'-Phosphat.

Tabelle C.79: Adapter, Primer und Sequenzierprimer [120, 84].

Original Label	ID aDNA lab	Sequenz (5' - 3')
Adapter ¹		
IS1_adapter.P5	IS1_adapter.P5	A*C*A*C*TCTTTCCCTACACGACGCTCTTCCG*A*T*C*T
IS2_adapter.P7	IS2_adapter.P7	G*T*G*A*CTGGAGTTCAGACGTGTGCTCTTCCG*A*T*C*T
IS3_adapter.P5+P7	IS3_adapter.P5+P7	A*G*A*T*CGGAA*G*A*G*C
Primer		
P7 Index Primer		
see	see	CAAGCAGAAGACGGCATAACGAGATxxxxxxxGTGACTGGAGTTCAGACGTGT
P5 Index Primer		
see	see	AATGATACGGCGACCACCGAGATCTACACxxxxxxxACACTCTTTCCCTACACGACGCTCTT
IS4_indPCR.P5	IS4	AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTT
Amplification primers		
IS5_reamp.P5	IS5	AATGATACGGCGACCACCGA
IS6_reamp.P7	IS6	CAAGCAGAAGACGGCATAACGA
IS7_short_amp.P5	IS7	ACACTCTTTCCCTACACGAC
IS8_short_amp.P7	IS8	GTGACTGGAGTTCAGACGTGT
Sequenzier Primer		
Read 1 (P7) [120]		ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Index Read (P7) [120]		GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
Read 2 (P5) [120]		GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
P5 index sequencing [84]		AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT

¹ Zu Beginn wurde der adaptermix nach Meyer und Kircher (2010) hybridisiert [120]. Dies wurde während der Projektzeit geändert indem der Adaptermix in bereits hybridisierter Form bestellt werden konnte (*Integrated DNA Technologies*) bezeichnet PTO Bindung.

Lebenslauf

Anschrift Susanne Kreutzer, Mühlehaldenstr. 3, 8953 Dietikon
Geburtsdatum 12.05.83
Geburtsort Biedenkopf, Hessen, Deutschland
Staatsangehörigkeit Deutsch

Beruflicher Werdegang

Ab 12/2017 *Functional Genomics Center Zurich* (NGS Experte)
08/2016 – 12/2017 Institut für Molekulare Pathologie, UniversitätsSpital Zürich
(Diagnostik)
04/2012 – 07/2015 Institut für Anthropologie, Johannes-Gutenberg Universität,
Mainz (Wissenschaftliche Mitarbeit)
06/2007 – 04/2011 Radiologie Dr. Buddenbrock, Mainz (MRT Assistenz)
01/2004 – 07/2005 SGS Institut Fresenius, Taunusstein (BSE Analytik)

Akademischer Werdegang

Ab 04/2012 Promotion im Fachbereich der Biologie (Johannes Gutenberg Uni-
versität, Mainz)
10/2005 - 07/2011 Diplomstudiengang Biologie, Johannes-Gutenberg Universität,
Mainz
Studienschwerpunkte: Molekulargenetik, Anthropologie, Rechts-
medizin, Pharmakologie & Toxikologie
Diplomarbeit: Etablierung eines Protokolls zur selektiven Anrei-
cherung mitochondrialer DNA aus Skelettfunden und anschließen-
der Sequenzierung anhand der Illumina Technologie
09/2003 – 07/2005 Europa-Fachhochschule Fresenius, Idstein (Staatlich geprüfte
Biologisch-Technische Assistentin)
07/2003 Einhardschule Seligenstadt, Hessen , Abitur

Publikationen

- Zuzana Hofmanová, Susanne Kreutzer *et al.* (2016) "Early farmers from across Europe directly descended from Neolithic Aegeans." PNAS. doi: 10.1073/pnas.1523951113
- Nuno Da Silva, Susanne Kreutzer, Christina Papageorgopoulou & Mathias Currat. (2016) "Effect of population structure and migration when investigating genetic continuity using ancient DNA" (submitted to Proceedings of the Royal Society B)
- Joachim Burger und Susanne Kreutzer. "Fossilien und Genome – 200.000 Jahre Menschheitsgeschichte." PdN BIOLOGIE in der Schule/Schnittstelle Anthropologie/Archäologie (2016)
- Farnaz Broushaki *et al.* (2016) Early Neolithic genomes from the eastern Fertile Crescent. SCIENCE doi: 10.1126/science.aaf7943
- Daniela Lenggenhager *et al.* (2017) "Visualization of Hepatitis E virus RNA and proteins in the human liver." Journal of Hepatology (2017).